Bottom-up rewriting is inverse recognizability preserving

Irène Durand and Géraud Sénizergues

LaBRI, Université Bordeaux 1
351 Cours de la libération, 33405 Talence cedex, France

Abstract. For the whole class of linear term rewriting systems, we define *bottom-up rewriting* which is a restriction of the usual notion of rewriting. We show that bottom-up rewriting effectively inverse-preserves recognizability and analyze the complexity of the underlying construction. The *Bottom-Up* class (BU) is, by definition, the set of linear systems for which every derivation can be replaced by a bottom-up derivation. Membership to BU turns out to be undecidable; we are thus lead to define more restricted classes: the classes $\mathsf{SBU}(k), k \in \mathbb{N}$ of *Strongly Bottom-Up*(k) systems for which we show that membership is decidable. We define the class of *Strongly Bottom-Up* systems by $\mathsf{SBU} = \bigcup_{k\in\mathbb{N}} \mathsf{SBU}(k)$. We give a polynomial sufficient condition for a system to be in SBU. The class SBU contains (strictly) several classes of systems which were already known to inverse preserve recognizability.

1 Introduction

An important concept in term rewriting is the notion of *preservation of recognizability* through rewriting. Each identification of a more general class of systems preserving recognizability, yields almost directly a new decidable call-by-need class [9], decidability results for confluence, accessibility, joinability. Also, recently, this notion has been used to prove termination of systems for which none of the already known termination techniques work [13]. Such a preservation property is also a tool for studying the recognizable/rational subsets of various monoids which are defined by a presentation $\langle X, \mathcal{R}\rangle$, where X is a finite alphabet and $\mathcal{R}$ a Thue system (see e.g. [17, 18]).

Many such known classes have been defined by imposing *syntactical* restrictions on the rewrite rules. For instance, in *growing* systems [14, 19] variables at depth strictly greater than 1 in the left-handside of a rule cannot appear in the corresponding right-handside. Finite-path Overlapping systems [25] are also defined by syntactic restrictions on the system. Previous works on semi-Thue systems proving a recognizability preservation property, were based on syntactic restrictions (see [2],[3], [1],[21],[23]).

Other works establish that some *strategies* i.e. restrictions on the derivations rather than on the rules, ensure preservation of recognizability. Various such strategies were studied in [11], [20],[24].

We rather follow here this second approach: we define a new rewriting strategy which we call *bottom-up rewriting* for linear term rewriting systems. The bottom-up derivations are, intuitively, those derivations in which the rules are applied from the bottom of the term towards the top (this set of derivations contains strictly the bottom-up derivations of [20] and the one-pass leaf-started derivations of [11]). An important feature of this strategy, as opposed to the ones quoted above, is that it allows *overlaps* between successive applications of rules.

We define *bottom-up*(k) derivations for $k \in \mathbb{N}$ ($\mathsf{bu}(k)$ derivations for short) where a certain amount (limited by k) of top-down sequences of rules is allowed. Our main result is that bottom-up rewriting inverse-preserves recognizability (Theorem 2). We use a simulation argument which reduces our statement to the preservation of recognizability by ground systems (this preservation property is shown in [7]). Our proof is constructive *i.e* gives an algorithm for computing an automaton recognizing the antecedents of a recognizable set of terms. We sketch an estimation of the complexity of the algorithm for $k = 1$.

We then define the class of Bottom-up systems (BU for short) consisting of the linear systems for which, there exists some fixed $k \geq 0$, such that every derivation between two terms can be replaced by a derivation which is $\mathsf{bu}(k)$. We show that membership to $\mathsf{BU}(k)$ is undecidable for $k \geq 1$ even for semi-Thue systems. We thus define the restricted class of *strongly bottom-up* systems for which we show decidable membership. We finally give a polynomial sufficient condition for a system to be in SBU.

Note that the class SBU is rather large since it contains strictly all the classes of semi-Thue systems quoted above (once translated as term rewriting systems where all symbols have arity 0 or 1), the linear growing systems of [14] and the class of LFPO^{-1} (i.e. the linear systems which belong to the class FPO^{-1} defined in [25]).

2 Preliminaries

Words We denote by A^* the set of finite words over the alphabet A. The *empty* word is denoted by ε. A word u is a prefix of a word v iff there exists some $w \in A^*$ such that $v = uw$. We denote by $u \preceq v$ the fact that u is a prefix of v. We then note $v \backslash u := w$.

Terms We assume the reader familiar with terms. We call *signature* a set of symbols with fixed arity. For every $m \in \mathbb{N}$, $\mathcal{F}_m$ denotes the subset of symbols of arity m. As usual, a *tree-domain* is a subset of $\mathbb{N}^*$, which is downwards closed for prefix ordering and left-brother ordering. Let us call $P' \subseteq P$ a *subdomain* of P iff, P' is a domain and, for every $u \in P, i \in \mathbb{N}$ $(u \cdot i \in P'$ & $u \cdot (i+1) \in P) \Rightarrow u \cdot (i+1) \in P'$. A *term* on a signature $\mathcal{F}$ is a partial map $t : \mathbb{N}^* \to \mathcal{F}$ whose domain is a tree-domain and which respects the arities. The domain of t is also called its set of *positions* and denoted by $\mathcal{P}\mathsf{os}(t)$. We write $\mathcal{P}\mathsf{os}^+(t)$ for $\mathcal{P}\mathsf{os}(t) \setminus \{\varepsilon\}$. If $u, v \in \mathcal{P}\mathsf{os}(t)$ and $u \preceq v$, we say that u is an *ancestor* of v in t. Given $v \in \mathcal{P}\mathsf{os}^+(t)$, its *father* is the position u such that $v = uw$ and $|w| = 1$. Given a term t and $u \in \mathcal{P}\mathsf{os}(t)$ the *subterm of* t *at* u is denoted by t/u and defined by $\mathcal{P}\mathsf{os}(t/u) = \{w \mid uw \in \mathcal{P}\mathsf{os}(t)\}$ and $\forall w \in \mathcal{P}\mathsf{os}(t/u)$, $t/u(w) = t(uw)$. We denote by $\mathcal{T}(\mathcal{F}, \mathcal{V})$ the set of first-order terms built upon a signature $\mathcal{F}$ and a denumerable set of variables $\mathcal{V}$. The set of variables of a term t is denoted by $\mathcal{V}\mathsf{ar}(t)$. The set of variable positions (resp. non variable positions) of a term t is denoted by $\mathcal{P}\mathsf{os}_{\mathcal{V}}(t)$ (resp. $\mathcal{P}\mathsf{os}_{\overline{\mathcal{V}}}(t)$). The *depth* of a term t is defined by: $dpt(t) := \sup\{|u| \mid u \in \mathcal{P}\mathsf{os}_{\overline{\mathcal{V}}}(t)\}$. We denote by $|t| := \mathsf{Card}(\mathcal{P}\mathsf{os}(t))$ the *size* of a term t. A term which does not contain twice the same variable is called *linear*. Given a linear term $t \in \mathcal{T}(\mathcal{F}, \mathcal{V})$, $x \in \mathcal{V}\mathsf{ar}(t)$, we shall denote by $\mathsf{pos}(t, x)$ the position of x in t. A term containing no variable is called *ground*. The set of ground terms is abbreviated to $\mathcal{T}(\mathcal{F})$ or $\mathcal{T}$ whenever $\mathcal{F}$ is understood. Among all the variables, there is a special one designated by $\square$. A term containing exactly one occurrence of $\square$ is called a *context*. A context is usually denoted as $C[]$. If u is the position of $\square$ in $C[]$, $C[t]$ denotes the term $C[]$ where t has been substituted at position u.

Term rewriting A *rewrite rule* is a pair $l \to r$ of terms in $\mathcal{T}(\mathcal{F}, \mathcal{V})$ which satisfies $\mathcal{V}\mathsf{ar}(r) \subseteq \mathcal{V}\mathsf{ar}(l)$. We call l (resp. r) the *left-handside* (resp. *right-handside*) of the rule (*lhs* and *rhs* for short). A rule is *linear* if both its left and right-handside are linear. A rule is *left-linear* if its left-handside is linear.

