



HAL
open science

k-anti-power uniform morphisms

Francis Wlazinski

► **To cite this version:**

| Francis Wlazinski. k-anti-power uniform morphisms. 2024. hal-04375742v2

HAL Id: hal-04375742

<https://hal.science/hal-04375742v2>

Preprint submitted on 7 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

k -anti-power uniform morphisms

Francis Wlazinski

March 7, 2024

Abstract

Words whose every $k(\geq 3)$ successive factors of the same length are all different, i.e., k -anti-power words are natural extensions of square-free words (every two successive factors of the same length are different). We give a way to verify whether a uniform morphism preserves k -anti-power words, i.e., the image by this morphism of all k -anti-power words are k -anti-power words.

A consequence of the existence of such morphisms is the possibility of generating an infinite k -anti-power word.

1 Preliminaries

Let us recall some basic notions of Combinatorics of words.

1.1 Words

An *alphabet* A is a finite set of symbols called *letters*. A *word* over A is a finite sequence of letters from A . The *empty word* ε is the empty sequence of letters. Equipped with the concatenation operation, the set A^* of words over A is a free monoid with ε as neutral element and A as set of generators. Since an alphabet with one element is limited interest to us, we always assume that the cardinality of considered alphabets is at least two. Given a non-empty word $u = a_1 \dots a_n$ with $a_i \in A$ for any integer i from 1 to n , the *length* of u denoted by $|u|$ is the integer n that is the number of letters of u . By convention, we have $|\varepsilon| = 0$. We denote by A^+ the set of words of positive length over A , i.e., $A^+ = A^* \setminus \{\varepsilon\}$.

Given two words w and u , we denote by $|w|_u$ the number of different words p such that pu is a prefix of w . For instance, if $w = abaababa$, we have $|w|_a = 5$, $|w|_{aba} = 3$.

An infinite word over A is a map from \mathbb{N} to A that is an infinite sequence of letters $a_1 \dots a_n \dots$ with $a_i \in A$. And $A^{\mathbb{N}}$ is the set of all infinite words over A .

The *mirror image* of u , denoted by \tilde{u} , is the word $a_n \dots a_2 a_1$. In the particular case of the empty word, $\tilde{\varepsilon} = \varepsilon$.

A word u is a *factor* of a word v if there exist two (possibly empty) words p and s such that $v = pus$. We denote $\text{Fcts}(v)$ the set of all factors of v . If $u \in \text{Fcts}(v)$, we also say that v

contains the word u (as a factor). If $p = \varepsilon$, u is a *prefix* of v . If $s = \varepsilon$, u is a *suffix* of v . If $u \neq v$, u is a *proper* factor of v . If u , p and s are non-empty, u is an *internal* factor of v . Furthermore, $\text{alph}(v) = \{u \in \text{Fcts}(v) \mid |u| = 1\}$ is the set of all the letters in v .

Let w be a non-empty word and let i, j be two integers such that $0 \leq i - 1 \leq j \leq |w|$. We denote by $w[i..j]$ the factor of w such that $|w[i..j]| = j - i + 1$ and $w = pw[i..j]s$ for two words s and p verifying $|p| = i - 1$. When $j > i$, $w[i..j]$ is simply the factor of w that starts at the i^{th} letter and ends at the j^{th} . Note that, when $j = i - 1$, we have $w[i..j] = \varepsilon$. When $i = j$, we also denote by $w[i]$ the factor $w[i..i]$ which is the i^{th} letter of w . In particular, $w[1]$ and $w[|w|]$ are respectively the first and the last letter of w .

Powers of a word are defined inductively by $u^0 = \varepsilon$, and for any integer $n \geq 1$, $u^n = uu^{n-1}$. Given an integer $k \geq 2$, since the case ε^k is of little interest, we call a k -*power* any word u^k with $u \neq \varepsilon$. And a word is k -*power-free* if it does not contain any k -power as factor. A *primitive* word is a word which is not a ℓ -power of another word whatever the integer $\ell \geq 2$.

Given two integers $p > q \geq 1$ and two words x and y with $xy \neq \varepsilon$, a word of the form $(xy)^\alpha x$ with $\frac{|(xy)^\alpha x|}{|xy|} = \alpha + \frac{|x|}{|xy|} = \frac{p}{q}$ is called a $\frac{p}{q}$ -*power*. For instance, the word *anchorman* is a $(1 + \frac{2}{7} =) \frac{9}{7}$ -power and the word *abaabaa* is a $(2 + \frac{1}{3} =) \frac{7}{3}$ -power. In particular, a $\frac{3}{2}$ -power is a word of the form xyx with $|x| = |y| \neq \varepsilon$. For instance, the word *antman* is a $\frac{3}{2}$ -power. A word is $\frac{p}{q}$ -*power-free* if it does not contain any ℓ -power as factor with $\ell \geq \frac{p}{q}$. The word *abcaba* is not $\frac{8}{5}$ -power-free. Indeed, it contains the word *abcab* which is a $\frac{5}{3}$ -power.

A *border* of a word w is a factor $v \notin \{\varepsilon, w\}$ of w that is both a prefix and a suffix of w . When v is the greatest border of w and $|w| > 2|v|$, we say that the *bordered* word w is a *sesquipower*.

The two following propositions are well known. The first one is about bordered words.

Proposition 1.1 [8] *Let A be an alphabet and u, v, w three non-empty words over A .*

If $vt = uv$ then there exist two words r and s over A and an integer n such that $v = (rs)^n r$, $u = rs$ and $t = sr$.

Lemma 1.2 [6, 7] *If a non-empty word v is an internal factor of vv , i.e., if there exist two non-empty words x and y such that $vv = xvy$, then there exist a non-empty word t and two integers $i, j \geq 1$ such that $x = t^i$, $y = t^j$, and $v = t^{i+j}$.*

Given an integer $k \geq 2$ and an integer $j \geq 1$, a (k, j) -*anti-power sequence* or simply a k -*anti-power* [5] is a concatenation of k consecutive pairwise different words of the same length j .

