数理解析研究所講究録
第 1964 巻 2015 年 176-179

REGULARITY OF ITERATIVE HAIRPIN COMPLETIONS OF CROSSING (2,2)-WORDS*

Kayoko Shikishima-Tsuji
Center for Liberal Arts Education and Research,
Tenri University

May 30, 2015

Abstract

The hairpin completion is a formal operation inspired from DNA biochemistry. It is known that the (one step) hairpin completion of regular language is linear context- free, but not regular in general. It is decidable whether the (one step) hairpin completion of regular language is regular. However, it is an open question whether the iterated hairpin completion of a regular language is regular, even if it is a singleton. If the word is a non-crossing α-word, there are results, but for crossing words there are no results. In this paper, we give necessary and sufficient conditions that the iterated hairpin completion of a given crossing (2, 2)-α-word in $\alpha\Sigma^{*}\bar{\alpha}$ is regular.

1 Introduction

The mathematical hairpin concept was introduced by PĂUN ET AL. in [13]. This operation is inspired by DNA biochemical processes which play an important role in DNA-computing. A DNA-strand is composed of nucleotides which differ from each other in their bases: A (adenine), G (guanine), C (cytosine), and T (thymine). By Watson-Crick base pairing, the complementary bases A and T (resp., G and C) can bond to each other via hydrogen bonds. Two strands can bond to each other if they have opposite orientation and they are base-wise complementary. When a single strand of DNA has a suffix $\overline{\alpha}$ which is the mirrored complement of a middle subword α of the strand, it bonds to the middle subword part and looks like a hairpin. Then a polymerase chain reaction extends the shorter end to generate a complete double strand.

A strand can be seen as a word and a set of strands as a formal language. In this paper we discuss the hairpin manipulations from a language theoretic point of view. The (one step) hairpin completion operation was investigated in [1, 2, 5, 8, 9, 10, 11, 12].

*This paper is extended abstract and the detailed version was submitted.

Most basic algorithmic questions about(one step) hairpin completion have been answered (see, e. g. CHEPTEA ET AL. [1] and DIEKERT ET AL. [2]).

However the situation is different for the iterated version. It is a difficult question whether or not the iterated hairpin completion of a regular language is regular, even if the regular language is a singleton.

In the case of a bounded hairpin completion, which is a weaker variant of the hairpin completion introduced by ITO ET AL. in [4], the iterated bounded hairpin completion of a regular language is regular.

When a given word is a *non crossing* α-word, there are necessary and sufficient conditions that the iterated hairpin completion of this word is regular (see L. KARI, S. KOPECKI AND S. SEKII [5]). In this paper, we give necessary and sufficient conditions that the iterated hairpin completion of a given *crossing* (2, 2)-α-word in $\alpha\Sigma^*\bar{\alpha}$ is regular.

2 Preliminaries

We assume the reader to be familiar with fundamental concepts from Formal Language Theory, such as the classes of the Chomsky hierarchy, finite automaton, regular expressions (e. g., see the textbook by HARRISON [3]), as well as fundamental concepts from combinatorics on words (e. g., see the LOTHAIRE textbook [7]).

Let Σ be a non-empty finite *alphabet* with *letters* as elements. A sequence of letters constitutes a *word* $w \in \Sigma^*$ and we denote by ε the *empty word.*

Words together with the operation of concatenation form a free monoid, which is usually denoted by Σ^* for an alphabet Σ. Any subset of Σ^* is called a language. By Σ^+ we denote the set of non-empty words over Σ.

Repeated concatenation of a word w with itself is denoted by w^i for integers $i \geq 0$ with i representing a *power.* Furthermore, w is said to be *primitive* if there exists no non-empty word u such that $w = u^j$ for some integer $j > 1$. Otherwise, we call w a *repetition* and the smallest such u its *root* (note that in this case the word u is primitive).

The *length* of a finite word w is the number of not necessarily distinct symbols it consists of and is denoted by $|w|$. The ith symbol we write as $w[i]$ and use the notation $w[i..j]$ to refer to the part of a word starting at the ith and ending at the jth position.

A word u is a *factor* of w if there exist integers i, j with $1 \leq i, j \leq |w|$ such that $u = w[i..j]$. We say that u is a *prefix* of w whenever we can fix $i = 1$ and denote this by $u \leq_p w$. If $j < |w|$, then the prefix is called *proper. Suffixes* are the corresponding concept, reading at the end of the word to the front. A word w has a positive integer h as a *period* if for all i, j such that $i \equiv j (\mathrm{mod} h)$ we have $w[i] = w[j]$, whenever both $w[i]$ and $w[j]$ are defined. The number of occurrences of a word u in a word w as a factor is denoted by $|w|_u$.

Let $\bar{\ }$ be an *antimorphic involution,* i.e. $\bar{\ }: \Sigma^* \to \Sigma^*$ is a function, such that for $\bar{\bar{a}} = a$ for all $a \in \Sigma$, and $\overline{uv} = \bar{v}\bar{u}$ for all $u, v \in \Sigma^+$. A word w is said to be a *pseudopalindrome* if $w = \overline{w}$.

Throughout this paper, let k be a fixed integer. For a word $w = \gamma\alpha\beta\overline{\alpha}$ where $\alpha \in \Sigma^k$ and γ, $\beta \in \Sigma^*$, we define *right k-hairpin completion* of w as $\gamma\alpha\beta\bar{\alpha}\bar{\gamma}$. If the number k is obvious, we can omit it. By the notation $w \rightarrow_r w'$, we mean that w' is a right k-hairpin completion of w. The *left k-hairpin completion* is defined analogously. By the notation $w \rightarrow_l w'$, we mean that u is *left k-hairpin completion* of w. The relation of *k-hairpin completion* $\rightarrow$ is defined as $\rightarrow_r$ or $\rightarrow_l$. We denote the reflexive transitive closure of $\rightarrow$, $\rightarrow_r$ and $\rightarrow_l$ by $\rightarrow^*$, $\rightarrow_r^*$ and $\rightarrow_r^*$, respectively.

