Automated Theorem Proving

Prof. Dr. Jasmin Blanchette, Yiming Xu, PhD,
Tanguy Bozec, and Lydia Kondylidou
based on exercises by Dr. Uwe Waldmann

Winter Term 2025/26

Exercises 9: Rewrite Systems

Exercise 9.1: Let $\Sigma = (\Omega, \emptyset)$ with $\Omega = \{b/0,\ f/1,\ g/1\}$. Let E be the set of (implicitly universally quantified) equations $\{f(g(f(x))) \approx b\}$.

Give one possible derivation for the statement $E \vdash f(g(b)) \approx b$.

Proposed solution.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{E \vdash f(g(f(b)) \approx b}}{E \vdash g(f(g(f(b)))) \approx g(b)}}{E \vdash f(g(f(g(f(b))))) \approx f(g(b))}}{E \vdash f(g(b)) \approx f(g(f(g(f(b)))))} \qquad \dfrac{}{E \vdash f(g(f(g(f(b))))) \approx b}}{E \vdash f(g(b)) \approx b}$$

Note that the "Instance" rule (which is used to derive the two leaf formulas) does not have a premise.

Exercise 9.2: Let $\Sigma = (\Omega, \emptyset)$, let $\Omega = \{f/1,\ b/0,\ c/0,\ d/0\}$. Let E be the set of equations $\{f(b) \approx d,\ f(c) \approx d,\ f(f(x)) \approx f(x)\}$. Let X be a countably infinite set of variables.

(a) Show that $f(d) \leftrightarrow_E^* d$.

(b) Sketch what the universe of $\mathrm{T}_\Sigma(\emptyset)/E$ looks like. How many elements does it have?

(c) Determine for each of the following equations whether it holds in $\mathrm{T}_\Sigma(X)/E$ and whether it holds in $\mathrm{T}_\Sigma(\emptyset)/E$. Give a very brief explanation.

$$f(b) \approx b \qquad (1)$$

$$\forall y\ f(f(f(y))) \approx f(f(y)) \qquad (2)$$

$$\forall x\, \forall y\ f(x) \approx f(y) \qquad (3)$$

Proposed solution. (a) $f(d) \leftarrow_E f(f(c)) \rightarrow_E f(c) \rightarrow_E d$.

(b) The universe of $\mathrm{T}_\Sigma(\emptyset)/E$ consists of the congruence classes of $\mathrm{T}_\Sigma(\emptyset)$ w.r.t. $\leftrightarrow_E^*$. Since every ground term except b and c can be rewritten to d using E, there are three such congruence classes, namely $[b] = \{b\}$, $[c] = \{c\}$, and $[d] = \mathrm{T}_\Sigma(\emptyset) \setminus \{b, c\}$.

(c) By Birkhoff's Theorem, an equation $\forall \vec{x}\,(s \approx t)$ holds in $\mathrm{T}_\Sigma(X)/E$ if and only if $s \leftrightarrow_E^* t$. Therefore, (2) holds in $\mathrm{T}_\Sigma(X)/E$, and (1) and (3) do not hold. (It is not possible to rewrite $f(b)$ to b or $f(x)$ to $f(y)$ using $\leftrightarrow_E$.)

For $\mathcal{T} = \mathrm{T}_\Sigma(\emptyset)/E$, we observe that for every assignment β, $\mathcal{T}(\beta)(f(b)) = [d]$ and $\mathcal{T}(\beta)(b) = [b]$, therefore (1) does not hold in $\mathrm{T}_\Sigma(\emptyset)/E$. On the other hand, for every assignment β, we have $\mathcal{T}(\beta)(f(f(f(y)))) = \mathcal{T}(\beta)(f(f(y))) = [d]$ and $\mathcal{T}(\beta)(f(y)) = \mathcal{T}(\beta)(f(x)) = [d]$, therefore both (2) and (3) hold in $\mathrm{T}_\Sigma(\emptyset)/E$.

Exercise 9.3: Let $\Sigma = (\Omega, \emptyset)$ with $\Omega = \{f/1,\ b/0,\ c/0,\ d/0\}$. Let E be the set of (implicitly universally quantified) equations $\{f(f(x)) \approx b\}$.

(a) Show that $b \leftrightarrow_E^* f(b)$. How does the rewrite proof look?

(b) Is the universe of the initial E-algebra $\mathrm{T}_\Sigma(\emptyset)/E$ finite or infinite? If it is finite, how many elements does it have?

