A decision procedure for term algebras with queues

TATIANA RYBINA and ANDREI VORONKOV
The University of Manchester

In software verification it is often required to prove statements about heterogeneous domains containing elements of various sorts, such as counters, stacks, lists, trees and queues. Any domain with counters, stacks, lists, and trees (but not queues) can be easily seen a special case of the term algebra, and hence a decision procedure for term algebras can be applied to decide the first-order theory of such a domain. We present a quantifier-elimination procedure for the first-order theory of term algebras extended with queues. The complete axiomatization and decidability of this theory can be immediately derived from the procedure.

Categories and Subject Descriptors: D.2.4 [**Software/Program Verification**]: Formal Methods; E.1 [**Data Structures**]: Lists, Stacks, and Queues; F.4.1 [**Mathematical Logic**]: Computational Logic; I.2.3 [**Artificial Intelligence**]: Deduction and Theorem Proving

General Terms: Algorithms, Theory, Verification

Additional Key Words and Phrases: Term algebras, queues, trees, words

1. INTRODUCTION

Our motivation for investigating decidability of term algebras with queues comes from software and hardware verification. In framework of Rainbow project (see Visser et al. [1998]), we had to prove correctness of some protocols for infinite-state systems, containing a queue of messages. The correctness problem was reduced to proving first-order statements about such queues. When we replaced the queue by a stack, we could use standard decision procedures for term algebras, since stacks can be easily interpreted in a many-sorted version of term algebras. However, such a replacement changes the protocol and does not prove correctness of the original protocol. When we looked into computer science and logical literature, to our surprise we could not find any decision procedure for the first-order theory of queues. We have found that queues have attracted a lot of attention in verification [Bjørner 1998; 2000]. However, Bjørner [1998; Bjørner [2000] presented a decision procedure only for the existential theory of queues (though in a much richer signature than ours), which was not enough for our purpose: some statements, like absence of deadlocks, required truly first-order formulas with quantifier alternations.

In this article we prove decidability of a first-order theory of term algebras extended with queues. This result implies decidability of the first-order theory of essentially any domain that may contain the following kinds of objects: queues,

Both authors were supported by grants from EPSRC. Author's address: Department of Computer Science, University of Manchester, Manchester M13 9PL, UK.
Permission to make digital/hard copy of all or part of this material without fee for personal or classroom use provided that the copies are not made or distributed for profit or commercial advantage, the ACM copyright/server notice, the title of the publication, and its date appear, and notice is given that copying is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or a fee.
© 2001 ACM 1529-3785/01/0400-TBD $5.00

lists, trees, stacks, finite domains, counters, or embedded domains, like queues of trees or queues of queues, since all these kinds of objects except for queues can be represented as sorts in a many-sorted term algebra.

Independently, our results are of general interest in logic, since they give quantifier elimination and decision procedures for the first-order theory of words with two operations: left and right multiplication by a single letter.

Our proof is based on a quantifier elimination procedure that extends standard quantifier elimination procedures for term algebras [Maĺcev 1961; Maher 1988; Comon and Lescanne 1989; Kunen 1987; Belegradek 1988a; Hodges 1993]. More precisely, we use the technique of decomposition functions introduced independently by Kunen [1987] and by Belegradek [1988a], and explained in Hodges [1993]. Our quantifier elimination for queues is based on handling quantifier-free formulas expressing periodicity of queues. Such formulas also play a major role in Bjørner [2000].

For verification purposes, quantifier elimination is even more attractive than a mere decidability result, since quantifier elimination allows one to "simplify" formulas into equivalent quantifier-free formulas (though the size of the quantifier-free formula may be huge). So, one can "reuse" the same subformulas occurring in verification statements by running the quantifier elimination procedure on each subformula once, instead of rerunning a decision procedure from scratch.

Our article is structured as follows. In Section 2 we define the structure $\mathcal{Q}(\mathcal{S})$ of term algebras with queues. Section 3 is devoted to the proof of quantifier elimination for the first-order theory of $\mathcal{Q}(\mathcal{S})$. The quantifier elimination procedure and the proof of this section provide an axiomatization for term algebras with queues, and this axiomatization is presented in Section 4. In Section 5 we discuss related papers, mostly on term algebras, and some possible results that can be proved for the first-order theories of queues.

2. TERM ALGEBRAS WITH QUEUES

2.1 Sorts

We assume a finite set of *basic sorts*. We define *sorts* as follows:

1. Every basic sort is a sort;
2. If α is a sort, then $queue(\alpha)$ is a sort, called a *queue sort*.

We call a *function type* any expression of the form $\alpha_1 \times \ldots \times \alpha_n \rightarrow \beta$, where $\alpha_1, \ldots, \alpha_n, \beta$ are sorts, $n \geq 0$. When $n = 0$, we will write β instead of $\rightarrow \beta$. A *signature* $\mathcal{S}$ is a pair consisting of a *finite* set of function symbols, and a function : mapping every function symbol f to a function type $\alpha_1 \times \ldots \times \alpha_n \rightarrow \beta$, called the *type of* f, where β is a basic sort. The number n is called the *arity* of f. Function symbols of arity 0 are called *constants*. *We assume that every signature contains at least one constant.*

Though we deal with algebras, i.e., structures with no relation symbols besides the equality $=$, we will sometimes need to extend them by relation symbols of different types. We call a *relation type* any expression of the form $\alpha_1 \times \ldots \times \alpha_n$, where $\alpha_1, \ldots, \alpha_n$ are sorts, $n \geq 0$. In this article we understand validity and equivalence as validity and equivalence in $\mathcal{Q}(\mathcal{S})$. Likewise, tautologies are formulas valid in $\mathcal{Q}(\mathcal{S})$.

— Sort (Natural numbers) :	— Sort (Integers) :
1. 0 : nat;	1. pos : nat → int;
2. successor : nat → nat	2. neg : neg → int
— Sort (Negative numbers) :	— Sort (Stacks):
1. −1 : neg;	1. empty : stack;
2. predecessor : neg → neg	2. push : int, stack → stack

$$\begin{array}{rcl}\mathbf{next}(x,y) & \Leftrightarrow & \forall z\ (x = \mathbf{pos}(z) \rightarrow \\ & & \quad y = \mathbf{pos}(\mathbf{successor}(z))) \wedge \\ & & (x = \mathbf{neg}(-1) \rightarrow y = \mathbf{pos}(0)) \wedge \\ & & \forall z\ (x = \mathbf{neg}(\mathbf{predecessor}(z)) \rightarrow \\ & & \quad y = \mathbf{neg}(z))\end{array}$$

Fig. 1. Example: defining several domains and the integer successor function in term algebras.

2.2 Structure $\mathcal{Q}(\mathcal{S})$

Every signature $\mathcal{S}$ defines a unique structure $\mathcal{Q}(\mathcal{S})$, called a *term algebra with queues* as follows.

Universe of $\mathcal{Q}(\mathcal{S})$. The *universe of* $\mathcal{Q}(\mathcal{S})$ is defined inductively as follows.

1. For every constant $a : \beta$, a is an element of $\mathcal{Q}(\mathcal{S})$ of the sort β.
2. For every elements $e_1, \ldots, e_n$ of the same sort α, where $n \geq 0$, the sequence $e_1 \ldots e_n$ is an element of $\mathcal{Q}(\mathcal{S})$ of the sort $queue(\alpha)$. Elements of any sort $queue(\alpha)$ will be called *queues*. The empty sequence is called the *empty queue*, and denoted ε_α. Note that there exist empty queues of different sorts. Alternatively, we can characterize the domain of $queue(\alpha)$ as the set of words on α.
3. If $f : \alpha_1 \times \ldots \times \alpha_n \rightarrow \beta$ and $e_1, \ldots, e_n$ are elements of $\mathcal{Q}(\mathcal{S})$ of the sorts $\alpha_1, \ldots, \alpha_n$, respectively, then $f(e_1, \ldots, e_n)$ is an element of $\mathcal{Q}(\mathcal{S})$ of the sort β.

We consider two elements equal only if they have the same sort and coincide as expressions. For example, $f(a) \neq f(b)$ if a and b are different constants. Two queues $e_1 \ldots e_n$ and $d_1 \ldots d_m$ are equal if and only if $n = m$ and $e_1 = d_1, \ldots, e_n = d_n$. The *length* of a queue $q = e_1 \ldots e_n$ is a number of elements of q, i.e., n.

To give the reader the idea of how different domains can be specified in term algebras with sorts, in Figure 1 we define several sorts, including integers and stacks of integers, and write down a first-order formula **next** expressing the successor function on integers.

Using different signatures we obtain different domains, in which sorts can be empty or nonempty, finite, or infinite.

LEMMA 2.1. *Given a signature $\mathcal{S}$ and a sort β, one can effectively recognize whether the domain of β is empty, finite nonempty or infinite. Moreover, there exists a computable mapping D from the set of sorts to the set of finite subsets of $\mathcal{Q}(\mathcal{S})$ such that $D(\beta)$ is the domain of β, whenever the domain of β is finite.*

PROOF. We will only sketch the proof. It is not hard to show that nonemptiness of the domain of a sort α could be checked using the following inductive rules: the domain is nonempty if the nonemptiness can be derived using these rules.

1. If α is a basic sort and there is a constant $c : \alpha$, then the domain of α is nonempty.
2. If α is a basic sort, the domains of $\alpha_1, \ldots, \alpha_n$ are nonempty, and there exists a function symbol $f : \alpha_1 \times \ldots \times \alpha_n \to \alpha$, then the domain of α is nonempty.
3. If the domain of α is nonempty, then the domain of $queue(\alpha)$ is nonempty.

Infiniteness of the domain of a sort α can be checked as follows. Let G be the directed graph whose nodes are the sorts with nonempty domains and where the edges are defined as follows:

1. If there exists a function symbol $f : \alpha_1 \times \ldots \times \alpha_n \to \alpha$, then the graph contains all edges from α_i to α, for $i = 1, \ldots, n$.
2. If α is a queue sort, then the graph contains the edge from α to α.

Then the infiniteness of the domain of a sort can be checked using the following inductive rules:

1. If G contains a path from α to α, then the domain of α is infinite.
2. If the domain of α is infinite and G contains an edge from α to β, then the domain of β is infinite.

We leave it to the reader to check how all elements of a finite domain can be effectively found. $\square$

We are not interested in having empty sorts (and all the resulting complications of first-order logic with function symbols and empty sorts), so we assume that the signature $\mathcal{S}$ is such that every sort defined in it is nonempty.

Language of $\mathcal{Q}(\mathcal{S})$. The *language of* $\mathcal{Q}(\mathcal{S})$, denoted $\mathcal{L}_{\mathcal{Q}(\mathcal{S})}$, is a first-order language using, in addition to function symbols in $\mathcal{S}$, the equality relation symbol $=$ and the following symbols:

1. for every sort α, the constant $\varepsilon_\alpha : queue(\alpha)$;
2. for every sort α, the function symbols $ladd_\alpha : \alpha \times queue(\alpha) \to queue(\alpha)$ and $radd_\alpha : \alpha \times queue(\alpha) \to queue(\alpha)$.

As usual, we assume an infinite set of variables of each sort α. Intuitively, *ladd* and *radd* are the left- and right-insert operations on queues respectively. We denote by $\top$ and $\bot$ the logical connectives "true" and "false."

Semantics of $\mathcal{Q}(\mathcal{S})$. Now we define the interpretation for each symbol of the language $\mathcal{L}_{\mathcal{Q}(\mathcal{S})}$.

For every function symbol $f : \alpha_1 \times \ldots \times \alpha_n \to \beta$ of $\mathcal{S}$, the interpretation $\underline{f}$ of this symbol in $\mathcal{Q}(\mathcal{S})$ is defined as follows: for all elements $e_1, \ldots, e_n$ of the sorts $\alpha_1, \ldots, \alpha_n$ we have $\underline{f}(e_1, \ldots, e_n) = f(e_1, \ldots, e_n)$.

Likewise, the interpretation $\underline{\varepsilon_\alpha}$ of the constant ε_α is the empty queue of the sort $queue(\alpha)$, i.e., ε_α itself.

