Discrete Mathematics and Theoretical Computer Science DMTCS vol. **26:1**, 2014, 159–178

Canonical forms for free κ-semigroups

José Carlos Costa

CMAT & Dep. Matemática e Aplicações, Universidade do Minho, Braga, Portugal

received 29th May 2013, revised 6th Jan. 2014, accepted 28th Feb. 2014.

The implicit signature κ consists of the multiplication and the $(\omega - 1)$-power. We describe a procedure to transform each κ-term over a finite alphabet A into a certain canonical form and show that different canonical forms have different interpretations over some finite semigroup. The procedure of construction of the canonical forms, which is inspired in McCammond's normal form algorithm for ω-terms interpreted over the pseudovariety $\mathbf{A}$ of all finite aperiodic semigroups, consists in applying elementary changes determined by an elementary set Σ of pseudoidentities. As an application, we deduce that the variety of κ-semigroups generated by the pseudovariety $\mathbf{S}$ of all finite semigroups is defined by the set Σ and that the free κ-semigroup generated by the alphabet A in that variety has decidable word problem. Furthermore, we show that each ω-term has a unique ω-term in canonical form with the same value over $\mathbf{A}$. In particular, the canonical forms provide new, simpler, representatives for ω-terms interpreted over that pseudovariety.

Keywords: Pseudovariety, implicit signature, κ-term, word problem, McCammond's normal form, finite semigroup, κ-semigroup, regular language

1 Introduction

A κ-term is a formal expression obtained from the letters of an alphabet A using two operations: the binary, associative, concatenation and the unary $(\omega - 1)$-power. Instead of working only with κ-terms, we will operate in a larger set $T_A^{\bar{\kappa}}$ of terms, called $\bar{\kappa}$-terms in [4] (in which $\bar{\kappa}$ is called the *completion* of κ), obtained from A using the binary concatenation and the unary $(\omega + q)$-power for each integer q. Any $\bar{\kappa}$-term can be given a natural interpretation on each finite semigroup S: the concatenation is interpreted as the semigroup multiplication while the $(\omega + q)$-power is the unary operation which sends each element s of S to: s^ω, the unique idempotent power of s, when $q = 0$; $s^\omega s^q$, denoted $s^{\omega+q}$, when $q > 0$; the inverse of $s^{\omega-q}$ in the maximal subgroup containing s^ω, when $q < 0$. For a class $\mathcal{C}$ of finite semigroups and $\bar{\kappa}$-terms α and β, we say that $\mathcal{C}$ satisfies the $\bar{\kappa}$-identity $\alpha = \beta$, and write $\mathcal{C} \models \alpha = \beta$, if α and β have the same interpretation over every semigroup of $\mathcal{C}$. The $\bar{\kappa}$-word (resp. κ-word) problem for $\mathcal{C}$ consists in deciding, given a $\bar{\kappa}$-identity (resp. a κ-identity) $\alpha = \beta$, whether $\mathcal{C} \models \alpha = \beta$. The κ-word problem for $\mathcal{C}$ is certainly a subproblem of the $\bar{\kappa}$-word problem for $\mathcal{C}$. Conversely, each finite semigroup verifies $x^{\omega+q} = x^{\omega-1}x^{q+1}$ for $q \geq 0$ and $x^{\omega+q} = (x^{\omega-1})^{-q}$ for $q < 0$. This means that for each $\bar{\kappa}$-term there exists a well determined κ-term with the same interpretation over every finite semigroup. As a consequence, the word problems for κ-terms and for $\bar{\kappa}$-terms over $\mathcal{C}$ are equivalent problems.

A pseudovariety of semigroups is a class of finite semigroups closed under taking subsemigroups, homomorphic images and finite direct products. We also remember that κ and $\bar{\kappa}$ are instances of so called implicit signatures [7], that is, sets of implicit operations on finite semigroups containing the multiplication. A motivation to prove the decidability of the σ-word problem, for an implicit signature σ and a pseudovariety $\mathbf{V}$, is that this is one of the properties required for $\mathbf{V}$ to be a *σ-tame* pseudovariety. The tameness property of pseudovarieties was introduced by Almeida and Steinberg [7] with the purpose of solving the decidability problem for iterated semidirect products of pseudovarieties. Although that objective has not yet been reached, tameness has proved to be of interest to solve membership problems involving other types of operators [5]. For pseudovarieties of aperiodic semigroups it is common to use the signature ω, consisting of the multiplication and the ω-power. A solution to the ω-word problem has been obtained for the pseudovariety $\mathbf{A}$ of all finite aperiodic semigroups [15, 16] as well as for some of its most important subpseudovarieties such as $\mathbf{J}$ of $\mathcal{J}$-trivial semigroups [1], $\mathbf{LSl}$ of local semilattices [10] and $\mathbf{R}$ of $\mathcal{R}$-trivial semigroups [9]. For non-aperiodic examples, in which the ω-power is not enough, we refer to the pseudovarieties $\mathbf{CR}$ of completely regular semigroups [8] and $\mathbf{LG}$ of local groups [12] for which the κ-word problem is solved.

In this paper, we study the $\bar{\kappa}$-word problem (and so, equivalently, the κ-word problem) for the pseudovariety $\mathbf{S}$ of all finite semigroups. A positive solution to this problem has been announced and outlined by Zhil'tsov in [17] but, unfortunately, the author died without publishing a full version of that note. Our approach is completely independent and consists of three stages. First, we declare some elements of $T_A^{\bar{\kappa}}$ to be in a certain *canonical form*. Next, we show that an arbitrary $\bar{\kappa}$-term can be algorithmically transformed into one in canonical form and with the same value over $\mathbf{S}$. Finally, we prove that distinct canonical forms have different interpretations on some finite semigroup. This shows that for each $\bar{\kappa}$-term there is exactly one in canonical form with the same value over $\mathbf{S}$. To test whether a $\bar{\kappa}$-identity $\alpha = \beta$ holds over $\mathbf{S}$, it then suffices to verify if the canonical forms of the $\bar{\kappa}$-terms α and β are equal, thus proving the decidability of the word problem for $\Omega_A^{\bar{\kappa}}\mathbf{S}$, the free $\bar{\kappa}$-semigroup on A, via the homomorphism of $\bar{\kappa}$-semigroups $\eta : T_A^{\bar{\kappa}} \to \Omega_A^{\bar{\kappa}}\mathbf{S}$ that sends each $a \in A$ to itself. The canonical forms we use, as well as the procedure of their construction, are close to the normal forms introduced by McCammond [15] for ω-terms over $\mathbf{A}$. For this reason, we adopt some of McCammond's terminology. The proof of correctness of our algorithm is achieved by associating to each $\bar{\kappa}$-term α a family of regular languages $L_{\mathtt{n},\mathtt{p}}(\alpha)$, where $\mathtt{n}$ and $\mathtt{p}$ are positive integers. The key property is that, if α and β are $\bar{\kappa}$-terms in canonical form such that $L_{\mathtt{n},\mathtt{p}}(\alpha) \cap L_{\mathtt{n},\mathtt{p}}(\beta) \neq \emptyset$ for large enough $\mathtt{n}$ and $\mathtt{p}$, then $\alpha = \beta$. This approach is similar to the one followed by Almeida and Zeitoun in collaboration with the author [6] to give an alternative proof of correctness over $\mathbf{A}$ of McCammond's normal form reduction algorithm for ω-terms.

Denote by T_A^{ω} the subset of $T_A^{\bar{\kappa}}$ formed by all ω-terms. The subset of the elements of T_A^{ω} that are in canonical form does not coincide with the set of McCammond's ω-terms in normal form. Although the notions of canonical form and (McCammond's) normal form for ω-terms are similar, our definition introduces an essential modification in the conditions of the normal form. This change makes in general a canonical form be shorter than its normal form. For instance, the ω-term $(a^{\omega}b^{\omega})^{\omega}$ is in canonical form while its normal form is $(a^{\omega}abb^{\omega}ba)^{\omega}a^{\omega}abb^{\omega}$. Moreover each subterm of a canonical form is also a canonical form, a property that is useful in inductive proofs and that fails for normal forms. Furthermore, we show that each ω-term has a unique representative in canonical form with the same interpretation over $\mathbf{A}$.

The paper is organized as follows. In Section 2, we review background material and set the basic notation for $\bar{\kappa}$-terms. We introduce the $\bar{\kappa}$-terms canonical form definition in Section 3 and prove some of their fundamental properties. Section 4 is devoted to the description of the algorithm to transform any given $\bar{\kappa}$-term into one in canonical form. The languages $L_{\mathtt{n},\mathtt{p}}(\alpha)$ and their basic characteristics are determined in Section 5. In Section 6, we complete the proof of the main results of the paper. Finally, Section 7 attests the uniqueness of canonical forms for ω-terms over $\mathbf{A}$.

2 Preliminaries

In this section we begin by briefly reviewing the main definitions and some facts about combinatorics on words and profinite semigroups. The reader is referred to [14, 2, 3] for further details about these topics. We then introduce a representation of $\bar{\kappa}$-terms as well-parenthesized words that extends the representation of ω-words used by McCammond and set up the basic terminology on these objects.

2.1 Words

Throughout the paper, we work with a finite alphabet A. The free semigroup (resp. the free monoid) generated by A is denoted by A^+ (resp. A^*). An element w of A^* is called a (finite) word and its length is represented by $|w|$. The empty word is denoted by 1 and its length is 0. The following result is known as Fine and Wilf's Theorem (see [14]).

Proposition 2.1 *Let $u, v \in A^+$. If two powers u^k and v^n of u and v have a common prefix of length at least $|u| + |v| - gcd(|u|, |v|)$, then u and v are powers of the same word.*

A word is said to be *primitive* if it cannot be written in the form u^n with $n > 1$. We say that two words u and v are *conjugate* if there exist words $w_1, w_2 \in A^*$ such that $u = w_1w_2$ and $v = w_2w_1$. Note that, if u is a primitive word and v is a conjugate of u, then v is also primitive. Let a total order be fixed on the alphabet A. A *Lyndon word* is a primitive word which is minimal in its conjugacy class, for the lexicographic order that extends to A^+ the order on A. For instance, with a binary alphabet $A = \{a, b\}$ such that $a < b$, the Lyndon words until length four are $a, b, ab, aab, abb, aaab, aabb, abbb$. Lyndon words are characterized as follows [13].

Proposition 2.2 *A word is a Lyndon word if and only if it is strictly less than each of its proper suffixes.*

In particular, any Lyndon word is *unbordered*, that is, none of its proper prefixes is one of its suffixes.

2.2 Pseudowords and σ-words

We denote by $\overline{\Omega}_A\mathbf{S}$ the free profinite semigroup generated by A, whose elements are called *pseudowords* (also known as *implicit operations*). The free semigroup A^+ embeds in $\overline{\Omega}_A\mathbf{S}$ and is dense in $\overline{\Omega}_A\mathbf{S}$. Given $x \in \overline{\Omega}_A\mathbf{S}$, the closed subsemigroup of $\overline{\Omega}_A\mathbf{S}$ generated by x contains a single idempotent denoted by x^ω, which is the limit of the sequence $x^{n!}$. More generally, for each $q \in \mathbb{Z}$, we denote by $x^{\omega+q}$ the limit of the sequence $x^{n!+q}$ (with $n! + q > 0$).

An *implicit signature* is a set σ of pseudowords containing the multiplication. A *σ-semigroup* is an algebra in the signature σ whose multiplication is associative. The σ-subsemigroup of $\overline{\Omega}_A\mathbf{S}$ generated by A is denoted by $\Omega_A^\sigma\mathbf{S}$ and its elements are called *σ-words*. It is well known that $\Omega_A^\sigma\mathbf{S}$ is the free σ-semigroup on A. In this paper, we are interested in the most commonly used implicit signature $\kappa = \{xy, x^{\omega-1}\}$, usually called the *canonical signature*, and in its extension $\bar{\kappa} = \{xy, x^{\omega+q} \mid q \in \mathbb{Z}\}$. Although κ is properly contained in $\bar{\kappa}$, for each $\bar{\kappa}$-term there exists a well determined κ-term with the same interpretation over $\mathbf{S}$, as observed above, since this pseudovariety verifies $x^{\omega+q} = x^{\omega-1}x^{q+1}$ $(q \geq 0)$ and

$x^{\omega+q} = (x^{\omega-1})^{-q}$ $(q < 0)$. This means that $\Omega^{\kappa}_{A}\mathbf{S} = \Omega^{\bar\kappa}_{A}\mathbf{S}$, whence the signatures κ and $\bar\kappa$ have the same *expressive power* over $\mathbf{S}$, and that the κ-word and the $\bar\kappa$-word problems over this class of semigroups are equivalent.

