

Formalized Signature Extension Results for Equivalence*

Alexander Lochmann¹, Fabian Mitterwallner¹, Aart Middeldorp¹

Department of Computer Science, University of Innsbruck, Austria
alexander.lochmann@student.uibk.ac.at,
{fabian.mitterwallner,aart.middeldorp}@uibk.ac.at

Abstract

Conversion equivalence and normalization equivalence are important properties of two rewrite systems. We investigate how many constants are needed to reduce these properties to their ground versions for linear variable-separated rewrite systems. Our results are implemented in the decision tool **FORT-h** and formalized in Isabelle/HOL. The latter enables the validation of the proofs produced by the former in the certifier **FORTify**.

1 Introduction

FORT-s is a tool to synthesize rewrite systems (TRSs) that satisfy a given property expressible in the first-order theory of rewriting. It is based on **FORT-h**, a tool that implements a decision procedure for the first-order theory of rewriting for the class of linear variable-separated TRSs, which comprises all left-linear right-ground TRSs. We refer to [3, 4, 6] for further details. It is of interest to synthesize TRSs that depend on one or more other TRSs. This can be done by passing additional TRSs to **FORT-s** in addition to a formula which references multiple systems. The additional systems are then also passed to the decision procedure. For example, if we want to transform the TRS \mathcal{R} consisting of the rules

$$a \rightarrow b \qquad f(a) \rightarrow b \qquad g(a, x) \rightarrow f(a)$$

into an equivalent complete (confluent and terminating) TRS, the command

```
fort-s "[0] (WCR & SN) & forall s, t ([0] s <->* t <=> [1] s <->* t)" file.trs
```

with `file.trs` containing \mathcal{R} (in COPS format) is executed. The latter is referred to by the index 1 in the formula whereas 0 refers to the TRS to be synthesized. The command returns the TRS \mathcal{S} consisting of the rules

$$a \rightarrow b \qquad f(b) \rightarrow g(a, a) \qquad g(b, b) \rightarrow a$$

The result is complete (as demanded by "[0] (WCR & SN)"), but not equivalent to \mathcal{R} ! The reason is that "`forall s, t ([0] s <->* t <=> [1] s <->* t)`" ensures equivalence on ground terms (since the decision procedure implemented in **FORT-h** is based on tree automata techniques) but this is not the same as equivalence on all terms; we have $g(a, x) \leftrightarrow_{\mathcal{R}}^* a$ but $g(a, x) \leftrightarrow_{\mathcal{S}}^* a$ does not hold.

In [2] we presented formalized results that allow reducing confluence-related properties (confluence **CR**, unique normal forms with respect to reduction **UNR** and conversion **UNC**, commutation **COM**) to properties on ground terms by adding fresh constants to the underlying signature. In this note we present similar results for two different notions of equivalence, conversion equivalence and normalization equivalence.

*This work was supported by the Austrian Science Fund (FWF) project P30301.

2 Preliminaries

In this paper we drop the usual constraints on TRSs by allowing terms on the right-hand sides of rules to contain variables not appearing on the left, and left-hand sides to be variables. A rule $\ell \rightarrow r$ is called *variable-separated* if $\text{Var}(\ell) \cap \text{Var}(r) = \emptyset$. A TRS is variable-separated if all its rules are variable-separated. Two TRSs \mathcal{R} and \mathcal{S} over a common signature \mathcal{F} are *conversion equivalent* (CE) if the relations $\leftrightarrow_{\mathcal{R}}^*$ and $\leftrightarrow_{\mathcal{S}}^*$ coincide on $\mathcal{T}(\mathcal{F}, \mathcal{V})$. They are *ground conversion equivalent* (GCE) if the relations coincide on the set $\mathcal{T}(\mathcal{F})$ of ground terms. We call \mathcal{R} and \mathcal{S} *normalization equivalent* (NE) if the relations $\rightarrow_{\mathcal{R}}^!$ and $\rightarrow_{\mathcal{S}}^!$ coincide on $\mathcal{T}(\mathcal{F}, \mathcal{V})$ and *ground normalization equivalent* (GNE) if this holds for $\mathcal{T}(\mathcal{F})$.