The interpretations $\underline{ladd_\alpha}$ and $\underline{radd_\alpha}$ of the function symbols $ladd_\alpha$ and $radd_\alpha$ are defined as follows. For all elements $e, e_1, \ldots, e_n$ of a sort α:

$$\underline{ladd_\alpha}(e, e_1 \ldots e_n) = e e_1 \ldots e_n;$$
$$\underline{radd_\alpha}(e, e_1 \ldots e_n) = e_1 \ldots e_n e.$$

In computer science literature, queues are usually defined as datatypes with three operations *push*, *top*, and *pop*, where *push* inserts an element into the queue from the right, *top* gives the leftmost element of the queue, and where *pop* removes the leftmost element. Using queues with the operations *ladd* and *radd* instead gives us a definitionally equivalent first-order language as we will now show. We write down explicit definitions of $push_\alpha$, top_α, and pop_α in terms of $ladd_\alpha$ and $radd_\alpha$ and vice versa.

$$\begin{aligned} push_\alpha(x,y) = z &\leftrightarrow radd_\alpha(x,y) = z; \\ top_\alpha(x) = y &\leftrightarrow \exists z\ ladd_\alpha(y,z) = x; \\ pop_\alpha(x) = y &\leftrightarrow \exists z\ ladd_\alpha(z,y) = x. \end{aligned}$$

The operations $ladd_\alpha$ and $radd_\alpha$ can be defined in terms of $push_\alpha$, top_α and pop_α as follows:

$$\begin{aligned} ladd_\alpha(x,y) = z &\leftrightarrow z \neq \varepsilon_\alpha \wedge top_\alpha(z) = x \wedge pop_\alpha(z) = y; \\ radd_\alpha(x,y) = z &\leftrightarrow push_\alpha(x,y) = z. \end{aligned}$$

In term algebras *without queues*, the interpretation $\underline{t}$ of every ground term t is t itself. Since the domain consists of the ground terms of $\mathcal{S}$, every element of the domain is the interpretation of a unique ground term. In term algebras *with queues*, every element of the domain is also the interpretation of some ground term, but this term is in general not unique. For example, the queue ab consisting of two constants a and b of a sort α is the interpretation of four different ground terms:

$$\begin{aligned} &ladd_\alpha(a, ladd_\alpha(b, \varepsilon_\alpha)); \\ &ladd_\alpha(a, radd_\alpha(b, \varepsilon_\alpha)); \\ &radd_\alpha(b, ladd_\alpha(a, \varepsilon_\alpha)); \\ &radd_\alpha(b, radd_\alpha(a, \varepsilon_\alpha)). \end{aligned}$$

LEMMA 2.2. *Each element of $\mathcal{Q}(\mathcal{S})$ is the interpretation of a ground term. For every element e of $\mathcal{Q}(\mathcal{S})$ there exists a unique ground term t such that e is the interpretation of t and t has no occurrences of $radd_\alpha$. There exists an algorithm that, given two ground terms s, t, verifies whether $\mathcal{Q}(\mathcal{S}) \models s = t$.*

Due to this lemma, we can (and will) identify elements of the universe of $\mathcal{Q}(\mathcal{S})$ with ground terms, and thus blur the distinction between syntax and semantics.

3. QUANTIFIER ELIMINATION

We say that a theory T *admits quantifier elimination* if for any formula $A(\bar{v})$ of the language of this theory, one can effectively find a quantifier-free formula $B(\bar{v})$ of the same language equivalent to $A(\bar{v})$ in T.

The first-order theory of term algebras in the language $\mathcal{L}_{\mathcal{Q}(\mathcal{S})}$ does not admit quantifier elimination. A simple example is the formula $A(x) = \exists y(x = f(y))$ expressing that the outermost function symbol of x is f. This formula is not equivalent to any quantifier-free formula of the language $\mathcal{L}_{\mathcal{Q}(\mathcal{S})}$. However, we can change $\mathcal{L}_{\mathcal{Q}(\mathcal{S})}$ into a definitionally equivalent language $\mathcal{L'}_{\mathcal{Q}(\mathcal{S})}$ which admits quantifier elimination.

We introduce the following function and relation symbols:

1. a unary relation symbol $\mathbf{Is}_f$ for each function symbol $f : \alpha_1 \times \ldots \times \alpha_n \to \beta$, $n \geq 1$ of $\mathcal{S}$, this symbol is of the type β;

2. a unary function symbol f_i, called a *decomposition function*, for each function symbol $f : \alpha_1 \times \ldots \times \alpha_n \to \beta$, $n \geq 1$ of $\mathcal{S}$ and each $i = 1, \ldots, n$, this symbol is of the type $\beta \to \alpha_i$;
3. unary function symbols

$$\begin{aligned} lhead_\alpha &: queue(\alpha) \ \to \ \alpha, \\ rhead_\alpha &: queue(\alpha) \ \to \ \alpha, \\ ltail_\alpha &: queue(\alpha) \ \to \ queue(\alpha), \\ rtail_\alpha &: queue(\alpha) \ \to \ queue(\alpha) \end{aligned}$$

for each sort α. The symbols *lhead* and *rhead* are called *left and right head functions*. The symbols *ltail* and *rtail* are called *left and right tail functions*.

For convenience, we will usually omit types in function and relation symbols and simply write *ladd*, *radd*, ... etc. The first-order language using the constants of $\mathcal{S}$ (including ε_α) and these new symbols will be denoted $\mathcal{L}'_{\mathcal{Q}(\mathcal{S})}$. Note that all function symbols of $\mathcal{L}'_{\mathcal{Q}(\mathcal{S})}$ are unary. In our proofs we will sometimes use formulas with symbols both in $\mathcal{L}_{\mathcal{Q}(\mathcal{S})}$ and $\mathcal{L}'_{\mathcal{Q}(\mathcal{S})}$.

Now, we define interpretations of the new function and relation symbols. We want to avoid having partial functions, so we will have to give artificial values to some functions on some arguments. To this end, we choose one element in each sort α and declare it the value of these terms. The fixed element will be denoted by fix_α. When α is clear from the context, we simply write fix. For simplicity, we assume that for any queue sort $queue(\alpha)$ the fixed element is the empty queue ε_α.

1. $\mathbf{Is}_f(x)$ denotes that the outermost function symbol of x is f, i.e.

$$\mathbf{Is}_f(x) \ \leftrightarrow \ \exists x_1, \ldots, x_n \ f(x_1, \ldots, x_n) = x.$$

2. f_i's are decomposition functions; they extract the ith component from the terms of the form $f(t_1, \ldots, t_n)$:

$$f_i(x) = y \ \leftrightarrow \ \forall x_1, \ldots, x_n (f(x_1, \ldots, x_n) = x \to y = x_i) \wedge (\neg \mathbf{Is}_f(x) \to y = \mathsf{fix}_\alpha).$$

The rest of the new functions are intended to decompose queues.

3. The functions *lhead* and *ltail* give the leftmost element of a queue, and the queue obtained by removing this element:

$$\begin{aligned} lhead(q) = y \ &\leftrightarrow \ \exists w (ladd(y, w) = q) \vee (q = \varepsilon \wedge y = \mathsf{fix}); \\ ltail(q) = w \ &\leftrightarrow \ \exists y (ladd(y, w) = q) \vee (q = \varepsilon \wedge w = \varepsilon). \end{aligned}$$

4. Likewise, the functions *rhead* and *rtail* give the rightmost element of a queue, and the queue obtained by removing this element:

$$\begin{aligned} rhead(q) = y \ &\leftrightarrow \ \exists w (radd(y, w) = q) \vee (q = \varepsilon \wedge y = \mathsf{fix}); \\ rtail(q) = w \ &\leftrightarrow \ \exists y (radd(y, w) = q) \vee (q = \varepsilon \wedge w = \varepsilon). \end{aligned}$$

We will leave out parentheses in successive applications of *lhead*, *rhead*, *ltail*, *rtail*; and the decomposition functions, for example, we write $f_i rhead_\alpha ltail_\alpha x$ instead of $f_i(rhead_\alpha(ltail_\alpha(x)))$. Uppercase letters L, M stand for sequences, maybe empty, of any unary function symbols of $\mathcal{L}'_{\mathcal{Q}(\mathcal{S})}$; we often call such sequences *prefixes*. For example, if $L = rtail_\beta f_i lhead_\alpha$ then Lx denotes the term $rtail_\beta f_i lhead_\alpha x$.

Let us note that the new collection of function symbols is definitionally equivalent to the collection of function symbols of $\mathcal{L}_{\mathcal{Q}(\mathcal{S})}$. We have already shown above how to express the symbols of the new language using first-order formulas of the old language; now we will express the old language symbols using the new ones. Indeed, it is not hard to argue that the following formulas are tautologies:

$$\begin{aligned} ladd(x,y) = z \ &\leftrightarrow\ z \neq \varepsilon \wedge lhead(z) = x \wedge ltail(z) = y; \\ radd(x,y) = z \ &\leftrightarrow\ z \neq \varepsilon \wedge rhead(z) = x \wedge rtail(z) = y; \\ f(x_1, \ldots, x_n) = y \ &\leftrightarrow\ \mathbf{Is}_f(y) \wedge \bigwedge_{i=1\ldots n} f_i(y) = x_i. \end{aligned}$$

To avoid some technical problems, we will impose some restrictions on the fixed elements of sorts. Let us define the *weight* of any element e, denoted $weight(e)$ by induction as follows.

$$weight(e) = \begin{cases} 1, & \text{if } e \text{ is a constant;} \\ 1 + \sum_i weight(t_i), & \text{if } e = f(t_1, \ldots, t_n),\ f \in \mathcal{S}; \\ 1 + \sum_i weight(t_i), & \text{if } e \text{ is a queue } t_1 \ldots t_n. \end{cases}$$

Now we impose the following requirement on the fixed element fix_α of any sort: fix_α must be an element of the sort α with the minimal weight. Let us note that this property is consistent with the requirement that $\mathsf{fix}_{queue(\alpha)} = \varepsilon_\alpha$, since the empty queue has the least weight 1. We need this requirement in order to make the following lemma hold:

LEMMA 3.1. *Let e be any element and g be a unary function of the new language $\mathcal{L}'_{\mathcal{Q}(\mathcal{S})}$. Then $weight(g(e)) \leq weight(e)$; moreover, if $weight(g(e)) = weight(e)$, then $g(e)$ is the fixed element of its sort.*

PROOF. Straightforward from the definition of the semantics of the new functions and the definition of $weight(e)$. $\square$

As a simple consequence, we obtain the following lemma.

LEMMA 3.2. *Let e be any element and L be any nonempty sequence of unary functions of the new language $\mathcal{L}'_{\mathcal{Q}(\mathcal{S})}$. If $Le = e$, then e is the fixed element of its sort.*

Our aim in this section is to prove the following theorem:

THEOREM 3.3 *(Quantifier Elimination). The first-order theory of the term algebra with queues in the language $\mathcal{L}'_{\mathcal{Q}(\mathcal{S})}$ admits quantifier elimination.*

The proof of this theorem will be based on a series of transformations of $A(\bar{v})$ into equivalent formulas of special form, until we obtain a quantifier-free formula. Let us now introduce a convenient definition.

Definition 3.4 *(Quantifier Elimination Reduction Class).* Let $\mathfrak{C}_1$ and $\mathfrak{C}_2$ be two classes of formulas. We say $\mathfrak{C}_1$ is a *quantifier elimination reduction class for* $\mathfrak{C}_2$ if we have: every formula $A \in \mathfrak{C}_2$ is equivalent to a quantifier-free formula if every formula of $\mathfrak{C}_1$ is equivalent to a quantifier-free formula. A class $\mathfrak{C}$ is called an *absolute quantifier elimination reduction class*, if it is a quantifier elimination reduction class for the class of all formulas.

Let us note three main properties of quantifier elimination reduction classes.

1. The first-order theory of term algebras admits quantifier elimination if and only if the class of all quantifier-free formulas of this theory is an absolute quantifier elimination reduction class.
2. If $\mathfrak{C}_1$ is a quantifier elimination reduction class for $\mathfrak{C}_2$, and $\mathfrak{C}_2$ is a quantifier elimination reduction class for $\mathfrak{C}_3$, then $\mathfrak{C}_1$ is a quantifier elimination reduction class for $\mathfrak{C}_3$.
3. If $\mathfrak{C}_1$ is a quantifier elimination reduction class for $\mathfrak{C}_2$, and $\mathfrak{C}_2$ is an absolute quantifier elimination reduction class, then $\mathfrak{C}_1$ is an absolute quantifier elimination reduction class.

