ELSEVIER

Theoretical Computer Science 218 (1999) 61–81

Theoretical Computer Science

On Abelian squares and substitutions

Arturo Carpi *

Istituto di Cibernetica del CNR, via Toiano 6, 80072 Arco Felice (NA), Italy

Abstract

We study substitutions preserving Abelian square-free words and the more general notion of substitutions with bounded Abelian squares. In particular, we prove the existence of algorithms deciding whether a substitution mapping each letter into a set of commutatively equivalent words belongs to one of these classes. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Abelian square; Free monoid; Avoidable regularity

1. Introduction

Repetitions are one of the main topics in the study of avoidable regularities in words. The reader is referred to [1] for a comprehensive summary.

In literature, among several generalizations of the notion of repetition, an important place is occupied by Abelian squares. An Abelian square is the concatenation of two words which are permutations of each other. It is easily seen that, on a 3-letter alphabet, any word of length 8 contains an Abelian square. On the contrary, the existence of an unending word on 4 letters without Abelian squares among its factors (*Abelian square-free word*), conjectured by Erdős [5] in 1961, was proved by Keränen [7] more than 30 years later. Unending Abelian square-free words on larger alphabets had been previously found [6, 8]. Abelian repetitions of larger exponent were considered in [4].

The Abelian square-free words exhibited in the quoted papers are obtained by iteration of morphisms preserving Abelian square-free words (*Abelian square-free morphisms*). An effective characterization of Abelian square-free morphisms defined on alphabets with 6 or more letters can be found in [2]. The more general notion of a finite substitution preserving Abelian square-free words (*Abelian square-free*

* E-mail: arturo@arturo.cib.na.cnr.it.

PII: S0304-3975(98)00250-3

substitution) was used by the author to prove the exponential growth of Abelian square-free words on 4 letters [3].

Till now, iteration of Abelian square-free substitutions seems to be the only known method to systematically produce Abelian square-free words.

The aim of this paper is the study of Abelian square-free substitutions. We introduce the more general notion of a finite substitution with bounded Abelian squares, i.e. of a finite substitution such that the length of the Abelian squares occurring in the images of Abelian square-free words is upper bounded by a constant.

In order to describe the main results of this paper, we need to introduce some definitions. They will be repeated, in a more formal way, in the next section.

Two words are said to be *commutatively equivalent* if they are permutations of each other, like e.g. *abcab* and *baabc*. A finite substitution is said to be *commutatively functional* if any letter is mapped onto a set of commutatively equivalent words. The class of commutatively functional substitutions obviously contains all the morphisms of free monoids, but also the Abelian square-free substitutions considered in [3]. A finite substitution is said to be *commutatively injective* if words which are not commutatively equivalent are mapped into words which are not commutatively equivalent. To any commutatively injective finite substitution σ, we can effectively associate a finite set $\mathscr{L}_\sigma$ (see Eq. (11)), whose elements are tuples constituted by 3 letters and a vector.

The main result of this paper is an effective characterization of commutatively functional Abelian square-free substitutions. More precisely, we prove that a commutatively functional substitution σ is Abelian square-free if and only if it satisfies the following three conditions:

1. It is commutatively injective.
2. No 'short' Abelian square occurs in the images of Abelian square-free words.
3. The set $\mathscr{L}_\sigma$ does not intersect a certain finite set T.

Here, by 'short' Abelian square we mean that at least one of its two commutatively equivalent halves occurs in an image of a letter.

There is a difference between the cases where the considered finite substitution is defined on 4 letters or more. Indeed, if σ is defined on more than 4 letters, then we can explicitly list the elements of the finite set T quoted above. Thus, the previous characterization gives an algorithm to decide whether a commutatively functional substitution is Abelian square-free. If, on the contrary, σ is defined on 4 letters, then we are not able to give explicitly the elements of T. Thus, although an algorithm exists to decide whether a commutatively functional substitution, defined on 4 letters, is Abelian square-free, we do not know how such a decision should be taken.

If we consider finite substitutions which are not commutatively functional, then conditions 1–3 above do not characterize Abelian square-free substitutions. However, we show that any finite substitution satisfying conditions 1–3 is Abelian square-free.

We give also an effective characterization of commutatively functional substitutions with bounded Abelian squares. Indeed, we show that a commutatively functional substitution σ has bounded Abelian squares if and only if it is commutatively injective and $\mathscr{L}_\sigma$ does not intersect a certain finite set.