A binary predicate P on terms over a given signature \mathcal{F} is closed under *multi-hole contexts* if $P(C[s_1, \dots, s_n], C[t_1, \dots, t_n])$ holds whenever C is a multi-hole context over \mathcal{F} with $n \geq 0$ holes and $P(s_i, t_i)$ holds for all $1 \leq i \leq n$. In the results presented in the next section we make use of the following result from [2]. Here $\rightarrow_{\mathcal{A}}^{*\epsilon*}$ abbreviates $\rightarrow_{\mathcal{A}}^* \cdot \rightarrow_{\mathcal{A}}^{\epsilon} \cdot \rightarrow_{\mathcal{A}}^*$, so $s \rightarrow_{\mathcal{A}}^{*\epsilon*} t$ if $s \rightarrow_{\mathcal{A}}^* t$ contains a root step.

Lemma 1. *Let \mathcal{A} be a TRS over some signature \mathcal{F} and let P be a binary predicate that is closed under multi-hole contexts over \mathcal{F} . If $s \rightarrow_{\mathcal{A}}^{*\epsilon*} t \implies P(s, t)$ for all terms s and t then $s \rightarrow_{\mathcal{A}}^* t \implies P(s, t)$ for all terms s and t . \square*

Rewrite sequences involving root steps play an important role for linear variable-separated TRSs since they permit the use of different substitutions for the left and right-hand side of the employed rewrite rule, due to variable separation. We also make use of [2, Lemma 8].

Lemma 2. *Let \mathcal{R} be a linear TRS over a signature \mathcal{F} that contains a constant c which does not appear in \mathcal{R} . If $s \rightarrow_{\mathcal{R}}^* t$ with $c \in \text{Fun}(s) \setminus \text{Fun}(t)$ then $s[u]_p \rightarrow_{\mathcal{R}}^* t$ using the same rewrite rules at the same positions, for all terms u and positions $p \in \text{Pos}(s)$ such that $s|_p = c$. \square*

3 Results

The results in this section are formalized in Isabelle/HOL [1]. Similar to the example in the introduction, the following example shows that the two equivalence properties are not equivalent to their ground versions.

Example 3. The linear variable-separated TRSs

$$\mathcal{R}: \quad f(x) \rightarrow a \qquad \mathcal{S}: \quad f(a) \rightarrow a \quad f(f(a)) \rightarrow a$$

over the signature $\mathcal{F} = \{f, a\}$ are neither normalization equivalent nor conversion equivalent as can be seen from $f(x) \rightarrow_{\mathcal{R}}^! a$ and $f(x) \not\rightarrow_{\mathcal{S}}^! a$. Since every ground term rewrites in \mathcal{R} and \mathcal{S} to the unique ground normal form a , the TRSs are ground normalization equivalent as well as ground conversion equivalent. However, adding a single fresh constant c to the signature is sufficient to reproduce the counterexample: $f(c) \rightarrow_{\mathcal{R}}^! a$ and $f(c) \not\rightarrow_{\mathcal{S}}^! a$. So the TRSs are neither ground normalization equivalent nor ground conversion equivalent over the extended signature $\mathcal{F} \uplus \{c\}$.

In later proofs we will limit the rewrite sequences under consideration to those containing root steps by instantiating Lemma 1

- with $P_1(s, t): s \rightarrow_{\mathcal{S} \cup \mathcal{R}}^* t$ and $\mathcal{R} \cup \mathcal{R}^-$ for \mathcal{A} in proofs related to CE, and
- with $P_2(s, t): t \in \text{NF}_{\mathcal{R}} \implies s \rightarrow_{\mathcal{S}}^* t$ and \mathcal{R} for \mathcal{A} in proofs related to NE.

Note the identity $\rightarrow_{\mathcal{R} \cup \mathcal{R}^-} = \leftrightarrow_{\mathcal{R}}$ in the first case. We also use the symmetric instances, with \mathcal{R} and \mathcal{S} switching places, for both properties. Both P_1 and P_2 are closed under multi-hole contexts. By considering only sequences containing root steps we can use different substitutions on the left and right of the sequence, due to variable-separation. These substitutions will usually introduce fresh constants in the terms. We will also use Lemma 2 in subsequent proofs to remove these additional constants from rewrite sequences as follows. Let σ_c denote the substitution mapping all variables to c . If $s\sigma_c \rightarrow_{\mathcal{R}}^* t$ then $s \rightarrow_{\mathcal{R}}^* t$ by repeatedly applying Lemma 2 (to each occurrence of c in $s\sigma_c$), assuming c appears neither in \mathcal{R} nor in t .