A *term rewriting system* (*system* for short) is a pair $(\mathcal{R}, \mathcal{F})$ where $\mathcal{F}$ is a signature and $\mathcal{R}$ a set of rewrite rules built upon the signature $\mathcal{F}$. When $\mathcal{F}$ is clear from the context or contains exactly the symbols in $\mathcal{R}$, we may omit $\mathcal{F}$ and write simply $\mathcal{R}$. We call *size* of the set of rules $\mathcal{R}$ the number $\|\mathcal{R}\| := \sum_{l \to r \in \mathcal{R}} |l| + |r|$. Rewriting is defined as usual. A system is *linear* (resp. *left-linear*) if each of its rules is linear (resp. left-linear). A system $\mathcal{R}$ is *growing* [14] if every variable of a right-handside is at depth at most 1 in the corresponding left-handside.

Automata We shall consider exclusively bottom-up term (tree) automata [4] (which we abbreviate to $f.t.a$). An automaton $\mathcal{A}$ is given by a 4-tuple $(\mathcal{F}, Q, Q_f, \Gamma)$ where $\mathcal{F}$ is the signature, Q the set of states, Q_f the set of final states, Γ the set of transitions. The *size* of $\mathcal{A}$ is defined by: $\|\mathcal{A}\| := \mathsf{Card}(\Gamma) + \mathsf{Card}(Q)$. The set of rules Γ can be viewed as a rewriting system over the signature $\mathcal{F} \cup Q$. We then denote by $\to_\Gamma$ or by $\to_\mathcal{A}$ (resp. by $\to^*_\Gamma$ or by $\to^*_\mathcal{A}$) the one-step rewriting relation (resp. the rewriting relation) generated by Γ. Given an automaton $\mathcal{A}$, let $L(\mathcal{A})$ be the set of terms accepted by $\mathcal{A}$. A set of terms T is *recognizable* if there exists a term automaton $\mathcal{A}$ such that $T = L(\mathcal{A})$. The following technical normal form for $f.t.a$ will be useful in our proofs.

Definition 1. *A n.f.t.a.* $\mathcal{A} = (\mathcal{F}, Q, Q_f, \Gamma)$ *is called* standard *iff* $\mathcal{F}_0 \subseteq Q$ *and*
1- Every rule of $\mathcal{A}$ *has the form* $f(q_1, \ldots, q_m) \to q$ *with* $m \geq 1, f \in \mathcal{F}_m$, $q_i, q \in Q$
2- For every $m \geq 1, f \in \mathcal{F}_m$, $q_1, \ldots, q_m \in Q$ *there exists a unique* $q \in Q$ *such that* $f(q_1, \ldots, q_m) \to_\mathcal{A} q$.

Automata and rewriting Given a system $\mathcal{R}$ and a set of terms T, we define
$(\to^*_\mathcal{R})[T] = \{s \in \mathcal{T}(\mathcal{F}) \mid \exists t \in T, s \to^*_\mathcal{R} t\}$ and
$[T](\to^*_\mathcal{R}) = \{s \in \mathcal{T}(\mathcal{F}) \mid \exists t \in T, t \to^*_\mathcal{R} s\}$.
A system $\mathcal{R}$ is *recognizability preserving* (resp. *inverse recognizability preserving*) if $[T](\to^*_\mathcal{R})$ (resp. $(\to^*_\mathcal{R})[T]$) is recognizable for every recognizable T.

Lemma 1. *Let* $\mathcal{A}$ *be a standard* $f.t.a$ *over the signature* $\mathcal{F}$. *Let* $t, t_1, t_2 \in \mathcal{T}(\mathcal{F} \cup Q)$. *If* $t \to^*_\mathcal{A} t_1, t \to^*_\mathcal{A} t_2$ *and* $\mathcal{P}\mathsf{os}(t_1) = \mathcal{P}\mathsf{os}(t_2)$, *then* $t_1 = t_2$.

Definition 2 ($\mathcal{A}$-reduct). *Let* $\mathcal{A}$ *be a standard* $f.t.a$ *over the signature* $\mathcal{F}$. *Let* $t \in \mathcal{T}(\mathcal{F} \cup Q)$ *and let* P *be a subdomain of* $\mathcal{P}\mathsf{os}(t)$. *We define* $\mathsf{Red}(t, P) = t'$ *as the unique element of* $\mathcal{T}(\mathcal{F} \cup Q)$ *such that*

1- $t \rightarrow_{\mathcal{A}}^{*} t'$
2 -$\mathcal{P}\mathsf{os}(t') = P$

Lemma 2. *Let* $\mathcal{A}$ *be a standard* $f.t.a$ *over the signature* $\mathcal{F}$*. Let* $t, t_1, t_2 \in \mathcal{T}(\mathcal{F} \cup Q)$*. If* $t \rightarrow_{\mathcal{A}}^{*} t_1, t \rightarrow_{\mathcal{A}}^{*} t_2$ *and* $\mathcal{P}\mathsf{os}(t_1) \subseteq \mathcal{P}\mathsf{os}(t_2)$*, then* $t_2 \rightarrow_{\mathcal{A}}^{*} t_1$*.*

3 Bottom-up rewriting

In order to define *bottom-up rewriting*, we need some marking tools. In the following we assume that $\mathcal{F}$ is a signature. We shall illustrate many of our definitions with the following system

Example 1. $\mathcal{R}_1 = \{\mathsf{f}(x) \rightarrow \mathsf{g}(x), \mathsf{g}(\mathsf{h}(x)) \rightarrow \mathsf{i}(x), \mathsf{i}(x) \rightarrow \mathsf{a}\}$

3.1 Marking

We mark the symbols of a term using natural integers (as done in [12] for example).

Definition 3. *We define the (infinite)* signature of marked symbols*:* $\mathcal{F}^{\mathbb{N}} := \{f^i \mid f \in \mathcal{F}, i \in \mathbb{N}\}$*.*
For every integer $k \geq 0$ *we note:* $\mathcal{F}^{\leq k} := \{f^i \mid f \in \mathcal{F}, 0 \leq i \leq k\}$*. The operation* $\mathsf{m}()$ *returns the mark of a marked symbol:* $\mathsf{m}(f^i) = i$*.*

Definition 4. *The terms in* $\mathcal{T}(\mathcal{F}^{\mathbb{N}}, \mathcal{V})$ *are called* marked *terms.*

The operation $\mathsf{m}()$ extends to marked terms: if $t \in \mathcal{V}, \mathsf{m}(t) = 0$, otherwise, $\mathsf{m}(t) = \mathsf{m}(t(\varepsilon))$. For every $f \in \mathcal{F}$, we identify f^0 and f; it follows that $\mathcal{F} \subset \mathcal{F}^{\mathbb{N}}$, $\mathcal{T}(\mathcal{F}) \subset \mathcal{T}(\mathcal{F}^{\mathbb{N}})$ and $\mathcal{T}(\mathcal{F}, \mathcal{V}) \subset \mathcal{T}(\mathcal{F}^{\mathbb{N}}, \mathcal{V})$.

Example 2. $\mathsf{m}(\mathsf{a}^2) = 2, \mathsf{m}(\mathsf{i}(\mathsf{a}^2)) = 0, \mathsf{m}(\mathsf{h}^1(\mathsf{a})) = 1, \mathsf{m}(\mathsf{h}^1(x)) = 1, \mathsf{m}(x) = 0.$

Definition 5. *Given* $t \in \mathcal{T}(\mathcal{F}^{\mathbb{N}}, \mathcal{V})$ *and* $i \in \mathbb{N}$*, we define the marked term* t^i *whose marks are all equal to* i *(except at variables) by:*

$$\mathcal{P}\mathsf{os}(t^i) := \mathcal{P}\mathsf{os}(t), \;\; t^i(u) = t(u)^i \textit{ if } t(u) \notin \mathcal{V}, \;\; t^i(u) = t(u) \textit{ if } t(u) \in \mathcal{V}.$$

This marking extends to sets of terms S ($S^i := \{t^i \mid t \in S\}$) and substitutions σ ($\sigma^i : x \mapsto (x\sigma)^i$). We use the notation: $\mathsf{mmax}(t) := \max\{\mathsf{m}(t/u) \mid u \in \mathcal{P}\mathsf{os}(t)\}$. In the sequel, given a term $t \in \mathcal{T}(\mathcal{F}, \mathcal{V})$, $\overline{t}$ will always refer to a term of $\mathcal{T}(\mathcal{F}^{\mathbb{N}}, \mathcal{V})$ such that $\overline{t}^0 = t$. The same rule will apply to substitutions and contexts. Note that there are several possible $\overline{t}$ for a single t.