An interesting property concerning commutatively functional substitutions with bounded Abelian squares defined on 5 or more letters is also established. It ensures that, in order to check the Abelian square-freeness of an image of an Abelian square-free word, it is sufficient to check the Abelian square-freeness of its factors which are images of words of length 2.

A fundamental role in the proof of the aforementioned results is played by the existence, under suitable conditions, of Abelian square-free words on 4 letter of the form $aubvc$, where a,b,c are assigned letters and u,v are words whose Parikh vectors have an assigned difference. Words of this kind are obtained by applying an interesting property of consecutive factors of the infinite sequence of Keränen, namely that the difference between their Parikh vectors diverges with their lengths.

Some proofs required the aid of machine computation. The simple needed codes were written in Modula-2 and compiled in the MacMeth environment on a Power Macintosh 7500/100.

The paper is organized as follows: the basic definitions necessary for our purpose are presented in the next section. Section 3 is devoted to the construction of Abelian square-free words with a particular structure. In Section 4, we present the main results, concerning Abelian square-free substitutions and substitutions with bounded Abelian squares.

2. Preliminaries

Let A be a finite set or *alphabet*. We denote, respectively, by A^* and $\mathbb{Z}^A$ the *free monoid* and the *free* $\mathbb{Z}$*-module* generated by A. The elements of A are called *letters* and those of A^* *words*. In particular, the neutral element of A^* or *empty word* is denoted by ε.

We say that a word $u \in A^*$ is a *factor* (respectively, a *prefix*, a *suffix*) of another word $v \in A^*$ if there exist $x,y \in A^*$ such that $v = xuy$ (respectively, $v = uy$, $v = xu$). A factor u of v is *proper* if $u \neq v$.

The sets of the factors, prefixes, suffixes of a word w are denoted, respectively, by $\mathrm{Fact}(w)$, $\mathrm{Pref}(w)$, $\mathrm{Suff}(w)$. More generally, if $L \subseteq A^*$, then $\mathrm{Fact}(L) = \bigcup_{w \in L} \mathrm{Fact}(w)$, $\mathrm{Pref}(L) = \bigcup_{w \in L} \mathrm{Pref}(w)$, $\mathrm{Suff}(L) = \bigcup_{w \in L} \mathrm{Suff}(w)$.

The *mirror image* of a word $w = a_1 a_2 \dots a_n$, ($a_i \in A$, $1 \leqslant i \leqslant n$) is the word $w^M = a_n a_{n-1} \dots a_1$.

The $\mathbb{Z}$-module $\mathbb{Z}^A$ is partially ordered by the relation $\leqslant$, defined by

$$\mathbf{u} \leqslant \mathbf{u}' \quad \text{if } \mathbf{u}' - \mathbf{u} \text{ has non-negative components.}$$

The $\mathbb{Z}$-submodule generated by a subset V of $\mathbb{Z}^A$ will be denoted by $\langle V \rangle$.

The number of occurrences of a letter $a \in A$ in a word $w \in A^*$ is denoted by $|w|_a$. $|w| = \sum_{a\in A} |w|_a$ is the *length* of w. The set of the words of A^* of length n is denoted by A^n.

The application $\psi : A^* \to \mathbb{Z}^A$ defined by $\psi(w) = (|w|_a)_{a\in A}$ is a monoid morphism of A^* into the additive structure of $\mathbb{Z}^A$. $\psi(w)$ is said to be the *Parikh vector* of the word w. The kernel congruence of ψ is denoted by $\sim$ and is said to be the *commutative equivalence* on A^*.

For any $\mathbf{u} \in \mathbb{Z}^A$ we denote by $\ell(\mathbf{u})$ the sum of the components of $\mathbf{u}$ and by $\|\mathbf{u}\|$ its Euclidean norm. Remark that $\ell(\psi(w)) = |w|$, for any $w \in A^*$, and $\psi(A^*) = \{\mathbf{u} \in \mathbb{Z}^A \mid \mathbf{u} \geqslant 0\}$.

For any $w, w' \in A^*$, we set $\Delta(w, w') = \|\psi(w') - \psi(w)\|$. The function Δ is a pseudometrics on A^* and one has $\Delta(w, w') = 0$ if and only if $w \sim w'$.