For a language $L \subseteq \Sigma^+$, we define the *right k-hairpin completion of L* and the *left k-hairpin completion of L*, respectively, as follows: $RH_k(L) = \{w' | \exists w \in L,\ w \rightarrow_r w'\}$, $LH_k(L) = \{w' | \exists w \in L,\ w \rightarrow_l w'\}$. We also define the *k-hairpin completion of L* by $H_k(L) = RH_k(L) \cup LH_k(L)$. The *iterated k-hairpin completion $H_k^*(L)$ of L* is defined inductively, as follows: let $H_k^0(L) = L$, $H_k^n(L) = H_k(H_k^{n-1}(L))$ for every positive integer n, and $H_k^*(L) = \bigcup_{n \geq 0} H_k^n(L)$.

3 Main Results

We will restrict out attention to words of the form $\alpha\Sigma^*\bar{\alpha}$ where $|\alpha| = k$. This does not restrict generality, because, if any word $w \in \alpha\Sigma^*\beta$ and $|\alpha| = |\beta| = k$ has a one-step hairpin completion w' of w, then w' is in $\alpha\Sigma^*\bar{\alpha}$ or in $\bar{\beta}\Sigma^*\beta$. Regularity of the iterated hairpin completion of the word w depends on regularity of $H_k^*(w')$. In the rest of this paper, we investigate regularity of $H_k^*(w)$ where a given word w is in $\alpha\Sigma^*\bar{\alpha}$.

A word $w \in \Sigma^*$ is said to be an *(m, n)-α-word* (or simply an *(m, n)-word*) if $[w]_\alpha = m$ and $[w]_{\bar{\alpha}} = n$. We say that an (m,n)-α-word w is *non-α-crossing* if the rightmost occurrence of α precedes the leftmost occurrence of $\bar{\alpha}$ on w. Otherwise, the word w is *α-crossing*.

As an (m,n)-α-word in $\alpha\Sigma^*\bar{\alpha}$ is always non-crossing for $m = 1$ or $n = 1$, then if an (m,n)-α-word in $\alpha\Sigma^*\bar{\alpha}$ is crossing, we have $m, n \geq 2$.

We have the following theorem:

Theorem 1 *Let $x, y \in \alpha\Sigma^*\bar{\alpha}$ be $(1,1)$-words and w be a crossing $(2,2)$-word in $x\Sigma^* \cap \Sigma^*y$. Then the iterated k-hairpin completion $H_k^*(w)$ of w is regular, if and only if, w is a pseudopalindrome or x and y are both pseudopalindromes.*

References

[1] D. Cheptea, C. Martín-Vide and V. Mitrana, A new operation on words suggested by DNA biochemistry: Hairpin completion, *Transgressive Computing* (2006) 216-228.

[2] V. Diekert, S. Kopecki and V. Mitrana, On the Hairpin Completion of Regular Languages, ICTAC, Ed. by M. Leucker and C. Morgan, (Springer, 2009)170-184.

[3] M.A. Harrison, Introduction to Formal Language Theory, (Addison-Wesley, Reading, Massachusetts, 1978).

[4] M. Ito, P. Leupold, F. Manea and V. Mitrana, Bounded hairpin completion, *Information and Computation* **209** (2011) 471-485.

[5] L. Kari, S. Kopecki and S. Seki, Iterated hairpin completions of non-crossing words, SOFSEM, Ed. by M. Bieliková, G. Friedrich, G. Gottlob, S. Katzenbeisser and G. Turán, (Springer, 2012)337-348.

[6] S. Kopecki, On iterated hairpin completion, *Theoretical Computer Science* **412** (2011) 3629-3638.

[7] M. Lothaire, Combinatorics on Words, (Cambridge University Press, 1997).

[8] F. Manea, C. Martín-Vide and V. Mitrana, On some algorithmic problems regarding the hairpin completion, *Discrete Applied Mathematics* **157** (2009) 2143-2152.

[9] F. Manea, C. Martín-Vide and V. Mitrana, Hairpin lengthening, CiE, Ed. by F. Ferreira, B. Löwe, E. Mayordomo and L.M. Gomes, (Springer, 2010)296–306.

[10] F. Manea and V. Mitrana, Hairpin Completion Versus Hairpin Reduction, CiE, Ed. by S. B. Cooper, B. Löwe and A. Sorbi, (Springer, 2007)532-541.

[11] F. Manea, V. Mitrana and T. Yokomori, Two complementary operations inspired by the DNA hairpin formation: Completion and reduction, *Theoretical Computer Science* **410** (2009) 417-425.

[12] F. Manea, V. Mitrana and T. Yokomori, Some Remarks on the Hairpin Completion, *International Journal on Foundations of Computer Science* **21** (2010) 859-872.

[13] G. Păun, G. Rozenberg and T. Yokomori, Hairpin Languages, *International Journal of Foundations of Computer Science* (2001) 837-847.

Center for Liberal Arts Education and Research
Tenri Univertsity
1050 Somanouchi
Tenri, Nara 632-8510
Japan
E-mail address: tsuji@sta.tenri-u.ac.jp

天理大学・総合教育研究センター　辻　佳代子