For instance, if $A = \{a, b\}$, the words $u = aababab$ and $v = ababababaaa$ are respectively $(4, 2)$ -anti-power and $(4, 3)$ -anti-power sequences. But the prefix *abababab* of v is not a 4-anti-power sequence: it is even a 4-power.

Given an integer $k \geq 1$, if $\text{Card}(A) = \alpha \geq 2$ then there exist α^n different words in A^* of length $n \geq 1$. Among the words of length $k \times n$, there are α^n different k -powers (of length $k \times n$) and $A_{\alpha^n}^k = \frac{(\alpha^n)!}{(\alpha^n - k)!}$ different (k, n) -anti-power sequences if $\alpha^n \geq k$ and 0 otherwise.

It particularly means that there exists an integer k_0 such that there are no (k', n) -anti-power sequences over A for any $k' \geq k_0$.

For any alphabet A with $\text{Card}(A) = \alpha \geq 2$ and for any integer $k \geq 2$, there exists a smallest integer p_0 such that $\alpha^{p_0} \geq k$. And, if $p \geq p_0$, the set of (k, p) -anti-power sequences is greater than the set of k -powers of length $p \times k$.

A *2-anti-power word* is simply a square-free word. Given an integer $k \geq 3$, a word w is a *k-anti-power word* if it is a $(k-1)$ -anti-power word and, when $|w| \geq k$, if any factor of w of length $k \times j$ for every integer $1 \leq j \leq \lfloor \frac{|w|}{k} \rfloor$ is a (k, j) -anti-power sequence. As precised, w can contain a factor of length k only if $|w| \geq k$. But, by the choosen definition, a word of length n with $2 \leq n < k$ is a k -anti-power word if and only if it is a n -anti-power word. Of course, such a word exists only if $\text{Card}(A) = \alpha \geq n$.

In other words, for any integer $2 \leq \ell \leq k$, if we choose ℓ successive factors of the same length of a k -anti-power word, they are all different.

A word w is not a k -anti-power word ($k \geq 2$) if and only if it contains a factor of the form $U = u_1 u_2 \dots u_\ell$ for an integer $2 \leq \ell \leq k$ and for some non-empty words u_1, u_2, \dots, u_ℓ of the same length and such that $u_1 = u_\ell$. It means that U is a $\frac{\ell}{\ell-1}$ -power. So, a word w that is not a k -anti-power word contains at least a $\frac{k}{k-1}$ -power (when $\ell = k$). The converse is false. For instance, the word $abcdeabc$ is a $\frac{8}{5} (> \frac{4}{3})$ power. But it is a 4-anti-power word over a five-letter alphabet.

Let $\mathcal{A}_{A,k}$, or \mathcal{A}_k if no doubt exists for the alphabet A , be the set of all k -anti-power words over A . By definition, we have $\mathcal{A}_k \subset \mathcal{A}_{k-1} \dots \subset \mathcal{A}_2$. So let $\mathcal{B}_{A,k}$, or \mathcal{B}_k , be the set $\mathcal{A}_{k-1} \setminus \mathcal{A}_k$.

Remark 1.3 *We have $w \in \mathcal{A}_k$ if and only if $\tilde{u} \in \mathcal{A}_k$.*

Lemma 1.4 [5]

Let v be a border of a word w and let u be the word such that $w = uv$. If $|w| > \ell|u|$ for an integer ℓ then u^ℓ is a prefix of w .

More precisely, by Proposition 1.1, if v is a border of a word $w = uv (\neq \varepsilon)$, there exist two words r and s over A and an integer $n \geq 0$ such that $u = rs$ and $w = (rs)^{n+1}r$. When $|w| \geq 2|u|$, we get that w is a $\frac{p}{q}$ -power with $\frac{p}{q} \geq 2$. So, Proposition 1.1 and Lemma 1.4 mean that a bordered word w that contains a border of length at least $\frac{|w|}{2}$ is not a k -anti-power word for all integers $k \geq 2$. So, we will only be interested in sesquipowers, i.e., $\frac{p}{q}$ -powers with $1 < \frac{p}{q} < 2$.

Lemma 1.5 *Let A be an alphabet, let $d \geq 1$ be an integer and let $X, Y \in A^*$ such that $XY \neq \varepsilon$. If $d|XY|$ and $|Y| \leq (d-1)|X|$ then $XYX \notin \mathcal{A}_k$ for all integers $k > d$.*

Proof.

In fact, we will show that $XYX \notin \mathcal{A}_{d+1}$.

If $d = 1$ then $Y = \varepsilon$, $X \neq \varepsilon$ and $XYX = X^2 \notin \mathcal{A}_2$.

If $d \geq 2$, we have $|XY| > \frac{d-1}{d}|XY| \geq |Y|$. Let X_2 be the suffix of X such that $|X_2Y| = \frac{d-1}{d}|XY|$ and let X_1 be the non-empty word such that $X = X_1X_2$. Since $|X_1| = |XY| - |X_2Y| = \frac{1}{d}|XY|$, the word $[X_1]X_2Y[X_1] \notin \mathcal{A}_{d+1}$. \square

Taking $d = \gcd(|XY|; k)$, we get:

Corollary 1.6 *Let A be an alphabet and let $X, Y \in A^*$ such that $XY \neq \varepsilon$.*

For any integer $k \geq 2$, if $\gcd(|XY|; k) \neq k$ and if $|Y| \leq (\gcd(|XY|; k) - 1)|X|$, then $XYX \notin \mathcal{A}_k$.