From hereon, we will work with the signature $\bar\kappa$ and denote by T_A the set of all $\bar\kappa$-terms. We do not distinguish between $\bar\kappa$-terms that only differ in the order in which multiplications are to be carried out. Sometimes we will omit the reference to the signature $\bar\kappa$ simply referring to an element of T_A as a term. For convenience, we allow the empty term which is identified with the empty word.

2.3 Notation for $\bar\kappa$-terms

McCammond [15] represents ω-terms over A as nonempty well-parenthesized words over the alphabet $A \uplus \{(,)\}$, which do not have $(\,)$ as a factor. For instance, the ω-term $(a^{\omega}ba(ab)^{\omega})^{\omega}$ is represented by the parenthesized word $((a)ba(ab))$. Following this idea, we represent $\bar\kappa$-terms over A as nonempty well-parenthesized words over the alphabet $A_{\mathbb{Z}} = A \uplus \{\overset{q}{(}\,,\overset{q}{)} : q \in \mathbb{Z}\}$, which do not have $\overset{q}{(}\overset{q}{)}$ as a factor. Every $\bar\kappa$-term over A determines a unique well-parenthesized word over $A_{\mathbb{Z}}$ obtained by replacing each subterm $(*)^{\omega+q}$ by $\overset{q}{(}*\overset{q}{)}$, recursively. Recall that the *rank* of a term α is the maximum number $\mathrm{rank}(\alpha)$ of nested parentheses in it. For example, the $\bar\kappa$-term $(a^{\omega-1}ba(ab)^{\omega})^{\omega+5}$ has rank 2 and is represented by $\overset{5}{\Big(}\overset{-1}{(}a\overset{-1}{)}ba\overset{0}{(}ab\overset{0}{)}\overset{5}{\Big)}$, where the rank 2 parentheses are shown in larger size for a greater clarity in the representation of the term. Conversely, the $\bar\kappa$-term associated with such a word is obtained by replacing each matching pair of parentheses $\overset{q}{(}*\overset{q}{)}$ by $(*)^{\omega+q}$. We identify T_A with the set of these well-parenthesized words over $A_{\mathbb{Z}}$. Throughout the rest of the paper, we will usually refer to a $\bar\kappa$-term meaning its associated word over $A_{\mathbb{Z}}$. Notice that, while the set $A_{\mathbb{Z}}$ is infinite, each term uses only a finite number of its symbols.

2.4 Lyndon terms

Since $\bar\kappa$-terms are represented as well-parenthesized words over $A_{\mathbb{Z}}$, each definition on words extends naturally to $\bar\kappa$-terms. In particular, a term is said to be *primitive* if it cannot be written in the form α^n with $\alpha \in T_A$ and $n > 1$, and two terms α and β are *conjugate* if there exist terms $\gamma_1, \gamma_2 \in T_A$ such that $\alpha = \gamma_1\gamma_2$ and $\beta = \gamma_2\gamma_1$. In order to describe the canonical form for $\bar\kappa$-terms, we need to fix a representative element in each conjugacy class of a primitive term. For that, we extend the order on A to $A_{\mathbb{Z}}$ by letting $\overset{p}{(} < \overset{q}{(} < x < \overset{q}{)} < \overset{p}{)}$ for all $x \in A$ and $p, q \in \mathbb{Z}$ with $p < q$. A *Lyndon term* is a primitive term that is minimal, with respect to the lexicographic ordering, in its conjugacy class. For instance, aab, $\overset{-1}{(}aa\overset{-1}{)}b\overset{2}{(}aa\overset{2}{)}b$ and $\overset{-2}{(}a\overset{-2}{)}\overset{0}{\big(}\overset{-1}{(}a\overset{-1}{)}ab\overset{0}{\big)}$ are Lyndon terms.

2.5 Portions of a $\bar\kappa$-term

Terms of the form $\overset{q}{(}\delta\overset{q}{)}$ will be called *limit terms*, and δ and q will be called, respectively, its *base* and its *exponent*. Consider a rank $i+1$ $\bar\kappa$-term

$$\alpha = \gamma_0\overset{q_1}{(}\delta_1\overset{q_1}{)}\gamma_1\cdots\overset{q_n}{(}\delta_n\overset{q_n}{)}\gamma_n, \tag{2.1}$$

with $\mathrm{rank}(\gamma_j) \le i$ and $\mathrm{rank}(\delta_k) = i$. The number n, of limit terms of rank $i+1$ that are subterms of α, will be called the *lt-length* of α. The $\bar\kappa$-terms γ_j and δ_k in (2.1) will be called the *primary subterms* of α and each δ_k will in addition be said to be a *base* of α. The factors of α of the form $\overset{q_k}{(}\delta_k\overset{q_k}{)}\gamma_k\overset{q_{k+1}}{(}\delta_{k+1}\overset{q_{k+1}}{)}$ are called *crucial portions* of α. The prefix $\gamma_0\overset{q_1}{(}\delta_1\overset{q_1}{)}$ and the suffix $\overset{q_n}{(}\delta_n\overset{q_n}{)}\gamma_n$ of α will be called respectively the *initial* *portion* and the *final portion* of α. The product $\overset{q_n}{(}\delta_n\overset{q_n}{)}\gamma_n\gamma_0\overset{q_1}{(}\delta_1\overset{q_1}{)}$ of the final and initial portions will be called the *circular portion* of α. Notice that the circular portion of α is a crucial portion of α^2 and that, if α is not a primitive term, then its circular portion is a crucial portion of α itself. A term $\gamma_0\delta_1^{j_1}\gamma_1\cdots\delta_n^{j_n}\gamma_n$, obtained from α by replacing each subterm $\overset{q_k}{(}\delta_k\overset{q_k}{)}$ by $\delta_k^{j_k}$ with $j_k \geq 1$, is a rank i term called an *expansion* of α. When each exponent j_k is greater than or equal to a given positive integer p, the expansion is called a *p-expansion* of α. The notion of expansion is extended to a rank 0 term β by declaring that β is its own unique expansion.

3 Canonical forms for $\bar{\kappa}$-terms

In this section, we give the definition of canonical form $\bar{\kappa}$-terms and identify the reduction rules that will be used, in Section 4, to transform each $\bar{\kappa}$-term α into a canonical form α', with rank$(\alpha') \leq$ rank(α). A consequence of Theorem 6.1 below is that the $\bar{\kappa}$-term α' is unique and so we call it *the canonical form of* α.

3.1 Canonical form definition

The canonical form for $\bar{\kappa}$-terms is defined recursively as follows. *Rank 0 canonical forms* are the words from A^*. Assuming that rank i canonical forms have been defined, a *rank $i+1$ canonical form ($\bar{\kappa}$-term)* is a $\bar{\kappa}$-term α of the form

$$\alpha = \gamma_0\overset{q_1}{(}\delta_1\overset{q_1}{)}\gamma_1\cdots\overset{q_n}{(}\delta_n\overset{q_n}{)}\gamma_n, \tag{3.1}$$

where the primary subterms γ_j and δ_k are $\bar{\kappa}$-terms such that the following conditions hold:

$(cf.1)$ the 2-expansion $\gamma_0\delta_1^2\gamma_1\cdots\delta_n^2\gamma_n$ of α is a rank i canonical form;

$(cf.2)$ each base δ_k of α is a Lyndon term of rank i;

$(cf.3)$ no δ_k is a suffix of γ_{k-1};

$(cf.4)$ no δ_k is a prefix of some term $\gamma_k\delta_{k+1}^{\ell}$ with $\ell \geq 0$.

For instance, the rank 1 terms $ab\overset{0}{(}abb\overset{0}{)}ab\overset{-2}{(}a\overset{-2}{)}$ and $\overset{-1}{(}b\overset{-1}{)}\overset{4}{(}a\overset{4}{)}b\overset{1}{(}ab\overset{1}{)}$ as well as the rank 2 terms $\overset{1}{(}a\overset{1}{)}\overset{-3}{(}\overset{0}{(}b\overset{0}{)}\overset{1}{(}a\overset{1}{)}\overset{-3}{)}\overset{0}{(}b\overset{0}{)}\overset{2}{(}a\overset{2}{)}b$ and $\overset{2}{(}\overset{-1}{(}ab\overset{-1}{)}\overset{-1}{(}a\overset{-1}{)}b\overset{0}{(}a\overset{0}{)}b\overset{2}{)}\overset{0}{(}\overset{-1}{(}a\overset{-1}{)}b\overset{0}{(}a\overset{0}{)}b\overset{0}{)}$ are in canonical form. We say that a $\bar{\kappa}$-term is in *semi-canonical form* if it verifies condition $(cf.1)$ of the canonical form definition. Of course, all canonical forms and all rank 1 terms are in semi-canonical form. The term $\overset{0}{(}a\overset{0}{)}\overset{-1}{(}\overset{0}{(}b\overset{0}{)}\overset{0}{(}a\overset{0}{)}\overset{0}{(}b\overset{0}{)}\overset{0}{(}a\overset{0}{)}\overset{-1}{)}\overset{0}{(}b\overset{0}{)}\overset{0}{(}a\overset{0}{)}\overset{0}{(}b\overset{0}{)}\overset{0}{(}\overset{0}{(}a\overset{0}{)}\overset{0}{(}b\overset{0}{)}\overset{0}{)}$ constitutes an example of a semi-canonical form of rank 2 that is not in canonical form. Notice that the exponents q_k do not intervene in conditions $(cf.1)$–$(cf.4)$, which means that α being or not in (semi-)canonical form is independent of the q_k. That is, if a $\bar{\kappa}$-term α of the form (3.1) is in (semi-)canonical form, then any $\bar{\kappa}$-term obtained from α by replacing the exponent q_k $(k = 1, \ldots, n)$ by some q'_k is also in (semi-)canonical form.

As one may note, the canonical form definition for crucial portions does not coincide with the one that McCammond [15] imposed on crucial portions of ω-terms in normal form. While McCammond's definition is symmetric relative to the limit terms of the crucial portion and in some cases forces the central factor to have some copies of the bases adjacent to them, we choose to let each limit term absorb all adjacent occurrences of its base (even when they overlap the limit term on its right side), a strategy already used by the author in [10] to solve the ω-word problem for the pseudovariety $\mathbf{LSl}$. This way the canonical form definition for crucial portions looses symmetry but determines shorter canonical forms. For instance, the $\bar{\kappa}$-terms $\alpha_1 = (a^\omega b^\omega)^\omega$ and $\alpha_2 = a^{\omega-1}abb^{\omega-2}ba(a^{\omega-2}abb^{\omega-2}ba)^{\omega-2}a^{\omega-2}abb^{\omega-1}$ have the same interpretation over $\mathbf{S}$. With the above canonical form definition, α_1 is the canonical form of both α_1 and α_2, while with McCammond's alternative, α_2 would be their common canonical form. Moreover these canonical forms have the following nice property that fails, in part, for McCammond's normal forms.

Proposition 3.1 *The following conditions are equivalent for a term α:*

(a) *The term α is in (semi-)canonical form.*

(b) *Every subterm of α is in (semi-)canonical form.*

(c) *The initial portion, the final portion and all of the crucial portions of α are in (semi-) canonical form.*

Proof: The proof is made by induction on the rank of α. For rank$(\alpha) = 0$, the result holds trivially. Let now rank$(\alpha) = i + 1$ and suppose, by the induction hypothesis, that the proposition holds for $\bar{\kappa}$-terms of rank at most i.