Proposed solution. (a) The shortest rewrite proof has the form $b \leftarrow_E f(f(f(t))) \rightarrow_E f(b)$, where the term t can be chosen arbitrarily. (There are also more complicated rewrite proofs that consist of more than two steps.)

(b) The universe of $\mathrm{T}_\Sigma(\emptyset)/E$ consists of five congruence classes, namely $[c] = \{c\}$, $[d] = \{d\}$, $[f(c)] = \{f(c)\}$, $[f(d)] = \{f(d)\}$, and $[b]$. The last class contains all remaining ground terms, that is, b, $f(b)$, and all terms of the form $f(f(t))$ with $t \in \mathrm{T}_\Sigma(\emptyset)$.

Exercise 9.4: Let $\Sigma = (\Omega, \emptyset)$ be a first-order signature with $\Omega = \{f/1,\ b/0,\ c/0,\ d/0\}$. Let E be the set of Σ-equations

$$\{\forall x\,(f(x) \approx b),\ \ c \approx d\},$$

let $X = \{x, y, z\}$ be a set of variables. For any $t \in \mathrm{T}_\Sigma(X)$, let $[t]$ denote the congruence class of t w.r.t. E. Let $\mathcal{T} = \mathrm{T}_\Sigma(X)/E$, let $U_\mathcal{T}$ be the universe of $\mathcal{T}$, and let $\beta : X \to U_\mathcal{T}$ be the assignment that maps every variable to $[c]$. Determine for each of the following statements whether they are true or false:

(1) $[z]$ is a finite set of Σ-terms.

(2) $[f(z)]$ is a finite set of Σ-terms.

(3) $[c]$ is a set of ground Σ-terms.

(4) $[f(c)]$ is a set of ground Σ-terms.

(5) $U_{\mathcal{T}}$ is finite.

(6) $[b] \in U_{\mathcal{T}}$.

(7) $\{x, y\} \in U_{\mathcal{T}}$.

(8) $\mathcal{T}(\beta)(\forall z\,(z \approx f(x))) = 1$.

Proposed solution. The elements of the universe of $\mathcal{T}$ are the congruence classes of $\mathrm{T}_\Sigma(\{x, y, z\})$ with respect to E. There are five congruence classes, namely $\{x\}$, $\{y\}$, $\{z\}$, $\{c, d\}$, and a fifth class that contains all terms of $\mathrm{T}_\Sigma(\{x, y, z\})$ with f or b at the root. Consequently, we obtain:

(1) True. $[z] = \{z\}$.

(2) False. $[f(z)]$ contains b, $f(b)$, $f(f(b))$, $\ldots$.

(3) True. $[c] = \{c, d\}$.

(4) False. $[f(c)]$ contains, e.g., $f(z)$.

(5) True. See above.

(6) True. $[b]$ is a congruence class.

(7) False. $\{x, y\}$ is not a congruence class.

(8) False. $z \approx f(x)$ is false for $\beta[z \mapsto [c]]$.

Exercise 9.5: Let $\Sigma = (\Omega, \emptyset)$ be a first-order signature with $\Omega = \{f/2,\ b/0,\ c/0,\ d/0\}$. Let E be the set of Σ-equations

$$\{\forall x\,(f(x, c) \approx b),\ \ c \approx d\},$$

let $X = \{x, y, z\}$ be a set of variables. For any $t \in \mathrm{T}_\Sigma(X)$, let $[t]$ denote the congruence class of t w.r.t. E. Let $\mathcal{T} = \mathrm{T}_\Sigma(X)/E$ let $U_{\mathcal{T}}$ be the universe of $\mathcal{T}$, and let $\beta : X \to U_{\mathcal{T}}$ be the assignment that maps every variable to $[c]$. Determine for each of the following statements whether they are true or false:

(1) $[c]$ is a finite set of Σ-terms.

(2) $[f(c, c)]$ is a set of ground Σ-terms.

(3) $[x]$ is an element of the universe of $\mathcal{T}$.

(4) $\{b, f(x, c)\}$ is a congruence class w.r.t. E.

(5) $f(c, b) \in [f(d, b)]$.

(6) $f_{\mathcal{T}}([y], [d]) = [f(z, c)]$.

(7) $\mathcal{T}(\beta)(y \approx d) = 1$.

(8) $\mathcal{T}(\beta)(\forall z\,(z \approx c)) = 1$.

Proposed solution. (1) True. $[c] = \{c, d\}$.