Remark 1.7 *When $d = 2$ in Lemma 1.5, we get that if $|XY|$ is even and if $|Y| \leq |X|$ then $XYX \notin \mathcal{A}_k$ for all integers $k \geq 3$.*

Lemma 1.8 *Let A be the alphabet $\{a_1, a_2, \dots, a_k, \dots\}$ and let $k \geq 3$ be an integer. The word $[a_1a_2 \dots a_{k-1}]a_k[a_1a_2 \dots a_{k-1}]$ is a k -anti-power word if and only if k is prime.*

Proof.

Let $w_{A,k}$ be the word $[a_1a_2 \dots a_{k-1}]a_k[a_1a_2 \dots a_{k-1}]$.

For any integer $2 \leq \ell < k$, by Corollary 1.6, $w_{A,k}$ is a ℓ -anti-power word only if $\gcd(k; \ell) = 1$. Since $\mathcal{A}_k \subset \mathcal{A}_\ell$, the word $w_{A,k}$ is a k -anti-power word only if k is prime.

Let us now assume that k is prime and, by contradiction, let us assume that $w_{A,k} \notin \mathcal{A}_k$. It means that $w_{A,k}$ contains a factor of the form $U = u_1u_2 \dots u_\ell$ for an integer $2 \leq \ell \leq k$ and for some words u_1, u_2, \dots, u_ℓ of the same length $\Lambda \geq 1$ with $u_1 = u_\ell$. Let i_0 be the integer such that a_{i_0} is the first letter of u_1 and let u'_1 be the word $u_1 = a_{i_0}u'_1$. So $a_{i_0}u'_1u_2 \dots u_{\ell-1}a_{i_0}u'_1$ is a factor of $w_{A,k}$. Since a_{i_0} appears only twice in $w_{A,k}$, we get that $k = |a_{i_0}u'_1u_2 \dots u_{\ell-1}| = (\ell - 1) \times \Lambda$. Since k is a prime number, we get that $\ell = 2$ and $\Lambda = k$. It follows that $|u_1u_\ell| = 2k > |w_{A,k}|$: this is impossible. \square

Lemma 1.9 *Let A be an alphabet and let $X, Y \in A^*$ such that $|X|_a \times |Y|_a = 0$ for all letters $a \in A$. If $X \in \mathcal{A}_k$ and $Y \in \mathcal{A}_k$ and $|Y| > (k - 2)|X|$ then $XYX \in \mathcal{A}_k$ for any integer $k \geq 2$.*

Proof.

By contradiction, let us assume that $XYX \notin \mathcal{A}_k$ with $X \neq \varepsilon$. It means that XYX contains a factor of the form $U = u_1u_2 \dots u_\ell$ for an integer $2 \leq \ell \leq k$ and for some words u_1, u_2, \dots, u_ℓ of the same length $\Lambda \geq 1$ with $u_1 = u_\ell$.

Let a be the first letter of u_1 and let u'_1 be the word such that $u_1 = au'_1$.

If $|X|_a = 0$ then $au'_1u_2 \dots a \in \text{Fcts}(Y)$. We get that $U \in \text{Fcts}(Y)$: a contradiction with the hypothesis that $Y \in \mathcal{A}_k$. Thus, $|Y|_a = 0$ and it implies that $au'_1 = u_\ell \in \text{Fcts}(X)$. Since $U \notin \text{Fcts}(X)$, it follows that $|Y| \leq |u_2 \dots u_{\ell-1}| = (\ell - 2) \times \Lambda \leq (k - 2) \times |X|$: a contradiction with the hypotheses. \square

An *infinite k -anti-power word* is an infinite word whose all finite factors are k -anti-power words. Obviously, the first question is whether such a word exists.

If $A = \{a, b\}$, the only 2-anti-power words are aba , bab and their factors. But, for any $k \geq 3$, the only k -anti-power words are a , b , ab and ba .

If $\text{Card}(A) \geq 3$, there exist infinite 2-anti-power (square-free) words [1, 10, 11].

If $k = 3$ and $A = \{a, b, c\}$, the only 3-anti-power words are $abcab$, the exchange of letters of this word and their factors. Let us note that the word $abcab$ is not $\frac{3}{2}$ -power-free.

A $\frac{3}{2}$ -power-free word contain neither a factor of the form xyx with $|x| = |y|$, nor a factor of the form xx . Thus a $\frac{3}{2}$ -power-free word is a 3-anti-power word (but the converse does not hold). Thus a Dejean's word [4, 3, 9] over a four-letter alphabet, which does not contain any ℓ -power with $\ell > \frac{7}{5}$ -power-free, is a 3-anti-power word.

More generally, if $k \geq 3$ is a prime number and if A is a k -letter alphabet $\{a_1, a_2, \dots, a_k\}$, the word $a_1 a_2 \dots a_{k-1} a_k a_1 a_2 \dots a_{k-1}$ is a k -anti-power word. But it contains a $\frac{2k-1}{k}$ ($> \frac{k}{k-1}$) power. Furthermore, a Dejean's word over a $(k+1)$ -letter alphabet that avoids any power greater than $\frac{k+1}{k}$ ($\leq \frac{k}{k-1}$) is a k -anti-power word.

Remark 1.10 *If we had chosen not to add that a k -anti-power word must be a $(k-1)$ -anti-power word, we would have, for instance, that, for $A = \{a; b; c\}$, the word $abcabcab$ would have been a 3-anti-power word but not a 2-anti-power word.*

More precisely, without the condition that a k -anti-power word w must be a $(k-1)$ -anti-power word, we only could say that all prefixes and all suffixes of w of length between $k-1$ and $\left\lfloor \frac{(k-1)|w|}{k} \right\rfloor$ are $(k-1)$ -anti-power words.