To show the implication $(a)\Rightarrow(b)$, assume that α is in semi-canonical form and that it has the form (3.1). Consider a subterm β of α and let us prove that β is in semi-canonical form. Suppose first that rank$(\beta) = i + 1$. In this case β is of the form $\beta = \gamma'_{j-1} \overset{q_j}{(}\, \delta_j \overset{q_j}{)}\, \gamma_j \cdots \overset{q_k}{(}\, \delta_k \overset{q_k}{)}\, \gamma'_k$ where $1 \leq j \leq k \leq n$, γ'_{j-1} is a suffix of γ_{j-1} and γ'_k is a prefix of γ_k. The 2-expansion $\beta_1 = \gamma'_{j-1}\delta_j^2\gamma_j \cdots \delta_k^2\gamma'_k$ of β is a subterm of the 2-expansion $\alpha_1 = \gamma_0\delta_1^2\gamma_1 \cdots \delta_n^2\gamma_n$ of α. As α is in semi-canonical form, α_1 is in canonical form. Now, since α_1 is rank i and β_1 is a subterm of α_1, we infer from the induction hypothesis that β_1 is in canonical form. Hence, β is in semi-canonical form. Note that assuming further that α is in canonical form, i.e., that α verifies conditions $(cf.2)$–$(cf.4)$, it follows that also β verifies those conditions whence it is in canonical form. Suppose now that rank$(\beta) \leq i$. Then β is a subterm of some primary subterm of α, whence it is a subterm of the rank i canonical form α_1. By the induction hypothesis, it follows that β is also in semi-canonical form in this case (and it is in canonical form when α is in canonical form).

The implication $(b)\Rightarrow(c)$ is obvious, while $(c)\Rightarrow(a)$ follows easily from the hypothesis (c) and from the induction hypothesis. $\square$

We say that a $\bar{\kappa}$-term α is in *circular canonical form* if α^2 is in canonical form. The following observation is an immediate, trivially verifiable, consequence of Proposition 3.1.

Corollary 3.2 *Let α be a $\bar{\kappa}$-term.*

(a) *The term α is in circular canonical form if and only if both α and its circular portion are in canonical form.*

(b) *If α is in circular canonical form then any conjugate of α is also in circular canonical form.*

(c) *If α is in semi-canonical form then every base of α is in circular canonical form and the other primary subterms of α are in canonical form; more generally, for any subterm β of α, every base of β is in circular canonical form and the other primary subterms of β are in canonical form.*

(d) *If α is in canonical form and it is not a primitive term, then α is in circular canonical form.*

3.2 Rewriting rules for $\bar{\kappa}$-terms

The procedure to transform an arbitrary $\bar{\kappa}$-term into its canonical form, while retaining its value on finite semigroups, consists in applying elementary changes resulting from reading in either direction the $\bar{\kappa}$-identities of the following set Σ (where $n, p, q \in \mathbb{Z}$ with $n > 0$):

$$\begin{cases} (\alpha^{\omega+p})^{\omega+q} = \alpha^{\omega+pq}, \\ (\alpha^{n})^{\omega+q} = \alpha^{\omega+nq}, \\ \alpha^{\omega+p}\alpha^{\omega+q} = \alpha^{\omega+p+q}, \\ \alpha\alpha^{\omega+q} = \alpha^{\omega+q+1} = \alpha^{\omega+q}\alpha, \\ (\alpha\beta)^{\omega+q}\alpha = \alpha(\beta\alpha)^{\omega+q}. \end{cases}$$

The types of changes are therefore given by the following rewriting rules for terms

1. $\overset{q}{(}\overset{p}{(}\alpha\overset{p}{)}\overset{q}{)} \rightleftarrows \overset{pq}{(}\alpha\overset{pq}{)}$
2. $\overset{q}{(}\alpha^{n}\overset{q}{)} \rightleftarrows \overset{nq}{(}\alpha\overset{nq}{)}$
3. $\overset{p}{(}\alpha\overset{p}{)}\overset{q}{(}\alpha\overset{q}{)} \rightleftarrows \overset{p+q}{(}\alpha\overset{p+q}{)}$

4_L. $\alpha\overset{q}{(}\alpha\overset{q}{)} \rightleftarrows \overset{q+1}{(}\alpha\overset{q+1}{)}$

4_R. $\overset{q}{(}\alpha\overset{q}{)}\alpha \rightleftarrows \overset{q+1}{(}\alpha\overset{q+1}{)}$

5. $\overset{q}{(}\alpha\beta\overset{q}{)}\alpha \rightleftarrows \alpha\overset{q}{(}\beta\alpha\overset{q}{)}$

We call the application of a rule of type 1–4 from left to right (resp. from right to left) a *contraction* (resp. an *expansion*) of that type. An application of a rule of type 5, in either direction, will be called a *shift*. We say that terms α and β are *equivalent*, and denote $\alpha \sim \beta$, if there is a derivation from α to β (that is, there is a finite sequence of contractions, expansions and shifts that starts in α and ends in β).

Example 3.3 *Consider the rank 2 canonical form $\delta = b^5 a\overset{3}{(}\overset{0}{(}b\overset{0}{)}a\overset{3}{)}\overset{-5}{(}b\overset{-5}{)}$. The rank 3 term $\alpha = \overset{-2}{(}\delta\overset{-2}{)}$ can be rewritten as follows*

$$\begin{aligned} \alpha &\to \overset{-2}{(}b^5 a\overset{3}{(}\overset{-5}{(}b\overset{-5}{)}b^5 a\overset{3}{)}\overset{-5}{(}b\overset{-5}{)}\overset{-2}{)} \to \overset{-2}{(}\overset{3}{(}b^5 a\overset{-5}{(}b\overset{-5}{)}\overset{3}{)}b^5 a\overset{-5}{(}b\overset{-5}{)}\overset{-2}{)} \to \overset{-2}{(}\overset{4}{(}b^5 a\overset{-5}{(}b\overset{-5}{)}\overset{4}{)}\overset{-2}{)} \to \overset{-8}{(}b^5 a\overset{-5}{(}b\overset{-5}{)}\overset{-8}{)} \\ &\to \overset{-9}{(}b^5 a\overset{-5}{(}b\overset{-5}{)}\overset{-9}{)}b^5 a\overset{-5}{(}b\overset{-5}{)} \to b^5 a\overset{-9}{(}\overset{-5}{(}b\overset{-5}{)}b^5 a\overset{-9}{)}\overset{-5}{(}b\overset{-5}{)} \to b^5 a\overset{-9}{(}\overset{0}{(}b\overset{0}{)}a\overset{-9}{)}\overset{-5}{(}b\overset{-5}{)} = \alpha'. \end{aligned}$$

The first step in this derivation is an expansion of type 4_R, the second is a shift, the third step is a contraction of type 4_R, the fourth is a contraction of type 1, the fifth step is an expansion of type 4_R, the sixth is a shift, and the final step is a contraction of type 4_R.

Notice that in the above example δ is a term of the form $\varepsilon_1\overset{3}{(}\beta\overset{3}{)}\varepsilon_2$ such that $\varepsilon_2\varepsilon_1 \sim \beta$. Moreover $\overset{-2}{(}\varepsilon_1\overset{3}{(}\beta\overset{3}{)}\varepsilon_2\overset{-2}{)} = \alpha \sim \alpha' = \varepsilon_1\overset{-9}{(}\beta\overset{-9}{)}\varepsilon_2$ and $-9 = (3+1)(-2) - 1$. The example illustrates the following observation.

Fact 3.4 *If δ is a term of the form $\delta = \varepsilon_1\overset{p}{(}\beta\overset{p}{)}\varepsilon_2$ with $\varepsilon_2\varepsilon_1 \sim \beta$, then $\overset{q}{(}\delta\overset{q}{)} \sim \varepsilon_1\overset{r}{(}\beta\overset{r}{)}\varepsilon_2$ where $r = (p+1)q - 1$.*

Proof: The sequence of equivalences

$$\overset{q}{(}\varepsilon_1\overset{p}{(}\beta\overset{p}{)}\varepsilon_2\overset{q}{)} \sim \overset{q}{(}\varepsilon_1\overset{p}{(}\varepsilon_2\varepsilon_1\overset{p}{)}\varepsilon_2\overset{q}{)} \sim \overset{q}{(}\overset{p}{(}\varepsilon_1\varepsilon_2\overset{p}{)}\varepsilon_1\varepsilon_2\overset{q}{)} \sim \overset{q}{(}\overset{p+1}{(}\varepsilon_1\varepsilon_2\overset{p+1}{)}\overset{q}{)} \sim \overset{r+1}{(}\varepsilon_1\varepsilon_2\overset{r+1}{)} \sim \varepsilon_1\overset{r}{(}\varepsilon_2\varepsilon_1\overset{r}{)}\varepsilon_2 \sim \varepsilon_1\overset{r}{(}\beta\overset{r}{)}\varepsilon_2$$

can be easily deduced from the hypothesis $\varepsilon_2\varepsilon_1 \sim \beta$ and the reduction rules. $\square$

Since all $\bar{\kappa}$-identities of Σ are easily shown to be valid in $\mathbf{S}$, if $\alpha \sim \beta$ then $\mathbf{S} \models \alpha = \beta$. We will prove below that the converse implication also holds. We do this by transforming each $\bar{\kappa}$-term into an equivalent canonical form and by showing that, if two given canonical forms are equal over $\mathbf{S}$ then they are precisely the same $\bar{\kappa}$-term. This solves the $\bar{\kappa}$-word problem for $\mathbf{S}$.

4 The canonical form algorithm

We describe an algorithm that computes the canonical form of any given $\bar{\kappa}$-term. The algorithm will be defined recursively on the rank of the given term. Recall first that all rank 0 terms are already in canonical form and so they coincide with their canonical form. Assuming that the method to determine the canonical form of any term of rank at most i was already defined, we show below how to reduce an arbitrary term of rank $i+1$ to its canonical form. The rank $i+1$ canonical form reduction algorithm consists of two major steps. The first step reduces the given term to a semi-canonical form and the second step completes the calculation of the canonical form. It will be convenient to start with the description of the second step since this will be used, in rank i, to define the first step in rank $i+1$. Notice that the first step in rank 1 is trivial since every rank 1 term is already in semi-canonical form.

4.1 Step 2

The procedure to compute the canonical form of an arbitrary rank $i+1$ term α_1 in semi-canonical form is the following.

2.1) Apply all possible rank $i+1$ contractions of type 2.

2.2) By means of a rank $i+1$ expansion of type 4, if necessary, and a rank $i+1$ shift, write each rank $i+1$ limit term in the form $\overset{q}{(}\delta\overset{q}{)}$ where δ is a Lyndon term.

2.3) Apply all possible rank $i+1$ contractions of type 4.

2.4) Apply all possible rank $i+1$ contractions of type 3.

2.5) Put each rank $i+1$ crucial portion $\overset{q_1}{(}\delta_1\overset{q_1}{)}\gamma\overset{q_2}{(}\delta_2\overset{q_2}{)}$ in canonical form as follows. By Step 2.3, δ_1 is not a prefix and δ_2 is not a suffix of γ. Let ℓ be the minimum nonnegative integer such that $|\gamma\delta_2^\ell| \geq |\delta_1|$. If δ_1 is not a prefix of $\gamma\delta_2^\ell$ then the crucial portion $\overset{q_1}{(}\delta_1\overset{q_1}{)}\gamma\overset{q_2}{(}\delta_2\overset{q_2}{)}$ is already in canonical form. Otherwise $\ell \neq 0$. In this case, apply ℓ rank $i+1$ expansions of type 4_L to the limit term on the right side of the crucial portion, followed by all possible, say n, rank $i+1$ contractions of type 4_R, thus obtaining a term $\overset{q_1+n}{(}\delta_1\overset{q_1+n}{)}\varepsilon\overset{q_2-\ell}{(}\delta_2\overset{q_2-\ell}{)}$ where ε is a proper suffix of δ_2 having not a prefix δ_1. In view of the following claim the step is complete.

Claim The crucial portion $\overset{q_1+n}{(}\delta_1\overset{q_1+n}{)}\varepsilon\overset{q_2-\ell}{(}\delta_2\overset{q_2-\ell}{)}$ is in canonical form.

Proof: To prove the claim it suffices to show that δ_1 is not a prefix of $\varepsilon\delta_2^k$ for all $k \geq 1$. Assume, by way of contradiction, that δ_1 is a prefix of some $\varepsilon\delta_2^k$. We have $\delta_1 = \gamma\delta_2^{\ell-1}\varepsilon_1$ and $\delta_1^{n-1} = \varepsilon_2$ with $\delta_2 = \varepsilon_1\varepsilon_2\varepsilon$ and ε_1 nonempty. Since ε_1 is a suffix of the Lyndon term δ_1, we have $\delta_1 \leq \varepsilon_1$ by Proposition 2.2, and since it is a prefix of δ_2, we have $\varepsilon_1 \leq \delta_2$. Therefore $\delta_1 \leq \delta_2$. Suppose that $\delta_1 = \delta_2$. In this case, $\delta_2 \leq \varepsilon_1$ and $\varepsilon_1 \leq \delta_2$, whence $\varepsilon_1 = \delta_2$. From $\delta_1 = \gamma\delta_2^{\ell-1}\varepsilon_1$ it then follows that γ is the empty word (and $\ell = 1$). This is not possible since Step 2.4 eliminated all crucial portions of the form $\overset{p}{(}\delta_1\overset{p}{)}\overset{q}{(}\delta_1\overset{q}{)}$. Thus $\delta_1 \neq \delta_2$ and so $\delta_1 < \delta_2$.

