Parallel Complexity of Term Rewriting Systems

Thaïs Baudon
ENS de Lyon & LIP (UMR CNRS/ENS Lyon/UCB Lyon1/INRIA), Lyon, France
Carsten Fuhs
Birkbeck, University of London, United Kingdom
Laure Gonnord
University of Lyon & LIP (UMR CNRS/ENS Lyon/UCB Lyon1/INRIA), Lyon, France

Abstract

In this workshop paper, we revisit the notion of parallel-innermost term rewriting. We provide a definition of parallel complexity and propose techniques to derive upper bounds on this complexity via the Dependency Tuple framework by Noschinski et al.

2012 ACM Subject Classification Theory of computation → Program verification, Rewrite systems; Software and its engineering → Automated static analysis, Formal software verification

Keywords and phrases Complexity analysis, Parallelism, Rewriting

1 Introduction

In this paper, we consider the problem of evaluating the potentiality of parallelisation in pattern-matching based recursive functions like the one depicted in Figure 1.

```rust
def size(&self) -> int {
    match self {
    &Tree::Node { v, ref left, ref right } => left.size() + right.size() + 1,
    &Tree::Empty => 0 ,
    }
}
```

Figure 1 Tree size computation in Rust

In this particular example, the recursive calls to `left.size()` and `right.size()` can be done in parallel. Building on previous work on parallel-innermost rewriting [6, 4], and first ideas about parallel complexity [1], we propose a new notion of Parallel Dependency Tuples that capture such a behaviour, and a method to compute parallel complexity bounds.

2 Parallel-innermost Term Rewriting

The following definitions are mostly standard [3].

**Definition 1 (Term rewrite system, innermost rewriting).** $\mathcal{T}(\Sigma, \mathcal{V})$ denotes the set of terms over a finite signature $\Sigma$ and the set of variables $\mathcal{V}$. For a term $t$, the set $\text{Pos}(t)$ of its positions is defined inductively as a set of strings of positive integers: (a) if $t \in \mathcal{V}$, then $\text{Pos}(t) = \{\varepsilon\}$, and (b) if $t = f(t_1, \ldots, t_n)$, then $\text{Pos}(t) = \{\varepsilon\} \cup \cup_{1 \leq i \leq n} (\pi | \pi \in \text{Pos}(t_i))$. The position $\varepsilon$ is called the root position of term $t$. The (strict) prefix order $\prec$ on positions is the strict partial order given by: $\pi \prec \tau$ iff there exists $\pi' \neq \varepsilon$ such that $\pi \pi' = \tau$. For $\pi \in \text{Pos}(t)$, $t|\pi$ is the subterm of $t$ at position $\pi$, and we write $t[s]|\pi$ for the term that results from $t$ by replacing the subterm $t|\pi$ at position $\pi$ by the term $s$.

* This work was supported by ANR project CODAS.
2 Parallel Complexity of Term Rewriting Systems

For a term \( t \), \( \mathcal{V}(t) \) is the set of variables in \( t \). If \( t \) has the form \( f(t_1, \ldots, t_n) \), \( \text{root}(t) = f \) is the root of \( t \). A term rewrite system (TRS) \( \mathcal{R} \) is a set of rules \( \{ r_1 \rightarrow t_1, \ldots, r_n \rightarrow t_n \} \) with \( r_i, t_i \in \mathcal{T}(\Sigma, \mathcal{V}) \), \( r_i \notin \mathcal{V} \), and \( \mathcal{V}(t_i) \subseteq \mathcal{V}(r_i) \) for all \( 1 \leq i \leq n \). The rewrite relation of \( \mathcal{R} \) is \( s \rightarrow t \) if there is a rule \( \ell \rightarrow r \in \mathcal{R} \), a position \( \pi \in \text{Pos}(s) \), and a substitution \( \sigma \) such that \( s = s[\ell \sigma]_{\pi} \) and \( t = s[r \sigma]_{\pi} \). Here, \( \sigma \) is called the matcher and the term \( \ell \sigma \) is called the redex of the rewrite step. If \( \ell \sigma \) has no proper subterm that is also a possible redex, \( \ell \sigma \) is an innermost redex, and the rewrite step is an innermost rewrite step denoted by \( s \xrightarrow{\ell \sigma} t \).

\[ \Sigma_d^R = \{ f | f(t_1, \ldots, t_n) \rightarrow r \in \mathcal{R} \} \text{ and } \Sigma_c^R = \Sigma \setminus \Sigma_d^R \text{ are the defined and constructor symbols of } \mathcal{R}. \]

We may omit the superscript and just write \( \Sigma_d \) and \( \Sigma_c \) if \( \mathcal{R} \) is not of importance or clear from the context. Finally, let \( \text{Pos}_d(t) = \{ \pi | \pi \in \text{Pos}(t), \text{root}(t) = r \in \Sigma_d \} \).

