Associative Anti-Unification

Gabriela de Souza Ferreira Supervisor: Prof. Dr. Mauricio Ayala Ricón

Universidade de Brasília

February 10, 2023

э

A B A A B A







3 Associative Anti-Unification



э

About the development of work

This presentation is based in my masters final work, Syntactc, Commutative and Associative Anti-Unification, that was presented in the final of the last year and was supervised by Daniele Nantes.





イロト イポト イヨト イヨト

This talk

- We will present the Anti-Unification Problem modulo empty (∅) and associative (A) theories;
- We will present algorithms AUnif_E based on simplification rules for each of these cases:
 - pointing out the different results obtained for each equational theory,
 - give examples;
- Analyse the termination, confluence and correctness properties of the anti-unification algorithms,
 - with an especial attention on the prove of completeness of AUnif_A, that is different from the original approach in [AEEM14].

イロト イボト イヨト イヨト

History

7 Transformational Systems and the Algebraic Structure of	8 A Note on Inductive Cananatics for	11 An Experiment in Automatic Induction
Asamic Formulas	Conton D. Flottlin Sciences a landow histogram and the boots Sciences of a science of the scienc	P. J. Propoleticane Becatement of Models Indiana and Percellus Discords of Matters (
PPP II statutes to report if staty, 'smaller that if a statute is the state is t		<text><text><text><text><text><text></text></text></text></text></text></text>



First notions:

- Popplestone [Pop70],
- Plotkin [Plo70],
- and Reynolds [Rey70],
- Machine Intelligence Journal.

Important results:

- Existence of the solutions of the Syntactic, Commutative and Associative Anti-Unification Problems [Baa91],
- Development of methods to solve these problems [AEEM14].

Syntax

Before define the Anti-Unification Problem we need to define some basic concepts **Finite Signature:** $\sigma = \Sigma_{\emptyset} \cup \Sigma_A \cup \Sigma_C$.

•
$$\Sigma_{\emptyset} = \{\underbrace{a:0, b:0, c:0, d:0}_{\text{constants}}, f:n,...\}$$
 without an equational theory.
• $\Sigma_A = \{h:2\}$ with the associative function symbol

$$A = \{h(x, h(y, z)) \approx h(h(x, y), z)\}.$$

э

Generalizer

Definition (Generalizer)

Given two terms $s, t \in T(\mathcal{X}, \Sigma_{\emptyset})$. A generalizer of s and t is a term $r \in T(\mathcal{X}, \Sigma)$ for which there exists a pair of substitutions $\overline{\theta} = (\theta_1, \theta_2)$ such that $r\theta_1 = s$ and $r\theta_2 = t$.

gen(s, t).

Definition (Least General Generalization)

Given a signature Σ and terms s and $t \in T(\mathcal{X}, \Sigma)$. We define the **the least** general generalization of s and t as the greatest lower bound generalizer of s and t. In other words:

$$\lg(s,t) = \{r \in \operatorname{gen}(s,t) \mid r' \leq r, \forall r' \in \operatorname{gen}(s,t)\}$$

イロト イポト イヨト イヨト

э

Example - Generalizers

s = f(f(a, b), a) and t = f(f(c, b), c)



イロト 不得 トイヨト イヨト

э

Comparing term structures

The least general generalizer of s and t is the generalizer which maintains more the structure of s and t as possible.



3 x 3

5 A T 5 A A

Syntactic Anti-Unification Problem (AUP)

Definition $(\mathcal{A}\langle s,t \rangle)$

- Given: terms s and $t \in T(\mathcal{X}, \Sigma_{\emptyset})$,
- Find: The least general generalizer of s and t.

Example

Let s = f(f(a, b), a) and t = f(f(c, b), c). Then, r = f(f(z, b), z) it is an solution for the AUP A(s, t).

• • = • • = •

A rule based anti-unification algorithm

Input: $\mathcal{A}\langle s, t \rangle$



Output: $x\sigma \in \lg(s, t)$

• • = • • = •

$AUnif_{\emptyset}$ rules

(Dec) : Decompose:

$$\langle P \cup \{x : f(\overline{s_n}) \triangleq f(\overline{t_n})\} \mid S \mid \sigma \rangle \Longrightarrow \quad \langle P \cup \begin{cases} x_1 : s_1 \triangleq t_1, \\ \vdots \\ x_n : s_n \triangleq t_n \end{cases} \mid S \mid \sigma \{x \mapsto f(\overline{x_n})\} \rangle$$

where x_1, \ldots, x_n are fresh variables.