Speaking in new terms, we will provide a sequence of absolute quantifier elimination reduction classes.

We call a *literal* any atomic formula or its negation. Consider the class of formulas of the language $\mathcal{L}'_{\mathcal{Q}(\mathcal{S})}$ of the form

$$\exists x(A_1 \wedge \ldots \wedge A_n), \tag{1}$$

where $n > 0$ and all A_i's are literals.

LEMMA 3.5. *The class of formulas (1) is an absolute quantifier elimination reduction class.*

The proof is standard and can be found in, for example, Rybina and Voronkov [2000].

Now it is enough to show how to eliminate quantifiers from formulas (1). Our proof will be similar to the proof of quantifier elimination for term algebras as presented in [Hodges 1993], but with one essential difference. First-order formulas of the theory of term algebras cannot express very "deep" properties of terms. For example, one cannot express that c is the only constant occurring in a term x. The first-order language of queues can express certain statements about elements in a queue at an arbitrary depth. For example, the following formula expresses that c is the only element in a queue q, i.e., $q = c \ldots c$:

$$ladd(c, q) = radd(c, q). \tag{2}$$

This property of the first-order language of queues will create some technical difficulties. Later, we will introduce a special notation that allows us to express formulas like (2).

Consider literals of the following forms:

$$\begin{array}{l} x = t,\ x = Lx,\ Lx = t,\ Lx = Mx,\ \mathbf{Is}_f(x),\quad \mathbf{Is}_f(Lx), \\ x \neq t,\ x \neq Lx,\ Lx \neq t,\ Lx \neq Mx,\ \neg\mathbf{Is}_f(x),\ \neg\mathbf{Is}_f(Lx), \end{array} \tag{3}$$

where L, M are *nonempty* and t is a term in which x does not occur.

LEMMA 3.6. *The class of formulas $\exists x(C)$, where C is a conjunction of literals of the form (3), is a quantifier elimination reduction class for the class of formulas of the forms (1). Thus, by Lemma 3.5, this class of formulas is an absolute quantifier elimination reduction class.*

PROOF. Consider any formula (1): $\exists x(A_1 \wedge \ldots \wedge A_n)$. If any A_i contains no occurrences of x, it can be moved outside of $\exists x$, obtaining an equivalent formula

$$A_i \wedge \exists x(A_1 \wedge \ldots A_{i-1} \wedge A_{i+1} \ldots \wedge A_n).$$

Therefore, we can assume that x occurs in each A_i. Since all function symbols of $\mathcal{L}'_{\mathcal{Q}(\mathcal{S})}$ are unary, it is not hard to argue that every A_i has either one of the forms listed in (3), or $x = x$ or $x \neq x$. It is straightforward how to get rid of the last two literals. □

In our proofs we will often use the following trick. Given a formula $C = \exists x((A_1 \vee \ldots \vee A_n) \wedge B)$, we will consider instead formulas $C_i = \exists x(A_i \wedge B)$. Note that $C \leftrightarrow C_1 \vee \ldots \vee C_n$, but all of the C_i's do not contain disjunction. We will refer to this trick as *disjunctive splitting*.

It remains to show how to eliminate the existential quantifier applied to a conjunction of literals of forms given in (3). Let us first get rid of some simple cases in (3), so that only the following kind of literals remain:

$$Lx = t,\ Lx = Mx,\ \mathbf{Is}_f(x),\ \mathbf{Is}_f(Lx),\ x \neq t,\ Lx \neq t,\ Lx \neq Mx, \tag{4}$$

where L, M are *nonempty* and t is a term in which x does not occur.

LEMMA 3.7. *The class of formulas $\exists x(C)$, where C is a conjunction of literals of the form (4), is a quantifier elimination reduction class for the class of formulas $\exists x(D)$, where D is a conjunction of formulas (3). Thus, by Lemma 3.6, this class of formulas is an absolute quantifier elimination reduction class.*

PROOF. We have to show how to get rid of the literals $x = t$, $x = Lx$, $x \neq Lx$, $\neg\mathbf{Is}_f(x)$ $\neg\mathbf{Is}_f(Lx)$ in the formula $\exists x(D)$.

Case $(x = t)$. Any formula $\exists x(x = t \wedge E(x))$ is equivalent to $E(t)$, so we can eliminate $\exists x$.

Case $(x = Lx)$. By Lemma 3.2, the $x = Lx$ implies $x = \mathsf{fix}$, hence $x = Lx$ is equivalent to $x = Lx \wedge x = \mathsf{fix}$. Using $x = \mathsf{fix}$ we can eliminate $\exists x$ as in the previous case.

Case $(x \neq Lx)$. Using elementary logic, this formula is equivalent to $(x = \mathsf{fix} \wedge x \neq Lx) \vee (x \neq \mathsf{fix} \wedge x \neq Lx)$. Using disjunctive splitting, we can consider the formulas $(x = \mathsf{fix} \wedge x \neq Lx)$ and $(x \neq \mathsf{fix} \wedge x \neq Lx)$ separately. With the first formula we can eliminate quantifiers as in the case $x = t$ above. By Lemma 3.2, the second formula is equivalent to $x \neq \mathsf{fix}$, so we get rid of $x \neq Lx$.

Case $(\neg\mathbf{Is}_f(x))$. Let β be the type of x. Let F be the set of all function symbols of types $\alpha_1 \times \ldots \times \alpha_k \rightarrow \beta$ and C be the set of all constants of the type β. It is not hard to argue that $\neg\mathbf{Is}_f(x)$ is equivalent to the formula

$$\bigvee_{g \in F - \{f\}} \mathbf{Is}_g(x) \vee \bigvee_{c \in C} x = c.$$

Then we can use disjunctive splitting.

Case $(\neg\mathbf{Is}_f(Lx))$. This case is analogous to the previous one. □

It remains to show how to eliminate the existential quantifier applied to a conjunction of literals of forms given in (4). Our next step is to get rid of the prefixes L in terms Lx, after this the proof will be relatively straightforward. In the case of quantifier elimination for term algebras, there is a simple way to get rid of such prefixes, but for term algebras with queues variables x of queue sorts will cause some complications. Another complication is caused by the fact that an application of *lhead* or *rhead* to a term of a queue sort may give a term of a basic sort, and vice versa, the application of an decomposition function f_i to a term of a basic sort may give a term of a queue sort. So we will now split the prefixes L to *segments*, so that within each segment the terms have the same queue sort, or otherwise only basic sorts.

Let L be a sequence of unary functions of the new language $\mathcal{L}'_{\mathcal{Q}(\mathcal{S})}$ and $L = L_1 M L_2$, where L_1, L_2 may be empty. We say that M is a *segment* in L in one of the following cases:

1. M is a sequence of the decomposition functions f_i; the last symbol in L_1 and the first symbol in L_2 are *not* decomposition functions (in this case we can say that M is a subsequence consisting of decomposition functions that can be extended neither to the left nor to the right).
2. M is a sequence consisting of a head function followed by zero or more tail functions and the first symbol of L_2 is not a tail function.
3. M is a sequence consisting of one or more tail functions; the first symbol of L_2 is not a tail function, and L_1 is empty.

Segments are illustrated on the following example:

$$\underbrace{\mathit{ltail}\ \mathit{rtail}}_{\text{segment 1}}\ \underbrace{f_1\ h_2\ f_1}_{\text{segment 2}}\ \underbrace{\mathit{lhead}}_{\text{segment 3}}\ \underbrace{\mathit{rhead}\ \mathit{rtail}\ \mathit{rtail}}_{\text{segment 4}}.$$

Let A be any quantifier-free formula and x a variable. We call the *segment depth* of x in A the maximal number of segments in all prefixes L such that Lx occurs in A. If the segment depth is 0, no function symbol is applied to x in A at all.

Recall that our next aim is to eliminate the existential quantifier $\exists x$ applied to a conjunction of literals of forms given in (4). Our next step would be to show how to decrease the segment depth of x. We begin with a special case of the segment depth ≤ 1 and then note that the proof we have used for this case can be generalized to decrease an arbitrary segment depth. We will consider two separate cases, depending on whether x is of basic sort or not.

LEMMA 3.8. *Consider the class $\mathfrak{C}$ of formulas of the form*

$$\exists x(A_1(x) \wedge \ldots \wedge A_n(x)), \tag{5}$$

where x is of a basic sort β, each A_i is of the form (4), and the segment depth of x in this formula is at most 1. The class of quantifier-free formulas is a quantifier elimination reduction class for $\mathfrak{C}$.

PROOF. We assume that the domain of the sort β is infinite; otherwise by Lemma 2.1 we can effectively find terms $s_1, \ldots, s_k$ such that every element of the sort β is equal to one of $s_1, \ldots, s_k$, and replace (5) by the equivalent quantifier-free formula

$$\bigvee_{j=1\ldots k} \bigwedge_{i=1\ldots n} A_i(s_j).$$

Let us call the *height* of x in a formula of the language $\mathcal{L}'_{\mathcal{Q}(\mathcal{S})}$ the maximal number of symbols in sequences L such that Lx occurs in (5). If the height of x in (5) is 0 and (5) contains no occurrences of $\mathbf{Is}_f(x)$, then each A_i is simply of the form $x \neq t_i$ and (5) is

$$\exists x(x \neq t_1 \wedge \ldots \wedge x \neq t_n).$$

Since the domain of the sort β is infinite, this formula is valid and can be replaced by $\top$.

So we assume that either the height of x in (5) is positive or that (5) contains an occurrence of $\mathbf{Is}_f(x)$. Let F be the set of all function symbols of $\mathcal{L}_{\mathcal{Q}(\mathcal{S})}$ of sorts $\ldots \to \beta$, and C be the set of all constants of $\mathcal{L}_{\mathcal{Q}(\mathcal{S})}$ of the sort β. Then the following formula B is a tautology:

$$\bigvee_{f \in F} \mathbf{Is}_f(x) \vee \bigvee_{c \in C} x = c,$$

so the formula $\exists x(A_1(x) \wedge \ldots \wedge A_n(x))$ is equivalent to $\exists x(B \wedge A_1(x) \wedge \ldots \wedge A_n(x))$. Using disjunctive splitting and equivalences $\exists x(x = c \wedge A_1(x) \wedge \ldots \wedge A_n(x)) \leftrightarrow A_1(c) \wedge \ldots \wedge A_n(c)$ we can assume that at least one of A_i's is $\mathbf{Is}_f(x)$. Moreover, we can assume that there exists exactly one such A_i, because $\mathbf{Is}_f(x) \wedge \mathbf{Is}_g(x)$ is unsatisfiable for $f \neq g$. So our formula is of the form

$$\exists x(\mathbf{Is}_f(x) \wedge A_1(x) \wedge \ldots \wedge A_n(x)). \tag{6}$$

Suppose that the arity of f is m. Then this formula is equivalent to

$$\exists x_1 \ldots x_m(A_1(f(x_1, \ldots, x_m)) \wedge \ldots \wedge A_n(f(x_1, \ldots, x_m))).$$

We want to get rid of all occurrences of the function symbol f. We change each occurrence of $f_i f(x_1, \ldots, x_m)$ to x_i, and each occurrence of $g_i f(x_1, \ldots, x_m)$ for $g \neq f$ to fix. Formulas $f(x_1, \ldots, x_m) \neq t$ are changed into $\neg\mathbf{Is}_f(t) \vee (\mathbf{Is}_f(t) \wedge (f_1 t \neq x_1 \vee \ldots \vee f_m t \neq x_m))$; then we apply disjunctive splitting. After this transformation we obtained a disjunction of formulas

$$\exists x_1 \ldots \exists x_m(B_1 \wedge \ldots \wedge B_k), \tag{7}$$

where each B_i has one of the forms (4) but with x replaced by some x_j, or the form $x_j = t$. If any equation $x_j = t$ appears, we eliminate $\exists x_j$ and replace x_j by t everywhere, so we may assume that (7) contains no $x_j = t$. Now all atoms in (7) are of the forms (4), but with x replaced by x_j's. If x_m is of a basic sort, we can apply the procedure again to the subformula $\exists x_m(B_1 \wedge \ldots \wedge B_k)$. Termination is guaranteed by the following observation: either (7) contains fewer equations than (5) (in the case when at least one $x_j = t$ was eliminated), or the number of equations remains the same but the height of each x_j in (7) is strictly less than the height of

x in (5). If x_m is of a queue sort, then the height of x_m in $\exists x_m(B_1 \wedge \ldots \wedge B_k)$ is 0, since otherwise the segment depth of x would be more than 1. Without loss of generality assume that each B_i contains x_m. Then each B_i is of a form $x_m \neq t_i$ for some term t_i not containing x_m. Since every queue sort is infinite, the formula $\exists x_m(x_m \neq t_1 \wedge \ldots \wedge x_m \neq t_k)$ is valid, so $\exists x_m$ can be eliminated. □

We want to prove a statement similar to Lemma 3.8, but for queue sorts.

