Over-approximating Descendants by Synchronized Tree Languages

Yohan Boichut¹, Jacques Chabin¹, and Pierre Réty¹

¹ LIFO - Université d’Orléans, B.P. 6759, 45067 Orléans cedex 2, France
   {yohan.boichut, jacques.chabin, pierre.rety}@univ-orleans.fr

Abstract

Over-approximating the descendants (successors) of an initial set of terms by a rewrite system is used in verification. The success of such verification methods depends on the quality of the approximation. To get better approximations, we are going to use non-regular languages. We present a procedure that always terminates and that computes an over-approximation of descendants, using synchronized tree-(tuple) languages expressed by logic programs.

1998 ACM Subject Classification F.4.2

Keywords and phrases Rewriting systems, non-regular approximations, logic programming, tree languages, descendants

Digital Object Identifier 10.4230/LIPIcs.xxx.yyy.p

1 Introduction

Given an initial set of terms \(I\), computing the descendants (successors) of \(I\) by a rewrite system \(R\) is used in the verification domain, for example to check cryptographic protocols or Java programs [2, 8, 10, 9]. Let \(R^*(I)\) denote the set of descendants of \(I\), and consider a set \(Bad\) of undesirable terms. Thus, if a term of \(Bad\) is reached from \(I\), i.e. \(R^*(I) \cap Bad \neq \emptyset\), it means that the protocol or the program is flawed. In general, it is not possible to compute \(R^*(I)\) exactly. Instead, we compute an over-approximation \(App\) of \(R^*(I)\) (i.e. \(App \supseteq R^*(I)\)), and check that \(App \cap Bad = \emptyset\), which ensures that the protocol or the program is correct.

Most often, \(I\), \(App\) and \(Bad\) have been considered as regular tree languages, recognized by finite tree automata. In the general case, \(R^*(I)\) is not regular, even if \(I\) is. Moreover, the expressiveness of regular languages is poor, and the over-approximation \(App\) may not be precise enough, and we may have \(App \cap Bad \neq \emptyset\) whereas \(R^*(I) \cap Bad = \emptyset\). In other words, the protocol is correct, but we cannot prove it. Some work has proposed CEGAR-techniques (Counter-Example Guided Approximation Refinement) in order to conclude as often as possible [2, 4, 6]. However, in some cases, no regular over-approximation works, whatever the quality of the approximation is [5].

To overcome this theoretical limit, we want to use more expressive languages to express the over-approximation, i.e. non-regular ones. However, to be able to check that \(App \cap Bad = \emptyset\), we need a class of languages closed under intersection and whose emptiness is decidable. Actually, since we still assume that \(Bad\) is regular, closure under intersection with a regular language is enough. The class of context-free tree languages has these properties, and an over-approximation of descendants using context-free tree languages has been proposed in [14]. This class of languages is quite interesting, however it cannot express relations (or countings) in terms between independent branches, except if there are only unary symbols and constants. For example, let \(R = \{f(x) \rightarrow c(x, x)\}\) and the infinite set \(I = \{f(t)\}\) where...
Over-approximating Descendants by Synchronized Tree Languages

2 Preliminaries

Consider a finite ranked alphabet Σ and a set of variables Var. Each symbol f ∈ Σ has a unique arity, denoted by ar(f). The notions of first-order term, position, substitution, are defined as usual. Given σ and σ’ two substitutions, σ ◦ σ’ denotes the substitution such that for any variable x, σ ◦ σ’(x) = σ(σ’(x)). TΣ denotes the set of ground terms (without variables) over Σ. For a term t, Var(t) is the set of variables of t, Pos(t) is the set of positions of t. For p ∈ Pos(t), t(p) is the symbol of Σ ∪ Var occurring at position p in t, and t(p) is the subterm of t at position p. The term t[t]p is obtained from t by replacing the subterm at position p by t’. PosVar(t) = {p ∈ Pos(t) | t(p) ∈ Var}, PosNonVar(t) = {p ∈ Pos(t) | t(p) ∉ Var}. Note that if p ∈ PosNonVar(t), t[p] = f(t1, . . . , tn), and i ∈ {1, . . . , n}, then pi is the position of ti in t. For p, p’ ∈ Pos(t), p < p’ means that p occurs in t strictly above p’. Let t, t’ be terms, t is more general than t’ (denoted t ≤ t’) if there exists a substitution ρ s.t. ρ(t) = t’. Let σ, σ’ be substitutions, σ is more general than σ’ (denoted σ ≤ σ’) if there exists a substitution ρ s.t. ρ ◦ σ = σ’. A rewrite rule is an oriented pair of terms, written l → r. We always assume that l is not a variable, and Var(r) ⊆ Var(l). A rewrite system R is a finite set of rewrite rules. Lhs stands for left-hand-side, rhs for right-hand-side. The rewrite relation →R is defined as follows: t →R t’ if there exist a position p ∈ PosVar(t), a rule l → r ∈ R, and a substitution θ s.t. t[p] = θ(l) and t’ = t[θ(r)]p. →R denotes the reflexive-transitive closure of →R. t’ is a descendant of t if t →R t’. If E is a set of ground terms, R∗(E) denotes the set of descendants of elements of E.

In the following, we consider the framework of pure logic programming, and the class of synchronized tree-tuple languages defined by CS-clauses [16, 17]. Given a set Pred of predicate symbols; atoms, goals, bodies and Horn-clauses are defined as usual. Note that
both goals and bodies are sequences of atoms. We will use letters \( G \) or \( B \) for sequences of atoms, and \( A \) for atoms. Given a goal \( G = A_1, \ldots, A_k \) and positive integers \( i, j \), we define \( G|_i = A_i \) and \( G|_{i,j} = (A_i)|_j = t_j \) where \( A_i = P(t_1, \ldots, t_n) \).

\[\begin{align*}
\text{Definition 1.} & \quad \text{Let } B \text{ be a sequence of atoms. } B \text{ is flat if for each atom } P(t_1, \ldots, t_n) \text{ of } B, \text{ all terms } t_1, \ldots, t_n \text{ are variables. } B \text{ is linear if each variable occurring in } B \text{ (possibly at sub-term position) occurs only once in } B. \text{ Note that the empty sequence of atoms (denoted by } \emptyset \text{) is flat and linear. }
\end{align*}\]

A CS-clause\(^1\) is a Horn-clause \( H \leftarrow B \) s.t. \( B \) is flat and linear. A CS-program \( \text{Prog} \) is a logic program composed of CS-clauses.

Given a predicate symbol \( P \) of arity \( n \), the tree-(tuple) language generated by \( P \) is \( L(P) = \{t \in (T_\Sigma)^n \mid P(t) \in \text{Mod}(\text{Prog})\} \), where \( T_\Sigma \) is the set of ground terms over the signature \( \Sigma \) and \( \text{Mod}(\text{Prog}) \) is the least Herbrand model of \( \text{Prog} \). \( L(P) \) is called Synchronized language.