(Sol): Solve: If $root(s) \neq root(t)$ and there is no constraint $\{y : s \triangleq t\} \in S$

$$\langle P \cup \{x : s \triangleq t\} \mid S \mid \sigma \rangle \Longrightarrow \langle P \mid S \cup \{x : s \triangleq t\} \mid \sigma \rangle$$

(*Rec*): Recover: If $root(s) \neq root(t)$

$$P \cup \{x : s \triangleq t\} \mid S \cup \{y : s \triangleq t\} \mid \sigma \rangle \Longrightarrow \langle P \mid S \cup \{y : s \triangleq t\} \mid \sigma \{x \mapsto y\} \rangle$$

メロト メポト メヨト メヨト

Example

Consider the AUP $\mathcal{A}\langle s,t\rangle$ for s = f(f(a,b),a) and t = f(f(c,b),c).

$$\langle \{x : f(f(a, b), a) \triangleq f(f(c, b), c)\} \mid \emptyset \mid id \rangle$$

$$(Dec) \downarrow \qquad (Dec) \downarrow \qquad (Sol) \downarrow \qquad (Gec) \downarrow \qquad$$

Gabriela Ferreira (PPGMAT)

3

13/39

Example - Continuation

Consider the AUP $\mathcal{A}(s,t)$ for s = f(f(a,b),a) and t = f(f(c,b),c).

$$\langle \{x: f(f(a,b),a) \triangleq f(f(c,b),c)\} \mid \emptyset \mid id \rangle \stackrel{*}{\Longrightarrow}_{\operatorname{AUnif}_{\emptyset}} \langle \emptyset \mid \{x_{2}: a \triangleq c\} \mid \underbrace{\sigma_{1}\{x_{4} \mapsto b\}}_{\sigma_{4}} \rangle$$

Then, $\operatorname{AUnif}_{\emptyset}(s, t)$ gives $x\sigma_4 = f(f(x_2, b), x_2)$.



э

Properties of AUnif_Ø

Let $\mathcal{A}\langle s,t
angle$ be an AUP,

- Termination $AUnif_{\emptyset}$ terminates
- \bullet Confluence $\mathtt{AUnif}_{\emptyset}$ has a unique normal form except for variable renaming,
- Correctness $r \in \lg(s, t)$ iff there exists a derivation

$$\langle \{x: s \triangleq t\} \mid \emptyset \mid id \rangle \stackrel{*}{\Longrightarrow}_{\texttt{AUnif}_{\emptyset}} \langle \emptyset \mid S \mid \sigma \rangle$$

such that $x\sigma \equiv r$.

Therefore $\mathcal{A}\langle s,t\rangle$ always will have a solution that is unique, except for variable renaming.

オロト オポト オモト オモト

The Associative Anti-Unification Problem

Definition (Associative Anti-Unification Problem - AUP_A)

- Given: Two terms s and $t \in T(\mathcal{X}, \Sigma_{\emptyset \cup A})$,
- Find: The set $\lg_A(s, t)$.



The AUP_A for s and t is denoted by $\mathcal{A}_A(s, t)$.

Flattening

Given *h* the associative function symbol with $n \le 2$ arguments, flattened terms are canonical forms w.r.t. the set of rules given by the following rule schema

$$h(x_1,\ldots,h(t_1,\ldots,t_n),\ldots,x_n) \longrightarrow h(x_1,\ldots,t_1,\ldots,t_n,\ldots,x_n)$$

э

イロト 不得下 イヨト イヨト

AUnif_A: simplification rules

- (Dec): Decompose $(f \in \Sigma_{\emptyset} \text{ and } x \in \mathcal{X})$
- (Sol): Solve
- (Rec): Recover
- (A-Dec): Associative Decompose
 - (A-Left)
 - (A-Right)

э

イロト イボト イヨト イヨト

Associative for left and right sides

(A-Left) Associative-Left Decompose $\langle P \cup \{x : h(s_1, \ldots, s_n) \triangleq h(t_1, \ldots, t_m)\} \mid S \mid \sigma \rangle$

$$\implies \langle P \cup \begin{cases} x_1 : h(s_1, \dots, s_k) \triangleq t_1 \\ x_2 : h(s_{k+1}, \dots, s_n) \triangleq h(t_2, \dots, t_m) \end{cases} \mid S \mid \sigma\{x \mapsto h(x_1, x_2)\} \rangle$$

with $k \leq n-1$.