Consider the set of literals

$$Lq = t,\ Lq = Mq,\ q \neq t,\ Lq \neq t,\ Lq \neq Mq,\ \mathbf{Is}_f(Lq), \tag{8}$$

where L, M are *nonempty* and t is a term in which q does not occur.

LEMMA 3.9. *Consider the class $\mathfrak{C}$ of formulas of the form*

$$\exists q(A_1(q) \wedge \ldots \wedge A_n(q)),$$

where q is of a queue sort $queue(\alpha)$, each A_i is of the form (8), and the segment depth of q in C is at most 1. The class of quantifier-free formulas is a quantifier elimination reduction class for $\mathfrak{C}$.

This lemma requires more elaborate reasoning and will be proved after a series of other statements. We will explain the proof informally. The proof is carried out using the case analysis on the length of the queue q. For "short" queues, we will eliminate $\exists q$ immediately; for "sufficiently long" ones we will do the same trick as for the basic sorts, when we substituted $f(x_1, \ldots, x_n)$ for x, but we will substitute a "sufficiently long" term for q. To understand what is "sufficiently long," we will introduce the notions of left and right depths of q in a formula. For long enough queues we have to understand how to deal with quantifier-free formulas like (2), which express "deep" properties of queues. As we will see, all these deep properties of queues can be formulated as statements about periodicity of queues. So we will introduce special formulas $*_{m,n}$ expressing periodicity of queues, and prove several properties of these formulas. After having done all this preliminary work, quantifier elimination will be more or less straightforward.

Now we will introduce notation for "short" and "long" queues. For any terms $t_1, \ldots, t_n : \alpha$, where $n \geq 0$, we denote by $|t_1 \ldots t_n|$ the term

$$ladd(t_1, ladd(t_2, \ldots ladd(t_n, \varepsilon) \ldots)).$$

Evidently, this term denotes the queue consisting of elements $t_1, \ldots, t_n$. Note that $||$ is ε. Likewise, for terms $t_1, \ldots, t_n, s_1, \ldots, s_m : \alpha$ and $w : queue(\alpha)$ we denote by $|t_1 \ldots t_n \underline{w} s_1 \ldots s_m|$ the term

$$ladd(t_1, ladd(t_2, \ldots ladd(t_n, radd(s_m, radd(s_{m-1}, \ldots radd(s_1, w) \ldots))) \ldots)).$$

Evidently, this term denotes the queue obtained from w by adding the sequence $t_1, \ldots, t_n$ on the left and the sequence $s_1, \ldots, s_m$ on the right.

For each positive integer m, we denote by $length(q) < m$ the following formula

$$q = \varepsilon \vee \bigvee_{k=1\ldots m-1} \exists x_1 \ldots \exists x_k (q = |x_1 \ldots x_k|).$$

The following lemma asserts that every queue is either "short" (of length less than $l + r - 1$) or "sufficiently long" (of length at least $l + r$).

LEMMA 3.10. *For every nonnegative integers l, r such that $l+r > 0$ the following formula is a tautology:*

$$(length(q) < l + r) \lor \exists z_1 \dots \exists z_l \exists u_1 \dots \exists u_r \exists w (q = |z_1 \dots z_l \underline{w} u_1 \dots u_r|). \quad (9)$$

Let us now continue the proof of Lemma 3.9.

PROOF OF LEMMA 3.9 (continued). Recall that we are eliminating quantifiers in a formula C of the form

$$\exists q (A_1(q) \land \dots \land A_n(q)), \quad (10)$$

where q is of a queue sort $queue(\alpha)$, each A_i is of the form (8), and the segment depth of q in C is at most 1.

We define the *left (right) depth* of q in (10) as the maximal number of $lhead_\alpha$'s and $ltail_\alpha$'s (respectively, $rhead_\alpha$'s and $rtail_\alpha$'s) in all prefixes L such that Lq appears in (10). For example, the left depth of q in the formula $lhead_\alpha ltail_\alpha q = x \land q \neq w$ is 2 and the right depth is 0.

Let the left depth of q in (10) be l and the right depth be r. Using Lemma 3.10 and disjunctive splitting on formula (9), we can replace (10) by a disjunction of formulas of the following two forms:

$$\exists q \exists x_1 \dots \exists x_k (q = |x_1 \dots x_k| \land A_1(q) \land \dots \land A_n(q)), \quad (11)$$

$$\begin{aligned}\exists q \exists z_1 \dots \exists z_l \exists w \exists u_1 \dots \exists u_r (q = |z_1 \dots z_l \underline{w} u_1 \dots u_r| \land \\ A_1(q) \land \dots \land A_n(q)).\end{aligned} \quad (12)$$

So it remains to show how to eliminate quantifiers in such formulas.

Any formula (11) is equivalent to

$$\exists x_1 \dots \exists x_k (A_1(|x_1 \dots x_k|) \land \dots \land A_n(|x_1 \dots x_k|)).$$

Recall that every $A_i(q)$ is the form (8). Simplify all literals $A_i(|x_1 \dots x_k|)$ using the following valid equations:

$$\begin{array}{ll} lhead(|y_1 \dots y_m|) = y_1, & ltail(|y_1 \dots y_m|) = |y_2 \dots y_m|, \\ rhead(|y_1 \dots y_m|) = y_m, & rtail(|y_1 \dots y_m|) = |y_1 \dots y_{m-1}|, \\ ltail(||) = ||, & lhead(||) = \mathsf{fix}, \\ rtail(||) = ||, & rhead(||) = \mathsf{fix}. \end{array}$$

Then we obtain a formula of the form

$$\exists x_1 \dots \exists x_k (B_1 \land \dots \land B_n),$$

where the B_i's are formulas $p = s$, $p \neq s$, $\mathbf{Is}_f(p)$, $|p_1 \dots p_l| = |s_1 \dots s_u|$, and $|p_1 \dots p_l| \neq |s_1 \dots s_u|$, where all p's and s's are either x_i's or terms in which x_i's do not occur. Formulas $|p_1 \dots p_l| = |s_1 \dots s_u|$ for $l \neq u$ can be replaced by $\perp$, and for $l = u$ by $p_1 = s_1 \land \dots \land p_l = s_l$, and similarly for formulas $|p_1 \dots p_l| \neq |s_1 \dots s_u|$. After this transformation and disjunctive splitting we can assume that all B_i's are formulas $p = s$, $p \neq s$, $\mathbf{Is}_f(p)$; then we can proceed as in the proof of Lemma 3.8 to get rid of all quantifiers $\exists x_i$.

Now it remains to show how to eliminate quantifiers in (12). Formula (12) is equivalent to

$$\exists z_1 \ldots \exists z_l \exists w \exists u_1 \ldots \exists u_r (A_1(|z_1 \ldots z_l \underline{w} u_1 \ldots u_r|) \wedge \ldots \wedge A_n(|z_1 \ldots z_l \underline{w} u_1 \ldots u_r|)).$$

Recall that every $A_i(q)$ is of the form (8). Simplify all literals $A_i(|z_1 \ldots z_l \underline{w} u_1 \ldots u_r|)$ using the following valid equations:

$$\begin{aligned} \mathit{lhead}(|y_1 \ldots y_m \underline{w} v_1 \ldots v_k|) &= y_1, \\ \mathit{ltail}(|y_1 \ldots y_m \underline{w} v_1 \ldots v_k|) &= |y_2 \ldots y_m \underline{w} v_1 \ldots v_k|, \\ \mathit{rhead}(|y_1 \ldots y_m \underline{w} v_1 \ldots v_k|) &= v_k, \\ \mathit{rtail}(|y_1 \ldots y_m \underline{w} v_1 \ldots v_k|) &= |y_1 \ldots y_m \underline{w} v_1 \ldots v_{k-1}|. \end{aligned}$$

Note that because of the choice of l and r as the left and the right depth of q in (10), this transformation will remove *all* applications of *lhead*, *rhead*, *ltail*, *rtail* to w from the formula. Further simplify the obtained formula using the following equivalences:

$$\begin{aligned} &|y_1 \ldots y_m \underline{w} v_1 \ldots v_k| = |p_1 \ldots p_l \underline{w} q_1 \ldots q_u| \leftrightarrow \\ &\quad y_1 = p_1 \wedge |y_2 \ldots y_m \underline{w} v_1 \ldots v_k| = |p_2 \ldots p_l \underline{w} q_1 \ldots q_u|, \\ &|y_1 \ldots y_m \underline{w} v_1 \ldots v_k| = |p_1 \ldots p_l \underline{w} q_1 \ldots q_u| \leftrightarrow \\ &\quad v_k = q_u \wedge |y_1 \ldots y_m \underline{w} v_1 \ldots v_{k-1}| = |p_1 \ldots p_l \underline{w} q_1 \ldots q_{u-1}|, \end{aligned}$$

and similarly for $\neq$ instead of $=$. Using disjunctive splitting, when necessary, we obtain a formula of the form

$$\exists x_1 \ldots \exists x_k \exists w (B_1 \wedge \ldots \wedge B_n),$$

where the B_i's are formulas of the following forms

$$\begin{array}{cccc} x_i = x_j, & x_i = t, & \mathbf{Is}_f(x_i), & w = t, \\ x_i \neq x_j, & x_i \neq t, & & w \neq t \\ \multicolumn{2}{c}{|y_1 \ldots y_m \underline{w}| = |\underline{w} q_1 \ldots q_u|,} & \multicolumn{2}{c}{|y_1 \ldots y_m \underline{w} v_1 \ldots v_k| = t,} \\ \multicolumn{2}{c}{|y_1 \ldots y_m \underline{w}| \neq |\underline{w} q_1 \ldots q_u|,} & \multicolumn{2}{c}{|y_1 \ldots y_m \underline{w} v_1 \ldots v_k| \neq t,} \end{array}$$

where y_i, v_i, q_i are variables from the set $\{x_1, \ldots, x_k\}$, and where t are terms not containing any x_i's and w.

In the equalities $|y_1 \ldots y_m \underline{w}| = |\underline{w} q_1 \ldots q_u|$ and $|y_1 \ldots y_m \underline{w}| \neq |\underline{w} q_1 \ldots q_u|$ we can assume $m = u$; otherwise they can be replaced by $\perp$ and $\top$, respectively.

Equalities $|y_1 \ldots y_m \underline{w} v_1 \ldots v_k| = t$ can be dealt with in the standard way: we replace any equation $\mathit{ladd}(y_i, s) = t$ by $t \neq \varepsilon \wedge \mathit{lhead}(t) = y_i \wedge \mathit{rtail}(t) = s$, and similar for *radd* and for literals $|y_1 \ldots y_m \underline{w} v_1 \ldots v_k| \neq t$.

Equalities $w = t$ can be dealt with as in Lemma 3.7, by removing them and replacing w by t.

So we can assume that all B_i's have one of the following forms

$$x_i = x_j, \quad x_i = t, \quad x_i \neq x_j, \quad x_i \neq t, \quad \mathbf{Is}_f(x_i), \tag{13}$$

$$|y_1 \ldots y_m \underline{w}| = |\underline{w} q_1 \ldots q_m|, \quad |y_1 \ldots y_m \underline{w}| \neq |\underline{w} q_1 \ldots q_m|, \quad w \neq t. \tag{14}$$

The variable w does not occur in equations (13) and the variables x_j's in these equations are of segment depth 0. They can be dealt with as in Lemma 3.7 after we eliminate $\exists w$, so we will only consider the case when all B_i's have one of the forms in (14).