Finite automata and marked terms. Given a finite tree-automaton $\mathcal{A} = (\mathcal{F}, Q, Q_f, \Gamma)$ we extend it over the signature $\mathcal{F}^{\mathbb{N}}$, by setting

$$\Gamma^{\mathbb{N}} := \{(f^j(q_1^{j_1}, \ldots, q_m^{j_m}) \rightarrow q^j) \mid (f(q_1, \ldots, q_m \rightarrow q) \in \Gamma, j, j_1, \ldots, j_m \in \mathbb{N}\},$$

and $\mathcal{A}^{\mathbb{N}} := (\mathcal{F}^{\mathbb{N}}, Q^{\mathbb{N}}, Q_f^{\mathbb{N}}, \Gamma^{\mathbb{N}})$. The binary relation $\rightarrow_{\mathcal{A}^{\mathbb{N}}}$ is an extension of $\rightarrow_{\mathcal{A}}$ to $\mathcal{T}(\mathcal{F}^{\mathbb{N}}) \times \mathcal{T}(\mathcal{F}^{\mathbb{N}})$. We will often note simply $\rightarrow_{\mathcal{A}}$ what should be denoted $\rightarrow_{\mathcal{A}^{\mathbb{N}}}$.

$\mathbb{N}$ *acts on marked terms.* We define a right-action $\odot$ of the monoid $(\mathbb{N}, \max, 0)$ over the set $\mathcal{F}^{\mathbb{N}}$ which just consists in applying the operation $\max$ on every mark *i.e* for every $\bar{t} \in \mathcal{T}(\mathcal{F}^{\mathbb{N}}), n \in \mathbb{N}$,

$$\mathcal{P}\mathsf{os}(\bar{t}\odot n) := \mathcal{P}\mathsf{os}(\bar{t}), \;\; \forall u \in \mathcal{P}\mathsf{os}(\bar{t}), \mathsf{m}((\bar{t}\odot n)/u) := \max(\mathsf{m}(\bar{t}/u), n), \;\; (\bar{t}\odot n)^0 = \bar{t}^0$$

Lemma 3. *Let $\mathcal{A}$ be some finite tree automaton over $\mathcal{F}$, $\bar{s}, \bar{t} \in \mathcal{T}((\mathcal{F} \cup Q)^{\mathbb{N}})$ and $n \in \mathbb{N}$. If $\bar{s} \to_{\mathcal{A}}^* \bar{t}$ then $(\bar{s} \odot n) \to_{\mathcal{A}}^* (\bar{t} \odot n)$.*

Marked rewriting We define here the rewrite relation $\circ\!\!\to$ between marked terms. Figure 1 illustrates this definition.

For every linear marked term $\bar{t} \in \mathcal{T}(\mathcal{F}^{\mathbb{N}}, \mathcal{V})$ and variable $x \in \mathcal{V}\mathsf{ar}(\bar{t})$, we define:

$$M(\bar{t}, x) := \sup\{\mathsf{m}(\bar{t}/w) \mid w < \mathsf{pos}(\bar{t}, x)\} + 1. \tag{1}$$

Let $\mathcal{R}$ be a left-linear system, $\overline{s} \in \mathcal{T}(\mathcal{F}^{\mathbb{N}})$. Let us suppose that $\overline{s} \in \mathcal{T}(\mathcal{F}^{\mathbb{N}})$ decomposes as

$$\overline{s} = \overline{C}[\bar{l}\overline{\sigma}]_v, \quad \text{with} \quad (l, r) \in \mathcal{R}, \tag{2}$$

for some marked context $\overline{C}[]_v$ and substitution $\overline{\sigma}$. We define a new marked substitution $\overline{\overline{\sigma}}$ (such that $\overline{\overline{\sigma}}^0 = \overline{\sigma}^0$) by: for every $x \in \mathcal{V}\mathsf{ar}(r)$,

$$x\overline{\overline{\sigma}} := (x\overline{\sigma}) \odot M(\overline{C}[\bar{l}], x). \tag{3}$$

We then write $\overline{s} \circ\!\!\to \bar{t}$ where

$$\overline{s} = \overline{C}[\bar{l}\overline{\sigma}], \;\; \bar{t} = \overline{C}[r\overline{\overline{\sigma}}]. \tag{4}$$

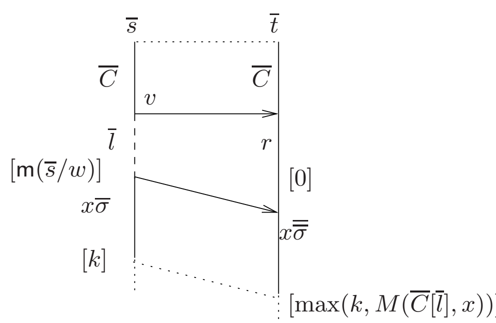


Fig. 1. A marked rewriting step

More precisely, an ordered pair of marked terms $(\overline{s}, \bar{t})$ is linked by the relation $\circ\!\!\to$ iff, there exists $\overline{C}[]_v, (l, r), \bar{l}, \overline{\sigma}$ and $\overline{\overline{\sigma}}$ fulfilling equations (2-4). A mark k will roughly mean that there were k successive applications of rules, each one with a leaf of the left-handside at a position strictly greater than a leaf of the previous right-handside.

The map $\overline{s} \mapsto \overline{s}^0$ (from marked terms to unmarked terms) extends into a map from marked derivations to unmarked derivations: every

$$\overline{s}_0 = \overline{C}_0[\overline{l}_0\overline{\sigma}_0]_{v_0} \circ\!\!\longrightarrow \overline{C}_0[r_0\overline{\overline{\sigma}}_0]_{v_0} = \overline{s}_1 \circ\!\!\longrightarrow \ldots \circ\!\!\longrightarrow \overline{C}_{n-1}[r_{n-1}\overline{\overline{\sigma}}_{n-1}]_{v_{n-1}} = \overline{s}_n \tag{5}$$

is mapped to the derivation

$$s_0 = C_0[l_0\sigma_0]_{v_0} \to C_0[r_0\sigma_0]_{v_0} = s_1 \to \ldots \to C_{n-1}[r_{n-1}\sigma_{n-1}]_{v_{n-1}} = s_n. \tag{6}$$

The context $\overline{C}_i$, the rule (l_i, r_i) and the substitution $\overline{\sigma}_i$ completely determines $\overline{s}_{i+1}$. Thus, for every fixed pair $(\overline{s_0}, s_0)$, this map is a bijection from the set of derivations (5) starting from $\overline{s_0}$, to the set of derivations (6) starting from s_0.

Example 3. With the system $\mathcal{R}_1$ of Example 1 we get the following marked derivation:

$$\mathsf{f}(\mathsf{h}(\mathsf{f}(\mathsf{h}(\mathsf{a})))) \circ\!\!\longrightarrow \mathsf{f}(\mathsf{h}(\mathsf{g}(\mathsf{h}^1(\mathsf{a}^1)))) \circ\!\!\longrightarrow \mathsf{f}(\mathsf{h}(\mathsf{i}(\mathsf{a}^2))) \circ\!\!\longrightarrow \mathsf{f}(\mathsf{h}(\mathsf{a}) \circ\!\!\longrightarrow$$
$$\mathsf{g}(\mathsf{h}^1(\mathsf{a}^1)) \circ\!\!\longrightarrow \mathsf{i}(\mathsf{a}^2) \circ\!\!\longrightarrow \mathsf{a}$$

From now on, each time we deal with a derivation $s \to^* t$ between two terms $s, t \in \mathcal{T}(\mathcal{F}, \mathcal{V})$, we may implicitly decompose it as (6) where n is the length of the derivation, $s = s_0$ and $t = s_n$.