The following definition describes the different kinds of CS-clauses that can occur.

\[\begin{align*}
\text{Definition 2.} & \quad \text{A CS-clause } P(t_1, \ldots, t_n) \leftarrow B \text{ is:} \\
= & \quad \text{empty if } \forall i \in \{1, \ldots, n\}, t_i \text{ is a variable.} \\
= & \quad \text{normalized if } \forall i \in \{1, \ldots, n\}, t_i \text{ is a variable or contains only one occurrence of function-symbol. A CS-program is normalized if all its clauses are normalized.} \\
= & \quad \text{preserving if } \var{P(t_1, \ldots, t_n)} \subseteq \var{B}. \text{ A CS-program is preserving if all its clauses are preserving.} \\
= & \quad \text{synchronizing if } B \text{ is composed of only one atom.}
\end{align*}\]

\[\begin{align*}
\text{Example 3.} & \quad \text{The CS-clause } P(x, y, z) \leftarrow G(x, y, z) \text{ is empty, normalized, and preserving} \\
& \quad \text{(} x, y, z \text{ are variables). The CS-clause } P(f(x), y, g(x, z)) \leftarrow G(x, y) \text{ is normalized and non-preservation. Both clauses are synchronizing.}
\end{align*}\]

Given a CS-program, we focus on two kinds of derivations: a classical one based on unification and a rewriting one based on matching and a rewriting process.

\[\begin{align*}
\text{Definition 4.} & \quad \text{Given a logic program } \text{Prog} \text{ and a sequence of atoms } G, \\
= & \quad G \text{ derives into } G' \text{ by a resolution step if there exist a clause}^2 \ H \leftarrow B \text{ in } \text{Prog} \text{ and an atom } A \in G \text{ such that } A \text{ and } H \text{ are unifiable by the most general unifier } \sigma \text{ (then } \sigma(A) = \sigma(H) \text{) and } G' = \sigma(G)[\sigma(A) \leftarrow \sigma(B)]. \text{ It is written } G \leadsto_\sigma G'. \\
= & \quad G \text{ rewrites into } G' \text{ if there exist a clause } H \leftarrow B \text{ in } \text{Prog}, \text{ an atom } A \in G, \text{ and a substitution } \sigma, \text{ such that } A = \sigma(H) \text{ (} A \text{ is not instantiated by } \sigma \text{) and } G' = G[A \leftarrow \sigma(B)]. \text{ It is written } G \rightarrow_\sigma G'.
\end{align*}\]

\[\begin{align*}
\text{Example 5.} & \quad \text{Let } \text{Prog} = \{P(x_1, g(x_2)) \leftarrow P'(x_1, x_2), \ P'(f(x_1), x_2) \leftarrow P''(x_1, x_2)\}, \text{ and consider } G = P(f(x), y). \text{ Thus, } P'(f(x), y) \leadsto_{\sigma_1} P'(f(x), x_2) \text{ with } \sigma_1 = [x_1/f(x), y/g(x_2)] \text{ and } P'(f(x), y) \rightarrow_{\sigma_2} P''(x, y) \text{ with } \sigma_2 = [x_1/x, x_2/y].
\end{align*}\]

We consider the transitive closure \( \leadsto^+ \) and the reflexive-transitive closure \( \leadsto^* \) of \( \leadsto \).

For both derivations, given a logic program \( \text{Prog} \) and three sequences of atoms \( G_1, G_2 \) and \( G_3 \):

\[\begin{align*}
= & \quad \text{if } G_1 \leadsto_{\sigma_1} G_2 \text{ and } G_2 \leadsto_{\sigma_2} G_3 \text{ then one has } G_1 \leadsto^*_{\sigma_2 \circ \sigma_1} G_3;
\end{align*}\]

---

\(^1\) In former papers, synchronized tree-tuple languages were defined thanks to sorts of grammars, called constraint systems. Thus “CS” stands for Constraint System.

\(^2\) We assume that the clause and \( G \) have distinct variables.
In the remainder of the paper, given a set of CS-clauses $Prog$ and two sequences of atoms $G_1$ and $G_2$, $G_1 \propto^*_{prog} G_2$ (resp. $G_1 \propto^*_\sigma G_2$) also denotes that $G_2$ can be derived (resp. rewritten) from $G_1$ using clauses of $Prog$.

It is well known that resolution is complete.

**Theorem 6.** Let $A$ be a ground atom. $A \in \text{Mod}(Prog)$ iff $A \propto^*_{prog} \emptyset$.

**Example 7.** Let $A = P(f(g(a)), g(a), c)$ and $A' = P'(f(g(a)), h(c))$ be two ground atoms. Let $Prog$ be the CS-program defined by:

$$Prog = \{ P(f(g(x)), y, c) \leftarrow P_1(x), P_2(g(y)). P_1(a) \leftarrow. P_2(g(x)) \leftarrow P_1(x). P'(f(x), u(z)) \leftarrow. \}$$

Thus, $A \in \text{Mod}(Prog)$ and $A' \notin \text{Mod}(Prog)$.

Note that for any atom $A$, if $A \rightarrow B$ then $A \propto B$. If in addition $Prog$ is preserving, then $\text{Var}(A) \subseteq \text{Var}(B)$. On the other hand, $A \propto_{\sigma} B$ implies $\sigma(A) \rightarrow B$. Consequently, if $A$ is ground, $A \propto B$ implies $A \rightarrow B$.

The following lemma focuses on a preserving property of the relation $\propto$.

**Lemma 8.** Let $Prog$ be a CS-program, and $G$ be a sequence of atoms. Let $|G|_\Sigma$ denote the number of occurrences of function-symbols in $G$. If $G$ is linear and $G \propto^* G'$, then $G'$ is also linear and $|G'|_\Sigma \leq |G|_\Sigma$. Consequently, if $G$ is flat and linear, then $G'$ is also flat and linear.

**Proof.** Let $G = A^1 \ldots A^k$ be a linear sequence of atoms and suppose that $G \propto_{\sigma} G'$. Then there exist an atom $A^i(t_1, \ldots, t_n)$ of $G$ and a CS-clause $A^i(t_1, \ldots, t_n) \leftarrow B$ in $Prog$ such that $G' = \sigma(G)[\sigma(A^i) \leftarrow \sigma(B)]$. As $G$ is linear and $\sigma$ is the most general unifier between $A^i(s_1, \ldots, s_n)$ and $A^i(t_1, \ldots, t_n)$, $\sigma$ does not instantiate variables from $A^i, \ldots, A^{i-1}, A^{i+1}, \ldots, A^k$. So $G' = A^1, \ldots, A^{i-1}, \sigma(B), A^{i+1}, \ldots, A^k$.

$G'$ is not linear only if $\sigma(B)$ is not linear. As $B$ is linear, $\sigma(B)$ is not linear would require that two distinct variables $x_1, x_2$ from $B$ are instantiated by two terms containing a same variable $y \in \text{Var}(\sigma(x_{j_1}) \cap \text{Var}(\sigma(x_{j_2})))$. Since $\sigma$ is the most general unifier, $x_{j_1}, x_{j_2}$ are also in $\text{Var}(A^i(t_1, \ldots, t_n))$ ($\sigma$ does not instantiate extra variables). Then $y$ occurs at least twice in $A^i(s_1, \ldots, s_n)$ (the atom of goal $G$), which is impossible since $G$ is linear. Consequently $G'$ is linear.

By contradiction: to obtain $|G'|_\Sigma > |G|_\Sigma$, we must have in $\sigma(B)$ a duplication of a non-variable subterm of $\sigma((A^i(s_1, \ldots, s_n))$ (because $B$ is flat), which is not possible because $B$ and $A^i(s_1, \ldots, s_n)$ are linear and $\sigma$ is the most general unifier.

The result trivially extends to the case of several steps $G \propto^* G'$.

**Example 9.** Let $Prog = \{ P(g(x), f(x)) \leftarrow P_1(x) \}$ and $G = P(g(f(y)), z)$). Then $G \propto G'$ with $G' = P_1(f(y))$, and $G'$ is linear. Moreover, $|G'|_\Sigma \leq |G|_\Sigma$ with $\Sigma = \{f^1, a^0\}$.

### 3 Computing Descendants

Given a CS-program $Prog$ and a left-linear rewrite system $R$, we propose a technique allowing us to compute a CS-program $Prog'$ such that $R^*(\text{Mod}(Prog)) \subseteq \text{Mod}(Prog')$. First of all, a notion of critical pairs is introduced in Section 3.1. Roughly speaking, this notion makes the detection of uncovered rewriting steps possible. Critical pair detection is at the heart of the technique. Thus, in Section 3.2 some restrictions are underlined on CS-programs in order to make the number of critical pairs finite. Moreover, when a CS-program does not fit these restrictions, we have proposed a technique in order to transform such a CS-program
into another one of the expected form (REMOVE CYCLES in Fig.1). The detected critical pairs lead to a set of CS-clauses to be added in the current CS-program. However, they may not be in the expected form i.e. normalized CS-clauses. Indeed, one of the restrictions set in Section 3.2 is that the CS-program has to be normalized. So, we propose in Section 3.3 an algorithm providing normalized CS-clauses from non-normalized ones. Finally, in Section 3.4, our main contribution, i.e. the computation of an over-approximating CS-program, is fully described.