We will now study properties of formulas $|y_1 \dots y_m \underline{w}| = |\underline{w} q_1 \dots q_m|$, and then return to the proof of Lemma 3.9 again. $\square$

For all sorts α and integers m, k with $m > k \geq 0$, denote by $*_{m,k,\alpha}(w, z_1, \dots, z_m)$ the following formula:

$$|z_1 \dots z_m \underline{w}| = |\underline{w} z_{k+1} \dots z_m z_1 \dots z_k|.$$

As usual, we will omit α and write $*_{m,k}$ instead. One can notice that the formula $*_{m,k}$ express some kind of periodicity properties of w. For example, the formula $*_{4,0}(w, a, b, a, b)$ expresses that w is obtained by repeating the sequence ab zero or more times.

We call a sequence of variables or elements $t_1 \dots t_m$ *periodic* if it has the form

$$\underbrace{s_1 \dots s_n \dots s_1 \dots s_n}_{k \text{ times}}, \tag{15}$$

for some $k \geq 2$. In this case we also write $t_1 \dots t_m = (s_1 \dots s_n)^k$. Any nonperiodic sequence of variables is called a *pattern*. A pattern $s_1, \dots, s_n$ is said to be the *pattern of a sequence* $t_1, \dots, t_m$ of variables, if $t_1, \dots, t_m$ has the form (15), for some $k \geq 1$. For example, the sequences x, y, x, y and z, z, z, z are periodic; their patterns are x, y and z respectively, but the sequence x, y, x is not periodic; and hence its pattern is x, y, x itself.

Let $e_1, \dots, e_n$ be a sequence of elements of $\mathcal{Q}(\mathcal{S})$ and $z_1, \dots, z_n$ be a sequence of variables. We call $e_1, \dots, e_n$ a *shadow* of $z_1, \dots, z_n$ if for every i, j, $e_i = e_j$ if and only if z_i coincides with z_j.

For every pattern $p_1, \dots, p_n$ denote by $\#^{p_1, \dots, p_n}(z_1, \dots, z_n)$ the formula

$$\bigwedge_{i<j, p_i \neq p_j} z_i \neq z_j \wedge \bigwedge_{i<j, p_i = p_j} z_i = z_j.$$

For example, we have

$$\#^{x,y,y,x}(z_1, z_2, z_3, z_4) \leftrightarrow z_1 \neq z_2 \wedge z_1 \neq z_3 \wedge z_2 \neq z_4 \wedge z_3 \neq z_4 \wedge z_1 = z_4 \wedge z_2 = z_3.$$

It is easy to see that the property "$e_1, \dots, e_n$ is a shadow of $p_1, \dots, p_n$" can be expressed as follows.

LEMMA 3.11. *Let $p_1, \dots, p_n$ be a pattern and $e_1, \dots, e_n$ be elements of $\mathcal{Q}(\mathcal{S})$. Then $e_1, \dots, e_n$ is a shadow of $p_1, \dots, p_n$ if and only if $\#^{p_1, \dots, p_n}(e_1, \dots, e_n)$ holds.*

Let $t_1, \dots, t_m$ be a sequence of terms or elements and $0 \leq k < m$. We call the *k-rotation* of this sequence the sequence

$$t_{k+1}, \dots, t_m, t_1, \dots, t_k.$$

We call a *rotation* of a sequence of terms or elements $t_1, \dots, t_m$ any k-rotation of this sequence, for $0 \leq k < m$. For example all rotations of the sequence x, y, z are x, y, z itself, y, z, x, and z, x, y. All rotations of x, y, x are x, y, x itself, y, x, x and x, x, y.

LEMMA 3.12. *Suppose $e_1, \dots, e_m$ and $d_1, \dots, d_m$ are elements, w is a queue, and $|e_1 \dots e_m \underline{w}| = |\underline{w} d_1 \dots d_m|$. Then $d_1, \dots, d_m$ is a rotation of $e_1, \dots, e_m$. Therefore, the following formula is valid:*

$$|z_1 \dots z_m \underline{w}| = |\underline{w} u_1 \dots u_m| \rightarrow \bigvee_{k=0 \dots m-1} \begin{array}{l} (z_1 = u_{k+1} \wedge \dots \wedge z_{m-k} = u_m \wedge \\ z_{m-k+1} = u_1 \wedge \dots \wedge z_m = u_k \wedge \\ *_{m,k}(w, z_1, \dots, z_m)). \end{array} \tag{16}$$

PROOF. Let $w = |x_1 \dots x_n|$. Consider two cases.

Case $(n \leq m)$. Evidently $|e_1 \dots e_m \underline{w}| = |\underline{w} d_1 \dots d_m|$ is equivalent to the following conjunction:

$$\begin{array}{l} e_1 = x_1 \wedge \dots \wedge e_n = x_n \wedge \\ e_{n+1} = d_1 \wedge \dots \wedge e_m = d_{m-n} \wedge \\ x_1 = d_{m-n+1} \wedge \dots \wedge x_n = d_m, \end{array}$$

which implies $|e_{n+1} \dots e_m e_1 \dots e_n| = |d_1 \dots d_m|$.

Case $(n > m)$. In this case we can represent n as $am + b$, where $b < m$, and w as $|x_1 \dots x_{am} y_1 \dots y_b|$. The equality $|e_1 \dots e_m \underline{w}| = |\underline{w} d_1 \dots d_m|$ can be rewritten as

$$\begin{array}{l} |e_1 \dots e_m x_1 \dots x_{(a-1)m} x_{(a-1)m+1} \dots x_{am} y_1 \dots y_b| \ = \\ |x_1 \dots x_m x_{m+1} \dots x_{am} y_1 \dots y_b d_1 \dots d_m|. \end{array}$$

By considering the first am elements and the rest of elements in the queues of this equation, we obtain

$$|e_1 \dots e_m x_1 \dots x_{(a-1)m}| = |x_1 \dots x_m x_{m+1} \dots x_{am}|; \tag{17}$$

$$|x_{(a-1)m+1} \dots x_{am} y_1 \dots y_b| = |y_1 \dots y_b d_1 \dots d_m|. \tag{18}$$

It is not hard to argue that (17) implies

$$\begin{array}{r} x_1 = e_1 \wedge \dots \wedge x_m = e_m \wedge \\ x_{m+1} = e_1 \wedge \dots \wedge x_{2m} = e_m \wedge \\ \dots \\ x_{(a-1)m+1} = e_1 \wedge \dots \wedge x_{am} = e_m. \end{array}$$

Therefore, (18) can be rewritten as

$$|e_1 \dots e_m y_1 \dots y_b| = |y_1 \dots y_b d_1 \dots d_m|.$$

Now $b < m$, so we can use the proof of the previous case to show that $d_1, \dots, d_m$ is a rotation of $e_1, \dots, e_m$. $\square$

In our proofs below we will use a standard result about word equations:

LEMMA 3.13. *Let X, Y be words and $XY = YX$. Then there exists a nonperiodic word S and integers $a, b \geq 0$ such that $X = S^a$ and $Y = S^b$.*

This result is proved as Propositions 1.2 of Hmelevskiĭ [1976, page 5].

Given an equality of the form $|e_1 \dots e_m \underline{q}| = |\underline{q} e_{k+1} \dots e_m e_1 \dots e_k|$, we now want to prove some statements about the sequence $e_1 \dots e_m$. Let us denote by $(e_1 \dots e_r)^*$ any sequence of the form

$$\underbrace{e_1 \dots e_r \quad \dots \quad e_1 \dots e_r}_{\text{0 or more times}}$$

LEMMA 3.14. *Let* $|e_1, \dots, e_m|$ *be a nonempty sequence of elements which is equal to its* k*-rotation for* $k > 0$*:*

$$|e_1 \dots e_m| = |e_{k+1} \dots e_m e_1 \dots e_k|. \tag{19}$$

Then the sequence $e_1 \dots e_m$ *is periodic.*

PROOF. Let $X = e_1 \dots e_k$ and $Y = e_{k+1} \dots e_m$. Then $XY = YX$. By Lemma 3.13 there exists a word S and integers $a, b \geq 0$ such that $X = S^a$ and $Y = S^b$. Note that X and Y are nonempty, so $a, b > 0$. We have $e_1 \dots e_m = XY = S^a S^b = S^{a+b}$, so $e_1 \dots e_m$ is periodic. □

One can prove an even stronger property: both $e_1 \dots e_m$ and $e_{k+1} \dots e_m$ are powers of $e_1 \dots e_r$, where $r = \gcd(m, k)$, see Rybina and Voronkov [2000].

LEMMA 3.15. *Let* $p_1, \dots, p_m$ *be a pattern,* $e_1, \dots, e_m$ *a sequence of elements of a sort* α*, and* q *a queue of the sort* $queue(\alpha)$*. Then*

$$\#^{p_1,\dots,p_m}(e_1, \dots, e_m) \wedge *_{m,k}(q, e_1, \dots, e_m)$$

holds if and only if $e_1, \dots, e_m$ *is a shadow of* $p_1, \dots, p_m$ *and*

$$q = \underbrace{e_1 \dots e_m \quad \dots \quad e_1 \dots e_m}_{\text{0 or more times}} e_1 \dots e_k. \tag{20}$$

PROOF. Let us first prove the "if" direction. Suppose that $e_1, \dots, e_m$ is a shadow of $p_1, \dots, p_m$ and

$$q = \underbrace{e_1 \dots e_m \quad \dots \quad e_1 \dots e_m}_{\text{0 or more times}} e_1 \dots e_k.$$

Then $\#^{p_1,\dots,p_m}(e_1, \dots, e_m)$ holds by Lemma 3.11, and evidently q is a solution to the equation $|e_1 \dots e_m \underline{q}| = |\underline{q} e_{k+1} \dots e_m e_1 \dots e_k|$.

Now we prove the "only if" direction. Suppose

$$\#^{p_1,\dots,p_m}(e_1, \dots, e_m) \wedge *_{m,k}(q, e_1, \dots, e_m).$$

Since $\#^{p_1,\dots,p_m}(e_1, \dots, e_m)$, by Lemma 3.11 and the definition of pattern the sequence $e_1, \dots, e_m$ is nonperiodic. Let $q = d_1 \dots d_n$. Then $*_{m,k}(q, e_1, \dots, e_m)$ implies

$$e_1 \dots e_m d_1 \dots d_n = d_1 \dots d_n e_{k+1} \dots e_m e_1 \dots e_k. \tag{21}$$

Let us now prove that

(22) for every $p \geq 0$ and $1 \geq l < m$ such that $mp + l \leq n$ we have $d_{mp+l} = e_l$.

The proof is by induction on m. When $m = 0$, $d_l = e_l$ follows immediately from (21). When $m > 0$, (21) implies $d_{mp+l} = d_{(m-1)p+l}$. By the induction hypothesis, $d_{(m-1)p+l} = e_l$; hence $d_{mp+l} = e_l$.

So (22) is proved. It is easy to see that (22) implies that for some l

$$q = d_1 \dots d_n = \underbrace{e_1 \dots e_m \quad \dots \quad e_1 \dots e_m}_{\text{0 or more times}} e_1 \dots e_l. \tag{23}$$

To conclude the proof it is enough to show $k = l$. Suppose, by contradiction, $k \neq l$. Using (23), we can rewrite (21) into

$$\begin{array}{l} e_1 \dots e_m \underbrace{e_1 \dots e_m \quad \dots \quad e_1 \dots e_m}_{\text{0 or more times}} e_1 \dots e_l = \\ \underbrace{e_1 \dots e_m \quad \dots \quad e_1 \dots e_m}_{\text{0 or more times}} e_1 \dots e_l e_{k+1} \dots e_m e_1 \dots e_k. \end{array}$$

In this equation, remove from both sequences all elements but the last m. The equation becomes

$$e_{l+1} \dots e_m e_1 \dots e_l = e_{k+1} \dots e_m e_1 \dots e_k. \tag{24}$$

Since $k \neq l$, then $k > l$ or $l > k$. Consider the case $k > l$; the other case is symmetric. Equation (24) implies that $e_{l+1} \dots e_m e_1 \dots e_l$ is equal to its $k - l$-rotation $e_{k+1} \dots e_m e_1 \dots e_k$. By Lemma 3.14, $e_{l+1} \dots e_m e_1 \dots e_l$ is periodic. But $e_1 \dots e_m$ is a rotation of a periodic sequence $e_{l+1} \dots e_m e_1 \dots e_l$, and hence it is periodic too. But we noted above that $e_1 \dots e_m$ is nonperiodic, so we obtain a contradiction. ☐

There is a shorter proof of this result using Proposition 1.23 of Hmelevskiĭ [1976]. However, the arguments of our proof will by used in the next lemma, for which we will need a more precise statement than the proposition of Hmelevskiĭ [1976].