3.2 Bottom-up derivations

Definition 6. *The marked derivation (5) is* weakly bottom-up *if, for every* $0 \le j < n$, $l_j \notin \mathcal{V} \Rightarrow \mathsf{m}(\overline{l_j}) = 0$, *and* $l_j \in \mathcal{V} \Rightarrow \sup\{\mathsf{m}(\overline{s_j}/u) \mid u < v_j\} = 0$.

Definition 7. *The derivation (6) is* weakly bottom-up *if the corresponding marked derivation (5) starting on the same term* $\overline{s} = s$ *is* weakly bottom-up *(following the above definition).*

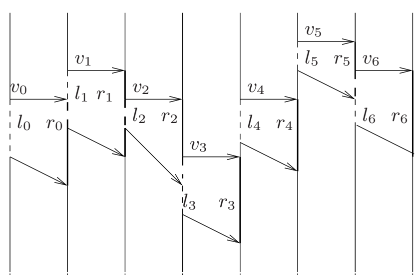


Fig. 2. A wbu-derivation

We shall abbreviate "weakly bottom-up" to wbu.

Lemma 4. *Let* $\mathcal{R}$ *be a linear system. If* $s \to^*_{\mathcal{R}} t$ *then there exists a* wbu-*derivation between* s *and* t.

The linear restriction cannot be relaxed: let $\mathcal{R} = \{\mathsf{f}(x) \to \mathsf{g}(x,x), \mathsf{a} \to \mathsf{b}\}$; we have $\mathsf{f}(\mathsf{a}) \to \mathsf{g}(\mathsf{a},\mathsf{a}) \to \mathsf{g}(\mathsf{b},\mathsf{a})$ but no wbu-derivation between $\mathsf{f}(\mathsf{a})$ and $\mathsf{g}(\mathsf{b},\mathsf{a})$.

Definition 8. *A marked term $\overline{s}$ is called* m-increasing *iff, for every $u, v \in \mathcal{P}\mathsf{os}(\overline{s}), u \preceq v \Rightarrow \mathsf{m}(\overline{s}/u) \leq \mathsf{m}(\overline{s}/v)$.*

Intuitively, in an m-*increasing* term, deeper positions carry the same or higher marks.

Lemma 5. *Let $\overline{s} \circ\!\!\to^* \overline{t}$ be a marked* wbu*-derivation between $\overline{s}, \overline{t} \in \mathcal{T}(\mathcal{F}^{\mathbb{N}})$. If $\overline{s}$ is* m-*increasing, then $\overline{t}$ is* m-*increasing too.*

We classify the derivations according to the maximal value of the marks. We abbreviate "bottom-up" as bu.

Definition 9. *A derivation is* $\mathsf{bu}(k)$ *(resp.* $\mathsf{bu}^-(k)$*) if it is* wbu *and, in the corresponding marked derivation $\forall i, 0 \leq i \leq n$, $\mathsf{mmax}(\overline{s_i}) \leq k$ (resp. $\forall i, 0 \leq i < n$, $\mathsf{mmax}(\overline{l_i}) < k$).*

Notation $\overline{s} \;{}_k\!\circ\!\!\to^*_{\mathcal{R}} \overline{t}$ means that there exists a wbu marked derivation from s to t where all the marks belong to $[0, k]$.
Notation $s \;{}_k\!\!\to^*_{\mathcal{R}} t$ means that there exists a $\mathsf{bu}(k)$ derivation from s to t.

Example 4. For the system $\mathcal{R}_0 = \{\mathsf{f}(\mathsf{f}(x)) \to \mathsf{f}(x)\}$ with the signature $\mathcal{F} = \{\mathsf{a}^{(0)}, \mathsf{f}^{(1)}\}$, although we may get a $\mathsf{bu}(k)$ derivation for a term of the form $\mathsf{f}(\ldots\mathsf{f}(\mathsf{a})\ldots)$ with $k+1$ f symbols: $\mathsf{f}(\mathsf{f}(\mathsf{f}(\mathsf{f}(\mathsf{a})))) \circ\!\!\to \mathsf{f}(\mathsf{f}^1(\mathsf{f}^1(\mathsf{a}^1))) \circ\!\!\to \mathsf{f}(\mathsf{f}^2(\mathsf{a}^2)) \circ\!\!\to \mathsf{f}(\mathsf{a}^3)$
we can always achieve a $\mathsf{bu}^-(1)$ derivation:
$\mathsf{f}(\mathsf{f}(\mathsf{f}(\mathsf{f}(\mathsf{a})))) \circ\!\!\to \mathsf{f}(\mathsf{f}(\mathsf{f}(\mathsf{a}^1))) \circ\!\!\to \mathsf{f}(\mathsf{f}(\mathsf{a}^1)) \circ\!\!\to \mathsf{f}(\mathsf{a}^1)$

3.3 Bottom-up systems

We introduce here a hierarchy of classes of rewriting systems, based on their ability to meet the bottom-up restriction over derivations.

Definition 10. *Let P be some property of derivations. A system $(\mathcal{R}, \mathcal{F})$ is called P if $\forall s, t \in \mathcal{T}(\mathcal{F})$ such that $s \to^*_{\mathcal{R}} t$ there exists a P-derivation from s to t.*

We denote by $\mathsf{BU}(k)$ the class of $\mathsf{BU}(k)$ systems, by $\mathsf{BU}^-(k)$ the class of $\mathsf{BU}^-(k)$ systems. One can check that, for every $k > 0, \mathsf{BU}(k-1) \subsetneq \mathsf{BU}^-(k) \subsetneq \mathsf{BU}(k)$. Finally, the class of *bottom-up systems*, denoted BU, is defined by: $\mathsf{BU} = \bigcup_{k \in \mathbb{N}} \mathsf{BU}(k)$.

Example 5. The system $\mathcal{R}_0 = \{\mathsf{f}(\mathsf{f}(x)) \to \mathsf{f}(x)\} \in \mathsf{BU}^-(1)$ and $\mathcal{R}_0$ is not growing.
The system $\mathcal{R}_1$ of Example 1 belongs to $\mathsf{BU}^-(2)$ and $\mathcal{R}_1$ is not growing.
The system $\mathcal{R}_2 = \{\mathsf{f}(x) \to \mathsf{g}(x), \mathsf{h}(\mathsf{g}(\mathsf{a})) \to \mathsf{a}\}$ is growing and belongs to $\mathsf{BU}^-(1)$.
The system $\mathcal{R}_3 = \{\mathsf{f}(x) \to \mathsf{g}(x), \mathsf{g}(\mathsf{h}(x)) \to \mathsf{a}\}$ is growing and belongs to $\mathsf{BU}(1)$.

4 Bottom-up rewriting is inverse recognizability preserving

Let us recall the following classical result about ground rewriting systems

Theorem 1 ([7, 6]). *Every ground system is inverse-recognizability preserving.*

The main theorem of this section (and of the paper) is the following extension of Theorem 1 to $\mathsf{bu}(k)$ derivations of linear rewriting systems

Theorem 2. *Let $\mathcal{R}$ be some linear rewriting system over the signature $\mathcal{F}$, let T be some recognizable subset of $\mathcal{T}(\mathcal{F})$ and let $k \geq 0$. Then, the set $({}_{k}\!\rightarrow^{*}_{\mathcal{R}})[T]$ is recognizable too.*

4.1 Construction

In order to prove Theorem 2 we are to introduce some technical definitions, and to prove some technical lemmas. Let us fix, from now on and until the end of the subsection, a linear system $(\mathcal{R}, \mathcal{F})$, a language $T \subseteq \mathcal{T}(\mathcal{F})$ recognized by a finite automaton over the extended signature $\mathcal{F} \cup \{\Box\}$, $\mathcal{A} = (\mathcal{F} \cup \{\Box\}, Q, Q_f, \Gamma)$ and an integer $k \geq 0$. In order to make the proofs easier, we assume in the first step of this subsection that:

$$\forall l \to r \in \mathcal{R}, l \notin \mathcal{V}, \text{ and } \quad \mathcal{A} \text{ is standard.} \tag{7}$$

Let us define the integer $\mathsf{d} := \max\{dpt(l) \mid l \to r \in \mathcal{R}\}$. We introduce a ground system $\mathcal{S}$ which will be enough, together with $\mathcal{A}$, for describing the set of terms which rewrite in $L(\mathcal{A})$.