(2) False. $f(y, c) \leftrightarrow_E b \leftrightarrow_E f(c, c)$ implies $f(y, c) \in [f(c, c)]$.

(3) True. The universe of $\mathcal{T} = \mathrm{T}_\Sigma(X)/E$ is the set of all E-congruence classes of terms in $\mathrm{T}_\Sigma(X)$, so it includes $[x]$.

(4) False. An E-congruence class contains *all* terms in $\mathrm{T}_\Sigma(X)$ that are E-equal to each other, so the E-congruence class of b and $f(x, c)$ contains, e.g., $f(c, c)$ and $f(f(y, y), c)$ as well.

(5) True. $f(c, b) \leftrightarrow_E f(d, b)$ implies $f(c, b) \in [f(d, b)]$.

(6) True. $f(y, d) \leftrightarrow_E f(y, c) \leftrightarrow_E b \leftrightarrow_E f(z, c)$ implies $f_\mathcal{T}([y], [d]) = [f(y, d)] = [f(z, c)]$.

(7) True. $\mathcal{T}(\beta)(y) = [c] = [d] = \mathcal{T}(\beta)(d)$, so $\mathcal{T}(\beta)(y \approx d) = 1$.

(8) False. For the modified assignment $\gamma = \beta[x \mapsto [b]]$, $\mathcal{T}(\gamma)(z) = [b] \neq [c] = \mathcal{T}(\gamma)(c)$.

Exercise 9.6 $(*)$**:** Find a signature Σ containing at least one constant symbol, a set E of Σ-equations, and two terms $s, t \in \mathrm{T}_\Sigma(X)$ such that

$$\mathrm{T}_\Sigma(\{x_1\})/E \models \forall \vec{x}\,(s \approx t),$$

but

$$\mathrm{T}_\Sigma(\{x_1, x_2\})/E \not\models \forall \vec{x}\,(s \approx t),$$

where $\vec{x}$ consists of all the variables occurring in s and t. (The variables in $\vec{x}$ need not be contained in $\{x_1, x_2\}$.)

Proposed solution. We take $\Sigma := (\{f/2,\ c/0\},\ \emptyset)$, $E := \{f(x, x) \approx c,\ f(x, c) \approx c,\ f(c, x) \approx c\}$ $s := f(x, y)$, and $t := c$.

Exercise 9.7: Let R be the following term rewrite system over $\Sigma = (\{f/1,\ g/2,\ h/1,\ c/0\}, \emptyset)$.

$$f(f(x)) \to h(h(x)) \qquad (1)$$
$$g(f(y), x) \to g(y, x) \qquad (2)$$
$$h(g(z, f(c))) \to f(z) \qquad (3)$$

Give all critical pairs between the three rules.

Proposed solution. There are three critical pairs:

Between (1) at position 1 and a renamed copy of (1):
mgu $\{x \mapsto f(x')\}$,
$h(h(f(x'))) \leftarrow f(f(f(x'))) \rightarrow f(h(h(x')))$,
critical pair: $\langle h(h(f(x'))), f(h(h(x')))\rangle$.

Between (2) at position 1 and a renamed copy of (1):
mgu $\{y \mapsto f(x')\}$,
$g(f(x'), x) \leftarrow g(f(f(x')), x) \rightarrow g(h(h(x')), x)$,
critical pair: $\langle g(f(x'), x), g(h(h(x')), x)\rangle$.

Between (3) at position 1 and (2):
mgu $\{z \mapsto f(y),\ x \mapsto f(c)\}$,
$f(f(y)) \leftarrow h(g(f(y), f(c))) \rightarrow h(g(y, f(c)))$,
critical pair: $\langle f(f(y)), h(g(y, f(c)))\rangle$.

Since there exists a nonjoinable critical pair, the system is not locally confluent.

Exercise 9.8: Let

$$\{f(b) \rightarrow f(c),\ f(c) \rightarrow f(d),\ f(d) \rightarrow f(b),\ f(x) \rightarrow x\}$$

be a rewrite system over $\Sigma = (\{f/1,\ b/0\, c/0,\ d/0\},\ \emptyset)$. Is it (a) terminating? (b) normalizing? (c) locally confluent? (d) confluent? Justify your answers.

Proposed solution. (a) No, the rewrite system is not terminating, due to the existence of infinite chains such as $f(b) \rightarrow f(c) \rightarrow f(d) \rightarrow f(b) \rightarrow \cdots$.