Figure 1 An overview of our contribution

3.1 Critical pairs

The notion of critical pair is at the heart of our technique. Indeed, it allows us to add CS-clauses into the current CS-program in order to cover rewriting steps. This notion is described in Definition 10.

Definition 10. Let Prog be a CS-program and \( l \rightarrow r \) be a left-linear rewrite rule. Let \( x_1, \ldots, x_n \) be distinct variables s.t. \( \{x_1, \ldots, x_n\} \cap \text{Var}(l) = \emptyset \). If there are \( P \) and \( k \) s.t. \( P(x_1, \ldots, x_{k-1}, l, x_{k+1}, \ldots, x_n) \rightarrow^\ast G \) where resolution is applied only on non-flat atoms, \( G \) is flat, and the clause \( P(t_1, \ldots, t_n) \leftarrow B \) used during the first step of this derivation satisfies \( t_k \) is not a variable\(^3\), then the clause \( \theta(P(x_1, \ldots, x_{k-1}, r, x_{k+1}, \ldots, x_n)) \leftarrow G \) is called critical pair.

Remark. Since \( l \) is linear, \( P(x_1, \ldots, x_{k-1}, l, x_{k+1}, \ldots, x_n) \) is linear, and thanks to Lemma 8 \( G \) is linear, then a critical pair is a CS-clause. Moreover, if \( \text{Prog} \) is preserving then a critical pair is a preserving CS-clause\(^4\).

Example 11. Let \( \text{Prog} \) be the normalized and preserving CS-program defined by:

\[
\text{Prog} = \{ P(c(x), c(x), y) \leftarrow Q(x, y), Q(a, b) \leftarrow, Q(c(x), y) \leftarrow Q(x, y) \}.
\]

and consider the left-linear rewrite rule: \( c(c(x')) \rightarrow h(h(x')) \). Recall that for all goals \( G, G' \), the step \( G \rightarrow G' \) means that \( G \sim \sigma G' \) where \( \sigma \) does not instantiate the variables of \( G \).

Thus \( P(c(c(x')), y', z') \sim \theta Q(c(x'), y) \rightleftarrows Q(x', y) \) where \( \theta = [x/c(x'), y'/c(c(x')), z'/y] \).

It generates the critical pair \( P(h(h(x'))), c(c(x')), y) \leftarrow Q(x', y) \). There are also two other critical pairs: \( P(c(c(x')), h(h(x')), y) \leftarrow Q(x', y) \) and \( Q(h(h(x')), y) \leftarrow Q(x', y) \).

\(^3\) In other words, the overlap of \( l \) on the clause head \( P(t_1, \ldots, t_n) \) is done at a non-variable position.

\(^4\) We have \( \theta(P(x_1, \ldots, x_{k-1}, l, x_{k+1}, \ldots, x_n)) \rightarrow^\ast G \), and since \( \text{Prog} \) is preserving \( \text{Var}(\theta(P(x_1, \ldots, x_{k-1}, l, x_{k+1}, \ldots, x_n))) \subseteq \text{Var}(G) \). Since \( \text{Var}(\theta) \subseteq \text{Var}(l) \) we have \( \text{Var}(\theta(P(x_1, \ldots, x_{k-1}, r, x_{k+1}, \ldots, x_n))) \subseteq \text{Var}(G) \).
However, some of the detected critical pairs are not so critical since they are already covered by the current CS-program. These critical pairs are said to be convergent.

**Definition 12.** A critical pair $H \leftarrow B$ is said convergent if $H \rightarrow^\ast_{Prog} B$.

**Example 13.** The three critical pairs detected in Example 11 are not convergent in $Prog$.

So, here we come to Theorem 14, i.e. the corner stone making our approach sound.

Indeed, given a rewrite system $R$ and CS-program $Prog$, if every critical pair that can be detected is convergent, then for any set of terms $I$ such that $I \subseteq Mod(Prog)$, $Mod(Prog)$ is an over-approximation of the set of terms reachable by $R$ from $I$.

**Theorem 14.** Let $Prog$ be a normalized and preserving CS-program and $R$ be a left-linear rewrite system.

If all critical pairs are convergent, then $Mod(Prog)$ is closed under rewriting by $R$, i.e. $(A \in Mod(Prog) \land A \rightarrow^\ast_{R} A') \implies A' \in Mod(Prog)$.

**Proof.** Let $A \in Mod(Prog)$ s.t. $A \rightarrow_{l\rightarrow r} A'$. Then $A[i] = C[\sigma(l)]$ for some $i \in \mathbb{N}$ and $A' = A[i \leftarrow C[\sigma(r)]$.

Since resolution is complete, $A \sim^\ast \emptyset$. Since $Prog$ is normalized and preserving, resolution consumes symbols in $C$ one by one, thus $G_0 = A \sim^\ast G_k \sim^\ast \emptyset$ and there exists an atom $A'' = P(t_1, \ldots, t_n)$ in $G_k$ and $j$ s.t. $t_j = \sigma(l)$ and the top symbol of $t_j$ is consumed during the step $G_k \sim G_{k+1}$. Consider new variables $x_1, \ldots, x_n$ s.t. $\{x_1, \ldots, x_n\} \cap Var(l) = \emptyset$, and let us define the substitution $\sigma'$ by $\forall i, \sigma'(x_i) = t_i$ and $\forall x \in Var(l), \sigma'(x) = \sigma(x)$. Then $\sigma'(P(x_1, \ldots, x_{j-1}, l, x_{j+1}, \ldots, x_n)) = A''$, and according to resolution (or narrowing) properties $P(x_1, \ldots, l, \ldots, x_n) \sim^\ast \emptyset$ and $\theta \leq \sigma'$.

This derivation can be decomposed into : $P(x_1, \ldots, l, \ldots, x_n) \sim^\ast_{\theta_1} G' \sim^\ast_{\theta_2} G \sim^\ast_{\theta_3} \emptyset$ where $\theta = \theta_1 \circ \theta_2 \circ \theta_3$, and s.t. $G'$ is not flat and $G$ is flat\(^5\). $P(x_1, \ldots, l, \ldots, x_n) \sim^\ast_{\theta_1} G' \sim^\ast_{\theta_2} G$ can be commuted into $P(x_1, \ldots, l, \ldots, x_n) \sim^\ast_{\theta_1} G' \sim^\ast_{\theta_2} G \sim^\ast_{\theta_3} \emptyset$. Let $B'$ be a flat goal. Then $\sigma'(P(x_1, \ldots, l, \ldots, x_n)) = A''$, and $B'$ is flat, $B'$ is not flat, and within $P(x_1, \ldots, l, \ldots, x_n) \sim^\ast_{\theta_1} G' \sim^\ast_{\theta_2} G \sim^\ast_{\theta_3} \emptyset$ $B$ resolution is applied only on non-flat atoms, and we have $\gamma_3 \circ \gamma_2 \circ \gamma_1 = \theta_2 \circ \theta_1$. Then $\gamma_2 \circ \gamma_1(P(x_1, \ldots, r, \ldots, x_n)) \leftarrow B$ is a critical pair. By hypothesis, it is convergent, then $\gamma_2 \circ \gamma_1(P(x_1, \ldots, r, \ldots, x_n)) \rightarrow^\ast B$. Note that $\gamma_3(G) \rightarrow^\ast G$ and recall that $\theta_3 \circ \gamma_3 \circ \gamma_2 \circ \gamma_1 = \theta_2 \circ \theta_1$. Then $\theta(P(x_1, \ldots, r, \ldots, x_n)) \rightarrow^\ast \theta_3(G) \rightarrow^\ast \emptyset$, and since $\theta \leq \sigma'$ we get $P(t_1, \ldots, \sigma(r), \ldots, t_n) = \sigma'(P(x_1, \ldots, r, \ldots, x_n)) \rightarrow^\ast \emptyset$. Therefore $A' \sim^\ast G_k[A'' \leftarrow P(x_1, \ldots, \sigma(r), \ldots, t_n)] \sim^\ast \emptyset$, hence $A' \in Mod(Prog)$.