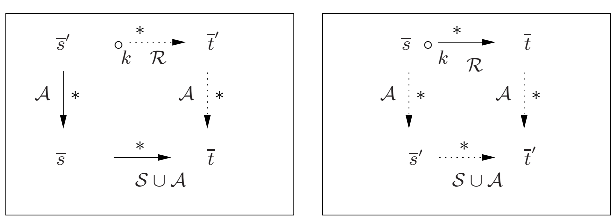
Definition 11. *We define $\mathcal{S}$ as the ground rewriting system over $\mathcal{T}((\mathcal{F} \cup Q)^{\leq k})$ consisting of all the rules of the form: $\bar{l}\overline{\tau} \to r\overline{\overline{\tau}}$ where $l \to r$ is a rule of $\mathcal{R}$*

$$\mathsf{m}(\bar{l}) = 0 \tag{8}$$

and $\overline{\tau}, \overline{\overline{\tau}} : \mathcal{V} \to \mathcal{T}((\mathcal{F} \cup Q)^{\leq k})$ are marked substitutions such that, $\forall x \in \mathcal{V}\mathsf{ar}(l)$

$$x\overline{\overline{\tau}} = (x\overline{\tau}) \odot M(\bar{l}, x), \quad dpt(x\overline{\tau}) \leq k \cdot \mathsf{d}. \tag{9}$$

Lemma 6 (lifting $\mathcal{S} \cup \mathcal{A}$ to $\mathcal{R}$).
Let $\overline{s}, \overline{s'}, \overline{t} \in \mathcal{T}((\mathcal{F} \cup Q)^{\leq k})$.
If $\overline{s}' \to^{*}_{\mathcal{A}} \overline{s}$ and $\overline{s} \to^{*}_{\mathcal{S} \cup \mathcal{A}} \overline{t}$ then, there exists a term $\bar{t}' \in \mathcal{T}((\mathcal{F} \cup Q)^{\leq k})$ such that $\overline{s}' \;{}_{k}\!\circ\!\!\to^{*}_{\mathcal{R}} \overline{t'}$ and $\overline{t'} \to^{*}_{\mathcal{A}} \overline{t}$.



Lifting $\mathcal{S} \cup \mathcal{A}$ Projecting $\mathcal{R}$

Fig. 3. Lemma 6 and 8

Proof. 1- Let us prove that the lemma holds for $\overline{s} \to_{\mathcal{S}\cup\mathcal{A}} \overline{t}$. Under the assumption that $\overline{s}' \to^*_{\mathcal{A}} \overline{s} \to_{\mathcal{A}} \overline{t}$, we can just choose $\overline{t}' = \overline{s}'$ and obtain the conclusion.
Suppose now that $\overline{s} \to_{\mathcal{S}} \overline{t}$. This means that

$$\overline{s} = \overline{C}[\overline{l}\overline{\tau}], \;\; \overline{t} = \overline{C}[r\overline{\overline{\tau}}]$$

for some rule $l \to r \in \mathcal{R}$, marked context $\overline{C}$, and marked substitutions $\overline{\tau}$, $\overline{\overline{\tau}}$, satisfying (8-9). Since $\overline{s}' \to^*_{\mathcal{A}} \overline{s}$, we must have $\overline{s}' = \overline{C}[\overline{l}\overline{\tau}']$ where, for every $x \in \mathcal{V}\mathsf{ar}(l)$, $x\overline{\tau}' \to^*_{\mathcal{A}} x\overline{\tau}$. Let us set

$$x\overline{\overline{\tau}}' := x\overline{\tau}' \odot M(\overline{l}, x), \;\; \overline{t}' := \overline{C}[r\overline{\overline{\tau}}'].$$

The relation $\overline{s}' \circ\!\!\longrightarrow_{\mathcal{R}} \overline{t}'$ holds (by definition of $\circ\!\!\longrightarrow$) and this one-step derivation is wbu (by condition (8)). Moreover, since every mark of $\overline{t}$ is $\leq k$ and every mark of $\overline{s}'$ is $\leq k$, we are sure that, for every $x \in \mathcal{V}\mathsf{ar}(l) \cap \mathcal{V}\mathsf{ar}(r)$, $M(\overline{l}, x) \leq k$ (because some non smaller mark occurs in $\overline{t}$) and $\mathsf{mmax}(x\overline{\tau}') \leq k$ (because this mark occurs in $\overline{s}'$), hence $\mathsf{mmax}(x\overline{\overline{\tau}}') \leq k$ which ensures that $\mathsf{mmax}(\overline{t}') \leq k$ and finally:

$$\overline{s}' \;{}_k\!\circ\!\!\longrightarrow_{\mathcal{R}} \overline{t}'.$$

By Lemma 3, for every $x \in \mathcal{V}\mathsf{ar}(l)$, $x\overline{\overline{\tau}}' = x\overline{\tau}' \odot M(\overline{l}, x) \to^*_{\mathcal{A}} x\overline{\tau} \odot M(\overline{l}, x) = x\overline{\overline{\tau}}$. Hence $\overline{t}' = \overline{C}[r\overline{\overline{\tau}}'] \to^*_{\mathcal{A}} \overline{C}[r\overline{\overline{\tau}}] = \overline{t}$.
2- The lemma can be deduced from point 1 above by induction on the integer n such that $\overline{s} \to^n_{\mathcal{S}\cup\mathcal{A}} \overline{t}$.

Definition 12 (Top domain of a term). *Let* $\overline{t} \in \mathcal{T}((\mathcal{F} \cup Q)^{\leq k}, \{\Box\})$*. We define the* top domain *of* $\overline{t}$*, denoted by* $\mathsf{Topd}(\overline{t})$ *as:* $u \in \mathsf{Topd}(\overline{t})$ *iff*
1- $u \in \mathcal{P}\mathsf{os}(\overline{t})$
2- $\forall u_1, u_2 \in \mathbb{N}^*$ *such that* $u = u_1{\cdot}u_2$*, either* $\mathsf{m}(\overline{t}/u_1) = 0$ *or* $|u_2| \leq (k{+}1{-}\mathsf{m}(\overline{t}/u_1))\mathsf{d}$.

We then define the *top* of a term $\overline{t}$, which is, intuitively, the only part of $\overline{t}$ which can be used in a marked derivation using marks not greater than k.

Definition 13 (Top of a term). *For every* $\overline{t} \in \mathcal{T}((\mathcal{F}\cup Q)^{\leq k}, \{\Box\})$, $\mathsf{Top}(\overline{t}) = \mathsf{Red}(\overline{t}, \mathsf{Topd}(\overline{t}))$.

Lemma 7 (projecting one step of $\mathcal{R}$ on $\mathcal{S} \cup \mathcal{A}$).
Let $\overline{s}, \overline{t} \in \mathcal{T}(\mathcal{F}^{\leq k})$ *such that:*
1- $\overline{s} \circ\!\!\longrightarrow_{\mathcal{R}} \overline{t}$*,*
2- The marked rule $(\overline{l}, r)$ *used in the above rewriting-step is such that* $\mathsf{m}(\overline{l}) = 0$*.*
3- $\overline{s}$ *is* m*-increasing.*
Then, $\mathsf{Top}(\overline{s}) \to^*_{\mathcal{A}}\to_{\mathcal{S}} \mathsf{Top}(\overline{t})$.