An *Abelian square* is any non-empty word of the form rr' with $r \sim r'$. A word is said to be *Abelian square-free* if none of its factors is an Abelian square. We shall denote respectively by $\mathscr{S}(A)$ and $\mathscr{F}(A)$ the set of the Abelian squares and of the Abelian square-free words on the alphabet A.

Let A, B be two finite alphabets. A *substitution* $\sigma : A^* \to B^*$ is any monoid morphism of A^* into the subset monoid of B^*. Its *domain* is the set $\mathrm{dom}(\sigma) = \{w \in A^* \mid \sigma(w) \neq \emptyset\}$. The substitution σ is *finite* if $\sigma(A)$ is a finite subset of B^*.

We shall often identify a monoid morphism $h : A^* \to B^*$, with the finite substitution h' defined by $h'(w) = \{h(w)\}$, $w \in A^*$.

We remark that, for any substitution $\sigma : A^* \to B^*$, if $u, u' \in A^*$ and $u \sim u'$, then there is a one-to-one correspondence associating any word of $\sigma(u)$ with a commutatively equivalent word of $\sigma(u')$.

A substitution $\sigma : A^* \to B^*$ is said to be *commutatively injective* if

$$(u, u' \in A^*,\ v \in \sigma(u),\ v' \in \sigma(u'),\ v \sim v') \Rightarrow u \sim u',$$

i.e. if words which are not commutatively equivalent are mapped into words which are not commutatively equivalent. We remark that it is effectively decidable whether a finite substitution $\sigma : A^* \to B^*$ is commutatively injective.

Indeed, the elements of $\psi(\sigma(u))$ $(u \in A^*)$ are the sums $\sum_{a\in A} \sum_{\mathbf{e}\in\psi(\sigma(a))} x_{a,\mathbf{e}}\mathbf{e}$, with $x_{a,\mathbf{e}} \geqslant 0$ and $\sum_{\mathbf{e}\in\psi(\sigma(a))} x_{a,\mathbf{e}} = |u|_a$. One derives that σ is commutatively injective if and only if the solutions of the linear system $\sum_{a\in A} \sum_{\mathbf{e}\in\psi(\sigma(a))} x_{a,\mathbf{e}}\mathbf{e} = \sum_{a\in A} \sum_{\mathbf{e}\in\psi(\sigma(a))} y_{a,\mathbf{e}}\mathbf{e}$ (in the semiring of non-negative integers) are also solutions of the linear system $\{\sum_{\mathbf{e}\in\psi(\sigma(a))} x_{a,\mathbf{e}} = \sum_{\mathbf{e}\in\psi(\sigma(a))} y_{a,\mathbf{e}},\ a \in A\}$, or, equivalently, if and only if the homogeneous linear system

$$\sum_{\mathbf{e}\in\psi(\sigma(a))} z_{a,\mathbf{e}} = 0, \quad a \in A,$$

linearly depends on the homogeneous linear system

$$\sum_{a\in A} \sum_{\mathbf{e}\in\psi(\sigma(a))} z_{a,\mathbf{e}}\mathbf{e} = 0.$$

With any commutatively injective substitution $\sigma : A^* \to B^*$, one can associate a linear application $\tilde{\sigma} : \langle \psi(\sigma(A)) \rangle \to \mathbb{Z}^A$, making the following diagram commutative:

$$\begin{tikzcd}
A^* \arrow[r, "\sigma"] \arrow[d, "\psi"'] & \sigma(A^*) \arrow[d, "\psi"] \\
\mathbb{Z}^A & \langle \psi(\sigma(A)) \rangle \arrow[l, "\tilde{\sigma}"']
\end{tikzcd}$$

Such a linear application is defined by $\tilde{\sigma}(\sum_{a\in A} \sum_{\mathbf{e}\in\psi(\sigma(a))} z_{a,\mathbf{e}}\mathbf{e}) = (\sum_{\mathbf{e}\in\psi(\sigma(a))} z_{a,\mathbf{e}})_{a\in A}$. The definition is consistent, in view of the previous remarks. In order to simplify notation, we shall look at $\tilde{\sigma}$ as a partially defined function of $\mathbb{Z}^B$.