Lemma 3.16. *Let $z_1, \dots, z_n$ be a periodic sequence of variables and $z_1, \dots, z_m$ be its pattern. Then the following formula is valid:*

$$\begin{array}{l} \#^{z_1,\dots,z_m}(z_1, \dots, z_m) \wedge *_{n,l}(q, z_1, \dots, z_n) \leftrightarrow \\ \#^{z_1,\dots,z_m}(z_1, \dots, z_m) \wedge *_{m,k}(q, z_1, \dots, z_m), \end{array} \tag{25}$$

where $k = (l \bmod m)$.

Proof. Let r be the number such that $n = mr$. Let $e_1, \dots, e_m$ be elements such that $\#^{z_1,\dots,z_m}(e_1, \dots, e_m)$. Extend $e_1, \dots, e_m$ to the periodic sequence $e_1, \dots, e_n$ by repeating the sequence $e_1, \dots, e_m$ r times. Take any queue q. It is enough to prove that $*_{n,l}(q, e_1, \dots, e_n)$ implies $*_{m,k}(q, e_1, \dots, e_m)$, and vice versa, $*_{m,k}(q, e_1, \dots, e_m)$ implies $*_{n,l}(q, e_1, \dots, e_n)$.

1. *We suppose $*_{n,l}(q, e_1, \dots, e_n)$ and prove $*_{m,k}(q, e_1, \dots, e_m)$.* Using the same arguments as in Lemma 3.15 (when we proved (23)) we can prove that

$$q = \underbrace{e_1 \dots e_m \quad \dots \quad e_1 \dots e_m}_{\text{0 or more times}} e_1 \dots e_p \tag{26}$$

for some $p < m$. By substituting q by the right-hand side of this equation in the formula $*_{n,l}(q, e_1, \dots, e_n)$ and repeating the arguments of Lemma 3.15 again (when we proved $k = l$ in that lemma), we can prove $p = k$. By Lemma 3.15, (26) implies $*_{m,p}(q, e_1, \ldots, e_m)$, and hence by $p = k$ we have $*_{m,k}(q, e_1, \ldots, e_m)$.

2. *We suppose* $*_{m,k}(q, e_1, \ldots, e_m)$ *and prove* $*_{n,l}(q, e_1, \ldots, e_n)$. By Lemma 3.15

$$q = \underbrace{e_1 \ldots e_m \quad \ldots \quad e_1 \ldots e_m}_{\text{0 or more times}} e_1 \ldots e_k.$$

It is not hard to show that for any such queue q we have $*_{n,l}(q, e_1, \ldots, e_n)$.

□

LEMMA 3.17. *Let* $z_1, \ldots, z_m$ *and* $u_1, \ldots, u_m$ *be two patterns and* $k \neq l$. *Then the following formula is valid:*

$$\begin{array}{l} \#^{z_1,\ldots,z_m}(z_1, \ldots, z_m) \wedge \#^{u_1,\ldots,u_m}(u_1, \ldots, u_m) \wedge \\ *_{m,k}(q, z_1, \ldots, z_m) \rightarrow \neg *_{m,l}(q, u_1, \ldots, u_m). \end{array} \tag{27}$$

PROOF. Take arbitrary elements $z_1, \ldots, z_m, u_1, \ldots, u_m, q$ such that

$$\begin{array}{l} \#^{z_1,\ldots,z_m}(z_1, \ldots, z_m) \wedge \#^{u_1,\ldots,u_m}(u_1, \ldots, u_m) \wedge \\ *_{m,k}(q, z_1, \ldots, z_m) \wedge *_{m,l}(q, u_1, \ldots, u_m). \end{array}$$

Denote the length of q by l_q. By Lemma 3.15,

$$\#^{z_1,\ldots,z_m}(z_1, \ldots, z_m) \wedge *_{m,k}(q, z_1, \ldots, z_m)$$

implies $k = (l_q \bmod m)$. Likewise, we obtain $l = (l_q \bmod m)$, which contradicts to $k \neq l$. □

LEMMA 3.18. *Let* $z_1, \ldots, z_m$ *and* $u_1, \ldots, u_n$ *be two patterns and* $m \neq n$. *Then the following formula is valid:*

$$\begin{array}{l} \#^{z_1,\ldots,z_m}(z_1, \ldots, z_m) \wedge \#^{u_1,\ldots,u_n}(u_1, \ldots, u_n) \wedge \\ *_{m,k}(q, z_1, \ldots, z_m) \wedge *_{n,l}(q, u_1, \ldots, u_n) \rightarrow \mathit{length}(q) < m + n. \end{array} \tag{28}$$

PROOF. We take any elements $z_1, \ldots, z_m, u_1, \ldots, u_n, q$ such that

$$\#^{z_1,\ldots,z_m}(z_1, \ldots, z_m) \wedge \#^{u_1,\ldots,u_n}(u_1, \ldots, u_n) \wedge *_{m,k}(q, z_1, \ldots, z_m) \wedge *_{n,l}(q, u_1, \ldots, u_n)$$

and prove that q has less than $m + n$ elements. Suppose, by contradiction, that q has at least $m + n$ elements. Assume $m > n$ (the case $n > m$ is symmetric). Since $\#^{z_1,\ldots,z_m}(z_1, \ldots, z_m) \wedge *_{m,k}(q, z_1, \ldots, z_m)$, by Lemma 3.15, q has the form

$$\underbrace{z_1 \ldots z_m \quad \ldots \quad z_1 \ldots z_m}_{\text{0 or more times}} z_1 \ldots z_k. \tag{29}$$

Likewise, since $\#^{u_1,\ldots,u_n}(u_1, \ldots, u_n) \wedge *_{n,l}(q, u_1, \ldots, u_n)$, q has the form

$$\underbrace{u_1 \ldots u_n \quad \ldots \quad u_1 \ldots u_n}_{\text{0 or more times}} u_1 \ldots u_l. \tag{30}$$

Let $r = (m \bmod n)$. Since (29) and (30) coincide, their first $m+n$ elements coincide, too. Then we have

$$z_1 \dots z_m z_1 \dots z_n = u_1 \dots u_n \dots u_1 \dots u_r. \tag{31}$$

This implies $z_1 \dots z_n = u_1 \dots u_n$. Substitute in (31) $u_1 \dots u_n$ for $z_1 \dots z_n$; we obtain

$$\dots u_1 \dots u_n = \dots u_{r+1} \dots u_n u_1 \dots u_r.$$

Thus, the sequence $u_1 \dots u_n$ is equal to its r-rotation. By Lemma 3.14, either $r = 0$ or $u_1 \dots u_n$ is periodic. But $\#^{u_1,\dots,u_n}(u_1, \dots, u_n)$ implies that $u_1 \dots u_n$ is not periodic, so $r = 0$. Recall that $r = (m \bmod n)$, $r = 0$; hence n divides m. But then (31) implies

$$z_1 \dots z_m = u_1 \dots u_n \dots u_1 \dots u_n,$$

and hence $z_1 \dots z_m$ is periodic. This contradicts to $\#^{z_1,\dots,z_m}(z_1, \dots, z_m)$. □

Lemma 3.19. *Let $z_1, \dots, z_m$ and $u_1, \dots, u_m$ be two patterns. Then the following formula is valid*

$$\begin{array}{l} \#^{z_1,\dots,z_m}(z_1, \dots, z_m) \land \#^{u_1,\dots,u_m}(u_1, \dots, u_m) \land \\ *_{m,k}(q, z_1, \dots, z_m) \land *_{m,k}(q, u_1, \dots, u_m) \ \rightarrow \\ q = |z_1, \dots, z_k| \lor (z_1 = u_1 \land \dots \land z_m = u_m). \end{array} \tag{32}$$

Proof. Suppose that $z_1, \dots, z_m, u_1, \dots, u_m$ and q are elements of the term algebra with queues such that

$$\begin{array}{l} \#^{z_1,\dots,z_m}(z_1, \dots, z_m) \land \#^{u_1,\dots,u_m}(u_1, \dots, u_m) \land \\ *_{m,k}(q, z_1, \dots, z_m) \land *_{m,k}(q, u_1, \dots, u_m) \end{array}$$

By Lemma 3.15, this implies

$$q = |(z_1 \dots z_m)^r z_1 \dots z_k| = |(u_1 \dots u_m)^r u_1 \dots u_k|,$$

for some r. If $r = 0$ then $q = |z_1 \dots z_k|$. If $r > 0$ then $q = |(z_1 \dots z_m)^r z_1 \dots z_k| = |(u_1 \dots u_m)^r u_1 \dots u_k|$, and hence $z_1 = u_1, \dots, z_m = u_m$. □

Now we can conclude the proof of Lemma 3.9.

Proof of Lemma 3.9 (continued). Recall that we have to show how to eliminate the quantifier in a formula C of the form $\exists w(B_1 \land \dots \land B_p)$, where all B_i's have one of the forms given in (14):

$$|y_1 \dots y_n \underline{w}| = |\underline{w} q_1 \dots q_n|, \quad |y_1 \dots y_n \underline{w}| \neq |\underline{w} q_1 \dots q_n|, \quad w \neq t.$$

By Lemma 3.12 one can assume that in all such formulas $q_1 \dots q_n$ is a rotation of $y_1 \dots y_n$, so all B_i's have one of the forms

$$|y_1 \dots y_n \underline{w}| = |\underline{w} y_{l+1} \dots y_n y_1 \dots y_l|, \quad |y_1 \dots y_n \underline{w}| \neq |\underline{w} y_{l+1} \dots y_n y_1 \dots y_l|, \quad w \neq t,$$

i.e., one of the forms

$$*_{n,l}(w, y_1, \dots, y_n), \quad \neg *_{n,l}(w, y_1, \dots, y_n), \quad w \neq t.$$

Using tautologies $y_i = y_j \lor y_i \neq y_j$ and disjunctive splitting, we can assume that for each pair i, j either $y_i = y_j$ or $y_i \neq y_j$ occurs in C. If $y_i = y_j$ occurs in C, replace in C y_i by y_j. Assume that $z_1, \ldots, z_m$ is the pattern of $y_1, \ldots, y_n$ after such a replacement. Since for each pair i, j either $y_i = y_j$ or $y_i \neq y_j$ occurs in C, it is not hard to argue that the formula $\#^{z_1,\ldots,z_m}(z_1, \ldots, z_m)$ occurs in C. So we can assume that all B_i's have one of the forms

$$\begin{array}{l}\#^{z_1,\ldots,z_m}(z_1, \ldots, z_m) \wedge *_{n,l}(w, y_1, \ldots, y_n),\\ \#^{z_1,\ldots,z_m}(z_1, \ldots, z_m) \wedge \neg *_{n,l}(w, y_1, \ldots, y_n),\\ w \neq t.\end{array}$$

By Lemma 3.16 one can replace in these formulas $y_1, \ldots, y_n$ by the nonperiodic sequence $z_1, \ldots, z_m$, n by m, and l by a suitable k. So we can assume that all B_i's have one of the forms

$$\#^{z_1,\ldots,z_m}(z_1, \ldots, z_m) \wedge *_{m,k}(w, z_1, \ldots, z_m), \tag{33}$$
$$\#^{z_1,\ldots,z_m}(z_1, \ldots, z_m) \wedge \neg *_{m,k}(w, z_1, \ldots, z_m), \tag{34}$$
$$w \neq t, \tag{35}$$

where the sequence $z_1, \ldots, z_m$ is nonperiodic.

Concerning formulas of the form (35), note the following.