Proof. (Sketch) By the hypotheses of the lemma

$$\overline{s} = \overline{C}[\overline{l}\overline{\sigma}], \;\; \overline{t} = \overline{C}[r\overline{\overline{\sigma}}]$$

for some $\overline{C}, \overline{\sigma}, \overline{l}, r, \overline{\overline{\sigma}}$ fulfilling (2-4) of marked rewriting and $\mathsf{m}(\overline{l}) = 0$. Let us then define a context $\overline{D}$ and marked substitutions $\overline{\tau}, \overline{\overline{\tau}}$ by: for every $x \in \mathcal{V}$

$$\overline{D}[] = \mathsf{Top}(\overline{C}[]). \tag{10}$$

$$x\overline{\overline{\tau}} = \mathsf{Top}(x\overline{\overline{\sigma}}), \quad x\overline{\tau} = \mathsf{Red}(x\overline{\sigma}, \mathcal{P}\mathsf{os}(x\overline{\overline{\tau}})). \tag{11}$$

We claim that

$$\mathsf{Top}(\overline{s}) \to_{\mathcal{A}}^{*} \overline{D}[\overline{l}\overline{\overline{\tau}}] \to_{\mathcal{S}} \overline{D}[r\overline{\overline{\tau}}] = \mathsf{Top}(\overline{t}). \tag{12}$$

(This claim is carefully checked in [10] and makes use of lemma 2).

Lemma 8 (projecting $\mathcal{R}$ on $\mathcal{S} \cup \mathcal{A}$).
Let $\overline{s}, \overline{t} \in \mathcal{T}(\mathcal{F}^{\leq k})$ and assume that $\overline{s}$ is m*-increasing. If $\overline{s}\ {}_{k}\!\circ\!\!\to_{\mathcal{R}}^{*}\ \overline{t}$ then, there exist terms $\overline{s}', \overline{t}' \in \mathcal{T}((\mathcal{F} \cup Q)^{\leq k})$ such that $\overline{s} \to_{\mathcal{A}}^{*} \overline{s}'$, $\overline{s}' \to_{\mathcal{S} \cup \mathcal{A}}^{*} \overline{t'}$ and $\overline{t} \to_{\mathcal{A}}^{*} \overline{t}'$.*

Proof. The marked derivation $\overline{s} \circ\!\!\to_{\mathcal{R}}^{*} \overline{t}$ is wbu, hence it can be decomposed into n successive steps where the hypothesis 2 of Lemma 7 is valid. Hypothesis 3 of Lemma 7 will also hold, owing to our assumption and to Lemma 5. We can thus deduce, inductively, from the conclusion of Lemma 7, that $\mathsf{Top}(\overline{s}) \to_{\mathcal{S}\cup\mathcal{A}}^{*} \mathsf{Top}(\overline{t})$. The choice $\overline{s}' := \mathsf{Top}(\overline{s}), \overline{t}' := \mathsf{Top}(\overline{t})$ fulfills the conclusion of the lemma.

Lemma 9. *Let $s \in \mathcal{T}(\mathcal{F})$. Then $s\ {}_{k}\!\!\to_{\mathcal{R}}^{*} T$ iff $s \to_{\mathcal{S}\cup\mathcal{A}}^{*} Q_f^{\leq k}$.*

Proof. ($\Rightarrow$): Suppose $s\ {}_{k}\!\!\to_{\mathcal{R}}^{*} t$ and $t \in T$. Let us consider the corresponding marked derivation

$$\overline{s}\ {}_{k}\!\circ\!\!\to_{\mathcal{R}}^{*} \overline{t} \tag{13}$$

where $\overline{s} := s$. Derivation (13) is wbu and lies in $\mathcal{T}(\mathcal{F}^{\leq k})$. Let us consider the terms $\overline{s}', \overline{t}'$ given by Lemma 8:

$$\overline{s} \to_{\mathcal{A}}^{*} \overline{s}' \to_{\mathcal{S}\cup\mathcal{A}}^{*} \overline{t}' \tag{14}$$

and $\overline{t} \to_{\mathcal{A}}^{*} \overline{t}'$. Since $\overline{t} \to_{\mathcal{A}}^{*} Q_f^{\leq k}$, by Lemma 2,

$$\overline{t}' \to_{\mathcal{A}}^{*} Q_f^{\leq k}. \tag{15}$$

Combining (14) and (15) we obtain that $s \to_{\mathcal{S}\cup\mathcal{A}}^{*} Q_f^{\leq k}$.
($\Leftarrow$): Suppose $s \to_{\mathcal{S}\cup\mathcal{A}}^{*} q^j \in Q_f^{\leq k}$. The hypotheses of Lemma 6 are met by $\overline{s} := s, \overline{s}' := s$ and $\overline{t} := q^j$. By Lemma 6 there exists some $\overline{t}' \in \mathcal{T}((\mathcal{F} \cup Q)^{\leq k})$ such that $s\ {}_{k}\!\circ\!\!\to_{\mathcal{R}}^{*} \overline{t'} \to_{\mathcal{A}}^{*} q^j \in Q_f^{\leq k}$. These derivations are mapped (by removal of the marks) into: $s\ {}_{k}\!\!\to_{\mathcal{R}}^{*} t' \to_{\mathcal{A}}^{*} q \in Q_f$, which shows that $t' \in T$ hence that $s\ {}_{k}\!\!\to_{\mathcal{R}}^{*} T$.

Proof. (of Theorem 2). By Lemma 9, $({}_{k}\!\!\to_{\mathcal{R}}^{*})[T] = (\to_{\mathcal{S}\cup\mathcal{A}}^{*})[Q_f^{\leq k}] \cap \mathcal{T}(\mathcal{F})$. The rewriting systems $\mathcal{S}$ and $\mathcal{A}$ being ground are inverse-recognizability preserving (Theorem 1). So $(\to_{\mathcal{S}\cup\mathcal{A}}^{*})[Q_f^{\leq k}]$ is recognizable and thus $({}_{k}\!\!\to_{\mathcal{R}}^{*})[T]$ is recognizable. In a second step one can extend Section 4.1 to the case where the restrictions (7) are not assumed anymore and consequently fully prove Theorem 2.

Corollary 1. *Every linear rewriting system of the class* BU *is inverse-recognizability preserving.*

4.2 Complexity

Upper-bounds We estimate here the complexity of the algorithm underlying our proof of Theorem 2.

Theorem 3. *Let $\mathcal{F}$ be a signature with symbols of arity ≤ 1, let $\mathcal{A}$ be some $n.f.t.a$ recognizing a language $T \subseteq \mathcal{T}(\mathcal{F})$ and let $\mathcal{R}$ be a finite rewriting system in $\mathsf{BU}^{-}(1)$. One can compute a $n.f.t.a$ $\mathcal{B}$ recognizing $(\rightarrow^{*}_{\mathcal{R}})[T]$ in time* $\mathrm{O}((\log(|\mathcal{F}|) \cdot \|\mathcal{A}\| \cdot \|\mathcal{R}\|)^3)$.

Our proof consists in reducing the above problem, via the computation of the ground system $\mathcal{S}$ of Section 4.1, to the computation of a set of descendants modulo some set of cancellation rules, which is achieved in cubic time in [2]. Since every left-basic semi-Thue system can be viewed as a $\mathsf{BU}^{-}(1)$ term rewriting system, Theorem 3 extends [2] (where a cubic complexity is proved for *cancellation systems* over a *fixed* alphabet) and improves [1] (where a degree 4 complexity is proved for *basic* semi-Thue systems).
Let us turn now to term rewriting systems over arbitrary signatures. Given a system $\mathcal{R}$ we define

$$\mathsf{A}(\mathcal{R}) := \max\{\mathsf{Card}(\mathcal{P}\mathsf{os}_{\mathcal{V}}(l)) \mid l \rightarrow r \in \mathcal{R}\}.$$

Theorem 4. *Let $(\mathcal{R}, \mathcal{F})$ a finite rewriting system in $\mathsf{BU}^{-}(1)$ and $\mathcal{A}$ be some $n.f.t.a$ over $\mathcal{F}$ recognizing a language $T \subseteq \mathcal{T}(\mathcal{F})$. One can compute a $n.f.t.a$ $\mathcal{B}$ recognizing $(\rightarrow^{*}_{\mathcal{R}})[T]$ in time polynomial w.r.t. $\|\mathcal{R}\| \cdot \|\mathcal{A}\|^{\mathsf{A}(\mathcal{R})}$.*

Our proof consists in computing the ground system $\mathcal{S}$ of Section 4.1 and to apply the result of [8], showing that the set of descendants of a recognizable set via a ground system $\mathcal{S}$ can be achieved in polynomial time.