(A-Right) Associative-Right Decompose

$$\langle P \cup \{x : h(s_1, \dots, s_n) \triangleq h(t_1, \dots, t_m)\} \mid S \mid \sigma \rangle \\ \Longrightarrow \langle P \cup \begin{cases} x_1 : s_1 \triangleq h(t_1, \dots, t_k) \\ x_2 : h(s_2, \dots, s_n) \triangleq h(t_{k+1}, \dots, t_m) \end{cases} \mid S \mid \sigma \{x \mapsto h(x_1, x_2)\} \rangle$$

Gabriela Ferreira (PPGMAT)	GTC-UnB	February 10, 2023	19/39
with and $1 < k \leq m - 1$.		₹ <i>•</i> १ ९ ९ ९	

Example: AUP_A .

Consider $\mathcal{A}_A(s,t)$ an AUP_A with s = h(h(a,a), h(b,c)) and t = h(h(b,b), c).

term	s =	h(h(a,a),h(b,c))	$s =_A$	h(a, h(a, h(b, c)))
term	t =	h(h(b, b), c)	t =	h(h(b, b), c)
generalizer	$r_1 =$	h(h(x,x),y)	$r_{2} =$	h(x, y)
term	s =	h(h(a,a),h(b,c))	$s =_A$	h(a, h(a, h(b, c)))
term	$t =_A$	h(b, h(b, c))	$t =_A$	h(b, h(b, c))
generalizer	$r_3 =$	h(x, h(b, c))	$r_4 =$	$h(\mathbf{x}, h(\mathbf{x}, \mathbf{y}))$

Notice that

- $r_2 <_A r_1, r_3$ and r_4 ;
- $r_1 \equiv_A r_4$.

3

イロト イボト イヨト イヨト

Example

To apply the rules of $AUnif_A$ to solve this problem we first put s and t in they flattened form.



э

イロト 不得下 イヨト イヨト

Example

$$(A-Left)_{k=1} \qquad (A-Left)_{k=2} \qquad (A-L$$

Gabriela Ferreira (PPGMAT)

3

◆□> ◆圖> ◆国> ◆国>

Therefore, AUnif_A gives:

•
$$x\sigma_3 = h(x_1, h(x_1, x_4)) \equiv_A r_1$$
,

•
$$x\sigma_4 = h(x_1, h(b, c)) \equiv_A r_3$$
,



Therefore, $AUnif_A$ is not confluent.

э

Properties of AUnif_A

Let $\mathcal{A}_A \langle s, t \rangle$ be an AUP_A.

- Terminates AUnif_A terminates.
- Sound If $\langle \{x : s \triangleq t\} \mid \emptyset \mid id \rangle \stackrel{*}{\Longrightarrow}_{\operatorname{AUnif}_C} \langle \emptyset \mid S \mid \sigma \rangle$ then $x\sigma \in \operatorname{gen}_{\mathcal{A}}(s, t)$.
- Complete If $r \in \lg_A(s, t)$, then there exists a derivation

$$\langle \{x: s \triangleq t\} \mid \emptyset \mid id \rangle \stackrel{*}{\Longrightarrow}_{\mathtt{AUnif}_{\mathcal{A}}} \langle \emptyset \mid S \mid \sigma \rangle$$

such that $x\sigma \equiv_A r$.

There was a problem in the original proof of Completeness of $AUnif_A$ in [AEEM14].

オロト オポト オモト オモト

Notions

Before explain this problem we need to establish some notions.

Definition (Associative pair of positions)



\sim			• • •	DDCM	
(aa)	hrie	la ⊢	erreira i	PPGM	
			aa. (,

Definition (Associative Pair of Subterms)

Let $s, t \in T(\mathcal{X}, \Sigma_{\emptyset \cup A})$ be terms in flattened form. The pair of terms (u, v) is called an **associative pair of subterms** of *s* and *t* iff

(Regular Subterms) For each pair of positions p ∈ pos(s) and p' ∈ pos(t) such that s|_p = u, t|_{p'} = v and (p, p') is an associative pair of positions of s and t. Or:

Definition (Associative pair of subterms)

2 (Associative pair of subterms) There are positions an associative pair of positions (p, p') such that



27 / 39

Example



Given terms s = h(a, b, c, d) and t = h(a', b', c', d'), it follows (h(a, b), h(a', b')) is an associative pair of s and t.