(b) Yes, the rewrite system is normalizing, because every term has a normal form. The normal form of b, c, and d is itself. The normal forms of any term of the form

$$\underbrace{f(f(\cdots(f}_{\geq 1\ f\text{'s}}(s))\cdots)),$$

where $s \in \{b, c, d\}$ are b, c, and d. For example, the normal form of b is b, the normal forms of $f(c)$ are b, c, and d, and the normal forms of $f(f(d))$ are b, c, and d.

(c) Yes, the rewrite system is locally confluent. There are three critical pairs:

Between the first rule at position ε and the fourth rule:
mgu $\{x \mapsto b\}$,
$f(c) \leftarrow f(b) \rightarrow b$,
critical pair: $\langle f(c), b\rangle$.
The pair is joinable: $f(c) \rightarrow f(d) \rightarrow f(b) \rightarrow b$.

Between the second rule at position ε and the fourth rule:
  mgu $\{x \mapsto c\}$,
  $f(d) \leftarrow f(c) \rightarrow c$,
  critical pair: $\langle f(d), c\rangle$.
The pair is joinable: $f(d) \rightarrow f(b) \rightarrow f(c) \rightarrow c$.

Between the third rule at position ε and the fourth rule:
  mgu $\{x \mapsto d\}$,
  $f(b) \leftarrow f(d) \rightarrow d$,
  critical pair: $\langle f(b), d\rangle$.
The pair is joinable: $f(b) \rightarrow f(c) \rightarrow f(d) \rightarrow d$.

(d) No, the system is not confluent. Consider the two chains $f(b) \rightarrow b$ and $f(b) \rightarrow f(c) \rightarrow d$. There is no way to join b and d, which are in normal form.

Exercise 9.9 (∗)**:** Let $\Sigma = (\Omega, \emptyset)$ with $\Omega = \{f/1, g/1, h/1, b/0, c/0\}$. Let R be the following term rewrite system over Σ:

$$\{g(f(x)) \rightarrow h(x),\ h(f(x)) \rightarrow g(x),\ g(b) \rightarrow c,\ h(c) \rightarrow b\}$$

Prove: If $s, t \in \mathrm{T}_\Sigma(X)$ and $R \models \forall \vec{x}\,(s \approx t)$, then there exists a rewrite derivation $s \leftrightarrow_R^* t$ with at most $|s| + |t| - 2$ rewrite steps.

Proposed solution. Since every application of a rule in R reduces the size of the term by 1, the rewrite system R is obviously terminating. It has no critical pairs, so it is locally confluent and, by termination, confluent. By Birkhoff's Theorem, $R \models \forall \vec{x}\,(s \approx t)$ if and only if $s \leftrightarrow_R^* t$. As R is confluent, $s \leftrightarrow_R^* t$ if and only if $s \rightarrow_R^* u \leftarrow_R^* t$ for some u. Since every R-rewrite step reduces the size of the term by 1, the derivation $s \rightarrow_R^* u$ can consist of at most $|s| - 1$ steps and the derivation $u \leftarrow_R^* t$ can consist of at most $|t| - 1$ steps; so we get a derivation $s \leftrightarrow_R^* t$ with at most $(|s| - 1) + (|t| - 1)$ rewrite steps.

Exercise 9.10 (∗)**:** Let $\Sigma = (\Omega, \emptyset)$ be a signature. Let R be a term rewrite system.

(a) Prove: If $s \rightarrow_R t$, then $\mathrm{var}(s) \supseteq \mathrm{var}(t)$.

(b) Prove: If $x \in X$ is a variable, $s \in \mathrm{T}_\Sigma(X)$ is a term such that $x \notin \mathrm{var}(s)$, and $R \models x \approx s$, then R is not confluent.

Proposed solution. (a) Assume that $s \to_R t$ using some rewrite rule $l \to r$ in R. Then $s = s[l\sigma]_p$ and $t = s[r\sigma]_p$. Since $\mathrm{var}(r) \subseteq \mathrm{var}(l)$, we obtain

$$\begin{aligned}\mathrm{var}(t) = \mathrm{var}(s[r\sigma]_p) &\subseteq \mathrm{var}(s) \cup \mathrm{var}(r\sigma)\\ &= \mathrm{var}(s) \cup \textstyle\bigcup_{x\in\mathrm{var}(r)} \mathrm{var}(x\sigma)\\ &\subseteq \mathrm{var}(s) \cup \textstyle\bigcup_{x\in\mathrm{var}(l)} \mathrm{var}(x\sigma)\\ &= \mathrm{var}(s) \cup \mathrm{var}(l\sigma) = \mathrm{var}(s).\end{aligned}$$

(b) First note that $s \to_R^* t$ implies $\mathrm{var}(s) \supseteq \mathrm{var}(t)$; this follows from part (a) by an obvious induction over the length of the rewrite derivation.

