数理解析研究所講究録
第 1915 巻 2014 年 113-117

Regularity of Iterative Hairpin Completion of crossing words.

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

1 Introduction

The mathematical hairpin notion introduced in [5] is a word in which some suffix is the mirrored complement of a middle factor of the word. This operation were introduced which are very similar in nature to it. Mirrored complementary sequences occur frequently in DNA and are often found at functionally interesting locations such as replication origins or operator sites. The most basic question about hairpin completion is "given a word, can we decide whether the iterated hairpin completion of the word is regular?" The situation is very complicated. In the special case when the word is non-crossing, some results were given ([1],[3]). In section 3, we add some results in the case when the word is crossing.

2 Preliminaries

We assume the reader to be familiar with basic concepts as alphabet, word, language and regular expression (for more details see [2]).

Words together with the operation of concatenation form a free monoid, which is usually denoted by Σ^* for an alphabet Σ. Repeated concatenation of a word w with itself is denoted by w^i for nonnegative integer i. The length of a finite word w is the number of not necessarily distinct symbols it consists of and is written by $|w|$.

Let θ be an *antimorphic involution*, i.e. $\theta : \Sigma^* \to \Sigma^*$ is a function, such that for $\theta(\theta(a)) = a$ for all $a \in \Sigma$, and $\theta(uv) = \theta(v)\theta(u)$ for all $u, v \in \Sigma^+$. Then, w is a *(θ-)pseudopalindrome* if $w = \theta(w)$. To make notation cleaner, we write $\bar{u}$ for $\theta(u)$, when θ is understood.

Throughout this paper, let k be a fixed integer. For a word $w=\gamma\alpha\beta\bar{\alpha}$ for some $\alpha \in \Sigma^k$ and $\gamma, \beta \in \Sigma^*$, we define *right (k-)hairpin completion* $\gamma\alpha\beta\bar{\alpha}\bar{\gamma}$ of $w=\gamma\alpha\beta\bar{\alpha}$ (with respect to α). By the notation $w \to_r u$ (or $w \,{}_k\!\!\to_r u$), we mean that u is right (k-)hairpin completion of w. The *left hairpin completion* is defined analogously. By the notation $w \to_l u$, we mean that u is left (k-)hairpin completion of w. The relation of *hairpin completion* $\to$ is defined as $\to_r$ or $\to_l$. By $\to^*$, $\to_r^*$ and $\to_l^*$, we denote the reflexive transitive closure of $\to$, $\to_r$ and $\to_l$, respectively.

For a language $L \subseteq \Sigma^+$, we define *the right (k-)hairpin completion* of *L* by $RH_k(L) = \{\gamma\alpha\beta\bar{\alpha}\bar{\gamma} \mid \gamma\alpha\beta\bar{\alpha} \in L, \gamma \in \Sigma^+, \beta \in \Sigma^*$ and $|\alpha| = k\}$ and *the left (k-)hairpin completion* of *L* by $LH_k(L) = \{\gamma\alpha\beta\bar{\alpha}\bar{\gamma} \mid \gamma\alpha\beta\bar{\alpha} \in L, \gamma \in \Sigma^+, \beta \in \Sigma^* \mid \alpha| = k\}$. *The (k-)hairpin completion* of *L* is $H_k(L) = RH_k(L) \cup LH_k(L)$. And we define the *iterated* (right or left) (*k*-)hairpin completion $H_k^*(L)$ (or $RH_k^*(L)$ or $LH_k^*(L)$) of *L*, inductively as follows:

$$H_k^0(L) = H_k(L),\ H_k^{n+1}(L) = H_k(H_k^n(L)) \text{ and } H_k^*(L) = \bigcup_{n\geq 0} H_k^n(L),$$

$$RH_k^0(L) = RH_k(L),\ RH_k^{n+1}(L) = RH_k(RH_k^n(L)) \text{ and } RH_k^*(L) = \bigcup_{n\geq 0} RH_k^n(L),$$

$$LH_k^0(L) = LH_k(L),\ LH_k^{n+1}(L) = LH_k(LH_k^n(L)) \text{ and } LH_k^*(L) = \bigcup_{n\geq 0} LH_k^n(L),$$

3 Iterated hairpin completion of crossing words

A word $w \in \Sigma^+$ is an *(m, n)-α-word* (or simply *(m, n)-word*) if the numbers of occurrence of α and $\bar{\alpha}$ in *w*, are *m* and *n* respectively. We say that an *(m, n)*-α-word *w* is *non-α-crossing* if the rightmost occurrence of α precedes the leftmost occurrence of α on *w*. Otherwise, the word is *α-crossing*.