For an infinite word, it does not change anything to add the condition that a k -anti-power word w must be a $(k-1)$ -anti-power word. Indeed, every factor of w whose length is a multiple of $k-1$ can be extended to a factor of w whose length is a multiple of k . Obviously, if these k factors are different, the same holds for $k-1$ ones.

Corollary 1.11 *Let $k \geq 2$ be an integer, let A be the alphabet $\{a_0, a_1, \dots, a_{k+3}, \dots\}$, and let w be a Dejean's word over $\{a_0, a_1, \dots, a_k\}$. The word $a_{k+1} a_{k+2} w a_{k+1} a_{k+2}$ is a k -anti-power word if at least one of these conditions holds :*

- i. $|w| \geq 2k - 3$.*
- ii. $|w| \geq k - 2$ and $|w|$ is not a multiple of 2*

The word $a_{k+1} a_{k+2} a_{k+3} w a_{k+1} a_{k+2} a_{k+3}$ is a k -anti-power word if at least one of these conditions holds :

- i. $|w| \geq 3k - 5$.*
- ii. $|w| \geq 2k - 4$ and $|w|$ is not a multiple of 3*
- iii. $|w| \geq k - 3$, $|a_{k+3} w|$ is not a multiple of 2 and $|w|$ is not a multiple of 3*

More generally, for any integer $p \geq 2$, let w_p be the word $a_{k+1} \dots a_{k+p} w a_{k+1} \dots a_{k+p}$. If $|w| + i$ is not a multiple of $p - i$ for every integer $0 \leq i \leq p - 2$ and if $|w| \geq k - p$ then w_p is a k -anti-power word.

Proof. It directly comes from Lemma 1.9 and the fact that the word Sw is a k -anti-power word for any suffix S of $a_{k+1} \dots a_{k+p}$. \square

1.2 Morphisms

Let A and B be two alphabets. A *morphism* f from A^* to B^* is a mapping from A^* to B^* such that $f(uv) = f(u)f(v)$ for all words u, v over A . When B has no importance, we say that f is a morphism on A or that f is defined on A .

Given an integer $L \geq 1$, f is L -uniform if $|f(a)| = L$ for every letter a in A . A morphism f is *uniform* if it is L -uniform for some integer $L \geq 1$.

Let $k \geq 2$ be an integer and let A and B be two alphabets. A morphism f from A^* to B^* is a k -anti-power morphism if and only if $f(w)$ is a k -anti-power word over B for every k -anti-power word w over A , i.e., $f(\mathcal{A}_{A,k}) \subset \mathcal{A}_{B,k}$. For instance, the *identity endomorphism* Id ($\forall a \in A, Id(a) = a$) is a k -anti-power morphism.

In particular, a 2-anti-power morphism is a square-free morphism. These last morphisms have been characterized in [2].

Given a morphism f on A , the *mirror morphism* \tilde{f} of f is defined for all words w over A , by $\tilde{f}(w) = f(\tilde{w})$. In particular, $\tilde{f}(a) = \tilde{f(a)}$ for every letter a in A .

Remark 1.12 *By remark 1.3, we get that f is a k -anti-power morphism if and only if \tilde{f} is a k -anti-power morphism.*

A morphism on A is called *prefix* (resp. *suffix*) if, for all different letters a and b in A , the word $f(a)$ is not a prefix (resp. not a suffix) of $f(b)$. A prefix (resp. suffix) morphism is non-erasing. A morphism is *bifix* if it is prefix and suffix.

Proofs of the following lemmas are left to the reader.

Lemma 1.13 *Let f be a bifix morphism on an alphabet A and let u, v, w , and t be words over A .*

The equality $f(u) = f(v)p$ with p be a prefix of $f(w)$ implies $u = vw'$ for a prefix w' of w such that $f(w') = p$.

And the equality $f(u) = sf(v)$ with s a suffix of $f(t)$ implies $u = t'v$ for a suffix t' of t such that $f(t') = s$.

Lemma 1.14 *Let f be a prefix morphism on an alphabet A , let u and v be words over A , and let a and b be letters in A . Furthermore, let p_1 (resp. p_2) be a prefix of $f(a)$ (resp. of $f(b)$). If $(p_1; p_2) \neq (\varepsilon; f(b))$ and if $(p_1; p_2) \neq (f(a); \varepsilon)$ then the equality $f(u)p_1 = f(v)p_2$ implies $u = v$ and $p_1 = p_2$.*

Lemma 1.15 *Let f be a suffix morphism on an alphabet A , let u and v be words over A , and let a and b be letters in A . Furthermore, let s_1 (resp. s_2) be a suffix of $f(a)$ (resp. of $f(b)$). If $(s_1; s_2) \neq (\varepsilon; f(b))$ and if $(s_1; s_2) \neq (f(a); \varepsilon)$ then the equality $s_1 f(u) = s_2 f(v)$ implies $u = v$ and $s_1 = s_2$.*

Taking $p_1 = p_2 = \varepsilon$ (resp. $s_1 = s_2 = \varepsilon$) in Lemma 1.14 (resp Lemma 1.15), we get that a prefix (resp. suffix) morphism is injective.

Definition 1.16 *A morphism f from A^* to B^* is a ps-morphism (Keränen [6] called f a ps-code) if and only if the equalities*

$$f(a) = ps, f(b) = ps' \text{ and } f(c) = p's$$

with $a, b, c \in A$ (possibly $c = b$) and $p, s, p', s' \in B^$ imply $b = a$ or $c = a$.*

Obviously, taking $c = b$, and $s = \varepsilon$ in a first time and $p = \varepsilon$ in a second time, we obtain that a ps-morphism is a bifix morphism.