Lower-bound

Theorem 5. *There exists a fixed signature $\mathcal{F}$ and two fixed recognizable sets T_1, T_2 over $\mathcal{F}$ such that: the problem to decide, for a given linear term rewriting system $(\mathcal{R}, \mathcal{F})$ in $\mathsf{BU}^{-}(1)$, whether $T_2 \cap (\rightarrow^{*}_{\mathcal{R}})[T_1] \neq \emptyset$, is NP-hard.*

The proof in [10] consists of a P-time reduction of the problem SAT to the above problem. This result shows that the exponential upper-bound in Theorem 4 cannot presumably be significantly improved.

5 Strongly Bottom-up systems

The following theorem is established in [10], even in the restricted case of semi-Thue systems.

Theorem 6. *For every $k \geq 1$, the problem to determine whether a finite linear term rewriting system $(\mathcal{R}, \mathcal{F})$ is $\mathsf{BU}(k)$ (resp. $\mathsf{BU}^{-}(k)$), is undecidable.*

We are thus lead to define some stronger but *decidable* conditions.

5.1 Strongly bottom-up systems

We abbreviate strongly bottom-up to sbu.

Definition 14. *A system* $(\mathcal{R},\mathcal{F})$ *is said* $\mathsf{SBU}(k)$ *iff*
for every $s \in \mathcal{T}(\mathcal{F}), \overline{t} \in \mathcal{T}(\mathcal{F}^{\mathbb{N}})$, $s \circ\!\!\rightarrow^*_{\mathcal{R}} \overline{t} \;\;\Rightarrow\; \overline{t} \in \mathcal{T}(\mathcal{F}^{\leq k})$.
We denote by $\mathsf{SBU}(k)$ *the class of* $\mathsf{SBU}(k)$ *systems and by* $\mathsf{SBU} = \bigcup_{k\in\mathbb{N}} \mathsf{SBU}(k)$ *the class of* strongly bottom-up *systems.*

The following lemma is obvious.

Lemma 10. *Every* $\mathsf{SBU}(k)$ *system is* $\mathsf{BU}(k)$.

This stronger condition over term rewriting systems is interesting because of the following property.

Proposition 1. *For every* $k \geq 0$*, it is decidable whether a finite term rewriting system* $(\mathcal{R},\mathcal{F})$ *is* $\mathsf{SBU}(k)$.

Proof. Note that every marked derivation starting from some $s \in \mathcal{T}(\mathcal{F})$ and leading to some $\overline{t} \in \mathcal{T}(\mathcal{F}^{\mathbb{N}}) \setminus \mathcal{T}(\mathcal{F}^{\leq k})$ must decompose as $s \;_{k+1}\!\circ\!\!\rightarrow^*_{\mathcal{R}} \;\overline{s}' \circ\!\!\rightarrow^*_{\mathcal{R}} \;\overline{t}$, with $\overline{s}' \in \mathcal{T}(\mathcal{F}^{\leq k+1}) \setminus \mathcal{T}(\mathcal{F}^{\leq k})$. A necessary and sufficient condition for $\mathcal{R}$ to be $\mathsf{SBU}(k)$ is thus that:

$$((\;_{k+1}\!\circ\!\!\rightarrow^*_{\mathcal{R}})[\mathcal{T}(\mathcal{F}^{\leq k+1}) \setminus \mathcal{T}(\mathcal{F}^{\leq k})]) \cap \mathcal{T}(\mathcal{F}) = \emptyset. \tag{16}$$

By Theorem 2 the left-handside of equality (16) is a recognizable set for which we can construct a $f.t.a$; we then just have to test whether this $f.t.a$ recognizes the empty set.

5.2 Sufficient condition

We show here that the LFPO^{-1} condition of [25] is a *sufficient* and tractable condition for the SBU property. Let us associate with every rewriting system a *graph* whose vertices are the rules of the system and whose arcs (R, R') express some kind of overlap between the right-handside of R and the left-handside of R'. Every arc has a *label* in $\{a, b, c, d\}$ indicating the category of overlap that occurs and a *weight* which is an integer (0 or 1). The intuitive meaning of the weight is that any derivation step using the corresponding overlap might increase some mark by this weight. (This graph is a slight modification of the sticking-out graph of [25]).

Definition 15. *Let* $s \in \mathcal{T}(\mathcal{F},\mathcal{V})$*,* $t \in \mathcal{T}(\mathcal{F},\mathcal{V}) \setminus \mathcal{V}$ *and* $w \in \mathcal{P}\mathsf{os}_{\mathcal{V}}(t)$*. We say that* s sticks out of t at w *if*
1- $\forall v \in \mathcal{P}\mathsf{os}(t)$ *s.t.* $\varepsilon \preceq v \prec w$*,* $v \in \mathcal{P}\mathsf{os}(s)$ *and* $s(v) = t(v)$*.*
2- $w \in \mathcal{P}\mathsf{os}(s)$ *and* $s/w \notin \mathcal{T}(\mathcal{F})$*.*
If in addition $s/w \notin \mathcal{V}$ *then* s strictly *sticks out of* t *at* w*.*

Definition 16. *Let* $\mathcal{R} = \{l_1 \to r_1, \ldots, l_n \to r_n\}$ *be a system. The* sticking-out graph *is the directed graph* $\mathsf{SG}(\mathcal{R}) = (V, E)$ *where* $V = \{1, \ldots, n\}$ *and* E *is defined as follows:*
a) if l_j *strictly sticks out of a subterm of* r_i *at* w*,* $i \xrightarrow{a} j \in E$*;*
b) if a strict subterm of l_j *strictly sticks out of* r_i *at* w*,* $i \xrightarrow{b} j \in E$*;*
c) if a subterm of r_i *sticks out of* l_j *at* w*,* $i \xrightarrow{c} j \in E$*;*
d) if r_i *sticks out of a strict subterm of* l_j *at* w*,* $i \xrightarrow{d} j \in E$*.*

Figure 4 shows all the possibilities in the four categories a, b, c, d. The *weight* of an arc

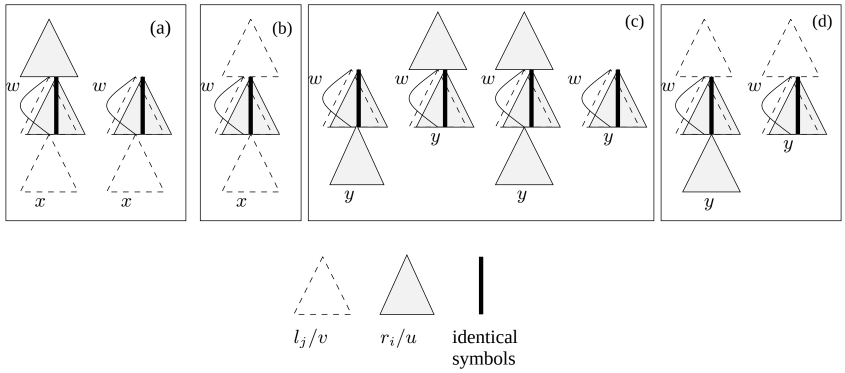


Fig. 4. Sticking-out cases

of $\mathsf{SG}(\mathcal{R})$ is defined by its label: if the label of the arc is a or b (resp. c or d) then its weight is 1 (resp. 0). The *weight of a path* in the graph is the sum of the weights of its arcs. The *weight of a graph* is the l.u.b. of the set of weights of all paths in the graph. The sticking-out graph of $\mathcal{R}_1$ of Example 1 is displayed in Figure 5. For an example with arities strictly greater than 1 see Example 6.