By trivial induction, the proof can be extended to the case of several rewrite steps. \hfill \blacksquare

If $Prog$ is not normalized, Theorem 14 does not hold.

**Example 15.** Let $Prog = \{ P(c(f(a))) \leftarrow \}$ and $R = \{ f(a) \rightarrow b \}$. All critical pairs are convergent since there is no critical pair. $P(c(f(a))) \in Mod(Prog)$ and $P(c(f(a))) \rightarrow_R P(c(b))$. However there is no resolution step issued from $P(c(b))$, then $P(c(b)) \notin Mod(Prog)$.

If $Prog$ is not preserving, Theorem 14 does not hold.

**Example 16.** Let $Prog = \{ P(c(x), c(x), y) \leftarrow Q(y). Q(a) \leftarrow \}$, and $R = \{ f(b) \rightarrow b \}$. All critical pairs are convergent since there is no critical pair. $P(c(f(b)), c(f(b)), a) \rightarrow_{Prog} Q(a) \rightarrow_{Prog} \emptyset$, then $P(c(f(b)), c(f(b)), a) \in Mod(Prog)$. On the other hand, $P(c(f(b)), c(f(b)), a) \rightarrow_R P(c(b), c(f(b)), a)$. However there is no resolution step issued from $P(c(b), c(f(b)), a)$, then $P(c(b), c(f(b)), a) \notin Mod(Prog)$.

\(^5\) Since $\emptyset$ is flat, a flat goal can always be reached, i.e. in some cases $G = \emptyset$. 
Unfortunately, for a given finite CS-program, there may be infinitely many critical pairs. In the following section, this problem is illustrated and some syntactical conditions on CS-program are underlined in order to avoid this critical situation.

3.2 Ensuring finitely many critical pairs

The following example illustrates a situation where the number of critical pairs is unbounded.

Example 17. Let $\Sigma = \{ f^{2}, c^{1}, d^{1}, s^{1}, a^{0} \}$ and $f(c(x), y) \rightarrow d(y)$ be a rewrite rule, and $\text{Prog} = \{ P_{0}(f(x, y)) \leftarrow P_{1}(x, y), P_{1}(x, s(y)) \leftarrow P_{1}(x, y), P_{1}(c(x), y) \leftarrow P_{2}(x, y), P_{2}(a, a) \leftarrow . \}$. Then $P_{0}(f(c(x), y)) \rightarrow P_{1}(c(x), y) \xrightarrow{s_{y/s}(y)} P_{1}(c(x), y) \xrightarrow{s_{y/s}(y)} \cdots P_{1}(c(x), y) \rightarrow P_{2}(x, y)$. Resolution is applied only on non-flat atoms and the last atom obtained by this derivation is flat. The composition of substitutions along this derivation gives $y/s^{n}(y)$ for some $n \in \mathbb{N}$. There are infinitely many such derivations, which generates infinitely many critical pairs of the form $P_{0}(d(s^{n}(y))) \leftarrow P_{2}(x, y)$.

This is annoying since the completion process presented in the following needs to compute all critical pairs. This is why we define sufficient conditions to ensure that a given finite CS-program has finitely many critical pairs.

Definition 18. $\text{Prog}$ is empty-recursive if there exist a predicate $P$ and distinct variables $x_{1}, \ldots, x_{n}$ s.t. $P(x_{1}, \ldots, x_{n}) \xrightarrow{\sigma^{+}} A_{1}, \ldots, P(x'_{1}, \ldots, x'_{n})\ldots A_{k}$ where $x'_{1}, \ldots, x'_{n}$ are variables and there exist $i, j$ s.t. $x'_{i} = \sigma(x_{i})$ and $\sigma(x_{j})$ is not a variable and $x'_{j} \in \text{Var}(\sigma(x_{j}))$.

Example 19. Let $\text{Prog}$ be the CS-program defined as follows:

$\text{Prog} = \{ P(x', s(y')) \leftarrow P(x', y'), P(a, b) \leftarrow . \}$

From $P(x, y)$, one can obtained the following derivation: $P(x, y) \xrightarrow{\sigma[x/x', y/s(y')] P(x', y')}$. Consequently, $\text{Prog}$ is empty-recursive since $\sigma = [x/x', y/s(y')]$ and $y'$ is a variable of $\sigma(y) = s(y')$.

The following lemma shows that the non empty-recursiveness of a CS-program is sufficient to ensure the finiteness of the number of critical pairs.

Lemma 20. Let $\text{Prog}$ be a normalized CS-program.

If $\text{Prog}$ is not empty-recursive, then the number of critical pairs is finite.

Remark. Note that the CS-program of Example 17 is normalized and has infinitely many critical pairs, however it is empty-recursive because $P_{1}(x, y) \xrightarrow{[x/x', y/s(y')]} P_{1}(x', y')$.

Proof. By contrapositive. Let us suppose there exist infinitely many critical pairs. So there exist $P_{1}$ and infinitely many derivations of the form (i) : $P_{1}(x_{1}, \ldots, x_{k-1}, l, x_{k+1}, \ldots, x_{n}) \xrightarrow{\sigma^{*}} G' \xrightarrow{\sigma} G$ (the number of steps is not bounded). As the number of predicates is finite and every predicate has a fixed arity, there exists a predicate $P_{2}$ and a derivation of the form (ii) : $P_{2}(t_{1}, \ldots, t_{p}) \xrightarrow{\sigma} G_{1}^{n} P_{2}(t'_{1}, \ldots, t'_{q}), G_{2}^{m}$ (with $k > 0$) included in some derivation of (i), strictly before the last step, such that :

1. $G_{1}^{n}$ and $G_{2}^{m}$ are flat.
2. $\sigma$ is not empty and there exists a variable $x$ in $P_{2}(t_{1}, \ldots, t_{p})$ such that $\sigma(x) = t$ and $t$ is not a variable and contains a variable $y$ that occurs in $P_{2}(t'_{1}, \ldots, t'_{q})$. Otherwise we could not have an infinite number of $\sigma$ necessary to obtain infinitely many critical pairs.
3. At least one term \( t'_j (j \in \{1, \ldots, p\}) \) is not a variable (only the last step of the initial derivation produces a flat goal \( G \)). As we use a CS-clause in each derivation step, we can assume that \( t'_j \) is a term among \( t_1, \ldots, t_n \) and moreover that \( t'_j = t_j \). This property does not necessarily hold as soon as \( P_2 \) is reached within (ii). We may have to consider further occurrences of \( P_2 \) so that each required term occurs in the required argument, which will necessarily happen because there are only finitely many permutations. So, for each variable \( x \) occurring in the non-variable terms, we have \( \sigma(x) = x \).

4. From the previous item, we deduce that the variable \( x \) found in item 2 is one of the terms \( t_1, \ldots, t_p \), say \( t_k \). We can assume that \( y \) is \( t'_k \).

If in the (ii) derivation we replace all non-variable terms by new variables, we obtain a new derivation : (iii) : \( P_2(x_1, \ldots, x_p) \leadsto G'_1 G'_2 P_3(x'_1, \ldots, x'_p) \) and there exists \( i, k \) such that \( \sigma(x_j) = x'_j \) (at least one non-variable term in the (ii) derivation), \( \sigma(x_k) = t_k \), and \( x'_k \) is a variable of \( t_k \). We conclude that \( \text{Prog} \) is empty-recursive.

Deciding the empty-recursiveness of a CS-program seems to be a difficult problem (undecidable?). Nevertheless, we propose a sufficient syntactic condition to ensure that a CS-program is not empty-recursive.

**Definition 21.** The clause \( P(t_1, \ldots, t_n) \leftarrow A_1, \ldots, Q(\ldots), \ldots, A_m \) is pseudo-empty over \( Q \) if there exist \( i, j \) s.t.