Suppose that γ is the empty word. Then $\delta_1 = \delta_2^{\ell-1}\varepsilon_1$ and so, as δ_2 cannot be a prefix of δ_1 (since in that case we would have $\delta_2 \leq \delta_1$, in contradiction with $\delta_1 < \delta_2$), $\ell = 1$ and $\delta_1 = \varepsilon_1$ with ε_1 a proper prefix of δ_2. Hence $|\delta_1| < |\delta_2|$ and so, from the initial assumption, $\delta_1 = \varepsilon\varepsilon_3$ for some nonempty proper prefix ε_3 of δ_2. As δ_1 is a prefix of δ_2, it follows that ε_3 is both a proper prefix and a suffix of δ_1. That is, δ_1 is a bordered word, which contradicts the fact of δ_1 being a Lyndon word. Consequently, we may assume that γ is not the empty word.

Suppose next that $n > 1$. Then ε_2 is nonempty and so $|\delta_1| < |\delta_2|$. Hence $\ell = 1$ and $\delta_1 = \gamma\varepsilon_1$ with $|\varepsilon_1| < |\delta_1|$, whence ε_1 is a proper suffix of δ_1 and a proper prefix of δ_2. In particular, by Proposition 2.2, $\delta_1 < \varepsilon_1$. On the other hand, as $\varepsilon_2\varepsilon$ is a proper suffix of δ_2, $\varepsilon_1 < \delta_2 < \varepsilon_2\varepsilon = \delta_1^{n-1}\varepsilon$ and thus, as $|\varepsilon_1| < |\delta_1|$, $\varepsilon_1 < \delta_1$. We reached a contradiction and so $n = 1$ and ε_2 is empty.

Suppose now that ε is the empty word. Then $\varepsilon_1 = \delta_2$ is a proper suffix of δ_1. From the initial assumption it then results that δ_2 is also a prefix of δ_1. This means that δ_1 is a bordered word, a condition that is impossible because δ_1 is a Lyndon word. Therefore ε is a nonempty suffix of the Lyndon word δ_2, whence $\delta_2 < \varepsilon$. But ε is a proper prefix of δ_1 by the initial assumption and so $\varepsilon < \delta_1$. It follows that $\delta_2 < \delta_1$ in contradiction with the above inequality $\delta_1 < \delta_2$. This shows that δ_1 cannot be a prefix of some $\varepsilon\delta_2^k$ and proves the claim. □

It is easy to verify that this procedure produces a rank $i + 1$ term α_2 in canonical form. Indeed, the reduction rules that are eventually used in the process are of type 2–5. Hence, the term α_2 is rank $i + 1$ since these rules do not change the rank of the original term α_1. Moreover, by Proposition 3.1, α_2 is in semi-canonical form since α_1 also is and the reduction rules are all applied in rank $i+1$ and do not change the initial, final and crucial portions of the 2-expansions of the term. On the other hand, Steps 2.1 and 2.2 guarantee that α_2 verifies condition $(cf.2)$ of the canonical form definition, while condition $(cf.3)$ is obtained in Step 2.3. Finally, α_2 satisfies $(cf.4)$ due to the application of Steps 2.3 to 2.5. For instance, applying the above algorithm to the rank 2 semi-canonical form

$$\alpha_1 = \overset{0}{(}a\overset{0}{)}\overset{-1}{\Big(}\overset{0}{(}b\overset{0}{)}\overset{0}{(}a\overset{0}{)}\overset{0}{(}b\overset{0}{)}\overset{0}{(}a\overset{0}{)}\overset{-1}{\Big)}\overset{0}{(}b\overset{0}{)}\overset{0}{(}a\overset{0}{)}\overset{0}{(}b\overset{0}{)}\overset{0}{\Big(}\overset{0}{(}a\overset{0}{)}\overset{0}{(}b\overset{0}{)}\overset{0}{\Big)}$$

one gets the following derivation

$$\begin{aligned}\alpha_1 \to\ & \overset{0}{(}a\overset{0}{)}\overset{-2}{\Big(}\overset{0}{(}b\overset{0}{)}\overset{0}{(}a\overset{0}{)}\overset{-2}{\Big)}\overset{0}{(}b\overset{0}{)}\overset{0}{(}a\overset{0}{)}\overset{0}{(}b\overset{0}{)}\overset{0}{\Big(}\overset{0}{(}a\overset{0}{)}\overset{0}{(}b\overset{0}{)}\overset{0}{\Big)} \to \overset{0}{(}a\overset{0}{)}\overset{0}{(}b\overset{0}{)}\overset{-2}{\Big(}\overset{0}{(}a\overset{0}{)}\overset{0}{(}b\overset{0}{)}\overset{-2}{\Big)}\overset{0}{(}a\overset{0}{)}\overset{0}{(}b\overset{0}{)}\overset{0}{\Big(}\overset{0}{(}a\overset{0}{)}\overset{0}{(}b\overset{0}{)}\overset{0}{\Big)}\\ \to\ & \overset{-1}{\Big(}\overset{0}{(}a\overset{0}{)}\overset{0}{(}b\overset{0}{)}\overset{-1}{\Big)}\overset{0}{(}a\overset{0}{)}\overset{0}{(}b\overset{0}{)}\overset{0}{\Big(}\overset{0}{(}a\overset{0}{)}\overset{0}{(}b\overset{0}{)}\overset{0}{\Big)} \to \overset{0}{\Big(}\overset{0}{(}a\overset{0}{)}\overset{0}{(}b\overset{0}{)}\overset{0}{\Big)}\overset{0}{\Big(}\overset{0}{(}a\overset{0}{)}\overset{0}{(}b\overset{0}{)}\overset{0}{\Big)} \to \overset{0}{\Big(}\overset{0}{(}a\overset{0}{)}\overset{0}{(}b\overset{0}{)}\overset{0}{\Big)}.\end{aligned}$$

The canonical form of α_1 is thus the rank 2 term $\alpha_2 = \overset{0}{\Big(}\overset{0}{(}a\overset{0}{)}\overset{0}{(}b\overset{0}{)}\overset{0}{\Big)}$.

4.2 Some preliminary remarks to the first step

The rank $i + 1$ canonical form reduction algorithm will be completed below with the description of Step 1. For now we present some preparatory results which will be useful for that purpose.

Lemma 4.1 *Let $i \geq 1$ and let α be a rank i crucial portion of the form $\overset{q_1}{(}\delta_1\overset{q_1}{)}\gamma\overset{q_2}{(}\delta_2\overset{q_2}{)}$, where the bases δ_1 and δ_2 are rank $i-1$ Lyndon terms in circular canonical form and rank$(\gamma) \leq i-1$. The canonical form α' of α is a rank i term that can be computed using Step 1 of rank at most $i-1$ and Step 2 of rank at most i of the canonical form reduction algorithm. Moreover, either*

(I) α' is a limit term $\overset{r_1}{(}\delta_1\overset{r_1}{)}$, in which case $\delta_1 = \delta_2$, the canonical form of γ is δ_1^p for some $p \geq 0$ and $r_1 = q_1 + q_2 + p$; or

(II) α' is a crucial portion of the form $\overset{r_1}{(}\delta_1\overset{r_1}{)}\varepsilon\overset{r_2}{(}\delta_2\overset{r_2}{)}$.

We then say that α is of type (I) or (II) depending on the condition (I) or (II) that the canonical form α' verifies.

Proof: We proceed by induction on i. Assume first that $i = 1$. Then α is a rank 1 term and so it is in semi-canonical form. The canonical form α' may therefore be obtained by the application to α of Step 2 of the rank 1 canonical form algorithm. Moreover, since δ_1 and δ_2 are Lyndon terms by hypothesis, Steps 2.1 and 2.2 of the algorithm do not apply. If a contraction of type 3 is applied in the process (necessarily in Step 2.4), then δ_1 and δ_2 are clearly the same word and so, by Step 2.3, $\gamma = \delta_1^p$ for some $p \geq 0$. Hence, $\alpha' = \overset{r_1}{(}\delta_1\overset{r_1}{)}$ with $r_1 = q_1 + q_2 + p$. If a contraction of type 3 is not applied, then α' is a crucial portion of the form $\overset{r_1}{(}\delta_1\overset{r_1}{)}\varepsilon\overset{r_2}{(}\delta_2\overset{r_2}{)}$ since the canonicalization process on Step 2.5 does not change the bases δ_1 and δ_2.

Let now $i > 1$ and suppose, by induction hypothesis, that the lemma holds for crucial portions of rank at most $i-1$. The term α' can be calculated as follows. First, use the rank j canonical form algorithm, where j is the rank of γ, to compute the canonical form γ' of γ. The application of two rank i expansions of type 4 then give the term $\overset{p_1}{(}\delta_1\overset{p_1}{)}\delta_1\gamma'\delta_2\overset{p_2}{(}\delta_2\overset{p_2}{)}$, where $p_1 = q_1 - 1$ and $p_2 = q_2 - 1$. By Proposition 3.1, to reduce $\delta_1\gamma'\delta_2$ to its canonical form β, it is sufficient to reduce at most two rank $i-1$ crucial portions to their canonical form. Indeed, when rank$(\gamma') < i-1$, at most the crucial portion $\pi_1\gamma'\pi_2$ is not in canonical form, where π_1 is the final portion of δ_1 and π_2 is the initial portion of δ_2. If rank$(\gamma') = i-1$, then at most the crucial portions $\pi_1\rho_1$ and $\rho_2\pi_2$ are not in canonical form, where ρ_1 and ρ_2 are respectively the initial and the final portions of γ'. By condition ($cf.2$) and Corollary 3.2 (c), the bases of those crucial portions are rank $i-2$ Lyndon terms in circular canonical form. Hence, by the induction hypothesis, they either reduce to a single rank $i-1$ limit term or to another rank $i-1$ crucial portion with the same bases. Therefore, the term $\delta_1^2\beta\delta_2^2$ is in canonical form by Proposition 3.1, whence $\overset{p_1}{(}\delta_1\overset{p_1}{)}\beta\overset{p_2}{(}\delta_2\overset{p_2}{)}$ is a rank i term in semi-canonical form. Since δ_1 and δ_2 are Lyndon terms by hypothesis, to reduce $\overset{p_1}{(}\delta_1\overset{p_1}{)}\beta\overset{p_2}{(}\delta_2\overset{p_2}{)}$ to its canonical form α' it suffices to apply to it Steps 2.3 to 2.5 of the rank i canonical form algorithm. As in the case $i = 1$ above, one deduces that α' is of one of the forms of the statement, thus completing the inductive step of the proof. □

The following is the analogue of Lemma 4.1 to initial and final portions. It has a similar proof and so we leave its verification to the reader.

Lemma 4.2 *Let $i \geq 1$ and let α be a rank i initial portion $\gamma\overset{q}{(}\delta\overset{q}{)}$ or final portion $\overset{q}{(}\delta\overset{q}{)}\gamma$, where the base δ is a rank $i-1$ Lyndon term in circular canonical form and rank$(\gamma) \leq i-1$. The canonical form α' of α can be computed using Step 1 of rank at most $i-1$ and Step 2 of rank at most i of the canonical form reduction algorithm. Moreover, α' is a rank i term respectively of the forms $\varepsilon\overset{r}{(}\delta\overset{r}{)}$ and $\overset{r}{(}\delta\overset{r}{)}\varepsilon$.*

As a consequence of Proposition 3.1 and Lemmas 4.1 and 4.2, we get the following property of the product of two canonical forms.