э

() < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < ()

Proof of Completeness given by [AEEM14]

Now we can explain the problem in the proof.

Lemma (c.f. Lemma 19 in [AEEM14])

Given flattened terms t and t' such that every symbol in t and t' is either free or associative, and a fresh variable x, then there is a sequence

$$\langle \{y: t \triangleq t'\} \mid \emptyset \mid id \rangle \stackrel{*}{\Longrightarrow}_{\texttt{AUnif}_A} \langle P \cup \{y: u \triangleq v\} \mid S \mid \sigma \rangle$$

such that there is no variable z such that $\{z : u \triangleq v\} \in S$ if and only if (u, v) is an associative pair of subterms of t and t'.

Counter example: Simplification tree

Let $\mathcal{A}_A(s, t)$ be an AUP_A, where s = h(a, b, c, d) and t = h(a', b', c', d'), as applying the simplification rules of AUnif_A to solve this problem we obtain the following simplification tree:



30 / 39

Counter example: Description of configurations

$$C_{1} = \langle \{x_{2} : h(b, c, d) \triangleq h(b', c', d')\} | \{x_{1} : a \triangleq a'\}| \{x \mapsto h(x_{1}, x_{2})\}\rangle,$$

$$C_{2} = \langle \{x_{2} : h(c, d) \triangleq h(b', c', d')\} | \{x_{1} : h(a, b) \triangleq a'\}| \{x \mapsto h(x_{1}, x_{2})\}\rangle,$$

$$C_{3} = \langle \{x_{1} : h(a, b, c) \triangleq a', x_{2} : d \triangleq h(b', c', d')\} | \emptyset| \{x \mapsto h(x_{1}, x_{2})\}\rangle,$$

$$C_{4} = \langle \{x_{2} : h(b, c, d) \triangleq h(c', d')\} | \{x_{1} : a \triangleq h(a', b')\}| \{x \mapsto h(x_{1}, x_{2})\}\rangle,$$

$$C_{5} = \langle \{x_{1} : a \triangleq h(a', b', c'), x_{2} : h(b, c, d) \triangleq d'\} | \emptyset| \{x \mapsto h(x_{1}, x_{2})\}\rangle,$$

$$C_{1.1} = \langle \{x_{3} : b \triangleq b', x_{4} : h(c, d) \triangleq h(c', d')\}| \{x_{1} : a \triangleq a'\}\{x \mapsto h(x_{1}, h(x_{3}, x_{4}))\}\rangle,$$

$$C_{1.2} = \langle \{x_{3} : h(b, c) \triangleq b', x_{4} : d \triangleq h(c, d')\}| \{x_{1} : a \triangleq a'\} | \{x \mapsto h(x_{1}, h(x_{3}, x_{4}))\rangle,$$

$$C_{2.1} = \langle \{x_{3} : c \triangleq b', x_{4} : d \triangleq h(c, d')\}| \{x_{1} : h(a, b) \triangleq a'\}| \{x \mapsto h(x_{1}, h(x_{3}, x_{4}))\rangle,$$

$$C_{4.1} = \langle \{x_{3} : b \triangleq c, x_{4} : h(c, d) \triangleq d'\}| \{x_{1} : a \triangleq h(a, b')\}| \{x \mapsto h(x_{1}, h(x_{3}, x_{4}))\rangle,$$

$$C_{4.2} = \langle \{x_{3} : h(b, c) \triangleq c', x_{4} : d \triangleq d'\}| \{x_{1} : a \triangleq h(a, b')\}| \{x \mapsto h(x_{1}, h(x_{3}, x_{4}))\rangle,$$

$$C_{4.1} = \langle \{x_{3} : b \triangleq c, x_{4} : h(c, d) \triangleq d'\}| \{x_{1} : a \triangleq h(a, b')\}| \{x \mapsto h(x_{1}, h(x_{3}, x_{4}))\rangle,$$

$$C_{4.2} = \langle \{x_{3} : h(b, c) \triangleq c', x_{4} : d \triangleq d'\}| \{x_{1} : a \triangleq h(a, b')\}| \{x \mapsto h(x_{1}, h(x_{3}, x_{4}))\rangle,$$

Gabriela Ferreira (PPGMAT)

31/39

Counter example: conclusion

- There is no configuration $\langle P \cup \{y : h(a, b) \triangleq h(a', b')\} \mid S \mid \sigma \rangle$ in the simplification tree of $AUnif_A(s, t)$;
- Lemma 19 in [AEEM14] does not hold!