(36) Every formula $\exists w(w \neq t_1 \wedge \ldots \wedge w \neq t_v)$ is valid.

Indeed, every queue sort has an infinite number of elements, so we can always find a w different from $t_1, \ldots, t_v$.

If at least two different B_i's are of the form (33), then C contains the subformula

$$\begin{array}{l}\#^{z_1,\ldots,z_m}(z_1, \ldots, z_m) \wedge *_{m,k}(w, z_1, \ldots, z_m) \wedge\\ \#^{u_1,\ldots,u_{m'}}(u_1, \ldots, u_{m'}) \wedge *_{m',k'}(w, u_1, \ldots, u_{m'}).\end{array} \tag{37}$$

If $m \neq m'$, then by Lemma 3.18 this formula implies $\mathit{length}(w) < m + m'$, so we can eliminate the quantifier $\exists w$ as for formula (11). If $m = m'$ but $k \neq k'$, then by Lemma 3.17 (37) is false, so we eliminate the quantifier again. If $m = m'$ and $k = k'$, then by Lemma 3.19 formula (37) implies

$$w = |z_1, \ldots, z_k| \vee z_1 = u_1 \wedge \ldots \wedge z_m = u_m.$$

Using disjunctive splitting, we can consider the cases $w = |z_1, \ldots, z_k|$ and $z_1 = u_1 \wedge \ldots \wedge z_m = u_m$ separately. If $w = |z_1, \ldots, z_k|$, we can again eliminate $\exists w$ as for formula (11). If $z_1 = u_1 \wedge \ldots \wedge z_m = u_m$, then (37) can be replaced by the equivalent formula

$$\begin{array}{l}\#^{z_1,\ldots,z_m}(z_1, \ldots, z_m) \wedge *_{m,k}(w, z_1, \ldots, z_m) \wedge\\ z_1 = u_1 \wedge \ldots \wedge z_m = u_m,\end{array}$$

which has less occurrences of subformulas of the form (33).

Thus, we can now assume that at least one of B_i's is either of the form (33) or of the form (34), but only at most one of B_i's is of the form (33). Consider two possible cases.

Case (all B_i's are of the form (34)). Then C is of the form

$$\exists w(\bigwedge_i (\#^{z_1,\ldots,z_{m_i}}(z_1, \ldots, z_{m_i}) \wedge \neg *_{m_i,k_i}(w, z_1, \ldots, z_{m_i})) \wedge w \neq t_1 \wedge \ldots \wedge w \neq t_v).$$

The variable w is of the sort $queue(\alpha)$ for some α. Note the following:

(38) *If the domain of α contains exactly one element, then $\neg *_{m,k} (w, z_1, \dots, z_m)$ is false for all m, k.*

So for sorts α with one element C is equivalent to $\perp$. Assume that the domain of α has more than one element. By Lemma 3.15, if $\#^{z_1,\dots,z_{m_i}}(z_1, \dots, z_{m_i})$ holds only when w has period m_i, then let m be the maximum of all m_i's. It is not hard to argue that there is an infinite number of queues having no period between 1 and m. Each of these queues satisfies

$$\bigwedge_i \#^{z_1,\dots,z_{m_i}}(z_1, \dots, z_{m_i}) \wedge \neg *_{m_i,k_i} (w, z_1, \dots, z_{m_i}).$$

Since there is an infinite number of such queues, at least one of them also satisfies $w \neq t_1, \dots, w \neq t_v$, so C can be replaced by $\top$.

Case (exactly one B_i is of the form (33)). We will now show how to get rid of all B_j's of the form (34). Suppose that at least one such B_j occurs in C; then C contains the subformula

$$\begin{array}{l} \#^{z_1,\dots,z_m}(z_1, \dots, z_m) \wedge *_{m,k}(w, z_1, \dots, z_m) \wedge \\ \#^{u_1,\dots,u_{m'}}(u_1, \dots, u_{m'}) \wedge \neg *_{m',k'} (w, u_1, \dots, u_{m'}). \end{array} \tag{39}$$

Consider the following three possible cases.

Subcase ($m \neq m'$). Using the tautology

$$length(w) < m + m' \vee \neg(length(w) < m + m')$$

and disjunctive splitting, we can assume that either $length(w) < m+m'$ or $\neg(length(w) < m + m')$ occurs in C. In the former case, we can eliminate the quantifier as before. In the latter case, (39) becomes

$$\begin{array}{l} \#^{z_1,\dots,z_m}(z_1, \dots, z_m) \wedge *_{m,k}(w, z_1, \dots, z_m) \wedge \\ \#^{u_1,\dots,u_{m'}}(u_1, \dots, u_{m'}) \wedge \neg *_{m',k'} (w, u_1, \dots, u_{m'}) \wedge \\ \neg(length(w) < m + m'). \end{array} \tag{40}$$

By Lemma 3.18, the formula

$$\begin{array}{l} \#^{z_1,\dots,z_m}(z_1, \dots, z_m) \wedge *_{m,k}(w, z_1, \dots, z_m) \wedge \\ \#^{u_1,\dots,u_{m'}}(u_1, \dots, u_{m'}) \wedge \\ \neg(length(w) < m + m') \end{array}$$

implies $\neg *_{m',k'} (w, z_1, \dots, z_{m'})$; hence $\neg *_{m',k'} (w, u_1, \dots, u_{m'})$ can be omitted from (40).

Subcase ($m = m'$ and $k \neq k'$). By Lemma 3.17

$$\begin{array}{l} \#^{z_1,\dots,z_m}(z_1, \dots, z_m) \wedge *_{m,k}(w, z_1, \dots, z_m) \wedge \\ \#^{u_1,\dots,u_{m'}}(u_1, \dots, u_{m'}) \rightarrow \neg *_{m',k'} (w, u_1, \dots, u_{m'}). \end{array}$$

Hence $\neg *_{m',k'} (w, z_1, \dots, z_{m'})$ can be omitted from (39).

Subcase ($m = m'$ and $k = k'$). In this case (39) is

$$\begin{array}{l} \#^{z_1,\ldots,z_m}(z_1,\ldots,z_m) \wedge *_{m,k}(w,z_1,\ldots,z_m) \wedge \\ \#^{u_1,\ldots,u_m}(u_1,\ldots,u_m) \wedge \neg *_{m,k}(w,u_1,\ldots,u_m). \end{array} \tag{41}$$

Consider the tautology

$$\begin{array}{l} w = |z_1,\ldots,z_k| \vee (z_1 = u_1 \wedge \ldots \wedge z_m = u_m) \vee \\ \neg(w = |z_1,\ldots,z_k| \vee (z_1 = u_1 \wedge \ldots \wedge z_m = u_m)). \end{array}$$

Using this tautology and disjunctive splitting we can assume that one of the formulas $w = |z_1,\ldots,z_k|$, or $z_1 = u_1 \wedge \ldots \wedge z_m = u_m$, or $\neg(w = |z_1,\ldots,z_k| \vee (z_1 = u_1 \wedge \ldots \wedge z_m = u_m)$ occurs positively in C. If $w = |z_1,\ldots,z_k|$ occurs positively in C, then we can eliminate $\exists w$ as before. If $z_1 = u_1 \wedge \ldots \wedge z_m = u_m$ occurs positively in C, then (41) can be replaced by $\perp$. The only remaining case is that

$$\neg(w = |z_1,\ldots,z_k| \vee (z_1 = u_1 \wedge \ldots \wedge z_m = u_m))$$

occurs positively in C. By Lemma 3.19

$$\begin{array}{l} \#^{z_1,\ldots,z_m}(z_1,\ldots,z_m) \wedge *_{m,k}(w,z_1,\ldots,z_m) \wedge \\ \neg(w = |z_1,\ldots,z_k| \vee (z_1 = u_1 \wedge \ldots \wedge z_m = u_m)) \wedge \\ \#^{u_1,\ldots,u_m}(u_1,\ldots,u_m) \rightarrow \neg *_{m,k}(w,u_1,\ldots,u_m). \end{array}$$

Hence $\neg *_{m,k}(w,z_1,\ldots,z_m)$ can be removed from (41).

In all three cases we could remove the subformula of the form (34). Therefore, we can assume that C has the form

$$\exists w(*_{m,k}(w,z_1,\ldots,z_m) \wedge w \neq t_1 \wedge \ldots \wedge w \neq t_v).$$

Using Lemma 3.15, one can see that there exists an infinite number of queues w satisfying $*_{m,k}(w,z_1,\ldots,z_m)$, so C is valid.

The proof of Lemma 3.9 is completed. □

LEMMA 3.20. *Let k be a nonnegative integer. The class of formulas $\exists xC$, where C is a conjunction of literals of the form (4), and where the segment depth of x in C is $\leq k$, is a quantifier elimination reduction class for the class of such formulas but with the segment depth of x in C being $\leq k+1$.*

PROOF. Consider any formula $\exists x(C)$. For every prefix Lx occurring in C such that L contains more than one segment do the following. Let $L = MN$, where N is the last segment in L. Introduce a new variable y and replace the formula $\exists x(C)$ by the $\exists y \exists x(Nx = y \wedge C')$, where C' is obtained from C by replacing Lx by My. It is not hard to argue that

$$\exists x(C) \leftrightarrow \exists y \exists x(Nx = y \wedge C').$$

Note that in the new formula $\exists y \exists x(Nx = y \wedge C')$ the segment depth of x is 1. Eliminate the quantifier $\exists x$ in $\exists x(Nx = y \wedge C')$ as in Lemmas 3.8 and 3.9. This elimination may introduce new variables z_i, but only of segment depth 0. By routine inspection of the proofs of 3.9 and 3.8, one can note that the elimination of $\exists x$ does not increase the segment depth of y. We can continue this procedure with y in place of x, until the segment depth of y becomes 1. Then x can be eliminated by Lemmas 3.8 and 3.9. □

Only a few final steps are to be made toward the proof of the Quantifier Elimination Theorem.

Proof of Theorem 3.3. By Lemma 3.9 it is enough to show how to eliminate the quantifier in $\exists x(A_1 \wedge \ldots \wedge A_n)$, where each A_i is of the form (4). By Lemma 3.20 we can decrease the segment depth of x to 0. When the segment depth is 0, the quantifier can be eliminated by Lemmas 3.8 and 3.9. □

As a consequence, we obtain the decidability of the first-order theory of the term algebras with queues.

Theorem 3.21. *The first-order theory of any term algebra with queues is decidable.*

Proof. Given any first-order sentence A of the theory of term algebras with queues, first find an equivalent sentence A' of the language $\mathcal{L}'_{\mathcal{Q}(\mathcal{S})}$. By Quantifier Elimination Theorem 3.3, one can effectively find a quantifier-free sentence A'' equivalent to A', and hence to A. Any quantifier-free sentence is a boolean combination of variable-free literals. Evidently, any literal of A'' can be effectively evaluated to $\top$ or $\perp$, and hence also A'' itself. □

4. AXIOMATIZATION

Our proof gives us the following axiomatization of term algebras with queues. Such an axiomatization can be extracted from our proof, and is presented below. This axiomatization lacks elegance compared to the axiomatizations of term algebras. It is likely that our axiomatization can be slightly simplified, but we doubt that a considerable simplification is possible, since we cannot see how to avoid all combinatorics related to the formulas $*_{m,n}$. The axioms are shown in Figure 2.

In this figure, the variables x, y, z can be of any sort; q, w are of a queue sort; u, v are of a basic sort; F is the set of all function symbols of arity ≥ 1; and C is the set of all constants. For every variable u of a sort α, F_u is the set of all function symbols of types $\ldots \to \alpha$, and C_u is the set of all constants of the sort α. We assume that all axioms respect sorts. The variables in each axiom except $(l1)$–$(l4)$ are pairwise distinct. For example, each variable x_i in $(ta2)$ is distinct from x_j, if $i \neq j$, and from each variable y_j. Further conditions on the axioms are as follows:

1. In $(ta1)$, f is a function symbol of $\mathcal{S}$; in $(ta2)$ f and g are two distinct function symbols of $\mathcal{S}$, possibly of arity 0; in $(ta4)$, x is a proper subterm of $t[x]$.
2. In $(l1)$, $z_1, \ldots, z_n$ is a periodic sequence of variables with the pattern $z_1, \ldots, z_m$ and $k = (l \bmod m)$; in $(l2)$ $k \neq l$; in $(l3)$ $l \neq m$.
3. In $(q9)$ $m \neq n$.
4. In $(u1)$ the domain of α contains exactly one element.