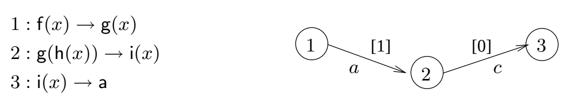


Fig. 5. The sticking-out graph of $\mathcal{R}_1$

Proposition 2. *Let $\mathcal{R}$ be a linear system. If $W(\mathsf{SG}(\mathcal{R})) = k$ then $\mathcal{R} \in \mathsf{SBU}(k+1)$.*

The proof consists of two steps. One first shows that the proposition holds for semi-Thue systems. In a second step one associates to every term-rewriting system $\mathcal{R}$ its *branch semi-Thue system* defined by: $T := \{u \to v \in \mathcal{F}^* \times \mathcal{F}^* \mid \exists l \to r \in \mathcal{R}, \exists x \in \mathcal{V}, ux \in \mathcal{B}(l), vx \in \mathcal{B}(r)\}$, where $\mathcal{B}(t)$ denotes the set of words over $\mathcal{F} \cup \mathcal{V}$ labeling the branches of t. The apparition of the mark $k+1$ in a marked derivation for $\mathcal{R}$ would imply the apparition of the mark $k+1$ in a marked derivation for T. But every path of weight $k+1$ in $\mathsf{SG}(T)$ leads to a path of weight k in $\mathsf{SG}(\mathcal{R})$.

Corollary 2 (sufficient condition). *Let $\mathcal{R}$ be a linear system. If $W(\mathsf{SG}(\mathcal{R}))$ is finite then $\mathcal{R} \in \mathsf{SBU}$.*

The above sufficient condition can be tested in P-time. An immediate consequence of Corollary 2 is the following.

Proposition 3. *1- Every right-ground system is* $\mathsf{SBU}(0)$.
2- Every inverse of a left-basic semi-Thue system is $\mathsf{SBU}(1)$.
3- Every growing linear system is $\mathsf{SBU}(1)$.
4- Every LFPO^{-1} *system is* SBU.

It can be checked, with ad hoc examples, that inclusions (2,3,4) above are strict.

Example 6. Let $\mathcal{R}_5 = \{\mathsf{f}(\mathsf{g}(x), \mathsf{a}) \to \mathsf{f}(x, \mathsf{b})\}$. $\mathcal{R}_5 \notin \mathsf{LFPO}^{-1}$ as $\mathsf{SG}(\mathcal{R})$ contains a loop a so a loop of weight $[1]$. It is easy to show by an ad-hoc proof that $\mathcal{R}_5 \in \mathsf{SBU}^{-}(1)$. However our sufficient condition is not able to capture $\mathcal{R}_5$.

6 Related work and perspectives

Related work Beside the references already mentioned, our work is also related to:
- $[16, 5]$ from which we borrow some framework concerning equivalence relations over derivations in order to give in [10] a criterium for the $\mathsf{BU}(k)$ property for STS,
- [15], where an undecidability result concerning a notion analogous to our SBU systems, makes us think that the SBU-condition (without any specified value of k) should be undecidable for semi-Thue systems, hence for term rewriting systems.

Perspectives Let us mention some natural perspectives of development for this work:
- it is tempting to extend the notion of *bottom-up* rewriting (resp. system) to non-linear systems. This class would extend the class of growing systems studied in [19]. Also allowing free variables in the right-handsides seems reasonable.
- a dual notion of *top-down* rewriting and a corresponding class of top-down systems should be defined. This class would presumably extend the class of Layered Transducing systems [22].

Some work in these two directions has been undertaken by the authors.

Acknowledgments: We thank the referees for their useful comments.

References

1. M. Benois. Descendants of regular language in a class of rewriting systems: algorithm and complexity of an automata construction. In *Rewriting techniques and applications (Bordeaux, 1987)*, volume 256 of *Lecture Notes in Comput. Sci.*, pages 121–132. Springer, Berlin, 1987.
2. M. Benois and J. Sakarovitch. On the complexity of some extended word problems defined by cancellation rules. *Inform. Process. Lett.*, 23(6):281–287, 1986.
3. R.V. Book, M. Jantzen, and C. Wrathall. Monadic Thue systems. *TCS 19*, pages 231–251, 1982.
4. H. Comon, M. Dauchet, R. Gilleron, F. Jacquemard, D. Lugiez, S. Tison, and M. Tommasi. Tree automata techniques and applications, 2002. Draft, available from `www.grappa.univ-lille3.fr/tata`.
5. R. Cremanns and F. Otto. Finite derivation type implies the homological finiteness condition FP_3. *J. Symbolic Comput.*, 18(2):91–112, 1994.
6. M. Dauchet, T. Heuillard, P. Lescanne, and S. Tison. Decidability of the confluence of finite ground term rewrite systems and of other related term rewrite systems. *Inf. Comput.*, 88(2):187–201, 1990.

7. M. Dauchet and S. Tison. The theory of ground rewrite systems is decidable. In *Fifth Annual IEEE Symposium on Logic in Computer Science (Philadelphia, PA, 1990)*, pages 242–248. IEEE Comput. Soc. Press, Los Alamitos, CA, 1990.
8. A. Deruyver and R. Gilleron. The reachability problem for ground TRS and some extensions. In *TAPSOFT '89 (Barcelona, 1989)*, volume 351 of *Lecture Notes in Comput. Sci.*, pages 227–243. Springer, Berlin, 1989.
9. I. Durand and A. Middeldorp. Decidable call-by-need computations in term rewriting. *Information and Computation*, 196:95–126, 2005.
10. I. Durand and G. Sénizergues. Bottom-up rewriting for words and terms. Manuscript, available at `http://dept-info.labri.u-bordeaux.fr/~ges`, 2007.
11. Z. Fülöp, E. Jurvanen, M. Steinby, and S. Vágvölgyi. On one-pass term rewriting. In *MFCS*, pages 248–256, 1998.
12. A. Geser, D. Hofbauer, and J. Waldmann. Match-bounded string rewriting systems. *Journal Applicable Algebra in Engineering, Communication and Computing*, 15(3-4):149–171, November, 2004.
13. A. Geser, D. Hofbauer, J. Waldmann, and H. Zantema. On tree automata that certify termination of left-linear term rewriting systems. In *Proceedings of the 16th International Conference on Rewriting Techniques and Applications*, volume 3467 of *Lecture Notes in Computer Science*. Springer Verlag, 2005.
14. F. Jacquemard. Decidable approximations of term rewriting systems. In *Proceedings of the 7th International Conference on Rewriting Techniques and Applications*, volume 1103 of *Lecture Notes in Computer Science*, pages 362–376, 1996.
15. T. Knapik and H. Calbrix. Thue specifications and their monadic second-order properties. *Fund. Inform.*, 39(3):305–325, 1999.
16. Y. Lafont and A. Prouté. Church-Rosser property and homology of monoids. *Math. Structures Comput. Sci.*, 1(3):297–326, 1991.
17. M. Lohrey and G. Sénizergues. Rational subsets of HNN-extensions. Manuscript, available at `http://dept-info.labri.u-bordeaux.fr/~ges`, 2005.
18. P.V. Silva M. Kambites and B. Steinberg. On the rational subset problem for groups. *J. of Algebra*, To appear.
19. T. Nagaya and Y. Toyama. Decidability for left-linear growing term rewriting systems. *Information and Computation*, 178(2):499–514, 2002.
20. P. Réty and J. Vuotto. Tree automata for rewrite strategies. *J. Symb. Comput.*, 40(1):749–794, 2005.
21. J. Sakarovitch. Syntaxe des langages de Chomsky, essai sur le déterminisme. *Thèse de doctorat d'état de l'université Paris VII*, pages 1–175, 1979.
22. H. Seki, T. Takai, Y. Fujinaka, and Y. Kaji. Layered transducing term rewriting system and its recognizability preserving property. In *Proceedings of the 13th International Conference on Rewriting Techniques and Applications*, volume 2378 of *Lecture Notes in Computer Science*. Springer Verlag, 2002.
23. G. Sénizergues. Formal languages & word-rewriting. In *Term rewriting (Font Romeu, 1993)*, volume 909 of *Lecture Notes in Comput. Sci.*, pages 75–94. Springer, Berlin, 1995.
24. F. Seynhaeve, S. Tison, and M. Tommasi. Homomorphisms and concurrent term rewriting. In *FCT*, pages 475–487, 1999.
25. T. Takai, Y. Kaji, and H. Seki. Right-linear finite-path overlapping term rewriting systems effectively preserve recognizability. *Scienticae Mathematicae Japonicae*, 2006. (to appear, preliminary version: IEICE Technical Report COMP98-45).