\( t_i \) is a variable,

\( \text{and} \ t_j \) is not a variable,

\( \text{and} \ \exists x \in \text{Var}(t_j), x \neq t_i \land \{x, t_i\} \subseteq \text{Var}(Q(\ldots)). \)

Roughly speaking, when making a resolution step issued from the flat atom \( P(y_1, \ldots, y_n) \), the variable \( y_i \) is not instantiated, and \( y_j \) is instantiated by something that is synchronized with \( y_j \) (in \( Q(\ldots) \)).

The clause \( H \leftarrow B \) is pseudo-empty if there exists some \( Q \) s.t. \( H \leftarrow B \) is pseudo-empty over \( Q \).

The CS-clause \( P(t_1, \ldots, t_n) \leftarrow A_1, \ldots, Q(x_1, \ldots, x_k), \ldots, A_m \) is empty over \( Q \) if for all \( x_i \), \( \exists j, t_j = x_i \) or \( x_i \not\in \text{Var}(P(t_1, \ldots, t_n)) \).

**Example 22.** The CS-clause \( P(x, f(x), z) \leftarrow Q(x, z) \) is both pseudo-empty (thanks to the second and the third argument of \( P \)) and empty over \( Q \) (thanks to the first and the third argument of \( P \)).

**Definition 23.** Using Definition 21, let us define two relations over predicate symbols.

\( P_1 \geq_{\text{Prog}} P_2 \) if there exists in \( \text{Prog} \) a clause empty over \( P_2 \) of the form \( P_1(\ldots) \leftarrow A_1, \ldots, P_2(\ldots), \ldots, A_m \). The reflexive-transitive closure of \( \geq_{\text{Prog}} \) is denoted by \( \geq^+_{\text{Prog}} \).

\( P_1 >_{\text{Prog}} P_2 \) if there exist in \( \text{Prog} \) predicates \( P'_1, P'_2 \) s.t. \( P'_1 \geq^+_{\text{Prog}} P'_2 \) and \( \text{Prog} \) a clause pseudo-empty over \( P'_2 \) of the form \( P'_1(\ldots) \leftarrow A_1, \ldots, P'_2(\ldots), \ldots, A_m \). The transitive closure of \( >_{\text{Prog}} \) is denoted by \( >^+_{\text{Prog}} \).

\( >_{\text{Prog}} \) is cyclic if there exists a predicate \( P \) s.t. \( P >^+_{\text{Prog}} P \).

**Example 24.** Let \( \Sigma = \{f^\land, h^\land, a^\land\} \) Let \( \text{Prog} \) be the following CS-program:

\( \text{Prog} = \{P(x, h(y), f(z)) \leftarrow Q(x, z), R(y). \ Q(x, g(y, z)) \leftarrow P(x, y, z). \ R(a) \leftarrow. \ Q(a, a) \leftarrow. \} \)

One has \( P >_{\text{Prog}} Q \) and \( Q >_{\text{Prog}} P \). Thus, \( >_{\text{Prog}} \) is cyclic.

The lack of cycles is the key point of our technique since it ensures the finiteness of the number of critical pairs.
Lemma 25. If \( \triangleright_{Prog} \) is not cyclic, then \( \text{Prog} \) is not empty-recursive, consequently the number of critical pairs is finite.

Proof. By contrapositive. Let us suppose that \( \text{Prog} \) is empty recursive. So there exist \( P \) and distinct variables \( x_1, \ldots, x_n \) s.t. \( P(x_1, \ldots, x_n) \sim^+_\sigma A_1, \ldots, P(x'_1, \ldots, x'_n), \ldots, A_k \) where \( x'_1, \ldots, x'_n \) are variables and there exist \( i, j \) s.t. \( x'_i = \sigma(x_i) \) and \( \sigma(x_j) \) is not a variable and \( x'_j \in \text{Var}(\sigma(x_j)) \). We can extract from the previous derivation the following derivation which has \( p \) steps (\( p \geq 1 \)). \( P(x_1, \ldots, x_n) = Q^0(x_1, \ldots, x_n) \sim_{\alpha_1} B_1^1 \ldots Q^1(x_1, \ldots, x_n) \ldots B_1^1 \sim_{\alpha_2} B_1^1 \ldots B_1^2 \ldots Q^2(x_1', \ldots, x'_n) \ldots B_1^2 \sim_{\alpha_3} \ldots \sim_{\alpha_p} B_1^p \ldots B_1^p \ldots Q^0(x'_1, \ldots, x'_n) \ldots B_1^p \ldots B_{k_1}^1 \) where \( Q^p(x_1, \ldots, x_n) = P(x'_1, \ldots, x'_n) \).

For each \( k, \alpha_k \circ \alpha_{k-1} \ldots \circ \alpha_1(x_i) \) is a variable of \( Q^k(x_1^k, \ldots, x_n^k) \) and \( \alpha_k \circ \alpha_{k-1} \ldots \circ \alpha_1(x_j) \) is either a variable of \( Q^k(x_1^k, \ldots, x_n^k) \) or a non-variable term containing a variable of \( Q^k(x_1^k, \ldots, x_n^k) \).

Each derivation step issued from \( Q^k \) uses either a clause pseudo-empty over \( Q^k+1 \) and we deduce \( Q^k \triangleright_{Prog} Q^k+1 \), or an empty clause over \( Q^k+1 \) and we deduce \( Q^k \leq_{Prog} Q^k+1 \). At least one step uses a pseudo-empty clause otherwise no variable from \( x_1, \ldots, x_n \) would be instantiated by a non-variable term containing at least one variable in \( x'_1, \ldots, x'_n \). We conclude that \( P = Q^0 \circ_{p_1} Q^1 \circ_{p_2} Q^2 \ldots \circ_{p_{n-1}} Q^p = P \) with each \( p_i \) is \( \triangleright_{Prog} \) or \( \leq_{Prog} \) and there exists \( k \) such that \( p_k \) is \( \triangleright_{Prog} \). Therefore \( P \triangleright_{Prog} P \), so \( \text{Prog} \) is cyclic. ▶

So, if a CS-program \( \text{Prog} \) does not involve \( \triangleright_{Prog} \) to be cyclic, then all is fine. Otherwise, we have to transform \( \text{Prog} \) into another CS-program \( \text{Prog}' \) such as \( \triangleright_{Prog}' \) is not cyclic and \( \text{Mod}(\text{Prog}) \subseteq \text{Mod}(\text{Prog}') \).

The transformation is based on the following observation. If \( \triangleright_{Prog} \) is cyclic, there is at least one pseudo-empty clause over a given predicate that participates in a cycle. Note that this remark can be checked in Example 24 where \( P(x, h(y), f(z)) \leftarrow Q(x, z), R(y) \) is a pseudo-empty clause over \( Q \) involving the cycle. To remove cycles, we transform some pseudo-empty clauses into clauses that are not pseudo-empty anymore. It boils down to unsynchronize some variables. The process is mainly described in Definition 28. Definitions 26 and 27 are intermediary definitions involved in Definition 28.

Definition 26 (simplify). Let \( H \leftarrow A_1, \ldots, A_n \) be a CS-clause, and for each \( i \), let us write \( A_i = P_i(\ldots) \). If there exists \( P_i \) s.t. \( L(P_i) = \emptyset \) then simplify \( (H \leftarrow A_1, \ldots, A_n) \) is the empty set, otherwise it is the set that contains only the clause \( H \leftarrow B_1, \ldots, B_m \) such that
- \( \{B_i \mid 0 \leq i \leq m\} \subseteq \{A_i \mid 0 \leq i \leq n\} \) and
- \( \forall i \in \{1, \ldots, n\}, (\neg \exists j, B_j = A_i) \Leftrightarrow \text{Var}(A_i) \cap \text{Var}(H) = \emptyset \).

In other words, simplify deletes unproductive clauses, or it removes the atoms of the body that contain only extra-variables.