Completeness of this axiomatization can be checked by routine inspection of our proofs.

5. RELATED WORK AND CONCLUDING REMARKS

Bjorner [1998; 2000] discusses the use of queues in verification and gives a decision procedure for the existential theory of queues. He uses a richer signature

1. Term algebra axioms:

$$f(x_1,\ldots,x_n)=f(y_1,\ldots,y_n)\rightarrow x_1=y_1\wedge\ldots\wedge x_n=y_n \quad (ta1)$$

$$f(x_1,\ldots,x_n)\neq g(y_1,\ldots,y_m) \quad (ta2)$$

$$\bigvee_{f\in F_u}\exists\bar{x}(u=f(\bar{x}))\vee\bigvee_{c\in C_u}u=c \quad (ta3)$$

$$x\neq t[x] \quad (ta4)$$

2. Axioms proved in Lemmas 3.16, 3.17, 3.18, and 3.19:

$$\begin{array}{l}\#^{z_1,\ldots,z_m}(z_1,\ldots,z_m)\wedge *_{n,l}(q,z_1,\ldots,z_n)\ \leftrightarrow\\ \#^{z_1,\ldots,z_m}(z_1,\ldots,z_m)\wedge *_{m,k}(q,z_1,\ldots,z_m)\end{array} \quad (l1)$$

$$\begin{array}{l}\#^{x_1,\ldots,x_m}(x_1,\ldots,x_m)\wedge\#^{y_1,\ldots,y_m}(y_1,\ldots,y_m)\wedge\\ *_{m,k}(q,x_1,\ldots,x_m)\rightarrow\neg *_{m,l}(q,y_1,\ldots,y_m)\end{array} \quad (l2)$$

$$\begin{array}{l}\#^{z_1,\ldots,z_m}(z_1,\ldots,z_m)\wedge\#^{u_1,\ldots,u_n}(u_1,\ldots,u_n)\wedge\\ *_{m,k}(q,z_1,\ldots,z_m)\wedge *_{n,l}(q,u_1,\ldots,u_n)\rightarrow \mathit{length}(q)<m+n\end{array} \quad (l3)$$

$$\begin{array}{l}\#^{z_1,\ldots,z_m}(z_1,\ldots,z_m)\wedge\#^{u_1,\ldots,u_m}(u_1,\ldots,u_m)\wedge\\ *_{m,k}(q,z_1,\ldots,z_m)\wedge *_{m,k}(q,u_1,\ldots,u_m)\ \rightarrow\\ q=|z_1,\ldots,z_k|\vee(z_1=u_1\wedge\ldots\wedge z_m=u_m)\end{array} \quad (l4)$$

3. Properties of queues used in the quantifier elimination procedure:

$$\varepsilon\neq \mathit{ladd}(x,w) \quad (q1)$$

$$\varepsilon\neq \mathit{radd}(x,w) \quad (q2)$$

$$w=\varepsilon\vee\exists x\exists w'(w=\mathit{ladd}(x,w')) \quad (q3)$$

$$w=\varepsilon\vee\exists x\exists w'(w=\mathit{radd}(x,w')) \quad (q4)$$

$$\mathit{ladd}(x_1,w_1)=\mathit{ladd}(x_2,w_2)\rightarrow x_1=x_2\wedge w_1=w_2 \quad (q5)$$

$$\mathit{radd}(x_1,w_1)=\mathit{radd}(x_2,w_2)\rightarrow x_1=x_2\wedge w_1=w_2 \quad (q6)$$

$$\mathit{ladd}(x,\mathit{radd}(y,w))=\mathit{radd}(y,\mathit{ladd}(x,w)) \quad (q7)$$

$$\mathit{ladd}(x_1,\ldots\mathit{ladd}(x_n,\varepsilon)\ldots)=\mathit{radd}(x_n,\ldots\mathit{radd}(x_1,\varepsilon)\ldots) \quad (q8)$$

$$|x_1\ldots x_m\underline{w}|\neq|\underline{w}y_1\ldots y_n| \quad (q9)$$

4. An axiom for any basic sort α whose domain has exactly one element:

$$\mathit{ladd}_\alpha(u,w)=\mathit{radd}_\alpha(u,w) \quad (u1)$$

Fig. 2. Axiomatization of the term algebra with queues.

than we do. The signature of queues in Bjørner [1998] contains the binary relation "subqueue;" with such relations the first-order theory of queues (and even the $\forall\exists$-theory) is undecidable. Formulas expressing periodicity of queues also play a major role in Bjørner [2000]. Enrichment of first-order theory of tree algebras with cyclic descriptions was also studied in Peltier [1997]. Comon et al. [1994] prove decidability of factors of term algebra whose equational theory is axiomatizable by a finite set of *shallow* equations. It seems that their technique is not applicable in our case, since our axiomatization contains an infinite number of equations $ladd(x_1, \ldots ladd(x_n, \varepsilon) \ldots) = radd(x_n, \ldots radd(x_1, \varepsilon) \ldots)$.

Maĺcev [1961] was the first to prove decidability of term algebras, he used quantifier elimination. Other quantifier elimination procedures can be found in Kunen [1987], Belegradek [1988a], Maher [1988], Comon and Lescanne [1989], Hodges [1993]. Each of Maĺcev [1961], Maher [1988], and Hodges [1993] gives an axiomatization of term algebras; Maher [1988] also provides such an axiomatization for rational trees. The technique of using decomposition functions f_i is due to Kunen [1987] and Belegradek [1988a].

The decision procedure of Maĺcev [1961] gives a nonelementary upper bound on the first-order theory of term algebras. The first-order theory of term algebras also has a hereditary nonelementary lower bound. This result follows from Compton and Henson [1990]: the theory of any pairing function (i.e., binary one-to-one function) has a hereditary nonelementary lower bound. Obviously, such a function is definable in term algebras having at least one binary function symbol in the signature. When the signature contains only unary function symbols, the first-order theory of term algebras is LATIME($2^{O(n)}$): the class of problems solvable by alternating Turing machines running in exponential time but only with a linear number of alternations; for details see Ferrante and Rackoff [1979], Berman [1980], Volger [1983b; Volger [1983a], and Vorobyov and Voronkov [1998]. Complexity results for some fragments of the first-order theory of term algebras are proved in Vorobyov and Voronkov [1998].

Evidently, if we consider the first-order theory of term algebras with queues, the nonelementary lower bound follows from the nonelementary lower bound for term algebras.

If we only consider queues over a finite domain, then our results prove the decidability of the first-order theory of words with the operations of the left and right multiplications by a letter. The weak monadic second-order theory of $0,1$-words with right multiplication by 0 and 1 and left multiplication by 0 is known to be undecidable [Läuchly and Savoiz 1987; Seese 1972] (see also Thomas [1990]).

If we only consider queues over a finite domain, then LATIME($2^{O(n)}$)-hardness for queues follows from such a lower bound for term algebras with unary function symbols, since we can define term algebras with unary function symbols in any theory of queues, even queues over the domain of two elements. It is interesting to find the precise upper bound for the theory of queues over a finite (or infinite) domain; one can prove the upper bound LATIME($2^{O(n)}$) if the key result of Ferrante and Rackoff [1979] could be proved for queues: quantifiers in any quantified formula can be bound by elements of the exponential height (see also Vorobyov and Voronkov [1998]). However, our proof does not imply such a bound on the number of elements in a queue. A related question is the complexity of any term algebra with queues and unary functions: one cannot directly represent a pairing function in such structures, but these structures contain quite complex objects, for example queues of queues.

REFERENCES

ReferencesRef

BELEGRADEK, O. 1988a. Model theory of locally free algebras (in Russian). In *Model Theory and its Applications*. Trudy Instituta Matematiki, vol. 8. Nauka, Novosibirsk, 3–24. English summary in [Belegradek 1988b].

BELEGRADEK, O. 1988b. Some model theory of locally free algebras (in Russian). In *6th Easter Conference on Model Theory*, B. Dahn and H. Wolter, Eds. Humboldt University, Berlin, 28–32.

BERMAN, L. 1980. The complexity of logical theories. *Theoretical Computer Science 11*, 71–77.

BJØRNER, N. 1998. Reactive verification with queues. In *ARO/ONR/NSF/DARPA Workshop on Engineering Automation for Computer-Based Systems*. Carmel, CA, 1–8.

BJØRNER, N. 2000. Integrating decision procedures for temporal verification. Ph.D. thesis, Stanford University.

COMON, H. AND LESCANNE, P. 1989. Equational problems and disunification. *Journal of Symbolic Computations 7,* 3,4, 371–425.

COMON, H., HABERSTRAU, M., AND JOUANNAUD, J.-P. 1994. Syntacticness, cycle-syntacticness and shallow theories. *Journal of Symbolic Computations 111,* 1, 154–191.

COMPTON, K. AND HENSON, C. 1990. A uniform method for proving lower bounds on the computational complexity of logical theories. *Annals of Pure and Applied Logic 48*, 1–79.

FERRANTE, J. AND RACKOFF, C. W. 1979. *The computational complexity of logical theories*. Lecture Notes in Mathematics, vol. 718. Springer-Verlag.

HMELEVSKIĬ, J. 1976. *Equations in free semigroups*. Proceedings of the Steklov Institute of Mathematics, vol. 107. American Mathematical Society. Translated from Russian, the Russian edition appeared in 1971.

HODGES, W. 1993. *Model theory*. Cambridge University Press.

KUNEN, K. 1987. Negation in logic programming. *Journal of Logic Programming 4*, 289–308.

LÄUCHLY, H. AND SAVOIZ, C. 1987. Monadic second-order definable relations on the binary tree. *Journal of Symbolic Logic 52*, 219–226.

MAHER, M. 1988. Complete axiomatizations of the algebras of finite, rational and infinite trees. In *Proc. IEEE Conference on Logic in Computer Science (LICS)*. 348–357.

MAĹCEV, A. 1961. On the elementary theories of locally free universal algebras. *Soviet Mathematical Doklady 2,* 3, 768–771. Reprinted in [Maĺcev 1971].

MAĹCEV, A. 1971. Axiomatizable classes of locally free algebras of various types. In *The Metamathematics of Algebraic Systems. Anatoliĭ Ivanovič Maĺcev. Collected Papers: 1936–1967*, B. Wells III, Ed. Vol. 66. North Holland, Chapter 23, 262–281.

PELTIER, N. 1997. Increasing model building capabilities by constraint solving on terms with integer exponents. *Journal of Symbolic Computations 24,* 1, 59–101.

RYBINA, T. AND VORONKOV, A. 2000. A decision procedure for term algebras with queues. Preprint CSPP-2, Department of Computer Science, University of Manchester. Feb.

SEESE, D. 1972. Entscheidbarkeits- und Definierbarkeitsfragen der Theorie "netzartiger" Graphen I. *Wiss. Zeitschr. der Humboldt-Univ. zu Berlin, Math.-Nat. Reihe 21*, 513–517.

THOMAS, W. 1990. Automata on infinite objects. In *Handbook of Theoretical Computer Science*, J. Van Leeuwen, Ed. Vol. B: Formal Methods and Semantics. Elsevier Science, Amsterdam, Chapter 4, 133–191.

VISSER, W., BARRINGER, H., FELLOWS, D., GOUGH, G., AND WILLIAMS, A. 1998. Efficient CTL model checking for analysis of Rainbow designs. In *Advances in Hardware Design and Verification, Procedings of CHARME'97*, H. Li and K. Probst, Eds. Chapman and Hall, Montreal.

VOLGER, H. 1983a. A new hierarchy of elementary recursive decision problems. *Methods of Operations Research 45*, 509–519.

Volger, H. 1983b. Turing machines with linear alternation, theories of bounded concatenation and the decision problem of first order theories (Note). *Theoretical Computer Science 23*, 333–337.

Vorobyov, S. and Voronkov, A. 1998. Complexity of nonrecursive logic programs with complex values. In *PODS'98*. ACM Press, Seattle, Washington, 244–253.

Received April 2000; revised October 2000; accepted October 2000