Lemma 4.3 *Let α and β be terms in canonical form, let $(\alpha\beta)'$ be the canonical form of $\alpha\beta$ and let α_1 be the final portion of α and β_1 be the initial portion of β.*

(a) *When rank$(\alpha) <$ rank(β), $(\alpha\beta)'$ is obtained by reducing the initial portion $\alpha\beta_1$ of $\alpha\beta$ to its canonical form. In particular, the lt-length of $(\alpha\beta)'$ is the lt-length of β.*

(b) *When rank$(\alpha) =$ rank(β), $(\alpha\beta)'$ is obtained by reducing the crucial portion $\alpha_1\beta_1$ of $\alpha\beta$ to its canonical form. In particular, if α and β have lt-length of m and n respectively, then the lt-length of $(\alpha\beta)'$ is either $m+n-1$ when $\alpha_1\beta_1$ is of type (I), or $m+n$ when $\alpha_1\beta_1$ is of type (II).*

(c) *When rank$(\alpha) >$ rank(β), $(\alpha\beta)'$ is obtained by reducing the final portion $\alpha_1\beta$ of $\alpha\beta$ to its canonical form. In particular, the lt-length of $(\alpha\beta)'$ is the lt-length of α.*

Proof: For (a), as rank$(\alpha) <$ rank(β), the initial portion of $\alpha\beta$ is $\alpha\beta_1$ and, as β is in canonical form, the base of β_1 is a Lyndon term in circular canonical form by condition $(cf.2)$ and Corollary 3.2 (c). By Proposition 3.1, since all crucial portions and the final portion of $\alpha\beta$ are from β and β is in canonical form, to obtain the canonical form of $\alpha\beta$ it is sufficient to reduce $\alpha\beta_1$ to its canonical form. Condition (a) then follows immediately from Lemma 4.2. The proof of the other conditions is similar. $\square$

Another consequence of the above lemmas is the following property of limit terms, which will be fundamental for the construction of the first step of the canonical form reduction algorithm.

Proposition 4.4 *Let $\pi = \overset{q}{(}\rho\overset{q}{)}$ be a rank $i+1$ limit term with $i \geq 1$ and base ρ in canonical form. Using the canonical form reduction algorithm of rank at most i, it is possible to derive from π a semi-canonical form π_1 such that:*

(a) *If ρ has lt-length 1 and its circular portion is of type (I), then either:*

(1) *ρ is of the form $\overset{q_1}{(}\delta_1\overset{q_1}{)}$ and $\pi_1 = \overset{qq_1}{(}\delta_1\overset{qq_1}{)}$; or*

(2) *ρ is of the form $\gamma_0\overset{q_1}{(}\delta_1\overset{q_1}{)}\gamma_1$ with $\gamma_1\gamma_0 \sim \delta_1$ and $\pi_1 = \gamma_0\overset{r}{(}\delta_1\overset{r}{)}\gamma_1$ where $r = q(q_1+1)-1$.*

In both cases π_1 is a rank i term in canonical form.

(b) *If ρ has lt-length greater than 1 or its circular portion is of type (II), then π_1 is a rank $i+1$ term of the form $\varepsilon_0\overset{r}{(}\beta\overset{r}{)}\varepsilon_1$ with rank$(\varepsilon_0) =$ rank$(\varepsilon_1) = i$.*

Proof: Let $\rho = \gamma_0\overset{q_1}{(}\delta_1\overset{q_1}{)}\gamma_1\cdots\overset{q_n}{(}\delta_n\overset{q_n}{)}\gamma_n$ be the canonical form for ρ and let $\alpha = \overset{q_n}{(}\delta_n\overset{q_n}{)}\gamma_n\gamma_0\overset{q_1}{(}\delta_1\overset{q_1}{)}$ be the circular portion of ρ. In order to prove (a), suppose that $n=1$ and that α is of type (I). Then $\rho = \gamma_0\overset{q_1}{(}\delta_1\overset{q_1}{)}\gamma_1$ and, by Lemma 4.1, $\alpha = \overset{q_1}{(}\delta_1\overset{q_1}{)}\gamma_1\gamma_0\overset{q_1}{(}\delta_1\overset{q_1}{)}$ reduces to a limit term $\overset{r_1}{(}\delta_1\overset{r_1}{)}$ and the canonical form of $\gamma_1\gamma_0$ is δ_1^p for some $p \geq 0$. If $p = 0$ then γ_0 and γ_1 are both the empty term and so $\pi = \overset{q}{(}\overset{q_1}{(}\delta_1\overset{q_1}{)}\overset{q}{)}$. In this case, applying a contraction of type 1 one gets a semi-canonical form $\pi_1 = \overset{qq_1}{(}\delta_1\overset{qq_1}{)}$, which is in fact the canonical form of π. Suppose now that $p \neq 0$. In this case the rank of $\gamma_1\gamma_0$ is the rank of δ_1, that is, rank$(\gamma_1\gamma_0) = i-1$. If rank$(\gamma_1) <$ rank(γ_0) then, by Lemma 4.3 (a), the lt-length of δ_1^p is the lt-length of γ_0. As δ_1 is not a suffix of γ_0 by condition $(cf.3)$ of the canonical form definition, we deduce that $p = 1$. The events $\text{rank}(\gamma_1) = \text{rank}(\gamma_0)$ and $\text{rank}(\gamma_1) > \text{rank}(\gamma_0)$ are treated analogously and give the same result $p = 1$. This shows that δ_1 is the canonical form of $\gamma_1\gamma_0$ and, so, that $\gamma_1\gamma_0 \sim \delta_1$. By Fact 3.4, π reduces to the term $\gamma_0 \overset{r}{(}\delta_1\overset{r}{)}\gamma_1$, where $r = q(q_1+1)-1$, which is obviously in canonical form. This completes the proof of (a).

For (b), suppose first $n > 1$. By Lemma 4.1, α reduces to a crucial portion of the form $\overset{r_n}{(}\delta_n\overset{r_n}{)}\sigma$, where σ is either the empty term (in which case $\delta_n = \delta_1$) or a term of the form $\varepsilon \overset{r_1}{(}\delta_1\overset{r_1}{)}$. Let $q' = q-1$ and $q'' = q-2$. The following derivation is now easily deduced

$$\begin{aligned}
\pi \;&\to\; \gamma_0 \overset{q_1}{(}\delta_1\overset{q_1}{)}\gamma_1 \cdots \overset{q_n}{(}\delta_n\overset{q_n}{)}\gamma_n \overset{q''}{(}\gamma_0 \overset{q_1}{(}\delta_1\overset{q_1}{)}\gamma_1 \cdots \overset{q_n}{(}\delta_n\overset{q_n}{)}\gamma_n \overset{q''}{)}\gamma_0 \overset{q_1}{(}\delta_1\overset{q_1}{)}\gamma_1 \cdots \overset{q_n}{(}\delta_n\overset{q_n}{)}\gamma_n \\
&\to\; \gamma_0 \overset{q_1}{(}\delta_1\overset{q_1}{)}\gamma_1 \cdots \overset{q_n}{(}\delta_n\overset{q_n}{)}\gamma_n\gamma_0 \overset{q_1}{(}\delta_1\overset{q_1}{)}\overset{q''}{(}\gamma_1 \cdots \overset{q_n}{(}\delta_n\overset{q_n}{)}\gamma_n\gamma_0 \overset{q_1}{(}\delta_1\overset{q_1}{)}\overset{q''}{)}\gamma_1 \cdots \overset{q_n}{(}\delta_n\overset{q_n}{)}\gamma_n \\
&\overset{*}{\to}\; \gamma_0 \overset{q_1}{(}\delta_1\overset{q_1}{)}\gamma_1 \cdots \overset{r_n}{(}\delta_n\overset{r_n}{)}\sigma \overset{q''}{(}\gamma_1 \cdots \overset{r_n}{(}\delta_n\overset{r_n}{)}\sigma \overset{q''}{)}\gamma_1 \cdots \overset{q_n}{(}\delta_n\overset{q_n}{)}\gamma_n \\
&\to\; \gamma_0 \overset{q_1}{(}\delta_1\overset{q_1}{)}\overset{q'}{(}\gamma_1 \cdots \overset{r_n}{(}\delta_n\overset{r_n}{)}\sigma \overset{q'}{)}\gamma_1 \cdots \overset{q_n}{(}\delta_n\overset{q_n}{)}\gamma_n.
\end{aligned}$$

This last term is clearly in semi-canonical form. On the other hand it verifies the properties of the term π_1 in (b). Suppose now that α is of type (II). If $q \notin \{-1, 1\}$, then q has some prime divisor p. Let $k = \frac{q}{p}$. Applying an expansion of type 1 to π one gets the term $\overset{k}{(}\rho^p\overset{k}{)}$. Hence, π reduces to the term $\overset{k}{(}\tau\overset{k}{)}$ where τ is the canonical form of ρ^p. By Lemma 4.3 (b), as α is of type (II), the lt-length of τ is exactly pn and thus greater than 1. Therefore, by the case $n > 1$ above, $\overset{k}{(}\tau\overset{k}{)}$, and on its turn π, reduces to a term π_1 as stated in (b). Finally, for $q \in \{-1, 1\}$, apply an expansion of type 4_R to π in order to obtain the term $\overset{q'}{(}\rho\overset{q'}{)}\rho$, where $q' = q-1 \in \{-2, 0\}$. By the previous cases, this term reduces to some term $\varepsilon_0 \overset{r}{(}\beta\overset{r}{)}\varepsilon_1\rho$ with $\varepsilon_0 \overset{r}{(}\beta\overset{r}{)}\varepsilon_1$ a rank $i+1$ semi-canonical form such that $\text{rank}(\varepsilon_0) = \text{rank}(\varepsilon_1) = i$. To complete the construction of π_1 in the current case it suffices to put $\varepsilon_1\rho$ in canonical form, thus showing that condition (b) holds. □

We are now ready to present the first step of the canonical form reduction algorithm.

4.3 Step 1

The procedure to compute an equivalent semi-canonical form α_1 of an arbitrary rank $i+1$ term α is as follows.

1.1) In case $i = 0$, declare α_1 to be α and stop (since every rank 1 term is already in semi-canonical form).

1.2) Apply the rank i canonical form reduction algorithm to each base of α. We note that, this way, some (or all) of the original rank $i+1$ limit terms may have been transformed into terms with strictly smaller rank. If the term obtained is rank $j+1$ with $j < i$, then go to the beginning of Step 1 and take i as j.

1.3) Replace each rank $i+1$ limit term π by the semi-canonical form π_1 given by Proposition 4.4. Once again, if the term obtained is no longer of rank $i+1$, then go to the beginning of Step 1.

1.4) Apply the rank i canonical form reduction algorithm to each primary subterm that does not occur as a base.

The $\bar{\kappa}$-term α_1 that emerges from this procedure is indeed a term in semi-canonical form. This is an immediate consequence of Proposition 4.4 (b) since the bases of α_1 are the bases β of the subterms $\pi_1 = \varepsilon_0 \overset{r}{(} \beta \overset{r}{)} \varepsilon_1$, introduced on Step 1.3, that come from that result and, so, are terms in circular canonical form. Moreover, as ε_0 and ε_1 are rank i terms in canonical form, the reduction made by Step 1.4 does not change the final portion of ε_0 neither the initial portion of ε_1, except for possible modifications of the exponents of the corresponding limit terms. As a result, the term obtained from α_1 by replacing each limit term $\overset{r}{(} \beta \overset{r}{)}$ by β^2 is in canonical form, so that α_1 is in semi-canonical form.

5 Languages associated with $\bar{\kappa}$-terms

An alternative proof of correctness of McCammond's normal form reduction algorithm for ω-terms over $\mathbf{A}$ was presented in [6] and is based on properties of certain regular languages $L_{\mathtt{n}}(\alpha)$ associated with ω-terms α, where $\mathtt{n}$ is a positive integer. Informally, the language $L_{\mathtt{n}}(\alpha)$ is obtained from α by replacing each ω-power by a power of exponent at least $\mathtt{n}$. The key property of the languages $L_{\mathtt{n}}(\alpha)$ is that they are star-free when α is in McCammond's normal form and $\mathtt{n}$ is sufficiently large. In this paper, similar languages will play a fundamental role in the proof of Theorem 6.1. Given a $\bar{\kappa}$-term α and a pair $(\mathtt{n}, \mathtt{p})$ of positive integers, we define below a language $L_{\mathtt{n},\mathtt{p}}(\alpha)$ whose elements, informally speaking, are obtained from α by recursively replacing each ω by an integer beyond $\mathtt{n}$ and congruent modulo $\mathtt{p}$ with that threshold. In particular, when the $\bar{\kappa}$-term α is an ω-term, $L_{\mathtt{n},1}(\alpha) = L_{\mathtt{n}}(\alpha)$. So, the above operators $L_{\mathtt{n}}$ associated with ω-terms constitute an instance of a more general concept of operators $L_{\mathtt{n},\mathtt{p}}$ associated with $\bar{\kappa}$-terms. Moreover, as we shall see below, the basic properties of the operators $L_{\mathtt{n}}$ presented in [6] extend easily to $L_{\mathtt{n},\mathtt{p}}$.