Definition 27 (unSync). Let \( P(t_1, \ldots, t_n) \leftarrow B \) be a pseudo-empty CS-clause.

\[ \text{unSync}(P(t_1, \ldots, t_n) \leftarrow B) = \text{simplify}(P(t_1, \ldots, t_n) \leftarrow \sigma_0(B), \sigma_1(B)) \] where \( \sigma_0, \sigma_1 \) are substitutions built as follows:

\[ \sigma_0(x) = \begin{cases} x & \text{if } \exists i, \ t_i = x \\ \text{a fresh variable otherwise} \end{cases} \quad \sigma_1(x) = \begin{cases} x & \text{if } \exists i, \ t_i \not\in \text{Var} \land x \in \text{Var}(t_i) \\ \neg(\exists j, \ t_j = x) & \text{a fresh variable otherwise} \end{cases} \]

Definition 28 (removeCycles). Let \( \text{Prog} \) be a CS-program.

\[ \text{removeCycles}(\text{Prog}) = \begin{cases} \text{Prog} & \text{if } \triangleright_{\text{Prog}} \text{ is not cyclic} \\ \text{removeCycles}(\text{unSync}(H \leftarrow B)) \cup \text{Prog}' & \text{otherwise} \end{cases} \]

where \( H \leftarrow B \) is a pseudo-empty clause involved in a cycle and \( \text{Prog}' = \text{Prog} \setminus \{H \leftarrow B\} \).
Lemma 30. Let Prog be a CS-program and Prog' = removeCycles(Prog). Then >_{Prog'} is not cyclic, and Mod(Prog) ⊆ Mod(Prog'). Moreover, if Prog is normalized and preserving, then so is Prog'.

Proof. Proof are detailed in [3].

At this point, given a CS-program Prog, if >_{Prog} is not cyclic then the number of critical pairs is finite. Otherwise, we transform Prog into another CS-program Prog' in such a way that >_{Prog'} is not cyclic and Mod(Prog) ⊆ Mod(Prog'). Since Prog' is not cyclic, the finiteness of the number of critical pairs is ensured.

3.3 Normalizing critical pairs

In Section 3.1, we have defined the notion of critical pair and we have shown in Theorem 14 that this notion is useful for a matter of rewriting closure. Moreover, as mentioned at the very beginning of Section 3, non-convergent critical pairs correspond to the CS-clauses that we would like to add in the current CS-program. Unfortunately, these CS-clauses are not necessarily in the expected form (normalized).

Definition 34 describes the normalization process that transforms a non-normalized CS-clause into several normalized ones. For example, consider the non-normalized CS-clause $P(f(g(x)), b) \leftarrow P'(x)$. We want to generate a set of normalized CS-clauses covering at least the same Herbrand model. The following set of CS-clauses \{ $P(f(x_1), b) \leftarrow P_{new_1}(x_1)$. $P_{new_1}(g(x_1)) \leftarrow P'(x_1)$ \} is a good candidate with $P_{new_1}$ a new predicate symbol.

Definition 31 introduces tools for manipulating parameters of predicates (tuple of terms). Definition 32 formalizes a way for cutting a clause head, at depth 1. An example is given after Definition 34.

Definition 31. A tree-tuple $(t_1, \ldots, t_n)$ is normalized if for all $i$, $t_i$ is a variable or contains only one function-symbol.

We define tuple concatenation by $(t_1, \ldots, t_n), (s_1, \ldots, s_k) = (t_1, \ldots, t_n, s_1, \ldots, s_k)$.

The arity of the tuple $(t_1, \ldots, t_n)$ is $ar(t_1, \ldots, t_n) = n$.

Definition 32. Consider a tree-tuple $\overrightarrow{t} = (t_1, \ldots, t_n)$. We define:

$\overrightarrow{t}^{cut} = (t_1^{cut}, \ldots, t_n^{cut})$, where $t_i^{cut} = \begin{cases} x_{i,1} & \text{if } t_i \text{ is a variable} \\ t_i & \text{if } t_i \text{ is a constant} \\ t_i(e)(x'_{i,1}, \ldots, x'_{i,ar(t_i(e))}) & \text{otherwise} \end{cases}$

and variables $x'_{i,k}$ are new variables that do not occur in $\overrightarrow{t}$. 

Example 29. Let Prog be the CS-program of Example 24. Since Prog is cyclic, let us compute removeCycles(Prog). The pseudo-empty CS-clause $P(x, h(y), f(z)) \leftarrow Q(x, z), R(y)$ is involved in the cycle. Consequently, unSync is applied on it. According to Definition 27, one obtains $\sigma_0$ and $\sigma_1$ where $\sigma_0 = [x/x, y/x_1, z/x_2]$ and $\sigma_1 = [x/x_3, y/y, z/z]$. Thus, one obtains the CS-clause $P(x, h(y), f(z)) \leftarrow Q(x, x_2), R(x_1), Q(x_3, z), R(y)$. Note that according to Definition 27, simplify has to be applied on the CS-clause above-mentioned. Following Definitions 26 and 28, one has to remove $P(x, h(y), f(z)) \leftarrow Q(x, z), R(y)$ from Prog and to add $P(x, h(y), f(z)) \leftarrow Q(x, x_2), Q(x_3, z), R(y)$ instead. Note that the atom $R(x_1)$ has been removed using simplify. Note also that there is no cycle anymore.

Lemma 30 describes that our transformation preserves at least and may extend the initial least Herbrand Model.

Definition 34 formalizes a way for cutting a clause head, at depth 1. An example is given after Definition 34.

Definition 32 introduces tools for manipulating parameters of predicates (tuple of terms).
for each $i$, $\overrightarrow{\text{Var}}(t_i^\text{cut})$ is the (possibly empty) tuple composed of the variables of $t_i^\text{cut}$ (taken in the left-right order).

$\overrightarrow{\text{Var}}(\overrightarrow{\text{cut}}) = \overrightarrow{\text{Var}}(t_1^\text{cut}) \ldots \overrightarrow{\text{Var}}(t_n^\text{cut})$ (concatenation of tuples).

for each $i$, $t_i^\text{rest}$ is the tree-tuple $t_i^\text{rest} = \begin{cases} (t_i) & \text{if } t_i \text{ is a variable} \\ \text{the empty tuple} & \text{if } t_i \text{ is a constant} \\ (t_{1|1}, \ldots, t_{1|\text{arity}(t_i)}) & \text{otherwise} \end{cases}$

$t_i^\text{rest} = (t_1^\text{rest} \ldots t_n^\text{rest})$ (concatenation of tuples).

Example 33. Let $\overrightarrow{t}$ be a tree-tuple such that $\overrightarrow{t} = (x_1, x_2, g(x_3, h(x_1)), h(x_4), b)$ where $x_i$’s are variables. Thus,

$\overrightarrow{t}^\text{cut} = (y_1, y_2, g(y_3, y_4), h(y_5), b)$ with $y_i$’s new variables;

$\overrightarrow{\text{Var}}(\overrightarrow{t}^\text{cut}) = (y_1, y_2, y_3, y_4, y_5)$;

$\overrightarrow{t}^\text{rest} = (x_1, x_2, x_3, h(x_1), x_4)$.

Note that $\overrightarrow{t}^\text{cut}$ is normalized, $\overrightarrow{\text{Var}}(\overrightarrow{t}^\text{cut})$ is linear, $\overrightarrow{\text{Var}}(\overrightarrow{t}^\text{cut})$ and $\overrightarrow{t}^\text{rest}$ have the same arity.

**Notation:** $\text{card}(S)$ denotes the number of elements of the finite set $S$.

Definition 34 (norm). Let $\text{Prog}$ be a normalized CS-program.

Let $\text{Pred}$ be the set of predicate symbols of $\text{Prog}$, and for each positive integer $i$, let $\text{Pred}_i = \{ P \in \text{Pred} \mid \text{ar}(P) = i \}$ where $\text{ar}$ means arity.

Let $\text{arity-limit}$ and $\text{predicate-limit}$ be positive integers s.t. $\forall P \in \text{Pred}$, $\text{arity}(P) \leq \text{arity-limit}$, and $\forall i \in \{1, \ldots, \text{arity-limit}\}$, $\text{card} (\text{Pred}_i) \leq \text{predicate-limit}$. Let $H \leftarrow B$ be a CS-clause.