The notion of parallel-innermost rewriting dates back at least to [6]. Informally, in a parallel-innermost rewrite step, all innermost redexes are rewritten simultaneously. This corresponds to executing all function calls in parallel on a machine with unbounded parallelism.

**Definition 2 (Parallel-innermost rewriting [4]).** A term \( s \) rewrites innermost in parallel to \( t \) with a TRS \( \mathcal{R} \), written \( s \xrightarrow{\ell^R} t \), if \( s \xrightarrow{\ell^R} t \), and either (a) \( s \xrightarrow{\ell^R} t \) with \( s \) an innermost redex, or (b) \( s = f(s_1, \ldots, s_n) \), \( t = f(t_1, \ldots, t_n) \), and for all \( 1 \leq k \leq n \) either \( s_k \xrightarrow{\ell^R} t_k \) or \( s_k = t_k \) is a normal form.

**Example 3 (size).** Consider the TRS \( \mathcal{R} \) with the following rules modelling the code of Figure 1.

\[
\begin{align*}
\text{plus}(\text{Zero}, y) & \rightarrow y & \text{size}(\text{Nil}) & \rightarrow \text{Zero} \\
\text{plus}(S(x), y) & \rightarrow S(\text{plus}(x, y)) & \text{size}(\text{Tree}(v, l, r)) & \rightarrow S(\text{size}(l), \text{size}(r))
\end{align*}
\]

Here \( \Sigma_d^R = \{ \text{plus}, \text{size} \} \) and \( \Sigma_c^R = \{ \text{Zero}, S, \text{Nil}, \text{Tree} \} \). We have the following parallel innermost rewrite sequence, where innermost redexes are underlined:

\[
\begin{align*}
\text{size}(\text{Tree}(\text{Zero}, \text{Nil}, \text{Tree}(\text{Zero}, \text{Nil}, \text{Nil}))) \\
\xrightarrow{\ell^R} S(\text{size}(\text{Tree}(\text{Zero}, \text{Nil}, \text{Nil}))) \\
\xrightarrow{\ell^R} S(\text{size}(\text{Tree}(\text{Zero}, \text{Nil}, \text{Nil}))) \\
\xrightarrow{\ell^R} S(\text{size}(\text{Tree}(\text{Zero}, \text{Nil}, \text{Nil})))
\end{align*}
\]

Here, in the second and in the third step, two innermost steps each are happening in parallel. A corresponding regular innermost rewrite sequence without parallel evaluation of redexes would have needed two more steps.

## 3 Finding Upper Bounds for Parallel Complexity

### 3.1 Notion of Parallel Complexity

We extend the notion of innermost runtime complexity to parallel-innermost rewriting.

**Definition 4 ((Parallel) Innermost Runtime Complexity).** The size \( |t| \) of a term \( t \) is \( |x| = 1 \) if \( x \in \mathcal{V} \) and \( |f(t_1, \ldots, t_n)| = 1 + \sum_{i=1}^n |t_i| \), otherwise. The derivation height of a term \( t \) w.r.t. a relation \( \rightarrow \) is the length of the longest sequence of \( \rightarrow \)-steps from \( t \): \( \text{dh}(t, \rightarrow) = \sup \{ e | \exists t' \in \mathcal{T}(\Sigma, \mathcal{V}), t \rightarrow^e t' \} \) where \( \rightarrow^e \) is the \( e \)th iterate of \( \rightarrow \). If \( t \) starts an infinite \( \rightarrow \)-sequence, we write \( \text{dh}(t, \rightarrow) = \omega \).