5.1 Expansions of $\bar{\kappa}$-terms

Let α be a $\bar{\kappa}$-term. Denote by $Q(\alpha)$ the set of all $q \in \mathbb{Z}$ for which there exists a subterm of α of the form $\beta^{\omega+q}$, that is,

$$Q(\alpha) = \{q \in \mathbb{Z} : \overset{q}{(} \text{occurs in the well-parenthesized word of } A_{\mathbb{Z}} \text{ representing } \alpha\}.$$

Now, let $\nu(\alpha)$ be the nonnegative integer

$$\nu(\alpha) = \max\{|q| : q \in Q(\alpha)\},$$

named the *scale of* α, and note the following immediate property of this parameter.

Remark 5.1 *If α' is either a subterm or an expansion of a $\bar{\kappa}$-term α, then $\nu(\alpha') \leq \nu(\alpha)$.*

Fix a pair of positive integers $(\mathtt{n}, \mathtt{p})$. Usually we will impose high lower bounds for such integers in order to secure the properties we need. For now, when the pair $(\mathtt{n}, \mathtt{p})$ is associated with a $\bar{\kappa}$-term α, we assume that $\mathtt{n}$ is greater than the scale $\nu(\alpha)$ of α. For each $q \in Q(\alpha)$, we let $\overline{q}$ be the set

$$\overline{q} = \{\mathtt{n} + j\mathtt{p} + q : j \geq 0\} \tag{5.1}$$

of positive integers with minimal element $\mathtt{n} + q$ and congruent mod $\mathtt{p}$.

The language $L_{\mathtt{n},\mathtt{p}}(\alpha)$ is formally defined as follows, by means of sequential expansions that unfold the outermost $(\omega + q)$-powers enclosing subterms of maximum rank.

Definition 5.2 (Word expansions) *When $\alpha \in A^*$, we let $E_{\mathtt{n},\mathtt{p}}(\alpha) = \{\alpha\}$. Otherwise, consider a rank $i+1$ $\bar{\kappa}$-term $\alpha = \gamma_0\delta_1^{\omega+q_1}\gamma_1\cdots\delta_r^{\omega+q_r}\gamma_r$, where rank$(\delta_k) = i$ and rank$(\gamma_j) \le i$ for all k and j. We let*

$$E_{\mathtt{n},\mathtt{p}}(\alpha) = \{\gamma_0\delta_1^{n_1}\gamma_1\cdots\delta_r^{n_r}\gamma_r :\ n_j \in \overline{q_j} \text{ for } j = 1,\ldots,r\}.$$

For a set K of $\bar{\kappa}$-terms, we let $E_{\mathtt{n},\mathtt{p}}(K) = \bigcup_{\beta\in K} E_{\mathtt{n},\mathtt{p}}(\beta)$. We then let

$$L_{\mathtt{n},\mathtt{p}}(\alpha) = E_{\mathtt{n},\mathtt{p}}^{\mathit{rank}(\alpha)}(\alpha),$$

where $E_{\mathtt{n},\mathtt{p}}^j$ is the j-fold iteration of the operator $E_{\mathtt{n},\mathtt{p}}$.

For example, let $\alpha = (ab)^{\omega-1}aaba^\omega b(ab)^{\omega+5}a$ and let $(\mathtt{n},\mathtt{p})$ be arbitrary. We have

$$L_{\mathtt{n},\mathtt{p}}(\alpha) = E_{\mathtt{n},\mathtt{p}}(\alpha) = \{(ab)^{\mathtt{n}+j\mathtt{p}-1}aaba^{\mathtt{n}+k\mathtt{p}}b(ab)^{\mathtt{n}+\ell\mathtt{p}+5}a : j,k,\ell \ge 0\}.$$

Consider now the rank 2 canonical form $\beta = (a^{\omega-1}b)^\omega a^{\omega+1}$. We have $L_{8,4}(\beta) = E_{8,4}^2(\beta)$ and $E_{8,4}(\beta) = \{(a^{\omega-1}b)^{8+4j}a^{\omega+1} : j \ge 0\}$. Hence

$$\begin{aligned} L_{8,4}(\beta) &= \textstyle\bigcup_{j\ge0} E_{8,4}((a^{\omega-1}b)^{8+4j}a^{\omega+1}) \\ &= \textstyle\bigcup_{j\ge0}\{a^{7+4k_1}ba^{7+4k_2}b\cdots a^{7+4k_{8+4j}}ba^{9+4k_{9+4j}} : k_1,\ldots,k_{9+4j} \ge 0\}. \end{aligned}$$

The next lemma is analogous to [6, Lemma 3.2] and presents some simple properties of the operators $E_{\mathtt{n},\mathtt{p}}$ and $L_{\mathtt{n},\mathtt{p}}$.

Lemma 5.3 *Let $(\mathtt{n},\mathtt{p})$ be a pair of positive integers. The following formulas hold, where we assume that $\mathtt{n}$ is greater than the scale of all $\bar{\kappa}$-terms involved:*

(a) *for $\bar{\kappa}$-terms α and β,*

$$E_{\mathtt{n},\mathtt{p}}(\alpha\beta) = \begin{cases} E_{\mathtt{n},\mathtt{p}}(\alpha)\, E_{\mathtt{n},\mathtt{p}}(\beta) & \text{if } \mathit{rank}(\alpha) = \mathit{rank}(\beta) \\ \alpha\, E_{\mathtt{n},\mathtt{p}}(\beta) & \text{if } \mathit{rank}(\alpha) < \mathit{rank}(\beta) \\ E_{\mathtt{n},\mathtt{p}}(\alpha)\, \beta & \text{if } \mathit{rank}(\alpha) > \mathit{rank}(\beta); \end{cases}$$

(b) *for a $\bar{\kappa}$-term α, $L_{\mathtt{n},\mathtt{p}}(\alpha) = L_{\mathtt{n},\mathtt{p}}(E_{\mathtt{n},\mathtt{p}}(\alpha))$;*

(c) *for sets U and V of $\bar{\kappa}$-terms, we have $L_{\mathtt{n},\mathtt{p}}(UV) = L_{\mathtt{n},\mathtt{p}}(U)\, L_{\mathtt{n},\mathtt{p}}(V)$;*

(d) *for a $\bar{\kappa}$-term $\alpha = \gamma_0\delta_1^{\omega+q_1}\gamma_1\cdots\delta_r^{\omega+q_r}\gamma_r$ with each rank$(\delta_k) = i$ and rank$(\gamma_j) \le i$,*

$$L_{\mathtt{n},\mathtt{p}}(\alpha) = L_{\mathtt{n},\mathtt{p}}(\gamma_0)\, L_{\mathtt{n},\mathtt{p}}(\delta_1^{\omega+q_1})\, L_{\mathtt{n},\mathtt{p}}(\gamma_1)\cdots L_{\mathtt{n},\mathtt{p}}(\delta_r^{\omega+q_r})\, L_{\mathtt{n},\mathtt{p}}(\gamma_r);$$

(e) *for a $\bar{\kappa}$-term α and an integer q, $L_{\mathtt{n},\mathtt{p}}(\alpha^{\omega+q}) = L_{\mathtt{n},\mathtt{p}}(\alpha)^{\mathtt{n}+q}(L_{\mathtt{n},\mathtt{p}}(\alpha)^{\mathtt{p}})^*$.*

Proof: The proof of each condition (a)–(d) is identical to the proof of the corresponding statement in [6, Lemma 3.2]. For (e), we only need to introduce minor changes. We have

$$\begin{aligned} L_{\mathtt{n},\mathtt{p}}(\alpha^{\omega+q}) &\underset{(b)}{=} L_{\mathtt{n},\mathtt{p}}(E_{\mathtt{n},\mathtt{p}}(\alpha^{\omega+q})) = \textstyle\bigcup_{j\ge0} L_{\mathtt{n},\mathtt{p}}(\alpha^{\mathtt{n}+j\mathtt{p}+q}) \\ &\underset{(c)}{=} \textstyle\bigcup_{j\ge0} L_{\mathtt{n},\mathtt{p}}(\alpha)^{\mathtt{n}+j\mathtt{p}+q} = L_{\mathtt{n},\mathtt{p}}(\alpha)^{\mathtt{n}+q}(L_{\mathtt{n},\mathtt{p}}(\alpha)^{\mathtt{p}})^*, \end{aligned}$$

thus completing the proof of the lemma. □

The following important property of the languages $L_{\mathtt{n},\mathtt{p}}(\alpha)$ can now be easily deduced.

Proposition 5.4 *Let α be a $\bar{\kappa}$-term in canonical form and let $(\mathtt{n},\mathtt{p})$ be a pair of positive integers with $\mathtt{n} > \nu(\alpha)$. Then $L_{\mathtt{n},\mathtt{p}}(\alpha)$ is a regular language.*

Proof: We proceed by induction on rank(α). For rank$(\alpha) = 0$ the result is clear since in this case $L_{\mathtt{n},\mathtt{p}}(\alpha) = \{\alpha\}$. Let now rank$(\alpha) = i+1$ with $i \geq 0$ and suppose, by the induction hypothesis, that the lemma holds for $\bar{\kappa}$-terms of rank at most i. Let $\alpha = \gamma_0\delta_1^{\omega+p_1}\gamma_1\cdots\delta_r^{\omega+p_r}\gamma_r$ be the canonical form expression for α, where rank$(\delta_k) = i$ and rank$(\gamma_j) \leq i$ for all k and j. Then, by Lemma 5.3,

$$L_{\mathtt{n},\mathtt{p}}(\alpha) = L_{\mathtt{n},\mathtt{p}}(\gamma_0)L_{\mathtt{n},\mathtt{p}}(\delta_1)^{\mathtt{n}+q_1}(L_{\mathtt{n},\mathtt{p}}(\delta_1)^{\mathtt{p}})^*L_{\mathtt{n},\mathtt{p}}(\gamma_1)\cdots L_{\mathtt{n},\mathtt{p}}(\delta_r)^{\mathtt{n}+q_r}(L_{\mathtt{n},\mathtt{p}}(\delta_r)^{\mathtt{p}})^*L_{\mathtt{n},\mathtt{p}}(\gamma_r).$$

By the induction hypothesis, each $L_{\mathtt{n},\mathtt{p}}(\gamma_j)$ and $L_{\mathtt{n},\mathtt{p}}(\delta_k)$ is a regular language, whence $L_{\mathtt{n},\mathtt{p}}(\alpha)$ is itself a regular language. This completes the inductive step and concludes the proof of the result. □

For instance, the language $L_{8,4}(\beta)$ associated with the above $\bar{\kappa}$-term $\beta = (a^{\omega-1}b)^{\omega}a^{\omega+1}$ admits the regular expression $L_{8,4}(\beta) = (a^7(a^4)^*b)^8((a^7(a^4)^*b)^4)^*a^9(a^4)^*$. Notice that, in this example, $\mathtt{n} = 8$ is a multiple of $\mathtt{p} = 4$ and so the sequence $\big((a^{k!-1}b)^{k!}a^{k!+1}\big)_k$ of A^+ is ultimately contained in $L_{8,4}(\beta)$. Thus,

$$\eta(\beta) \in \mathrm{cl}(L_{8,4}(\beta)) \tag{5.2}$$

where $\eta : T_A^{\bar{\kappa}} \to \Omega_A^{\bar{\kappa}}\mathbf{S}$ is the homomorphism of $\bar{\kappa}$-semigroups that sends each $x \in A$ to itself and $\mathrm{cl}(L_{8,4}(\beta))$ denotes the topological closure of the language $L_{8,4}(\beta)$ in $\overline{\Omega}_A\mathbf{S}$.

5.2 Schemes for canonical forms

We define the *length* of a $\bar{\kappa}$-term α as the length of the corresponding well-parenthesized word over $A_{\mathbb{Z}}$, and denote it $|\alpha|$. We now associate to each $\bar{\kappa}$-term α a parameter $\mu(\alpha)$, introduced in [6] for ω-terms. In case $\alpha \in A^+$, let $\mu(\alpha) = 0$. Otherwise, let

$$\mu(\alpha) = 2^{\mathrm{rank}(\alpha)}\max\{|\beta| : \beta \text{ is a crucial portion of } \alpha^2\}.$$

It is important to remark the following feature of this parameter, whose proof is an easy adaptation of [6, Lemma 3.5].