Function $\text{norm}_{\text{prog}}(H \leftarrow B)$

Res = $\text{Prog}$

If $H \leftarrow B$ is normalized

then Res = Res $\cup \{ H \leftarrow B \}$ (a)

else If $H \rightarrow_{\text{Res}} A$ by a synchronizing and non-empty clause

then (note that $A$ is an atom) Res = $\text{norm}_{\text{Res}}(A \leftarrow B)$ (b)

else let us write $H = P(\overrightarrow{t})$

If $\text{ar}(\overrightarrow{\text{Var}}(\overrightarrow{t}^\text{cut})) \leq \text{arity-limit}$

then let $c'$ be the clause $P(\overrightarrow{t}^\text{cut}) \leftarrow P'(\overrightarrow{\text{Var}}(\overrightarrow{t}^\text{cut}))$

where $P'$ is a new or an existing predicate symbol$^6$

Res = $\text{norm}_{\text{Res} \cup \{c'\}}(P'(\overrightarrow{t}^\text{rest}) \leftarrow B)$ (c)

else choose tuples $\overrightarrow{v_1}, \ldots, \overrightarrow{v_k}$ and tuples $\overrightarrow{t_1}, \ldots, \overrightarrow{t_k}$ s.t.

$\overrightarrow{v_1} \ldots \overrightarrow{v_k} = \overrightarrow{\text{Var}}(\overrightarrow{t}^\text{cut})$ and $\overrightarrow{t_1} \ldots \overrightarrow{t_k} = \overrightarrow{t}^\text{rest}$,

and for all $j$, $\text{ar}(\overrightarrow{v_j}) = \text{ar}(\overrightarrow{t_j})$ and $\text{ar}(\overrightarrow{v_j}) \leq \text{arity-limit}$

let $c'$ be the clause $P(\overrightarrow{t}^\text{cut}) \leftarrow P'_1(\overrightarrow{v_1}), \ldots, P'_k(\overrightarrow{v_k})$

where $P'_1, \ldots, P'_k$ are new or existing predicate symbols$^7$

Res = Res $\cup \{c'\}$

For $j$=1 to $k$ do Res = $\text{norm}_{\text{Res}}(P'_j(\overrightarrow{t_j}) \leftarrow B)$ EndFor (d)

EndIf

EndIf

return Res

---

$^6$ If $\text{card}(\text{Pred}_{\text{ar}(\overrightarrow{v_j})}(\text{Res})) < \text{predicate-limit}$, then $P'$ is new, otherwise $P'$ is arbitrarily chosen in $\text{Pred}_{\text{ar}(\overrightarrow{v_j})}(\text{Res})$.

$^7$ For all $j$, $P'_j$ is new if $\text{card}(\text{Pred}_{\text{ar}(\overrightarrow{v_j})}(\text{Res})) + j - 1 < \text{predicate-limit}$. 
Example 35. Consider the CS-program $Prog =$
\[
\{ P_0(f(x)) \leftarrow P_1(x), \; P_1(a) \leftarrow, \; P_0(u(x)) \leftarrow P_2(x), \; P_2(f(x)) \leftarrow P_3(x), \; P_3(v(x,x)) \leftarrow P_1(x) \}.
\]
Let $arity-limit = 1$ and $predicate-limit = 5$. Let $P_2(u(f(v(x,x)))) \leftarrow P_3(x)$ be a CS-clause to normalize. According to Definition 34, we are not in case (a) nor in (b), we are in case (c).
Then, according to Definition 32, $u(f(v(x,x)))^{cut} = u(x_1)$ with $x_1$ a new variable. Since for now the number of predicates with arity 1 is equal to $4 < predicate-limit$, a new predicate $P_4$ can be created and then one has to add the CS-clause $P_2(u(x_1)) \leftarrow P_4(x_1)$. Then we have to solve the recursive call $\text{norm}_{Prog \cup \{ P_2(u(x_1)) \leftarrow P_4(x_1) \}}(P_4(f(v(x,x)))) \leftarrow P_5(x)$. The same process is applied except for the creation of a new predicate, because $predicate-limit$ would be exceeded. Consequently, no new predicate with arity 1 can be generated. One has to choose an existing one. Let us try with $x_1$.

Let $P_3(v(x,x)) \leftarrow P_3(x)$. Then, according to Definition 32, $P_3(v(x,x)) \leftarrow P_3(x)$ has produced three new clauses, which are $P_2(u(x_1)) \leftarrow P_4(x_1), P_4(f(v(x,x))) \leftarrow P_5(x)$ and $P_3(v(x,x)) \leftarrow P_3(x)$.

Obviously, termination of $\text{norm}$ is guaranteed according to Lemma 36.

Lemma 36. Function $\text{norm}$ always terminates.

Proof. Consider a run of $\text{norm}_{Prog}(H \leftarrow B)$, and any recursive call $\text{norm}_{Prog}(H' \leftarrow B')$. We can see that $|H'|_{\Sigma} < |H|_{\Sigma}$. Consequently a normalized clause is necessarily reached, and there is no recursive call in this case.

Given a normalized CS-program $Prog$, Theorem 37 raises two important points:
1. given a non-normalized clause $H \leftarrow B$, one obtains $\delta H \rightarrow_{\text{norm}_{Prog}(H \leftarrow B)} B$, and 2. adding the CS-clauses provided by $\text{norm}$ into $Prog$ may increase the least Herbrand model of $Prog$.

Theorem 37. Let $c$ be a critical pair in $Prog$. Then $c$ is convergent in $\text{norm}_{Prog}(c)$.
Moreover for any CS-clause $c'$, we have $\text{Mod}(Prog \cup \{c'\}) \subseteq \text{Mod}(\text{norm}_{Prog}(c'))$.

Proof. The second item of the theorem is a consequence of the first item.
Let us now prove the first item. Let $c = (H \leftarrow B)$ and let us prove that $H \rightarrow_{\text{res}} B$. The proof is by induction on recursive calls to Function $\text{norm}$ (we write $\text{ind-hyp}$ for “induction hypothesis”). We consider items (a), (b),... in Definition 34:

(a) From Lemma 30.
(b) We have $H \rightarrow A \rightarrow^*_{\text{ind-hyp}} B$.
(c) $H = P(\overrightarrow{t}) \rightarrow c' P'(\overrightarrow{t'}_{\text{rest}}) \rightarrow_{\text{ind-hyp}} B$.
(d) $H = P(\overrightarrow{t}) \rightarrow c' (P_1'(\overrightarrow{t}^1_{\text{rest}}), ..., P_k'(\overrightarrow{t}^k_{\text{rest}})) \rightarrow^*_{\text{ind-hyp}} (B, ..., B)$ (up to variable renamings).

3.4 Completion
In Sections 3.1 and 3.3, we have described how to detect critical pairs and how to convert them into normalized clauses. Moreover, in a given finite CS-program the number of critical pairs is finite as shown in Section 3.2. Definition 38 explains precisely our technique for computing over-approximation using a CS-program completion.
Definition 38 (comp). Let $R$ be a left-linear rewrite system, and $Prog$ be a finite and normalized CS-program s.t.
- $R_{prog}$ is not cyclic (otherwise apply removeCycles to remove cycles),
- and $\forall P \in Pred$, $arity(P) \leq arity-limit$,
- and $\forall i \in \{1, \ldots, arity-limit\}$, $card(Pred_i) \leq predicate-limit$.
where $card(Pred_i)$ is the number of predicate symbols of arity $i$.

Function $\text{comp}_R(Prog)$

while there exists a non-convergent critical pair $H \leftarrow B$ do
    $Prog = \text{removeCycles}(\text{norm}_{\text{prog}}(H \leftarrow B))$
end while
return $Prog$

Theorem 39. Function $\text{comp}$ always terminates, and all critical pairs are convergent in $\text{comp}_R(Prog)$. Moreover $\text{Mod}(Prog) \subseteq \text{Mod}(\text{comp}_R(Prog))$.

Proof. Proofs are detailed in [3].