Let $w \in \alpha\Sigma^* \cap \Sigma^*\beta$ and $|\alpha| = |\beta| = k$. By one-step hairpin completion with respect to α and β, we get two words $w' \in \bar{\beta}\Sigma^* \cap \Sigma^*\beta$ and $w'' \in \alpha\Sigma^* \cap \Sigma^*\bar{\alpha}$. Regularity of the iterative hairpin completion of *w* is depend on those of $H_k^*(w')$ and $H_k^*(w'')$. Our problem in this paper is "for a given word w, whether the iterated hairpin completion of w is regular?" In the rest of this paper, we may assume that the word w is in $\alpha\Sigma^* \cap \Sigma^*\bar{\alpha}$.

Let *w* be a crossing *(m, n)*-α-word. Then we have $m,n \geq 2$.

The following example shows that the iterated hairpin completion of a (3, 2)-α crossing word is not always regular ([4]).

Example 1. (Kopecki) We consider a crossing word $w = \alpha ba\bar{\alpha}\alpha c\bar{\alpha}$. Let $u = \bar{b}\bar{\alpha}$, $v = \alpha\bar{\alpha}b\bar{\alpha}$ and $R = wu^+v\bar{u}^+\overline{wu}^+\bar{w}$. Then, $R \cap H_k^*(\{w\}) = \{wu^r v\bar{u}^r \overline{wu}^r \bar{w} \mid r \geq 1\}$. Therefore, the iterated hairpin completion of a word is not always regular.

We have some results for the iterated hairpin completion of a (2, 2)-α crossing word.

Proposition 1. Let $w \in \alpha\Sigma^* \cap \Sigma^*\bar{\alpha}$ be (2, 2)-α-crossing-word such that $w = xvy$ where *x* and *y* are (1,1)-α-words in $\alpha\Sigma^*\bar{\alpha}$ and $v \in \Sigma^*$. If *x* and *y* are pseudo-palindromes, then $H_k^*(w)$ is regular.

Proof) Let $R=\{x(vx)^*(vy(\overline{v}x)^+(vx)^*)^*vy(\overline{v}x)^+\}$ and $L=\{(y\overline{v})^+xv((yv)^*(y\overline{v})^+xv)^*(yv)^*y\}$. Since these languages are regular and $H_k^*(w)=H_k^*(xvy\overline{v}x)\cup H_k^*(y\overline{v}xvy)\cup\{w\}$, we show that $H_k^*(w)=R\cup L\cup\{w\}$. First, we prove that $H_k^*(xvy\overline{v}x)=R$.

We show that $H_k^*(xvy\overline{v}x)\supset R$. Let $R_n=x(vx)^*(vy(\overline{v}x)^+(vx)^*)^n(vy)(\overline{v}x)^+$ for every nonnegative integer n. Then it is obvious that $R=\bigcup_{n\geq 0}R_n$. We show that $R_n\subset H_k^*(xvy\overline{v}x)$ by induction on $n\geq 0$. It is clear that $R_0=x(vx)^*vy(\overline{v}x)^+\subset H_k^*(xvy\overline{v}x)$ by the following:

$$xvy\overline{v}x\rightarrow_l^i x(vx)^i vy(\overline{v}x)\rightarrow_r^j x(vx)^i vy(\overline{v}x)^{j+1}.$$

Suppose that $R_n=x(vx)^*(vy(\overline{v}x)^+(vx)^*)^n(vy)(\overline{v}x)^+$ is contained in $H_k^*(xvy\overline{v}x)$ for some nonnegative integer n. Every word w in R_{n+1} is written by the form $x(vx)^r vy(vx)^{m_1}(vx)^{t_1}\cdots vy(\overline{v}x)^{m_{n+1}}(vx)^{t_{n+1}}vy(\overline{v}x)^s$ where $m_1,\cdots,m_{n+1}>0$, $t_1,\cdots,t_{n+1}\geq 0$, and $r,s\geq 0$. We show that the word $w_1=xvy(vx)^{m_1}(vx)^{t_1}\cdots vy(\overline{v}x)^{m_{n+1}}(vx)^{t_{n+1}}vy(\overline{v}x)$ is in $H_k^*(xvy\overline{v}x)$.