A term \( f(t_1, \ldots, t_k) \) is basic (for a TRS \( \mathcal{R} \)) if \( f \in \Sigma_d^R \) and \( t_1, \ldots, t_k \in \mathcal{T}(\Sigma_c^R, \mathcal{V}) \). \( \mathcal{T}_\text{basic}^R \) is the set of basic terms for a TRS \( \mathcal{R} \). For \( n \in \mathbb{N} \), we define the innermost runtime complexity function \( \text{irc}_R(n) = \sup \{ \text{dh}(t, \rightarrow) | t \in \mathcal{T}_\text{basic}^R, |t| \leq n \} \) and we introduce the parallel innermost runtime complexity function \( \text{irc}_R^\parallel(n) = \sup \{ \text{dh}(t, \rightarrow) | t \in \mathcal{T}_\text{basic}^R, |t| \leq n \} \).
In the following, given a TRS $\mathcal{R}$, our goal shall be to infer (asymptotic) upper bounds for $\text{ir} \mathcal{R}$ fully automatically. As usual for runtime complexity, we are considering only basic terms as start terms, corresponding to a defined function called on data objects as arguments. An upper bound for (sequential) $\text{ir} \mathcal{R}$ is also an upper bound for $\text{ir} \mathcal{R} \mathcal{R}$. We will introduce techniques to find upper bounds for $\text{ir} \mathcal{R} \mathcal{R}$ that are strictly tighter than these trivial bounds.

3.2 Complexity: the sequential case

We build on the Dependency Tuple framework [5], originally introduced to determine upper bounds for (sequential) innermost runtime complexity. A central idea is to group all function calls by a rewrite rule together rather than to regard them separately (as for termination [2]).

Definition 5 (Sharp Terms $\mathcal{T}^2$). For every $f \in \Sigma_d$, we introduce a fresh symbol $f^2$ of the same arity. For a term $t = f(t_1, \ldots, t_n)$ with $f \in \Sigma_d$, we define $\ell^2 = f^2(t_1, \ldots, t_n)$ and let $\mathcal{T}^2 = \{\ell^2 \mid \ell \in \mathcal{T}(\Sigma, \mathcal{V}), \text{root}(\ell) \in \Sigma_d\}$.

To compute an upper bound for sequential complexity, we “count” how often each rewrite rule is used. The idea is that the cost of the function call to the lhs of a rule is $1 + \text{sum}$ of the costs of all the function calls in the rhs, counted separately. To group $k$ function calls together, we use “compound symbols” $\text{Com}_k$, which intuitively represent the sum of the runtimes of their arguments. Then, we can use polynomial interpretations $\mathcal{P}$ with $\mathcal{P}(\text{Com}_k(x_1, \ldots, x_k)) = x_1 + \cdots + x_k$ for all $k$ to compute a complexity bound [5, Thm. 27].

Definition 6 (Dependency Tuple, DT [5]). A dependency tuple (DT) is a rule of the form $\mathsf{s} \mathsf{r} \rightarrow \mathsf{Com}_n(\mathsf{t}_1^1, \ldots, t_n)$ where $\mathsf{s}, \mathsf{t}_1^1, \ldots, \mathsf{t}_n \in \mathcal{T}^2$. Let $\ell \rightarrow r$ be a rule with $\text{Pos}_d = \{\pi_1, \ldots, \pi_n\}$ and $\pi_1 \prec \ldots \prec \pi_n$ for a total order $\prec$ on positions. Then $\mathcal{DT}(\ell \rightarrow r) = \mathcal{DT}(\ell \rightarrow r) + \mathcal{Com}_n(r_{\pi_1}^1, \ldots, r_{\pi_n}^1)$.

Example 7. For our running example, we get the following DTs:

<table>
<thead>
<tr>
<th>Function</th>
<th>DT</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathsf{size}^2(\mathsf{Tree}(v, l, r))$</td>
<td>$\mathcal{Com}_3(\mathsf{size}^2(l), \mathsf{size}^2(r), \mathsf{plus}^2(\mathsf{size}(l), \mathsf{size}(r)))$</td>
</tr>
</tbody>
</table>