This lemma is used to prove the completeness of $AUnif_A$ in [AEEM14]. In order to show that this property still holds, we replace the Lemma 19 in [AEEM14] for tree new lemmas (Lemma 4.2, 4.3 and 4.4) that will be stated in the following frames.

э

Lemma (4.2)

Let $\mathcal{A}_A\langle s, t \rangle$ an AUP_A . If (p, p') is an associative pair of positions of s and t, then there exists a derivation of the form $\langle \{x : s \triangleq t\} \mid \emptyset \mid id \rangle \stackrel{*}{\Longrightarrow}_{AUnif_A} \langle \{y : u \triangleq v\} \mid S \mid \sigma \rangle$ with $(s|_p, t|_{p'}) = (u, v)$.



• Relates the arguments of the flattened terms with the configurations of AUnif_A.

Lemma (4.3)

Let $A_A(s, t)$ be an AUP_A and (p, p') an associative pair of positions of s and t, such that

$$s|_{p} = h(s_{1}, \ldots, s_{k}, u_{1}, \ldots, u_{n}, s_{k+1}, \ldots, s_{q}),$$

 $t|_{p'} = h(t_{1}, \ldots, s_{k'}, v_{1}, \ldots, v_{m}, t_{k'+1}, \ldots, s_{q'}).$

If $(u, v) = (h(\overline{u_n}), h(\overline{v_m}))$ is an associative pair of subterms, then there exists derivations such that

1.
$$\langle \{x : s \triangleq t\} | \emptyset | id \rangle \stackrel{*}{\Longrightarrow}_{AUnif_A} \langle P \cup \{y : h(u_1, \dots, u_i) \triangleq v_1\} | S | \sigma \rangle$$
 with
 $1 \le i \le n - 1$, and
2. $\langle \{x : s \triangleq t\} | \emptyset | id \rangle \stackrel{*}{\Longrightarrow}_{AUnif_A} \langle P \cup \{y : u_1 \triangleq h(v_1, \dots, v_j)\} | S | \sigma \rangle$ with
 $1 < i < m - 1$.



• Relates the associative pairs of subterms with the configurations of the derivations of $AUnif_A(s, t)$.

э

Lemma (4.4)

Let $\mathcal{A}_A\langle s,t\rangle$ be an AUP_A. If there exists a sequence of the form

$$\langle \{x: s \triangleq t\} \mid \emptyset \mid id \rangle \stackrel{*}{\Longrightarrow}_{\texttt{AUnif}_A} \langle P \cup \{y: u \triangleq v\} \mid S \mid \sigma \rangle$$

then (u, v) is an associative pair of subterms of s and t.

э

• • = • • = •

Conclusion

We have verified that:

problem	algorithm	terminating	confluent	sound	complete
AUP	$\texttt{AUnif}_{\emptyset}$	\checkmark	\checkmark	\checkmark	\checkmark
AUP _A	$\texttt{AUnif}_{\mathcal{A}}$	\checkmark	×	\checkmark	\checkmark

- AUP always have a unique solution (except for variable renaming)
- AUP_A always have a finite and minimal set of solutions (but AUnif_A do not gives minimal solutions),

Future work

- To obtain measure that gives a maximal bound of the number of normal forms obtained by $\Longrightarrow_{AUnif_C}$ and $\Longrightarrow_{AUnif_A}$.
- To extend the study of AUP_A for high order context of Nominal Framework, extending the work by Baumgartner et. al. in [BKLV15].



María Alpuente, Santiago Escobar, Javier Espert, and José Meseguer.

A modular order-sorted equational generalization algorithm. *Inf. Comput.*, 235:98–136, 2014.

Franz Baader.

Unification, weak unification, upper bound, lower bound, and generalization problems. In *Proc. of RTA*, volume 488 of *LNCS*, pages 86–97. Springer, 1991.



Alexander Baumgartner, Temur Kutsia, Jordi Levy, and Mateu Villaret. Nominal Anti-Unification.

In Proc. of. RTA, volume 36 of (LIPIcs), pages 57-73, Dagstuhl, Germany, 2015.



Gordon D Plotkin.

A note on inductive generalization. Machine intelligence, 5:153–163, 1970.



RJ Popplestone.

An experiment in automatic induction. *Machine Intelligence*, 5:203–215, 1970.



John C Reynolds.

Transformational systems and algebraic structure of atomic formulas. *Machine intelligence*, 5:135–151, 1970.