A single fresh constant suffices to reduce conversion equivalence to ground conversion equivalence.

Theorem 4. *Linear variable-separated TRSs \mathcal{R} and \mathcal{S} over a common signature \mathcal{F} such that $\mathcal{T}(\mathcal{F}) \neq \emptyset$ are conversion equivalent if and only if \mathcal{R} and \mathcal{S} are ground conversion equivalent over the signature $\mathcal{F} \uplus \{c\}$.*

Proof. For the if direction we assume that \mathcal{R} and \mathcal{S} are ground conversion equivalent over $\mathcal{F} \uplus \{c\}$. Due to Lemma 1 (instantiated with P_1) and symmetry, it suffices to show the inclusion $\leftrightarrow_{\mathcal{R}}^{*\epsilon*} \subseteq \leftrightarrow_{\mathcal{S}}^*$ on terms in $\mathcal{T}(\mathcal{F}, \mathcal{V})$. Suppose $s \leftrightarrow_{\mathcal{R}}^{*\epsilon*} t$. Let $d \in \mathcal{F}$ be a constant, whose existence is guaranteed by the assumption $\mathcal{T}(\mathcal{F}) \neq \emptyset$, and consider the substitutions σ_c and σ_d mapping all variables to the constants c and d respectively. Closure under substitutions and variable separation yields $s\sigma_c \leftrightarrow_{\mathcal{R}}^{*\epsilon*} t\sigma_c$ and $s\sigma_c \leftrightarrow_{\mathcal{R}}^{*\epsilon*} t\sigma_d$. Ground conversion equivalence gives $s\sigma_c \leftrightarrow_{\mathcal{S}}^* t\sigma_c$ and $s\sigma_c \leftrightarrow_{\mathcal{S}}^* t\sigma_d$, and thus also $t\sigma_c \leftrightarrow_{\mathcal{S}}^* t\sigma_d$. Using Lemma 2 yields $s \leftrightarrow_{\mathcal{S}}^* t\sigma_d$ and $t \leftrightarrow_{\mathcal{S}}^* t\sigma_d$. Hence $s \leftrightarrow_{\mathcal{S}}^* t$ as desired.

For the only-if direction we assume that \mathcal{R} and \mathcal{S} are conversion equivalent over \mathcal{F} . Consider $s \leftrightarrow_{\mathcal{R}}^* t$ with $s, t \in \mathcal{T}(\mathcal{F} \uplus \{c\})$ and let $\phi_x^c(\cdot)$ be the function that replaces all occurrences of the constant c with the variable x in terms. Since the constant c does not appear in \mathcal{R} , we obtain $\phi_x^c(s) \leftrightarrow_{\mathcal{R}}^* \phi_x^c(t)$ from $s \leftrightarrow_{\mathcal{R}}^* t$. Conversion equivalence yields $\phi_x^c(s) \leftrightarrow_{\mathcal{S}}^* \phi_x^c(t)$. By choosing a variable $x \notin \text{Var}(s) \cup \text{Var}(t)$, the latter implies $s \leftrightarrow_{\mathcal{S}}^* t$ by closure under substitutions. \square

Two fresh constants are required to reduce normalization equivalence to its ground version.

Theorem 5. *Linear variable-separated TRSs \mathcal{R} and \mathcal{S} over a common signature \mathcal{F} are normalization equivalent if and only if \mathcal{R} and \mathcal{S} are ground normalization equivalent over $\mathcal{F} \uplus \{c, d\}$.*