Lemma 1.17 [6, 7] *If f is not a ps-morphism then f is not a k -power-free morphism for every integer $k \geq 2$.*

2 Main results

Proposition 2.1 *Let $k \geq 2$ be an integer, let A and B be two alphabets with $\text{Card}(A) = p \geq k$ and let f be a L -uniform morphism from A^* to B^* (with $L \geq 1$). If there exist five letters a, b, c, d and x (possibly equal) and four words p, s, π and σ such that s is a suffix of $f(a)$, p is a prefix of $f(b)$, σ is a suffix of $f(c)$, π is a prefix of $f(d)$, $(\pi; p) \neq (\varepsilon; f(b))$, $(\sigma; s) \neq (\varepsilon; f(a))$ and $sp = \sigma f(x)\pi$ then f is not a k -anti-power morphism.*

Proof.

In fact, we will show that there exists a square-free word of length at most two whose image by f contains a square (f is not a square-free morphism). Therefore, this word will be a k -anti-power-word for any $k \geq 2$ but its image will not be. This will mean that f is not a k -anti-power morphism.

Since $(\pi; p) \neq (\varepsilon; f(b))$, $(\sigma; s) \neq (\varepsilon; f(a))$ and $|sp| = L + |\sigma\pi|$, we get $|s| > |\sigma|$ and $|p| > |\pi|$.

Let s' be the non empty prefix of $f(x)$ such that $s = \sigma s'$ and let p' be the non empty suffix of $f(x)$ such that $p = p'\pi$.

If $x = a$ and $x = b$, then $f(x)$ is an internal factor of $f(xx)$. By Lemma 1.2, $f(x)$ is not primitive, i.e., f is not a k -anti-power morphism.

Therefore, at least one of the word ax or bx is not a square. But $f(ax)$ contains the square $s's'$ and $f(bx)$ contains the square $p'p'$. \square

Proposition 2.2 *Let $k \geq 2$ be an integer. Let A and B be two alphabets. Let f be a L -uniform morphism from A^* to B^* (with $L \geq 2$). Let us assume that there exists a prime number p such that $\text{Card}(A) \geq p \geq k$.*

If there exists an integer $2 \leq d < k$ such that $d|L$ then f is not a k -anti-power morphism.

Proof.

Let us assume that $A = \{a_1, a_2 \dots a_p, \dots\}$.

The word $w_{A,p} = a_1 a_2 \dots a_{p-1} a_p a_1 a_2 \dots a_{p-1} \in \mathcal{A}_p \subset \mathcal{A}_k \subset \mathcal{A}_{d+1}$. But, by Lemma 1.5, the word $[f(a_1)(a_2) \dots f(a_{p-1})] f(a_p) [f(a_1)(a_2) \dots f(a_{p-1})] \notin \mathcal{A}_{d+1}$, i.e., $f(w_{A,p}) \notin \mathcal{A}_k$. \square

Proposition 2.3 *Let $k \geq 3$ and $L \geq 2$ be two integers. Let A and B be two alphabets with $\text{Card}(A) \geq k$ and let f be L -uniform morphism from A^* to B^* . Let us assume that there exist three letters a, b and c such that $a \notin \{b, c\}$ and $f(bac)$ contains a square $T^2 (\neq \varepsilon)$.*

If there exists an integer $2 \leq d < k$ such that $d||T|$ then f is not a k -anti-power morphism.

Proof.

If $b \neq c$ then abc is a k -anti-power word and f is not a k -anti-power morphism.

From now, let us assume that $f(bab) = \pi T^2 \sigma$ (with $|T| \geq 1$).

If $|\pi| > |f(b)|$ (resp. $|\sigma| > |f(b)|$) then T^2 is factor of $f(ab)$ (resp. $f(ba)$) with ab (resp. ba) a k -anti-power word: f is not a k -anti-power morphism.

Thus there exist two words p and s such that $f(a) = ps$, $\pi T = f(b)p$ and $T\sigma = sf(b)$. Without loss of generality (using mirror image for instance), we may assume that $|p| \geq |s|$.

If $|T| = |L|$ then $T = sp$, $f(b) = ps$ and $f(ba)$ contains the square $(ps)^2$: f is not a k -anti-power morphism.

If $|T| > |L|$, let Y be non-empty word such that $T = sYp$. It follows that $f(b) = \pi sY = Yp\sigma$. By Lemma 1.1, there exist two words u and v and an integer n such that $Y = u(vu)^n$, $p\sigma = vu$ and $\pi s = uv$. Moreover, since $Y \neq \varepsilon$, we get that $uv \neq \varepsilon$. If $n \geq 1$ then $f(b)$ contains $(uv)^2$ that is f is not a k -anti-power morphism. From now, $u = Y$ and $f(b) = YvY$. If $|p| \geq |v|$ (resp. $|s| \geq |v|$) then p starts with v (resp. s ends with v) and $f(ba)$ (resp. fab) contains the square $(Yv)^2$ (resp. $(vY)^2$): f is not a k -anti-power morphism. When $|p| < |v|$ and $|s| < |v|$, the word v starts with p and ends with s . But $|v| < |f(b)| = |f(a)| = |ps|$. Thus there exist three words p' , s' and $Z \neq \varepsilon$ such that $v = p'Zs'$, $p = p'Z$ and $s = Zs'$. It means that $f(a)$ contains Z^2 : f is not a k -anti-power morphism.

If $0 < |T| < |L| (= |sp|)$, since T starts with s and ends with p , let p' , s' and $Y \neq \varepsilon$ be the words such that $p = Yp'$, $s = s'Y$ and $T = s'Yp'$. It follows that $f(a) = Yp's'Y$. Moreover, from the equality $f(bab) = \pi T^2 \sigma$, we get that $f(b)$ starts with p' and ends with s' . Thus, there exist a word B such that $|B| = 2|Y|$ and $f(b) = p'Bs'$.