Moreover, thanks to Theorem 14, $\text{Mod}(\text{comp}_R(Prog))$ is closed under rewriting by $R$. Then:

Corollary 40. If in addition $Prog$ is preserving, $R^*(\text{Mod}(Prog)) \subseteq \text{Mod}(\text{comp}_R(Prog))$.

4 Examples

In this section, our technique is applied on several examples. In Examples 41, 42 and 43, $I$ is the initial set of terms and $R$ is the rewrite system. Moreover, initially, we define a CS-program $Prog$ that generates $I$.

Example 41. In this example, we define $\Sigma$ as follows: $\Sigma = \{e^{1,2}, a^{0}\}$. Let $I$ be the set of terms $I = \{f(t) \mid t \in T_{\Sigma}\}$. Let $R$ be the rewrite system $R = \{f(x) \rightarrow b(x, x)\}$. Obviously, one can easily guess that $R^*(I) = \{b(t, t) \mid t \in T_{\Sigma}\} \cup I$. Note that $R^*(I)$ is not a regular, nor a context-free language [1, 13].

Initially, $Prog = \{P_0(f(x)) \leftarrow P_1(x), P_1(c(x, y)) \leftarrow P_1(x), P_1(y), P_1(a) \leftarrow \}$. Using our approach, the critical pair $P_0(b(x, x)) \leftarrow P_1(x)$ is detected. This critical pair is already normalized, then it is immediately added into $Prog$. Then, there is no more critical pair and the procedure stops. Note that we get exactly the set of descendants, i.e. $L(P_0) = R^*(I)$. So, given $t, t' \in T_{\Sigma}$ such that $t \neq t'$, one can show that $b(t, t') \not\in R^*(I)$.

The example right above shows that non-context-free descendants can be handled in a conclusive manner with our approach. Such example cannot be handled by [14] in an exact way, because they use context-free languages. Actually, the classes of languages covered by our approach and theirs are in some sense orthogonal. However, the examples below shows that our approach can also be relevant for other problems.

Example 42.

Let $I$ be the set of terms $I = \{f(a, a)\}$, and $R$ be the rewrite system $R = \{f(x, y) \rightarrow u(f(v(x), v(y)))\}$. Intuitively, the exact set of descendants is $R^*(I) = \{w^n(f(u^n(a), w^n(a))) \mid n \in \mathbb{N}\}$. We define $Prog = \{P_0(f(x, y)) \leftarrow P_1(x), P_1(y), P_1(a) \leftarrow \}$. We choose $\text{predicate-limit} = 4$ and $\text{arity-limit} = 2$. 

...
Over-approximating Descendants by Synchronized Tree Languages

First, the following critical pair is detected: \( P_0(u(f(v(x), w(y)))) \leftarrow P_1(x), P_1(y) \). According to Definition 34, the normalization of this critical pair produces three new CS-clauses: \( P_0(u(x)) \leftarrow P_2(x) \), \( P_2(f(x, y)) \leftarrow P_3(x, y) \) and \( P_3(v(x), w(y)) \leftarrow P_1(x), P_1(y) \). Adding these three CS-clauses into \( \text{Prog} \) produces the new critical pair \( P_2(u(f(v(x), w(y)))) \leftarrow P_3(x, y) \). This critical pair can be normalized without exceeding predicate-limit. So, we add: \( P_2(u(x)) \leftarrow P_4(x) \). \( P_4(f(x, y)) \leftarrow P_5(x, y) \). and \( P_5(v(x), w(y)) \leftarrow P_3(x, y) \).

Once again, a new critical pair has been introduced: \( P_4(u(f(v(x), w(y)))) \leftarrow P_5(x, y) \). Note that, from now, we are not allowed to introduce any new predicate of arity 1. Let us proceed the normalization of \( P_4(u(f(v(x), w(y)))) \leftarrow P_5(x, y) \) step by step. We choose to re-use the predicate \( P_4 \). Thus, we first generate the following CS-clause: \( P_4(u(x)) \leftarrow P_4(x) \). So, we have to normalize now \( P_4(f(v(x), w(y))) \rightarrow P_5(x, y) \). Note that \( P_4(f(v(x), w(y))) \rightarrow P_5(x, y) \) is added into \( \text{Prog} \).

Note that there is no critical pair anymore.

To summarize, we obtain the final CS-program \( \text{Prog}_f \) composed of the following CS-clauses:

\[
\begin{align*}
\text{Prog}_f = \{ & P_0(f(x, y)) \leftarrow P_1(x), P_1(y), P_1(a) \leftarrow P_1(u(x)) \leftarrow P_2(x) \\
& P_2(f(x, y)) \leftarrow P_3(x, y), P_2(v(x), w(y)) \leftarrow P_3(x, y), P_1(y), P_2(u(x)) \leftarrow P_2(x), \ \\
& P_3(f(x, y)) \leftarrow P_5(x, y), P_3(v(x), w(y)) \leftarrow P_5(x, y), P_4(v(x), w(y)) \leftarrow P_5(x, y), P_4(u(x)) \leftarrow P_4(x) \}
\end{align*}
\]

For \( \text{Prog}_f \), note that \( L(P_0) = \{ u^n(f(v^m(a), w^m(a))) \mid n, m \in \mathbb{N} \} \) and \( R^*(I) \subseteq L(P_0) \).

In Example 42, the approximation computed is still a non-regular language. Nevertheless, it is a strict over-approximation since a synchronization is broken between the three counters.

Let us also show the application of our technique on an example introduced in [5]. In [5] authors propose an example that cannot be handled by regular approximations. Example 43 shows that this limitation can now be overcome.

**Example 43.** Let \( I \) be the set of terms \( I = \{ f(a, a) \} \) and \( R \) be the rewrite system \( R = \{ f(x, y) \rightarrow f(g(x), g(y)), f(g(x), g(y)) \rightarrow f(x, y), f(a, g(a)) \rightarrow \text{error} \} \). Obviously, \( R^*(I) = \{ f(g^n(a), g^n(a)) \mid n \in \mathbb{N} \} \). Consequently, \( \text{error} \) is not a reachable term.

We start with the CS-program \( \text{Prog} = \{ P_0(f(x, y)) \leftarrow P_1(x), P_1(y), P_1(a) \leftarrow \} \). After applying Function \( \text{comp} \), we obtain the following CS-program for any predicate-limit \( \geq 2 \):

\[
\begin{align*}
\text{Prog}_f = \{ & P_0(f(x, y)) \leftarrow P_1(x), P_1(y), P_0(f(x, y)) \leftarrow P_2(x, y), P_0(f(x, y)) \leftarrow P_2(x, y), P_2(x, y) \leftarrow P_2(x, y), P_1(a) \leftarrow \}
\end{align*}
\]

Note that \( L(P_0) \) is exactly \( R^*(I) \). Note also that \( \text{error} \not \in L(P_0) \). Consequently, we have proved that \( \text{error} \) is not reachable from \( I \).

5 **Further Work**

We have presented a procedure that always terminates and that computes an over-approximation of the set of descendants, expressed by a synchronized tree language. This is the first attempt using synchronized tree languages. It could be improved or extended:

- In Definition 34, when predicate-limit is reached (in items (c) and (d)), an (several in item (d)) existing predicate of the right arity is chosen arbitrarily and re-used, instead of creating a new one. Of course, if there are several existing predicates of the right arity, the achieved choice affects the quality of the approximation. When using regular languages [8], a similar difficulty happens: to make the procedure terminate, it is sometimes necessary to choose and re-use an existing state instead of creating a new one. Some ideas have been proposed to make this choice in a smart way [11]. We are going to extend these ideas in order to improve the choice of existing predicates.
A similar problem arises when \textit{arity-limit} is reached (item (d)): a tuple is divided into several smaller tuples in an arbitrary way, and there may be several possibilities, which may affect the quality of the approximation.

To compute descendants, we have used synchronized tree languages, whereas context-free languages have been used in [14]. Each approach has advantages and drawbacks. Therefore, it would be interesting to mix the two approaches to get the advantages of both.

\section*{References}
\begin{itemize}
\end{itemize}