(1) When $m_1>t_{n+1}$, let $w'=xvy(\overline{v}x)^{m_1}(vx)^{t_1}\cdots vy(\overline{v}x)^{m_n}(vx)^{t_n}vy(\overline{v}x)$. We have

$$\begin{aligned}w'&=xvy(\overline{v}x)^{m_1}(vx)^{t_1}\cdots vy(\overline{v}x)^{m_n}(vx)^{t_n}vy(\overline{v}x)\rightarrow_r^{m_{n+1}-1}xvy(\overline{v}x)^{m_1}(vx)^{t_1}\cdots vy(\overline{v}x)^{m_n}(vx)^{t_n}vy(\overline{v}x)^{m_{n+1}}\\&=xvy\overline{v}(x\overline{v})^{t_{n+1}}\cdot x(\overline{v}x)^{m_1-t_{n+1}-1}(vx)^{t_1}\cdots vy(\overline{v}x)^{m_n}(vx)^{t_n}vy(\overline{v}x)^{m_{n+1}}\\&\rightarrow_r xvy\overline{v}(x\overline{v})^{t_{n+1}}\cdot x(\overline{v}x)^{m_1-t_{n+1}-1}(vx)^{t_1}\cdots vy(\overline{v}x)^{m_n}(vx)^{t_n}vy(\overline{v}x)^{m_{n+1}}\cdot(vx)^{t_{n+1}}vy\overline{v}x=w_1\end{aligned}$$

Since w' is in R_n, it is contained in $H_k^*(xvy\overline{v}x)$. Then the word w_1 is also in $H_k^*(xvy\overline{v}x)$.

(2) When $m_1\leq t_{n+1}$, let $w''=xvy(\overline{v}x)^{m_2}(vx)^{t_2}\cdots vy(\overline{v}x)^{m_{n+1}}(vx)^{t_{n+1}}vy(\overline{v}x)$.

We have

$$\begin{aligned}w''&=xvy(\overline{v}x)^{m_2}(vx)^{t_2}\cdots vy(\overline{v}x)^{m_{n+1}}(vx)^{t_{n+1}}vy(\overline{v}x)\\&\rightarrow_r^{t_1}x(vx)^{t_1}vy(\overline{v}x)^{m_2}(vx)^{t_2}\cdots vy(\overline{v}x)^{m_{n+1}}(vx)^{t_{n+1}}vy(\overline{v}x)\\&=x(vx)^{t_1}vy(\overline{v}x)^{m_2}(vx)^{t_2}\cdots vy(\overline{v}x)^{m_{n+1}}(vx)^{t_{n+1}-m_1+1}\cdot(vx)^{m_1-1}vy(\overline{v}x)\\&\rightarrow_r xvy\overline{v}(x\overline{v})^{m_1-1}\cdot x(vx)^{t_1}\cdot vy(\overline{v}x)^{m_2}(vx)^{t_2}\cdots vy(\overline{v}x)^{m_{n+1}}(vx)^{t_{n+1}-m_1+1}\cdot(vx)^{m_1-1}vy(\overline{v}x)=w_1\end{aligned}$$

Since w'' is in R_n, it is contained in $H_k^*(xvy\overline{v}x)$. Then the word w_1 is also in $H_k^*(xvy\overline{v}x)$.

It is clear that $w\in H_k^*(w_1)\subset H_k^*(H_k^*(xvy\overline{v}x))=H_k^*(xvy\overline{v}x)$. We proved that $H_k^*(xvy\overline{v}x)\supset R$.

To prove that $H_k^*(xvy\overline{v}x)\subset R$ is easy. For a nonnegative integer n, suppose that $H_k^n(xvy\overline{v}x)\subset R$. Then we have $H_k^{n+1}(xvy\overline{v}x)=H_k^1(H_k^n(xvy\overline{v}x))\subset H_k^1(R)\subset R$.