Assume that $x \in X$ is a variable, $s \in \mathrm{T}_\Sigma(X)$ is a term such that $x \notin \mathrm{var}(s)$, $R \models x \approx s$, and R is confluent. By Birkhoff's Theorem, $R \models x \approx s$ is equivalent to $x \leftrightarrow_R^* s$. Since confluence is equivalent to the Church–Rosser property, this implies that there exists a term t such that $x \to_R^* t$ and $s \to_R^* t$. Now note that the left-hand side of a rewrite rule cannot be a variable; therefore a variable x cannot be rewritten to any other term using $\to_R$. Consequently, $x = t$. But then $s \to_R^* x$, which implies that $\mathrm{var}(s) \supseteq \mathrm{var}(x) = \{x\}$, contradicting the assumption that $x \notin \mathrm{var}(s)$.

Exercise 9.11 ($*$)**:** Let $\Sigma = (\Omega, \emptyset)$ be a first-order signature, let E be a set of Σ-equations such that for every equation $s \approx s'$ in E neither s nor s' is a variable. For any term $t \in \mathrm{T}_\Sigma(X)$, let $[t]$ denote the congruence class of t w.r.t. E.

Prove or refute: For every variable $x \in X$ we have $[x] = \{x\}$.

Proposed solution. The statement holds. Proof: Assume that there is a variable $x \in X$ such that $[x] \neq \{x\}$. Since $x \in [x]$, this means that $[x]$ must contain some term t different from x. Therefore $E \vdash x \approx t$, and by Birkhoff's Theorem, this implies $x \leftrightarrow_E^* t$. Since t is different from x, we have $x \leftrightarrow_E^+ t$, and therefore $x \leftrightarrow_E t' \leftrightarrow_E^* t$ for some term t'. Consequently, $x \to_E t'$ or $t' \to_E x$. So some subterm of x must be equal to either $s\sigma$ or $s'\sigma$ for some equation $s \approx s'$ in E. This is impossible, though, since neither s nor s' is a variable.

(An alternative proof uses induction over the derivation tree for $E \vdash t \approx t'$ to show that no statement $E \vdash x \approx t$ with $t \neq x$ can be derived.)

Exercise 9.12 ($*$)**:** A friend asks you to proofread her master thesis. On page 15 of the thesis, your friend writes the following:

Lemma 5. Let $\succ$ be a well-founded ordering over a set A, let $\to$ be a binary relation such that $\to\ \subseteq\ \succ$. Let r be an element of A that is irreducible with respect to $\to$, and define $A_r = \{t \in A \mid t \to^* r\}$. If for every $u_0, u_1, u_2 \in A$ such that $u_1 \leftarrow u_0 \to u_2 \to^* r$ there exists a $u_3 \in A$ such that $u_1 \to^* u_3 \leftarrow^* u_2$, then for every $t_0 \in A_r$ and $t_1 \in A$, $t_0 \to^* t_1$ implies $t_1 \in A_r$.

Proof. We use well-founded induction over t_0 with respect to $\succ$. Let $t_0 \in A_r$ and $t_1 \in A$ such that $t_0 \to^* t_1$. If this derivation is empty, the result is trivial, so suppose that $t_0 \to t_1' \to^* t_1$. Since $t_0 \in A_r$ is reducible, it is different from r, hence there is a nonempty derivation $t_0 \to t_2 \to^* r$. By assumption, there exists a $t_3 \in A$ such that $t_1' \to^* t_3 \leftarrow^* t_2$. Now $t_0 \succ t_2$ and $t_2 \in A_r$, hence $t_3 \in A_r$ by the induction hypothesis, and thus $t_1' \in A_r$. Since $t_0 \succ t_1'$, we can use the induction hypothesis once more and obtain $t_1 \in A_r$ as required.

(1) Is the "proof" correct?

(2) If the "proof" is not correct:

   (a) Which step is incorrect?

   (b) Does the "theorem" hold? If yes, give a correct proof; otherwise, give a counterexample.

Proposed solution. Yes, the proof is correct.