The given polynomial interpretation, which orients all DTs with $\rightarrow$ and all rules from $\mathcal{R}$ with $\rightarrow$, proves $\text{ir} \mathcal{R}(n) \in \mathcal{O}(n^2)$: $\mathcal{P}(\mathsf{plus}^2(x_1, x_2)) = \mathcal{P}(\mathsf{size}(x_1)) = x_1$, $\mathcal{P}(\mathsf{size}^2(x_1)) = 2x_1 + x_1^2$, $\mathcal{P}(\mathsf{plus}^2(\mathsf{size}(x_1), \mathsf{size}(x_2))) = x_1 + x_2$, $\mathcal{P}(\mathsf{Tree}(x_1, x_2, x_3)) = 1 + x_2 + x_3$, $\mathcal{P}(\mathsf{S}(x_1)) = 1 + x_1$, $\mathcal{P}(\mathsf{Zero}) = \mathcal{P}(\mathsf{Nil}) = 1$.

3.3 Computing Upper Bounds for Parallel Rewriting

To find upper bounds for runtime complexity of parallel-innermost rewriting, we can reuse the notion of DTs from Def. 6 for sequential innermost rewriting along with existing techniques [5] and implementations. We illustrate this in the following example.

Example 8. In the recursive $\mathsf{size}$-rule, the two calls to $\mathsf{size}(l)$ and $\mathsf{size}(r)$ happen in parallel (this will be captured by the notion of structural independency). Thus, the cost for these two calls is not the sum, but the maximum of the calls. Regardless of which of these two calls has the higher cost, we still need to add the cost for the call to $\mathsf{plus}$, which starts evaluating only after both calls to $\mathsf{size}$ have finished. With $\sigma$ as the used matcher for the rule, we have:

$$\text{dh}(\mathsf{size}(\mathsf{Tree}(v, l, r)) \sigma, \|\rightarrow R) = 1 + \max(dh(\mathsf{size}(l) \sigma, \|\rightarrow R), dh(\mathsf{size}(r) \sigma, \|\rightarrow R)) + \mathsf{dh}(\mathsf{plus}(\mathsf{size}(l), \mathsf{size}(r)) \sigma, \|\rightarrow R)$$
Equivalently, we can “factor in” the cost of calling \texttt{plus} into the maximum function:

$$\text{dh}(\text{size}(\text{Tree}(v, l, r))[\sigma], \parallel\rightarrow R) = \max(1 + \text{dh}(\text{size}(l)[\sigma], \parallel\rightarrow R) + \text{dh}(\text{plus}(\text{size}(l), \text{size}(r))[\sigma], \parallel\rightarrow R),$$

$$1 + \text{dh}(\text{size}(r)[\sigma], \parallel\rightarrow R) + \text{dh}(\text{plus}(\text{size}(l), \text{size}(r))[\sigma], \parallel\rightarrow R))$$

Intuitively, this would correspond to evaluating \texttt{plus}(\text{size}(l), \text{size}(r)) twice, in two parallel threads of execution, which costs the same amount of time as evaluating \texttt{plus}(\text{size}(l), \text{size}(r)) once. We can represent this maximum of the execution times of two threads by introducing two DTs for our recursive size-rule:

$$\text{size}^2(\text{Tree}(v, l, r)) \rightarrow \text{Com}_2(\text{size}^2(l), \text{plus}^2(\text{size}(l), \text{size}(r)))$$

$$\text{size}^2(\text{Tree}(v, l, r)) \rightarrow \text{Com}_2(\text{size}^2(r), \text{plus}^2(\text{size}(l), \text{size}(r)))$$

To express the cost of a concrete rewrite sequence, we would non-deterministically choose the DT that corresponds to the “slower thread”.

In other words, the cost of the function call to the lhs of a rule is 1 + the sum of the costs of all the function calls in the rhs that are in structural dependency with each other. The actual cost of the function call to the lhs in a concrete rewrite sequence is the maximum of all the possible costs of such chains (in the prefix order \(<\) on positions) of structural dependency. Thus, \textit{structurally independent} function calls are considered in separate DTs, whose non-determinism models the parallelism of these function calls.

The notion of \textit{structural dependency} of functions calls is captured by Def. 9. Basically, it comes from the fact that a term cannot be evaluated before all its subterms have been reduced to normal forms (innermost rewriting/call by value).

\textbf{Definition 9 (Structural dependency).} The term \(t_1\) structurally depends on \(t_2\) if there exists a rule \(\ell \rightarrow r \in R\) such that \(r|_{\tau_1} = t_1\) and \(r|_{\tau_2} = t_2\) with \(\tau_1 < \tau_2\) in the prefix order \(<\).