We have $f(ab) = Y[p'][s'Y][p']Bs'$ and $f(ba) = p'B[s'Y][p']s'Y$. If there exist an integer $2 \leq d < k$ such that $d||T|$, i.e., $d||p's'Y|$, since either $|s'Y| \leq |p'| \leq (d-1)|p'|$ or $|p'| \leq |s'Y| \leq (d-1)|s'Y|$, by Lemma 1.5, $f(ab) \notin \mathcal{A}_{B,k}$ or $f(ba) \notin \mathcal{A}_{B,k}$: f is not a k -anti-power morphism.

\square

Remark 2.4 Let B_0 be the alphabet $\{a, b, c, d, e, f, g, h, i, j, k, l, m\}^*$.

Let f be a uniform morphism from $\{a, b\}^*$ to B_0^* such that $f(a) = abcdefghiab$, $f(b) = cdefjklmghi$.

Then $f(bab)$ contains the square $(ghiabcdef)^2$. Therefore, f is not a 2-anti-power morphism (is not a square-free morphism).

For any integer $k \geq 3$, the only k -anti-power words over $\{a, b\}$ are a , b , ab , and ba .

Moreover, $f(ab)$ and $f(ba)$ are 3-anti-power words: f is a 3-anti-power morphism.

But $f(ba)$ contains $ghiabcdefghi$. Thus f is not a k -anti-power morphism for any integer $k \geq 4$.

Let $A = \{a_1, a_2, \dots, a_\alpha\}$ be an alphabet. A *trivial* morphism from A^* to B^* is an injective 1-uniform morphism. That is, $f(a_i) = b_i$ for some letter $b_i \in B$ and $b_i \neq b_j$ when $i \neq j$.

Proposition 2.5 Let $k \geq 2$ be an integer and let A be an alphabet. Let f be a non-trivial L -uniform morphism defined on A . If $(\text{alph}(f(a)))_{a \in A}$ are not pairwise disjoint sets and if $L < k$ then f is not a k -anti-power morphism.

Proof.

Let assume that $\text{Card}(A) = \alpha$ and $A = \{a_1, a_2, \dots, a_\alpha\}$.

Since $(\text{alph}(f(a)))_{a \in A}$ are not pairwise disjoint sets, let x be a letter and let $\ell \neq m$ two integers such that $x \in \text{alph}(f(a_\ell)) \cap \text{alph}(f(a_m))$.

We have $f(a_\ell) = A'_\ell x A''_\ell$ and $f(a_m) = A'_m x A''_m$ for four words A'_ℓ , A''_ℓ , A'_m and A''_m such that $A'_\ell + A''_\ell = A'_m + A''_m = L - 1$.

So $f(a_\ell a_m) = A'_\ell x A''_\ell A'_m x A''_m$ and $f(a_m a_\ell) = A'_m x A''_m A'_\ell x A''_\ell$ with either $|xA''_\ell A'_m| \leq L \leq k - 1$ or $|xA'_m A''_\ell| \leq L \leq k - 1$. Since $a_\ell a_m$ and $a_m a_\ell$ are k -anti-power words, it means that f is not a k -anti-power morphism. \square

A morphism f on A is k -anti-power up to ℓ ($k, \ell \geq 2$) if and only if $f(w)$ is a k -anti-power word for every k -anti-power word w over A of length at most ℓ .

Remark 2.6 Let A and B be two alphabets with $\text{Card}(A) \geq k \geq 3$ and let f be a 1-uniform morphism. Then f is a k -anti-power morphism if and only if f is trivial. So f is a k -anti-power morphism if and only if it is a k -anti-power morphism up to 2.

Proposition 2.7 Let $k \geq 3$ be an integer and let A and B be two alphabets. Let us assume that there exists a prime number p such that $\text{Card}(A) \geq p \geq k$. Let f be a square-free L -uniform morphism from A^* to B^* .

Then f is a k -anti-power morphism if and only if it is a k -anti-power morphism up to $2p - 1$.

Proof.

By definition of k -anti-power morphisms, we only have to prove the "if" part of Proposition 2.7. By Lemma 1.17, f (square-free) is a ps-morphism and so injective.

We may assume that, if there exists an integer $d \geq 2$ such that $d|L$, then $d \geq k$. Indeed (see the Proof of Proposition 2.2), if $d|L$ and $d < k$ then there exists a word $w_{A,p} \in \mathcal{A}_{A,k}$ such that $|w_{A,p}| = 2p - 1$ and $f(w_{A,p}) \notin \mathcal{A}_{B,k}$: f is not a k -anti-power morphism. It particularly means that if $2 \leq \ell \leq k$, then $\gcd(L, \ell - 1) = 1$.

By contradiction, we assume that a shortest k -anti-power word w (not necessarily unique) such that $f(w)$ contains a non- k -anti-power satisfies $|w| \geq 2p$. We will show that this assumption leads to contradictions.

Since the length of w is minimal, we may assume that there exist an integer $2 \leq \ell \leq k$ and some words $p, s, u_1, u_2, \dots, u_\ell$ such that $f(w) = pu_1u_2 \dots u_\ell s = p \left(\prod_{q=1}^{\ell} u_q \right) s$ where p is a prefix of $f(w[1])$ different from $f(w[1])$ and s is a suffix of $f(w[|w|])$ different from $f(w[|w|])$. Moreover, the words $(u_j)_{j=1, \dots, \ell}$ have the same length $\Lambda (\geq 1)$ and $u_1 = u_\ell$.

Let $i_0 = 1$ and, for all integers $1 \leq j \leq \ell$, let i_j be the shortest integer such that $p \prod_{q=1}^j u_q$ is the prefix of $f(w[1..i_j])$. By definition, we have $i_\ell = |w|$.

If there exist two different integers j and j' such that $i_j = i_{j'}$ then we get that $\Lambda \leq L$. It implies that $f(w) = |p| + |s| + \ell \times \Lambda \leq (k + 2)L$, i.e., $|w| \leq k + 2 < 2p$: a contradiction.