As the proof of $H_k^*(y\overline{v}xvy)=L$ is similar, we omit it. □

The following example is that the iterated hairpin completion of a $(2,2)$-α crossing word w which is not satisfies the condition of Proposition 1, that is $w=xvy$ such that $y\neq\overline{y}$, is not always regular.

Example 2 We consider a (2, 2)-α-crossing word $w=\alpha b\bar{\alpha}c\alpha d\bar{\alpha}$ where $a,b,c,d\in\Sigma$, $a\neq\bar{a}, b=\bar{b}, c\neq\bar{c}, d\neq\bar{d}$ and $\alpha=aa$. For integers $i,j>0$, suppose a word $w_{i,j}=xcy(\bar{c}\,\bar{x})^i(cx)^j c\bar{y}\,\bar{c}\,x$ where $x=\alpha b\bar{\alpha}$ and $y=\alpha d\bar{\alpha}$ is in $H_k^*(w)$. Since $w=xcy$ appears only one time as a factor of $w_{i,j}$, we have $w_{i,j}\in RH_k^*(w)$.
We have $w=\alpha b\bar{\alpha}c\alpha d\bar{\alpha}=xcy \to_r^m \alpha b\bar{\alpha}c\alpha d\bar{\alpha}\cdot(\bar{c}\alpha\bar{b}\bar{\alpha})^m=xcy\cdot(\bar{c}\,x)^m \to_r^1 xcy\cdot(\bar{c}\,x)^{m+1}$ or $xcy(\bar{c}\,x)^m\cdot(cx)^s c\bar{y}\,\bar{c}\,x$ where $m>0$ and $m\geq s>0$. Let $S=xcy(\bar{c}\,x)^*(cx)^*c\bar{y}\,\bar{c}\,x$. It is easy to see that $H_k^*(w)\cap S = RH_k^*(w)\cap S=\{xcy(\bar{c}\,x)^i(cx)^j c\bar{y}\,\bar{c}\,x \mid i\geq j\geq 0\}$. Since S is regular and $H_k^*(w)\cap S$ is not regular, $H_k^*(w)$ is not regular.

We have the following corollaries by the proof of Proposition 1.

Corollary 1. Let $w\in\alpha\Sigma^*\cap\Sigma^*\bar{\alpha}$ be (3, 3)-α-crossing-word such that $w=xvy\bar{v}\,\bar{x}$ where x and y are (1,1)-α-words in $\alpha\Sigma^*\bar{\alpha}$ and $v\in\Sigma^*$. If x and y are pseudo-palindromes, then the iterated hairpin completion $H_k^*(w)$ of w is regular and $H_k^*(w)=x(vx)^*(vy(\bar{v}x)^+(vx)^*)^*vy(\bar{v}x)^+$.

Corollary 2. Let $w\in\alpha\Sigma^*\cap\Sigma^*\bar{\alpha}$ be (2, 2)-α-crossing-word such that $w=xvy$ where x and y are (1,1)-α-words in $\alpha\Sigma^*\bar{\alpha}$ and $v\in\Sigma^*$. If x, y and v are pseudo-palindromes, then the iterative hairpin completion $H_k^*(w)$ is regular and $H_k^*(w)=\{w\}\cup x(vx)^*(vy(vx)^+)^*vy(vx)^+ \cup (yv)^+xv((yv)^+xv)^*(yv)^*y$.

The following theorem is proved by the similar way of Proposition 1.

Theorem 1. Let w be (2, 2)-α-crossing-word such that $w=x\Sigma^+\cap\Sigma^+y$ where x and y are (1,1)-α-words in $\alpha\Sigma^*\cap\Sigma^*\bar{\alpha}$. If x and y are pseudo-palindromes, then $H_k^*(w)$ is regular.

Proof) If $|w|\geq|x|+|y|$, then we already proved in Proposition 1. If $|w|\leq|x|+|y|-2|\alpha|$, then w is non-crossing. We may assume that $|x|+|y|-2|\alpha|<|w|<|x|+|y|$. Since $\bar{\alpha}$ overlaps with α, there exist factors $u,v\in\Sigma^+$ of α such that $\alpha=vu$, $\bar{\alpha}\alpha=\bar{u}vu=\bar{u}\,\bar{v}u$ and $w=vx_0vy_0v$ where $x=vx_0v$, $y=vy_0v$, $x_0,y_0\in\Sigma^*$.