Proof. The only-if direction can be proved with the methods used in the proof of Theorem 4. For the if direction we assume that \mathcal{R} and \mathcal{S} are ground normalization equivalent over $\mathcal{F} \uplus \{c, d\}$, which implies $\text{NF}_{\mathcal{R}} = \text{NF}_{\mathcal{S}}$. Hence, it remains to show that $s \rightarrow_{\mathcal{R}}^{*\epsilon*} t$ with $t \in \text{NF}_{\mathcal{R}}$ implies $s \rightarrow_{\mathcal{S}}^* t$ due to Lemma 1 (instantiated with P_2) and symmetry. From $s \rightarrow_{\mathcal{R}}^{*\epsilon*} t$ we obtain $s\sigma_c \rightarrow_{\mathcal{R}}^* t\sigma_d$, as the involved root step allows independent substitutions on the left and right-hand sides. Moreover, $t\sigma_d \in \text{NF}_{\mathcal{R}}$, since d does not occur in \mathcal{R} . From ground normalization equivalence we obtain $s\sigma_c \rightarrow_{\mathcal{S}}^* t\sigma_d$. Finally, Lemma 2 allows the removal of the substitutions, resulting in the desired rewrite sequence $s \rightarrow_{\mathcal{S}}^* t$. \square

Contrary to Theorem 4 one fresh constant is not sufficient as shown in the following example.

Example 6. Consider the two linear variable-separated TRSs

$$\begin{array}{llll}
 \mathcal{R}: & a \rightarrow b & f(f(x, y), z) \rightarrow f(b, b) & f(b, x) \rightarrow f(b, b) \\
 & f(x, a) \rightarrow f(z, b) & & \\
 \mathcal{S}: & a \rightarrow b & f(f(x, y), z) \rightarrow f(b, b) & f(b, x) \rightarrow f(b, b) \\
 & f(b, a) \rightarrow f(z, b) & f(f(x, y), a) \rightarrow f(z, b) &
 \end{array}$$

They are not normalization equivalent since $f(x, a) \rightarrow_{\mathcal{R}}^! f(z, b)$ and $f(x, a) \not\rightarrow_{\mathcal{S}}^* f(z, b)$. The TRSs are however ground normalization equivalent over the signature $\mathcal{F} \uplus \{c\}$. First observe that the only ground normal forms reachable via a rewrite sequence involving a root step are b and $f(c, b)$. The normal form b is reached (using a root step) only from a , in both \mathcal{R} and \mathcal{S} . The normal form $f(c, b)$ can be reached from all ground terms of the shape $f(t, a)$. For \mathcal{R} this is obvious and for \mathcal{S} this can be seen by a case analysis on the root symbol of t . Adding a second constant d allows one to mimick the original counterexample since $f(c, a) \rightarrow_{\mathcal{R}}^! f(d, b)$ and $f(c, a) \not\rightarrow_{\mathcal{S}}^* f(d, b)$.

For left-linear right-ground TRSs, a single fresh constant is enough to reduce normalization equivalence to ground normalization equivalence.

Theorem 7. *Left-linear right-ground TRSs \mathcal{R} and \mathcal{S} over a common signature \mathcal{F} are normalization equivalent if and only if \mathcal{R} and \mathcal{S} are ground normalization equivalent over $\mathcal{F} \uplus \{c\}$.*

Proof. We mention the differences with the proof of Theorem 5. For the identity of $\text{NF}_{\mathcal{R}}$ and $\text{NF}_{\mathcal{S}}$ for arbitrary terms, a single constant suffices. If $s \rightarrow_{\mathcal{R}}^{**} t$ then t is ground. Hence $s\sigma_c \rightarrow_{\mathcal{R}}^* t$ and thus $s\sigma_c \rightarrow_{\mathcal{S}}^* t$ by ground normalization equivalence. Lemma 2 gives $s \rightarrow_{\mathcal{S}}^* t$. \square

Each additional constant increases the execution time of FORT-h significantly. Hence results that reduce the required number are of obvious interest. For example, ground TRSs need no additional constants for the properties described in this paper. In the remainder of this section we present results for TRSs over *monadic* signatures, which are signatures that consist of constants and unary function symbols. In [6, Lemma 6] it is shown that for left-linear right-ground TRSs and properties related to confluence, no additional constants are needed. The same holds for commutation, which is a new (and formalized) result.