Thus $1 = i_0 < i_1 < \dots < i_\ell = |w|$. If we denote a_j the letter $w[i_j]$ and x_j the word $w[i_{j-1} + 1..i_j - 1]$, we have $w = a_0 x_1 a_1 x_2 a_2 \dots x_\ell a_\ell$. For all integers $0 \leq j \leq \ell$, there exist some words p_j and s_j such that $f(a_j) = p_j s_j$ and $u_j = s_{j-1} f(x_j) p_j$. In particular, we have $p_0 = p$ and $s_\ell = s$. Moreover, by definition, the words $s_0, p_1, p_2, \dots, p_\ell$ are non empty.

Let us note that, for all integers $1 \leq j_1, j_2 \leq \ell$, we have $||x_{j_1}| - |x_{j_2}|| \leq 1$. Indeed, in the contrary, for instance if $|x_{j_1}| \geq |x_{j_2}| + 2$ (obviously with $j_1 \neq j_2$), we get that $\Lambda = |u_{j_1}| = |s_{j_1-1} f(x_{j_1}) p_{j_1}| > |f(x_{j_1})| \geq |f(x_{j_2})| + 2L$. Furthermore, $\Lambda = |u_{j_2}| = |s_{j_2-1} f(x_{j_2}) p_{j_2}| < |f(x_{j_2})| + 2L$: this is impossible. Since $\sum_{q=1}^{\ell} x_q \geq 2p - \ell - 1 \geq \ell - 1$, it also implies that $\text{card}\{q \mid 1 \leq q \leq \ell \text{ and } x_q = \varepsilon\} \leq 1$.

If $j_1 < j_2$ and $u_{j_1} = u_{j_2}$ with $(j_1, j_2) \neq (1, \ell)$ then $w[i_{j_1-1}..i_{j_2}]$ is a proper factor of w and $f(w[i_{j_1-1}..i_{j_2}])$ contains the non- $(j_2 - j_1 + 1)$ -anti-power word $\prod_{q=j_1}^{j_2} u_q$ that is a non- k -anti-power word: a contradiction with the hypothesis of the minimal length of w .

In particular, we have $s_0 f(x_1) p_1 = u_1 = u_\ell = s_{\ell-1} f(x_\ell) p_\ell$.

• *Case 1* : $|s_0| \neq |s_{\ell-1}|$ and $(s_0, s_{\ell-1}) \neq (f(a_0), \varepsilon)$

Since $\text{card}\{q \mid 1 \leq q \leq \ell \text{ and } x_q = \varepsilon\} \leq 1$, we have either $x_1 \neq \varepsilon$ or $x_\ell \neq \varepsilon$.

If $s_{\ell-1} \neq \varepsilon$ and $(0 <) |s_0| < |s_{\ell-1}|$, by a length criterion, we necessarily have $x_1 \neq \varepsilon$. Let χ be the first letter of x_1 and let x'_1 be the word such that $x_1 = \chi x'_1$. If $x_\ell = \varepsilon$, let $P = p_\ell$ and if $x_\ell \neq \varepsilon$, let $P = f(\gamma)$ where γ is the first letter of x_ℓ . In particular, we have P non-empty. Let π be the non empty prefix of $f(x'_1) p_1$ such that $s_0 f(\chi) \pi = s_{\ell-1} P$. By Proposition 2.1 (see also its proof), this last equation implies that f is not a square-free morphism: a contradiction with the hypotheses.

The case $s_{\ell-1} \neq \varepsilon$ with $|s_0| > |s_{\ell-1}|$ and the case $s_{\ell-1} = \varepsilon$ with $s_0 \neq f(a_0)$ are solved as previous one using Proposition 2.1.

• *Case 2* : $(s_0, s_{\ell-1}) = (f(a_0), \varepsilon)$

We get that $f(a_0x_1)p_1 = f(x_\ell)p_\ell$. By Lemma 1.14, we get $a_0x_1 = x_\ell$ and $p_1 = p_\ell$. It implies that $|s_1| = |s_\ell|$.

It follows that $(\ell - 1) \times \Lambda = \left| \prod_{q=2}^{\ell} u_q \right| = |s_1 f(x_2 a_2 x_3 a_3 \dots x_\ell) p_\ell| = |f(x_2 a_2 x_3 a_3 \dots x_\ell a_\ell)| = L \times |x_2 a_2 x_3 a_3 \dots x_\ell a_\ell| = L \times (\ell - 1) + L \times \sum_{q=2}^{\ell} |x_q|$. Since $\gcd(L, \ell - 1) = 1$, we get that $\sum_{q=2}^{\ell} |x_q|$ (≥ 1) is a multiple of $\ell - 1$. But $||x_{j_1}| - |x_{j_2}|| \leq 1$ for all integers $2 \leq j_1, j_2 \leq \ell$, so it implies that $|x_{j_1}| = |x_{j_2}| = |x_1| + 1$, $\Lambda = L \times (|x_1| + 2)$ and $|s_{j-1} p_j| = L$ for all $1 \leq j \leq \ell$.

For all integers $2 \leq q \leq \ell$, we have $|u_q| = |f(a_0x_1)p_1| = |f(x_\ell)p_\ell| = |s_{q-1} f(x_q) p_q|$ and so $|s_{q-1} p_q| = |p_\ell|$. It means that all the $(|s_q|)_{q=1..l}$ are equal to 0 and that all the $(|p_q|)_{q=1..l}$ are equal to L , i.e., $p_q = f(a_q)$.

In particular, we get that $f(a_1) = p_1 = p_\ell = f(a_\ell)$ and thus $a_1 = a_\ell$. It follows that $w = a_0 x_1 a_1 x_2 a_2 \dots x_{\ell-1} a_{\ell-1} x_1 a_1$ with $|a_q x_q| = |x_1 a_1|$ for all integers $q \in \llbracket 2; \ell - 1 \rrbracket$, i.e., w is not a k -anti-power word: a contradiction.