We thus revisit the notion of DTs, which now embed structural dependencies.

\textbf{Definition 10 (Parallel Dependency Tuples \(DT^\parallel\), PDTs).} For a rewrite rule \(\ell \rightarrow r\), we define the set of its Parallel Dependency Tuples (PDTs) \(DT^\parallel(\ell \rightarrow r)\): if \(\text{Pos}_d(r) = \emptyset\), then \(DT^\parallel(\ell \rightarrow r) = \{ \ell^2 \rightarrow \text{Com}_0 \}\); otherwise, \(DT^\parallel(\ell \rightarrow r) = \{ \ell^2 \rightarrow \text{Com}_k(r|_{\tau_1}, \ldots, r|_{\tau_k}) \mid k > 0, \tau_1 > \cdots > \tau_k \text{ is a maximal } \triangleright\text{-chain in } \text{Pos}_d(r) \}.\) For aTRS \(R\), let \(DT^\parallel(R) = \bigcup_{\ell \rightarrow r \in R} DT^\parallel(\ell \rightarrow r)\).

\textbf{Example 11.} For our recursive size-rule \(lhs \rightarrow rhs\), we have \(\text{Pos}_d(rhs) = \{1, 11, 12\}\). The two maximal \(\triangleright\text{-chains} are 11 > 12 > 1\). With \(rhs|_{11} = \text{plus}(\text{size}(l), \text{size}(r))\), \(rhs|_{11} = \text{size}(l)\), and \(rhs|_{12} = \text{size}(r)\), we get the PDTs from Ex. 8.

To connect PDTs with our parallel-innermost rewrite relation \(\parallel\rightarrow R\), we need the notion of \textit{chain tree}, which is an extension of dependency chains \([2]\), and its complexity.

\textbf{Definition 12 (Chain Tree, \textit{Cpx} [5]).} Let \(D\) be a set of DTs and \(R\) be a TRS. Let \(T\) be a (possibly infinite) tree whose nodes are labelled with a DT from \(D\) and a substitution.

Let the root node be labelled with \((s^2 \rightarrow \text{Com}_n(\ldots) \mid \sigma)\). Then \(T\) is a \((D, R)\)-chain tree for \(s^2\) if we have the following for all nodes of \(T\): if a node is labelled with \((u^2 \rightarrow \text{Com}_m(v^i_1, \ldots, v^i_m) \mid \mu)\), then \(u^\mu\) is in normal form w.r.t. \(R\). Moreover, if this node has the children \((p^2_1 \rightarrow \text{Com}_{m_1}(\ldots) \mid \delta_1), \ldots, (p^2_k \rightarrow \text{Com}_{m_k}(\ldots) \mid \delta_k)\), then there are pairwise different \(i_1, \ldots, i_k \in \{1, \ldots, m\}\) with \(v^i_{\delta_1} \xrightarrow{R} p^2_j \delta_j\) for all \(j \in \{1, \ldots, k\}\).

Let \(S \subseteq D\) and \(s^2 \in T^\parallel\). For a chain tree \(T\), \(|T|_S \in N \cup \{\omega\}\) is the number of nodes in \(T\) labelled with a DT from \(S\). We define \(\text{Cpx}_{(D, S, R)}(s^2) = \sup(|T|_S \mid T\text{ is a } (D, R)\text{-chain tree for } s^2\})\) for terms \(s^2\) without a \((D, R)\)-chain tree, we define \(\text{Cpx}_{(D, S, R)}(s^2) = 0\).

We can now make our main correctness claim:
Proposition 13 (Cplx bounds Derivation Height for $\parallel \rightarrow_R$). Let $\mathcal{R}$ be a TRS, let $t = f(t_1, \ldots, t_n) \in T(\Sigma, \Delta)$ such that all $t_i$ are in normal form. Then we have $\text{dh}(t, \parallel \rightarrow_R) \leq \text{Cplx}(\text{DT}(\mathcal{R}), \text{DT}(\mathcal{R}), \mathcal{R})(t^2)$. If $\mathcal{R}$ is confluent, then $\text{dh}(t, \parallel \rightarrow_R) = \text{Cplx}(\text{DT}(\mathcal{R}), \text{DT}(\mathcal{R}), \mathcal{R})(t^2)$.