Lemma 5.5 *If α' is an expansion of a $\bar{\kappa}$-term α, then $\mu(\alpha') \leq \mu(\alpha)$.*

Let α be a $\bar{\kappa}$-term in canonical form and let $(\mathtt{n},\mathtt{p})$ be a pair of integers. We say that $(\mathtt{n},\mathtt{p})$ is a *scheme for α* if the following conditions hold:

- $\mathtt{n}$ is a multiple of $\mathtt{p}$ such that $\mathtt{n} - \mathtt{p} > \mu(\alpha)$;
- $\mathtt{p} > 2\nu(\alpha)$.

The next result is an immediate consequence of Proposition 3.1, of Remark 5.1 and of Lemma 5.5.

Lemma 5.6 *Let α be a $\bar{\kappa}$-term in canonical form and let $(\mathtt{n},\mathtt{p})$ be a scheme for α. If α' is an expansion of α, then α' is in canonical form and $(\mathtt{n},\mathtt{p})$ is a scheme for α'.*

The following is a significant property of a scheme.

Lemma 5.7 *Let α and β be canonical forms with rank(α) = rank(β) and let $(\mathtt{n},\mathtt{p})$ be a scheme for both α and β. If $L_{\mathtt{n},\mathtt{p}}(\alpha) \cap L_{\mathtt{n},\mathtt{p}}(\beta) \neq \emptyset$, then $\alpha = \beta$.*

Proof: The proof is made by induction on the rank of α (and β). The case rank$(\alpha)=0$ is trivial. Indeed, in this case we have $L_{\mathtt{n},\mathtt{p}}(\alpha)=\{\alpha\}$ and $L_{\mathtt{n},\mathtt{p}}(\beta)=\{\beta\}$.

Let now rank$(\alpha)=i+1$ with $i\geq 0$ and suppose, by induction hypothesis, that the lemma holds for rank i canonical forms. Let $\alpha=\gamma_0\delta_1^{\omega+p_1}\gamma_1\cdots\delta_r^{\omega+p_r}\gamma_r$ and $\beta=\pi_0\rho_1^{\omega+q_1}\pi_1\cdots\rho_s^{\omega+q_s}\pi_s$ be the canonical form expressions for α and β and suppose that $w\in L_{\mathtt{n},\mathtt{p}}(\alpha)\cap L_{\mathtt{n},\mathtt{p}}(\beta)$. By Lemma 5.3 (b) there exist expansions $\alpha'\in E_{\mathtt{n},\mathtt{p}}(\alpha)$ of α and $\beta'\in E_{\mathtt{n},\mathtt{p}}(\beta)$ of β such that $w\in L_{\mathtt{n},\mathtt{p}}(\alpha')\cap L_{\mathtt{n},\mathtt{p}}(\beta')$. In particular, α' and β' are rank i canonical forms of the type $\alpha'=\gamma_0\delta_1^{m_1}\gamma_1\cdots\delta_r^{m_r}\gamma_r$, with each $m_\ell\in\overline{p_\ell}=\{\mathtt{n}+j\mathtt{p}+p_\ell : j\geq 0\}$, say $m_\ell=\mathtt{n}+j_\ell\mathtt{p}+p_\ell$ with $j_\ell\geq 0$, and $\beta'=\pi_0\rho_1^{n_1}\pi_1\cdots\rho_s^{n_s}\pi_s$, where each $n_\ell=\mathtt{n}+k_\ell\mathtt{p}+q_\ell$ with $k_\ell\geq 0$. Moreover, by Lemma 5.6, $(\mathtt{n},\mathtt{p})$ is a scheme for both α' and β'. Hence, the induction hypothesis entails the equality of the $\bar\kappa$-terms α' and β', that is,

$$\gamma_0\delta_1^{m_1}\gamma_1\cdots\delta_r^{m_r}\gamma_r=\pi_0\rho_1^{n_1}\pi_1\cdots\rho_s^{n_s}\pi_s. \tag{5.3}$$

As $(\mathtt{n},\mathtt{p})$ is a scheme for α and β, $m_\ell\geq\mathtt{n}+p_\ell\geq\mathtt{n}-\nu(\alpha)>\mathtt{n}-\mathtt{p}$ and, analogously, $n_\ell>\mathtt{n}-\mathtt{p}$ for every ℓ. On the other hand $\mathtt{n}-\mathtt{p}>\max\{\mu(\alpha),\mu(\beta)\}$. Thus, in particular, m_1 and n_1 are both greater than $\max\{|\delta_r^{\omega+p_r}\gamma_r\gamma_0\delta_1^{\omega+p_1}|,|\rho_s^{\omega+q_s}\pi_s\pi_0\rho_1^{\omega+q_1}|\}$. Hence, the terms $\delta_1^{m_1}$ and $\rho_1^{n_1}$, occurring on the opposite sides of equality (5.3), must overlap on a factor of length at least $|\delta_1|+|\rho_1|$. Therefore, by Fine and Wilf's Theorem, δ_1 and ρ_1 have conjugates that are powers of the same $\bar\kappa$-term, say σ. Since δ_1 and ρ_1 are Lyndon terms by condition $(cf.2)$ of the $\bar\kappa$-term canonical form definition, it follows that $\delta_1=\sigma=\rho_1$. Suppose, without loss of generality, that $|\gamma_0|\geq|\pi_0|$ and recall that any Lyndon term is unbordered. Then, as $\gamma_0\delta_1$ is a prefix of β', $\gamma_0\delta_1=\pi_0\delta_1^j$ for some $j\geq 1$. Hence $\gamma_0=\pi_0$ since α is in canonical form and condition $(cf.3)$ of the $\bar\kappa$-term canonical form definition states that δ_1 is not a suffix of γ_0. On the other hand, the canonical forms α and β verify condition $(cf.4)$. Thus, δ_1 is not a prefix of $\gamma_1\delta_2^{m_2}$ and ρ_1 is not a prefix of $\pi_1\rho_2^{n_2}$. Consequently, the equalities $\alpha'=\beta'$, $\gamma_0=\pi_0$ and $\delta_1=\rho_1$ and the fact that both m_2 and n_2 are greater than $\max\{\mu(\alpha),\mu(\beta)\}$ (and so greater than $|\delta_1|$) imply that $m_1=n_1$. As $\mathtt{n}$ is a multiple of $\mathtt{p}$ by the definition of a scheme, the positive integers m_1 and n_1 are congruent mod $\mathtt{p}$ with p_1 and q_1 respectively. Therefore $p_1=q_1$ once $\mathtt{p}>2\max\{\nu(\alpha),\nu(\beta)\}$. Iterating the above procedure, one deduces that, for every $1\leq\ell\leq\min\{r,s\}$, $\gamma_{\ell-1}=\pi_{\ell-1}$, $\delta_\ell=\rho_\ell$, $p_\ell=q_\ell$ and

$$\gamma_{\ell-1}\delta_\ell^{m_\ell}\gamma_\ell\cdots\delta_r^{m_r}\gamma_r=\pi_{\ell-1}\rho_\ell^{n_\ell}\pi_\ell\cdots\rho_s^{n_s}\pi_s.$$

By symmetry, we have further that $\gamma_r=\pi_s$. Since each m_k and each n_ℓ is greater than $\max\{\mu(\alpha),\mu(\beta)\}$, it is now straightforward to deduce that $r=s$. This shows that $\alpha=\beta$ and concludes the inductive step of the proof. $\square$

The following result is an extension of [6, Theorem 5.3] and it will be essential to prove Theorem 6.1.

Proposition 5.8 *Let α and β be canonical forms and let $(\mathtt{n},\mathtt{p})$ be a scheme for both α and β such that $\mathtt{n}-\mathtt{p}>\max\{|\alpha|,|\beta|\}$. If $L_{\mathtt{n},\mathtt{p}}(\alpha)\cap L_{\mathtt{n},\mathtt{p}}(\beta)\neq\emptyset$, then $\alpha=\beta$.*

Proof: Let $w\in L_{\mathtt{n},\mathtt{p}}(\alpha)\cap L_{\mathtt{n},\mathtt{p}}(\beta)$. Suppose that rank$(\alpha)>$ rank$(\beta)=i$ and let $j=$ rank$(\alpha)-i$. Hence, by the hypothesis $w\in L_{\mathtt{n},\mathtt{p}}(\alpha)$ and Lemma 5.3 (b), there is $\alpha'\in E^j_{\mathtt{n},\mathtt{p}}(\alpha)$ such that $w\in L_{\mathtt{n},\mathtt{p}}(\alpha')$. Moreover, α' is a rank i canonical form and $(\mathtt{n},\mathtt{p})$ is a scheme for α'. Therefore, by Lemma 5.7, $\alpha'=\beta$. This is however impossible since $|\beta|<\mathtt{n}-\mathtt{p}$, by hypothesis, and $\mathtt{n}-\mathtt{p}<|\alpha'|$, by the fact that $\alpha'\in E^j_{\mathtt{n},\mathtt{p}}(\alpha)$ and $\mathtt{n}-\mathtt{p}<\mathtt{n}-\nu(\alpha)\leq\mathtt{n}+q=\min\overline{q}$ for every $q\in Q(\alpha)$. By symmetry it follows that rank$(\alpha)=$ rank(β) and so, by Lemma 5.7, $\alpha=\beta$. $\square$

6 Main results

For $L \subseteq A^+$, let $\mathrm{cl}(L)$ be the topological closure of L in $\overline{\Omega}_A\mathbf{S}$ and notice that $\mathrm{cl}(L) \cap A^+ = L$ since any sequence of words converging to a word $w \in A^+$ is ultimately equal to w. We can now complete the proof of our central result.

Theorem 6.1 *Let α and β be $\bar{\kappa}$-terms in canonical form. If $\mathbf{S} \models \alpha = \beta$, then α and β are the same $\bar{\kappa}$-term.*

Proof: We adapt the corresponding proof for McCammond's normal forms, given in [6, Corollary 5.4]. Let $(\mathtt{n}, \mathtt{p})$ be a scheme for both α and β, with $\mathtt{n} - \mathtt{p} > \max\{|\alpha|, |\beta|\}$. The languages $L_{\mathtt{n},\mathtt{p}}(\alpha)$ and $L_{\mathtt{n},\mathtt{p}}(\beta)$ are regular by Proposition 5.4, whence $\mathrm{cl}(L_{\mathtt{n},\mathtt{p}}(\alpha))$ and $\mathrm{cl}(L_{\mathtt{n},\mathtt{p}}(\beta))$ are clopen subsets of $\overline{\Omega}_A\mathbf{S}$. On the other hand, since $\mathtt{n}$ is a multiple of $\mathtt{p}$ by the definition of a scheme, and as exemplified in (5.2), $\eta(\alpha) \in \mathrm{cl}(L_{\mathtt{n},\mathtt{p}}(\alpha))$ and $\eta(\beta) \in \mathrm{cl}(L_{\mathtt{n},\mathtt{p}}(\beta))$. As $\eta(\alpha) = \eta(\beta)$ by hypothesis, it follows that the intersection $\mathrm{cl}(L_{\mathtt{n},\mathtt{p}}(\alpha)) \cap \mathrm{cl}(L_{\mathtt{n},\mathtt{p}}(\beta))$ is a nonempty open set and so it contains some elements of the dense set A^+. Since $\mathrm{cl}(L_{\mathtt{n},\mathtt{p}}(\alpha)) \cap \mathrm{cl}(L_{\mathtt{n},\mathtt{p}}(\beta)) \cap A^+ = L_{\mathtt{n},\mathtt{p}}(\alpha) \cap L_{\mathtt{n},\mathtt{p}}(\beta)$, we deduce that $L_{\mathtt{n},\mathtt{p}}(\alpha) \cap L_{\mathtt{n},\mathtt{p}}(\beta) \neq \emptyset$. Hence $\alpha = \beta$ by Proposition 5.8. □

In particular, we derive from this result that the canonical form reduction algorithm applied to any $\bar{\kappa}$-term produces a unique $\bar{\kappa}$-term in canonical form. It also leads to an easy deduction of the main results of this paper.