• *Case 3* : $|s_0| = |s_{\ell-1}|$

As previous case, we obtain, as a first step, that Λ is a multiple of L . Indeed, we have $(\ell - 1) \times \Lambda = \left| \prod_{q=1}^{\ell-1} u_q \right| = |s_0 f(x_1 a_1 x_2 a_2 \dots x_{\ell-1}) p_{\ell-1}| = |f(x_1 a_1 x_2 a_2 \dots x_{\ell-1} a_{\ell-1})| = L \times |x_1 a_1 x_2 a_2 \dots x_{\ell-1} a_{\ell-1}| = L \times (\ell - 1) + L \times \sum_{q=1}^{\ell-1} |x_q|$. Since $\gcd(L, \ell - 1) = 1$, we get that $\sum_{q=1}^{\ell-1} |x_q|$ (≥ 1) is a multiple of $\ell - 1$. So $|x_j| = |x_1|$, $\Lambda = L \times (1 + |x_1|)$ and $|s_{j-1} p_j| = L$ for all $1 \leq j \leq \ell$.

As previous case, it means that all the $(|s_q|)_{q=0..l}$ are equal and, consequently, all the $(|p_q|)_{q=0..l}$ are also equal.

From the equalities $s_0 f(x_1) p_1 = s_{\ell-1} f(x_\ell) p_\ell$ and $|s_0| = |s_{\ell-1}|$, we get $s_{\ell-1} = s_0 (\neq \varepsilon)$. By Lemma 1.14, it also implies $x_\ell = x_1$ and $p_1 = p_\ell (\neq \varepsilon)$.

In particular, since all the $(|x_i|)_{i=1..l}$ are equal and since $w = a_0 x_1 a_1 x_2 a_2 \dots x_{\ell-1} a_{\ell-1} x_1 a_\ell$ is a k -anti-power word, we have $a_0 \neq a_{\ell-1}$ and $a_1 \neq a_\ell$.

If $a_0 = a_1$ then we get $p_0 = p_1 (= p_\ell)$, $f(a_0) = p_0 s_0$, $f(a_{\ell-1}) = p_{\ell-1} s_{\ell-1} = p_{\ell-1} s_0$ and $f(a_\ell) = p_\ell s_\ell = p_0 s_\ell$. It means that $f(a_{\ell-1} a_0 a_\ell)$ contains $(s_0 p_0)^2$ with $a_{\ell-1} a_0 a_\ell$ square-free since $a_0 \neq a_{\ell-1}$ and $a_0 = a_1 \neq a_\ell$: a contradiction with the hypothesis that f is a square-free morphism.

In the same way, if $a_{\ell-1} = a_\ell$, we get that $f(a_0 a_\ell a_1)$ contains $(s_\ell p_\ell)^2$ with $a_0 a_\ell a_1$ square-free. If $a_0 = a_\ell$, we get that $f(a_{\ell-1} a_\ell a_1)$ contains $(s_\ell p_\ell)^2$ with $a_0 a_\ell a_1$ square-free. And, if $a_1 = a_{\ell-1}$, we get that $f(a_0 a_1 a_\ell)$ contains $(s_1 p_1)^2$ with $a_0 a_1 a_\ell$ square-free. All these cases lead to the same contradiction.

Thus $a_0, a_1, a_{\ell-1}$ and a_ℓ are four different letters. It means that $a_0 a_1 a_{\ell-1} a_\ell$ is a k -anti-power word of length lower than $2p$. But $f(a_0 a_1 a_{\ell-1} a_\ell)$ contains the non-3-anti-power sequence $s_0 p_1 s_1 p_{\ell-1} s_{\ell-1} p_\ell = s_0 p_1 s_1 p_{\ell-1} s_0 p_1$: a (final) contradiction with the minimality of $|w|$. \square

References

- [1] J. Berstel. Axel Thue's papers on repetition in words: a translation. Technical Report 20, Laboratoire de Combinatoire et d'Informatique Mathématique, Université du Québec, Montréal, 1995.
- [2] M. Crochemore. Sharp characterizations of squarefree morphisms. *Theoretical Computer Science*, 18:221–226, 1982.
- [3] James D. Currie and Narad Rampersad. A proof of dejean's conjecture. *Mathematics of computation*, 80:1063–1070, 2011.
- [4] F. Dejean. Sur un théorème de Thue. *J. Comb. Theory*, 13:90–99, 1972. series A.
- [5] Gabriele Fici, Antonio Restivo, Manuel Silva, and Luca Q. Zamboni. Anti-powers in infinite words. *Journal of Combinatorial Theory, Series A*, 157:109 – 119, 2018.
- [6] V. Keränen. On the k -freeness of morphisms on free monoids. *Annales Academiae Scientiarum Fennicae* 61, Series A, 1986.
- [7] M. Leconte. *Codes sans répétition*. PhD thesis, LITP Université Paris 6, october 1985.
- [8] M. Lothaire. *Combinatorics on words*, volume 17 of *Encyclopedia of Mathematics*. Addison-Wesley, 1983. Reprinted in 1997 by Cambridge University Press in the Cambridge Mathematical Library, Cambridge, UK, 1997.
- [9] Michaël Rao. Last cases of dejean's conjecture. *Theoretical Computer Science*, 412(27):3010 – 3018, 2011. *Combinatorics on Words (WORDS 2009)*.
- [10] A. Thue. Über unendliche zeichenreihen. *Kristiania Videnskapsselskapets Skrifter Klasse I. Mat.-naturv*, 7:1–22, 1906.
- [11] A. Thue. Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. *Kristiania Videnskapsselskapets Skrifter Klasse I. Mat.-naturv*, 1:1–67, 1912.