Thus, in particular we can use polynomial interpretations in the DT framework for our PDTs to get upper bounds for $\text{irc}^\parallel_R$.

Example 14 (Ex. 8 continued). For our TRS $\mathcal{R}$ computing the size function on trees, we get the set $\text{DT}^\parallel(\mathcal{R})$ with the following PDTs:

\begin{align*}
\text{plus}^2(\text{Zero}, y) & \rightarrow \text{Com}_0, \quad \text{size}^2(\text{Nil}) \rightarrow \text{Com}_0 \\
\text{plus}^2(S(x), y) & \rightarrow \text{Com}_1(\text{plus}^2(x, y)), \quad \text{size}^2(\text{Tree}(v, l, r)) \rightarrow \text{Com}_2(\text{size}^2(l), \text{plus}^2(\text{size}(l), \text{size}(r))) \\
& \quad \text{size}^2(\text{Tree}(v, l, r)) \rightarrow \text{Com}_2(\text{size}^2(r), \text{plus}^2(\text{size}(l), \text{size}(r)))
\end{align*}

The interpretation $\text{Pol}$ from Ex. 7 implies $\text{irc}^\parallel_R(n) \in O(n^2)$. This bound is tight: consider $\text{size}(t)$ for a comb-shaped tree $t$ where the first argument of $\text{Tree}$ is always $\text{Zero}$ and the third is always $\text{Nil}$. The function $\text{plus}$, which needs time linear in its first argument, is called linearly often on data linear in the size of the start term. Due to the structural dependencies, these calls do not happen in parallel (so call $k + 1$ to $\text{plus}$ must wait for call $k$).

Example 15. Note that $\text{irc}^\parallel_R(n)$ can be asymptotically lower than $\text{irc}(n)$, for instance in:

\begin{align*}
\text{doubles}(\text{Zero}) & \rightarrow \text{Nil} \\
\text{doubles}(S(x)) & \rightarrow \text{Cons}(\text{d}(S(x)), \text{doubles}(x)) \\
\text{d}(\text{Zero}) & \rightarrow \text{Zero} \\
\text{d}(S(x)) & \rightarrow S(\text{d}(S(x)))
\end{align*}

The upper bound $\text{irc}^\parallel_{\mathcal{R}}(n) \in O(n^2)$ is tight: from a term $\text{doubles}(S(S(\ldots S(\text{Zero} \ldots))))$, we get linearly many calls of the linear-time function $\text{d}$ on arguments of size linear in the start term. However, the Parallel Dependency Tuples in this example are:

\begin{align*}
\text{doubles}^2(\text{Zero}) & \rightarrow \text{Com}_0, \quad \text{d}(\text{Zero}) \rightarrow \text{Com}_0 \\
\text{doubles}^2(S(x)) & \rightarrow \text{Com}_1(\text{d}^2(S(x))), \quad \text{d}(S(x)) \rightarrow \text{Com}_1(\text{d}^2(S(x))) \\
\text{d}(\text{Zero}) & \rightarrow \text{Com}_0 \\
\text{d}(S(x)) & \rightarrow \text{Com}_1(\text{d}^2(S(x)))
\end{align*}

Then the given polynomial interpretation, which orients all DTS with $\succ$ and all rules from $\mathcal{R}$ with $\succeq$, proves $\text{irc}^\parallel_{\mathcal{R}} \in O(n)$: $\text{Pol}(\text{doubles}^2(x_1)) = \text{Pol}(\text{d}(x_1)) = 2x_1, \text{Pol}(\text{d}(x_1)) = x_1, \text{Pol}(\text{d}(x_1)) = \text{Pol}(\text{Cons}(x_1, x_2)) = \text{Pol}(\text{Zero}) = \text{Pol}(\text{Nil}) = 1, \text{Pol}(\text{S}(x_1)) = 1 + x_1$.

Conclusion

We have come up with a notion of parallel runtime complexity and a concrete algorithm to compute upper bounds on this complexity on TRSs. Future work includes practical design of parallel rewriting engines as well as the formalisation of complexity w.r.t. term height (considering terms as trees), which seems to be more practical for our parallelisation needs.

References