Let $R=(vx_0)^+(vy_0(vx_0)^+)^*vy_0(vx_0)^+v$ and $L=v(y_0v)^+x_0v((y_0v)^+x_0v)^*(y_0v)^+$. Since $H_k^*(w)=H_k^*(vx_0vy_0vx_0v)\cup H_k^*(vy_0vx_0vy_0v)\cup\{w\}$, to prove the theorem we show that $R=H_k^*(vx_0vy_0vx_0v)$ and $L=H_k^*(vy_0vx_0vy_0v)$.

First, we prove that $R=H_k^*(vx_0vy_0vx_0)$. Let $R_n=(vx_0)^+(vy_0(vx_0)^+)^n vy_0(vx_0)^+v$ for every nonnegative integer n. Then it is obvious that $R=\bigcup_{n\geq 0}R_n$.

We show that $R_n\subset H_k^*(vx_0vy_0vx_0v)$ by induction on $n\geq 0$. It is clear that $R_0=(vx_0)^+vy_0(vx_0)^+v \subset H_k^*(vx_0vy_0vx_0v)$. Suppose that $R_n=(vx_0)^+(vy_0(vx_0)^+)^n vy_0(vx_0)^+v$ is contained in $R_n\subset H_k^*(vx_0vy_0vx_0v)$ for some nonnegative integer n. Every word w in R_{n+1} is written by the form $(vx_0)^r vy_0(vx_0)^{m_1}\cdots vy_0(vx_0)^{m_{n+1}}vy_0(vx_0)^s v$ where $m_1,\cdots,m_{n+1}>0$ and $r,s>0$. We show that the word $w_1=vx_0vy_0(vx_0)^{m_1}\cdots vy_0(vx_0)^{m_{n+1}}vy_0(vx_0)v$ is in $H_k^*(vx_0vy_0vx_0v)$.

Let $w'=vx_0vy_0(vx_0)^{m_1}\cdots vy_0(vx_0)^{m_n}vy_0(vx_0)v$. We have

$$
\begin{aligned}
w' &= vx_0vy_0(vx_0)^{m_1}\cdots vy_0(vx_0)^{m_n}vy_0(vx_0)v \ {}_k\!\!\rightarrow_r^{m_{n+1}-1} vx_0vy_0(vx_0)^{m_1}\cdots vy_0(vx_0)^{m_n}vy_0(vx_0)^{m_{n+1}}v \\
&= vx_0vy_0\cdot(vx_0)(vx_0)^{m_1-m_{n+1}}\cdots vy_0(vx_0)^{m_n}vy_0(vx_0)^{m_{n+1}}v \\
&\rightarrow_r vx_0vy_0\cdot(vx_0)^{m_1}\cdots vy_0(vx_0)^{m_n}vy_0(vx_0)^{m_{n+1}}vy_0vx_0v=w_1
\end{aligned}
$$

Since w' is in R_n, it is contained in $H_k^*(vx_0vy_0vx_0v)$. Then the word w_1 is also in $H_k^*(vx_0vy_0vx_0v)$ and then $R\subset H_k^*(vx_0vy_0vx_0v)$.

On the other hand, for a nonnegative integer n, suppose that $H_k^n(vx_0vy_0vx_0v)\subset R$, then we have $H_k^1(H_k^n(vx_0vy_0vx_0v))=H_k^1(H_k^n(vx_0vy_0vx_0v))\subset H_k^1(R)\subset R$.

As the proof of $H_k^*(vy_0vx_0vy_0v)=L$ is similar, we omit it. □

References.

[1] V. Diekert, S. Kopecki, and V. Mitrana. On the hairpin completion of regular languages. In ICTAC, vol. 5684 of Lect. Notes Comput. Sci., 170–184, 2009.

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

[3] Lila Kari, Steffen Kopecki, Shinnosuke Seki: Iterated Hairpin Completions of Non-crossing Words. SOFSEM 2012: 337-348

[4] S. Kopecki, On the iterated hairpin completion, Theoretical Computer Science 412 (2011) 3629.

[5] G. Paun, G. Rozenberg, and T. Yokomori. Hairpin languages. Int. J. Found. Comput. Sci., pages 837–847, 2001.

Center for Liberal Arts Education and Research
Tenri University
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