Theorem 8. *Right-ground TRSs \mathcal{R} and \mathcal{S} over a common monadic signature \mathcal{F} commute if and only if \mathcal{R} and \mathcal{S} ground commute.*

Proof. The only-if direction trivially holds. For the if direction we assume that \mathcal{R} and \mathcal{S} ground commute. Consider $s \rightarrow_{\mathcal{R}}^* t$ and $s \rightarrow_{\mathcal{S}}^* u$ for $s, t, u \in \mathcal{T}(\mathcal{F}, \mathcal{V})$. If $s = t$ or $s = u$, then $t \rightarrow_{\mathcal{S}}^* \cdot \mathcal{R}^* \leftarrow u$ obviously holds. So suppose $s \rightarrow_{\mathcal{R}}^+ t$ and $s \rightarrow_{\mathcal{S}}^+ u$. Since \mathcal{F} is monadic and \mathcal{R} and \mathcal{S} are right-ground, we infer that t and u are ground terms. Let $r \in \mathcal{T}(\mathcal{F})$ be an arbitrary ground term and let σ_r be the substitution which replaces all variables in $\text{Var}(s)$ by r . Since $t\sigma_r = t$, $u\sigma_r = u$ and \rightarrow^+ is closed under substitution, we obtain $s\sigma_r \rightarrow_{\mathcal{R}}^+ t$ and $s\sigma_r \rightarrow_{\mathcal{S}}^+ u$. Ground commutation yields the desired joining sequence $t \rightarrow_{\mathcal{S}}^* \cdot \mathcal{R}^* \leftarrow u$. \square

Note that Theorem 8 cannot be extended to linear variable-separated TRSs which require two constants even for monadic signatures, as seen by the TRS $\{a \rightarrow x\}$. It does not commute with itself, since $a \rightarrow x$ and $a \rightarrow y$, but it ground commutes with itself over the signatures $\{a\}$ and $\{a, c\}$.

Unlike commutation, the properties NE and CE require additional constants for TRSs over monadic signatures even for left-linear right-ground systems, as can be seen from Example 3. Nevertheless, we can reduce the number of constants to one if the signature is monadic, even if the restriction to left-linear right-ground TRSs is dropped. A key observation is that in non-empty rewrite sequences in a linear variable-separated TRS over a monadic signature fresh constants can be replaced by arbitrary terms.

Lemma 9. *Let \mathcal{R} be a variable-separated TRS over a monadic signature \mathcal{F} that contains a constant c which does not appear in \mathcal{R} . If $s \rightarrow_{\mathcal{R}}^+ t$ and $p \in \text{Pos}(s)$ such that $s|_p = c$ then $s[u]_p \rightarrow_{\mathcal{R}}^+ t$ using the same rewrite rules at the same positions, for all terms u . \square*

Table 1: Additional constants required to reduce a property P to ground P .

property	left-linear	right-ground TRSs	linear	variable-separated TRSs	
CE	<u>1</u>	(1)	<u>1</u>	(1)	(Theorem 4)
NE	<u>1</u>	(1)	<u>2</u>	(1)	(Theorems 5, 7, 10)
COM	1^\dagger	(0)	2^\dagger	(2)	(Theorem 8)
CR	$1^{*\dagger}$	(0)*	2^\dagger	(2)	
SCR	1^*	(0)*	<u>2</u>	(2)	
WCR	1^*	(0)*	<u>2</u>	(2)	
UNR	$1^{*\dagger}$	(0)*	2^\dagger	(2)	
UNC	$2^{*\dagger}$	(0)*	2^\dagger	(2)	
NFP	1^*	(0)*	<u>2</u>	(2)	

As variable-separated TRSs are closed under inverse we can immediately deduce that rewrite sequences of the shape $s\sigma_c \rightarrow_{\mathcal{R}}^+ t\sigma_c$ imply $s \rightarrow_{\mathcal{R}}^+ t$ for monadic systems. With this we are ready to prove our claim.

Theorem 10. *Variable-separated TRSs \mathcal{R} and \mathcal{S} over a common monadic signature \mathcal{F} are normalization equivalent if and only if \mathcal{R} and \mathcal{S} are ground normalization equivalent over $\mathcal{F} \uplus \{c\}$.*