Theorem 6.2 *The $\bar{\kappa}$-word problem for $\mathbf{S}$ is decidable. More precisely, given $\bar{\kappa}$-terms α and β, the canonical form reduction algorithm can be used to decide whether $\mathbf{S}$ satisfies $\alpha = \beta$.*

Proof: Let α' and β' be canonical forms obtained, respectively, from α and β by the application of the canonical form reduction algorithm. By construction of the algorithm, $\mathbf{S}$ verifies $\alpha = \alpha'$ and $\beta = \beta'$. In view of Theorem 6.1, to decide whether $\mathbf{S}$ verifies $\alpha = \beta$ it suffices therefore to verify whether α' and β' are the same $\bar{\kappa}$-term. □

Theorem 6.3 *The set Σ is a basis of $\bar{\kappa}$-identities for $\mathbf{S}^{\bar{\kappa}}$, the $\bar{\kappa}$-variety generated by all finite semigroups.*

Proof: Recall that the rewriting rules used in the canonical form reduction algorithm are determined by the $\bar{\kappa}$-identities of Σ. Hence, it suffices to prove that, for all $\bar{\kappa}$-terms α and β, $\mathbf{S} \models \alpha = \beta$ if and only if $\alpha \sim \beta$. That $\alpha \sim \beta$ implies $\mathbf{S} \models \alpha = \beta$ follows from the fact that $\mathbf{S}$ verifies all the $\bar{\kappa}$-identities of Σ. To show the reverse implication, suppose that $\mathbf{S} \models \alpha = \beta$ and let α' and β' be the canonical forms of α and β. As $\mathbf{S}$ verifies $\alpha = \alpha'$ and $\beta = \beta'$, it also verifies $\alpha' = \beta'$. By Theorem 6.1 we deduce that $\alpha' = \beta'$. Since $\alpha \sim \alpha'$ and $\beta \sim \beta'$ it follows by transitivity that $\alpha \sim \beta$. □

The instance of Theorem 6.1 in which α and β have rank at most 1 was proved, in a different way, by the author together with Nogueira and Teixeira in [11]. Moreover, we have shown in that paper that the pseudovariety $\mathbf{LG}$ does not identify different canonical forms of rank at most 1. It is, however, well-known that $\mathbf{LG}$ identifies the canonical forms $(a^\omega b)^\omega a^\omega$ and a^ω. This remark suggests the introduction of the notion of *$\bar{\kappa}$-index of a pseudovariety* $\mathbf{V}$, denoted $i_{\bar{\kappa}}(\mathbf{V})$, as: the least integer $j \geq 0$, whenever it exists, such that $\mathbf{V}$ identifies two different canonical forms of rank at most j; $+\infty$, otherwise. So, $i_{\bar{\kappa}}(\mathbf{LG}) = 2$ and $i_{\bar{\kappa}}(\mathbf{S}) = +\infty$. The pseudovarieties of $\bar{\kappa}$-index 0 are, by definition, the ones that verify some nontrivial identity. For easy examples of $\bar{\kappa}$-index 1, we may refer the pseudovarieties $\mathbf{G}$ of groups,

$\mathbf{N}$ of nilpotent semigroups and $\mathbf{A}$. As $\mathbf{N} \subseteq \mathbf{A}$, it follows that $i_{\bar{\kappa}}(\mathbf{V}) = 1$ for every aperiodic pseudovariety $\mathbf{LG}$ containing $\mathbf{N}$. For an integer $j > 2$, the author is not aware of examples of pseudovarieties having $\bar{\kappa}$-index j.

An unary implicit signature is a signature formed by unary non-explicit implicit operations together with multiplication. For instance, the signatures ω, κ and $\bar{\kappa}$ are unary. The above notion can be extended to any unary implicit signature σ, for which there is a natural definition of rank for σ-terms, as follows. For a pseudovariety $\mathbf{V}$, let

$$I_\sigma(\mathbf{V}) = \{j \geq 0 : \begin{array}{l} \text{there exist } \sigma\text{-terms } \alpha \text{ and } \beta \text{ with rank at most } j \\ \text{such that } \mathbf{V} \models \alpha = \beta \text{ and } \mathbf{S} \not\models \alpha = \beta\}. \end{array}$$

The *σ-index of* $\mathbf{V}$, denoted $i_\sigma(\mathbf{V})$, is defined to be $\min I_\sigma(\mathbf{V})$ when $I_\sigma(\mathbf{V})$ is non-empty and to be $+\infty$ otherwise.

7 Canonical representatives for ω-terms over $\mathbf{A}$

In this section, we explain how the above results can be adjusted in order to obtain canonical representatives for each class of ω-terms with the same interpretation on each finite aperiodic semigroup.

7.1 The canonical form algorithm for ω-terms over $\mathbf{A}$

In Section 4, we presented an algorithm that computes the canonical form of any given $\bar{\kappa}$-term. In particular, for an ω-term α the algorithm provides a unique $\bar{\kappa}$-term α' in canonical form such that $\mathbf{S} \models \alpha = \beta$ and, so, such that $\mathbf{A} \models \alpha = \beta$. As far as the ω-word problem over $\mathbf{A}$ is concerned, the trouble is that α' does not have to be an ω-term. In effect this is not a difficulty since in order to solve the word problem for κ-terms over $\mathbf{S}$, we also went outside the world of κ-terms. The real trouble is that ω-terms with the same value over $\mathbf{A}$ can have different canonical forms (when the ω-terms are different over $\mathbf{S}$). This is the case, for instance, of the ω-terms $a^\omega ab^\omega$ and $a^\omega bb^\omega$ whose canonical forms are respectively $a^{\omega+1}b^\omega$ and $a^\omega b^{\omega+1}$.

An algorithm that computes, for each ω-term α, a unique ω-term α' in canonical form with the same value over $\mathbf{A}$ can, however, be easily adapted from the algorithm in Section 4. The ω-term α' will then be called the *canonical form of α over* $\mathbf{A}$. For that, it suffices to replace everywhere in the algorithm each occurrence of a symbol $\overset{q}{(}$ or $\overset{q}{)}$ by, respectively, $\overset{0}{(}$ and $\overset{0}{)}$. This way all terms involved are ω-terms and this new algorithm preserves the value of the original ω-term α over $\mathbf{A}$. Indeed, the elementary changes are determined by the following rules for ω-terms, obtained from the rewriting rules for $\bar{\kappa}$-terms of Section 3 by the replacement of the symbols $\overset{q}{(}$ and $\overset{q}{)}$ by $\overset{0}{(}$ and $\overset{0}{)}$,

$$\begin{array}{llll}
1. & \overset{0}{(}\overset{0}{(}\alpha\overset{0}{)}\overset{0}{)} \rightleftarrows \overset{0}{(}\alpha\overset{0}{)} & 4_R. & \overset{0}{(}\alpha\overset{0}{)}\alpha \rightleftarrows \overset{0}{(}\alpha\overset{0}{)} \\
2. & \overset{0}{(}\alpha^n\overset{0}{)} \rightleftarrows \overset{0}{(}\alpha\overset{0}{)} & 4_L. & \alpha\overset{0}{(}\alpha\overset{0}{)} \rightleftarrows \overset{0}{(}\alpha\overset{0}{)} \\
3. & \overset{0}{(}\alpha\overset{0}{)}\overset{0}{(}\alpha\overset{0}{)} \rightleftarrows \overset{0}{(}\alpha\overset{0}{)} & 5. & \overset{0}{(}\alpha\beta\overset{0}{)}\alpha \rightleftarrows \alpha\overset{0}{(}\beta\alpha\overset{0}{)}
\end{array}$$

Actually, these are precisely the rules used in McCammond's algorithm. Our algorithm for ω-terms over $\mathbf{A}$ is essentially the same as McCammond's algorithm except in the procedure to put the crucial portions in canonical form (in view of their distinct definitions). For instance, the canonical form over $\mathbf{A}$ of the ω-terms $a^\omega ab^\omega$ and $a^\omega bb^\omega$ is $a^\omega b^\omega$, while their McCammond's normal form is $a^\omega abb^\omega$.

7.2 Star-freeness of the languages $L_{\mathtt{n},1}(\alpha)$

The star-freeness of the languages $L_{\mathtt{n}}(\alpha)$, for ω-terms α in McCammond's normal form and $\mathtt{n}$ large enough, was established in [6, Theorem 5.1]. For canonical forms an identical property holds.

Theorem 7.1 *Let α be an ω-term in canonical form and let $\mathtt{n} \geq \mu(\alpha)$. Then the language $L_{\mathtt{n},1}(\alpha)$ is star-free.*

The proof of this result can be obtained by a mere adjustment of the corresponding proof of [6, Theorem 5.1] and, so, we do not include it here. Actually, since each subterm of a canonical form is in canonical form as well, the arguments can be usually simplified. As a consequence of Theorem 7.1 and of Proposition 5.8, with $\mathtt{p} = 1$, one gets the following analogue of Theorem 6.1, that establishes the uniqueness of canonical forms for ω-terms over $\mathbf{A}$.

Theorem 7.2 *Let α and β be ω-terms in canonical form. If $\mathbf{A} \models \alpha = \beta$, then $\alpha = \beta$.*

Once again, we omit the proof of this result since it is identical to the proof of the corresponding result [6, Corollary 5.4] for McCammond's normal forms (and it is analogous to the one of Theorem 6.1).

Acknowledgements

I would like to thank Jorge Almeida for asking me about the decidability problem which is the object of this paper. I also benefited greatly from our joint work with Marc Zeitoun about the fascinating McCammond's normal forms.

This work was supported by the European Regional Development Fund, through the programme COMPETE, and by the Portuguese Government through FCT – *Fundação para a Ciência e a Tecnologia*, under the project PEst-C/MAT/UI0013/2011.

References

[1] J. Almeida, *Implicit operations on finite $\mathcal{J}$–trivial semigroups and a conjecture of I. Simon*, J. Pure Appl. Algebra **69** (1990), 205–218.

[2] ———, *Finite semigroups and universal algebra*, World Scientific, Singapore, 1995, English translation.

[3] ———, *Finite semigroups: an introduction to a unified theory of pseudovarieties*, Semigroups, algorithms, automata and languages (Coimbra, 2001), World Scientific, 2002, pp. 3–64.

[4] J. Almeida, J. C. Costa, and M. Zeitoun, *Factoriality and the Pin-Reutenauer procedure*, in preparation.

[5] J. Almeida, J. C. Costa, and M. Zeitoun, *Tameness of pseudovariety joins involving **R***, Monatsh. Math. **146** (2005), 89–111.

[6] J. Almeida, J. C. Costa, and M. Zeitoun, *McCammond normal forms for free aperiodic semigroups revisited*, Tech. Report CMUP 2012-3, Univ. Porto, 2012, submitted, available at `http://cmup.fc.up.pt/cmup/jalmeida/preprints/Almeida-Costa-Zeitoun-McCammond_normal_forms_revisited.pdf`.

[7] J. Almeida and B. Steinberg, *On the decidability of iterated semidirect products and applications to complexity*, Proc. London Math. Soc. **80** (2000), 50–74.

[8] J. Almeida and P. G. Trotter, *The pseudoidentity problem and reducibility for completely regular semigroups*, Bull. Austral. Math. Soc. **63** (2001), 407–433.

[9] J. Almeida and M. Zeitoun, *An automata-theoretic approach of the word problem for ω-terms over* $\boldsymbol{R}$, Theor. Comp. Sci. **370** (2007), 131–169.

[10] J. C. Costa, *Free profinite locally idempotent and locally commutative semigroups*, J. Pure Appl. Algebra **163** (2001), 19–47.

[11] J. C. Costa, C. Nogueira, and M. L. Teixeira, *Semigroup presentations for test local groups*, Semigroup Forum, to appear, preprint available at `http://arxiv.org/abs/1401.0637`.

[12] ______, *The word problem for κ-terms over the pseudovariety of local groups*, in preparation.

[13] J.P. Duval, *Factorizing words over an ordered alphabet*, J. Algorithms **4** (1983), 363–381.

[14] M. Lothaire, *Algebraic combinatorics on words*, Cambridge University Press, Cambridge, UK, 2002.

[15] J.P. McCammond, *Normal forms for free aperiodic semigroups*, Int. J. Algebra Comput. **11** (2001), 581–625.

[16] I.Y. Zhil'tsov, *On identities of finite aperiodic epigroups*, Tech. report, Ural State Univ., 1999.

[17] ______, *On identities of epigroups*, Doklady Math. **62** (2000), 322–324.