Proof. Note that TRSs over a monadic signature are necessarily linear. We mention the differences with the proof of Theorem 5. A single constant suffices to prove $\text{NF}_{\mathcal{R}} = \text{NF}_{\mathcal{S}}$. Consider a rewrite sequence $s \rightarrow_{\mathcal{R}}^{*\epsilon*} t$ with $t \in \text{NF}_{\mathcal{R}}$. Ground normalization equivalence and substitution closure yields $s\sigma_c \rightarrow_{\mathcal{S}}^+ t\sigma_c$. Furthermore, since the sequence $s \rightarrow_{\mathcal{R}}^{*\epsilon*} t$ is non-empty by definition, $s\sigma_c \notin \text{NF}_{\mathcal{R}} = \text{NF}_{\mathcal{S}}$ and thus $s\sigma_c \neq t\sigma_c$ as $t\sigma_c \in \text{NF}_{\mathcal{S}}$. Hence $s\sigma_c \rightarrow_{\mathcal{S}}^+ t\sigma_c$. Applying Lemma 9 twice allows us to replace all occurrences of c in $s\sigma_c$ and $t\sigma_c$ by the corresponding variables, resulting in $s \rightarrow_{\mathcal{S}}^* t$. \square

In Table 1 we summarize the results of this paper as well as the related results (the final six rows) from [2]. The numbers for TRSs over monadic signatures are given in parentheses. The underlined numbers are new results. The results marked with an asterisk are proved in [5], those marked with a dagger are formalized in [2].

4 Conclusion

In this paper we presented new signature extension results allowing us to reduce the problem of proving CE and NE to GCE and GNE respectively for linear variable-separated TRSs (Theorems 4 and 5). This is done by adding fresh constants to the signature. We also showed that the number of required fresh constants for reducing NE to GNE can be reduced for left-linear right-ground TRSs as well as for monadic systems (Theorem 10). The latter was also shown for the property COM (Theorem 8). All results are formalized in Isabelle/HOL [1] and implemented in the tools FORT-h, FORT-s, and the certifier FORTify. Binaries of the tools can be obtained from

[https://fortissimo.uibk.ac.at/fort\(ify\)/](https://fortissimo.uibk.ac.at/fort(ify)/)

The implemented results enable FORT-s to find an equivalent complete TRS of our leading example using the formula

"[0](WCR & SN) & {+1} forall s, t ([0] s <->* t <=> [1] s <->* t)"

The $\{+1\}$ instructs the decision procedure to add one fresh constant to the signature when evaluating the subformula for CE. Calling FORT-s with this formula on our leading example \mathcal{R} produces the TRS:

$$a \rightarrow b \qquad f(b) \rightarrow g(a, a) \qquad g(a, x) \rightarrow a$$

which is indeed complete and equivalent to \mathcal{R} on all terms (not just ground terms). For ease of use we also added the shorthands CE and NE to the formula language. When using these the tools FORT-h, FORT-s and FORTify add the appropriate amount of constants for any given input TRS.

References

- [1] Alexander Lochmann. Reducing rewrite properties to properties on ground terms. *Archive of Formal Proofs*, 2022. https://isa-afp.org/entries/Rewrite_Properties_Reduction.html, Formal proof development.
- [2] Alexander Lochmann, Fabian Mitterwallner, and Aart Middeldorp. Formalized signature extension results for confluence, commutation and unique normal forms. In *Proc. 10th International Workshop on Confluence*, pages 25–30, 2021.
- [3] Fabian Mitterwallner, Alexander Lochmann, Aart Middeldorp, and Bertram Felgenhauer. Certifying proofs in the first-order theory of rewriting. In *Proc. 27th International Conference on Tools and Algorithms for the Construction and Analysis of Systems*, volume 12652 of *Lecture Notes in Computer Science*, pages 127–144, 2021. doi:10.1007/978-3-030-72013-1_7.
- [4] Franziska Rapp and Aart Middeldorp. Automating the first-order theory of left-linear right-ground term rewrite systems. In *Proc. 1st FSCD*, volume 52 of *Leibniz International Proceedings in Informatics*, pages 36:1–36:12, 2016. doi:10.4230/LIPIcs.FSCD.2016.36.
- [5] Franziska Rapp and Aart Middeldorp. Confluence properties on open terms in the first-order theory of rewriting. In *Proc. 5th International Workshop on Confluence*, pages 26–30, 2016.
- [6] Franziska Rapp and Aart Middeldorp. FORT 2.0. In *Proc. 9th International Joint Conference on Automated Reasoning*, volume 10900 of *Lecture Notes in Artificial Intelligence*, pages 81–88, 2018. doi:10.1007/978-3-319-94205-6_6.