A substitution $\sigma : A^* \to B^*$ is said to be *commutatively functional* if $\mathrm{dom}(\sigma) = A^*$ and, for all $a \in A$, $v, v' \in \sigma(a)$, one has $v \sim v'$. Clearly, if σ is commutatively functional, then, for all $w \in A^*$, the words of $\sigma(w)$ are commutatively equivalent.

Now, we introduce the main objects of this paper. For any substitution $\sigma : A^* \to B^*$, the set $\mathscr{S}(\sigma) = \mathscr{S}(B) \cap \mathrm{Fact}(\sigma(\mathscr{F}(A)))$ is the set of the *Abelian squares* of σ. In other terms, $\mathscr{S}(\sigma)$ contains the Abelian squares occurring in the images of Abelian square-free words. A substitution $\sigma : A^* \to B^*$ is said to be *with bounded Abelian squares* if $\sigma(A) \nsubseteq \{\varepsilon\}$, and $\mathscr{S}(\sigma)$ is a finite set. If, moreover, $\mathscr{S}(\sigma)$ is empty, then σ is an *Abelian square-free* substitution.

3. Structure of Abelian square-free words

In this section, we shall exhibit some Abelian square-free words with some useful properties.

All through this section, unless differently stated, A will denote the 4-letter alphabet $\{0, 1, 2, 3\}$ and $h : A^* \to A^*$ the morphism of Keränen, defined by

$$\begin{aligned}
h(0) &= 0120232123203231301020103101213121021232021013\text{-} \\
&\quad\; 010203212320231210212320232132303132120, \\
h(1) &= \varphi(h(0)), \\
h(2) &= \varphi(h(1)), \\
h(3) &= \varphi(h(2)),
\end{aligned}$$

where '-' denotes the concatenation of words and φ is the 'circular' automorphism induced by $\varphi(0) = 1$, $\varphi(1) = 2$, $\varphi(2) = 3$, $\varphi(3) = 0$.

The morphism h is Abelian square-free [7]. Moreover, it is commutatively injective, and therefore we can consider the linear function $\tilde{h} : \langle \psi(h(A)) \rangle \to \mathbb{Z}^A$ such that $\tilde{h}\psi h = \psi$.

Our first goal is to prove that, by repeatedly applying the morphism of Keränen to an Abelian square-free word, one obtains a word where long consecutive factors are very far, from each other, with respect to the pseudo-metrics Δ (cf. Proposition 1). The proof is rather technical and requires several steps.

Remark that $\tilde{h}(\mathbf{u}) = \mathbf{u}m^{-1}$, where $m = (|h(a)|_b)_{a,b \in A}$. The matrix m is given by

$$m = \begin{pmatrix} 19 & 21 & 27 & 18 \\ 18 & 19 & 21 & 27 \\ 27 & 18 & 19 & 21 \\ 21 & 27 & 18 & 19 \end{pmatrix}.$$

One easily checks that $mm^{\mathrm{T}} = 49I + 1794U + 12V$, where

$$I = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad U = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}, \quad V = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix}$$

and T denotes matrix transposition. Consequently, for all $\mathbf{u} \in \mathrm{dom}(\tilde{h})$, one has

$$\begin{aligned} \|\mathbf{u}\| &= \|\tilde{h}(\mathbf{u})m\| \\ &= \sqrt{(\tilde{h}(\mathbf{u})m)(\tilde{h}(\mathbf{u})m)^{\mathrm{T}}} \\ &= \sqrt{49\tilde{h}(\mathbf{u})(\tilde{h}(\mathbf{u}))^{\mathrm{T}} + 1794\tilde{h}(\mathbf{u})U(\tilde{h}(\mathbf{u}))^{\mathrm{T}} + 12\tilde{h}(\mathbf{u})V(\tilde{h}(\mathbf{u}))^{\mathrm{T}}} \\ &\geqslant 7\|\tilde{h}(\mathbf{u})\| \end{aligned} \tag{1}$$

for U, V are associated with positive semi-defined quadratic forms. Moreover, since the sum of the elements of any row of m is 85, one has

$$\ell(\mathbf{u}) = \ell(\tilde{h}(\mathbf{u})m) = 85\ell(\tilde{h}(\mathbf{u})). \tag{2}$$

Now consider the set

$$\mathscr{L}_A = A^3 \times \mathbb{Z}^A = \{(a_0, a_1, a_2, \mathbf{u}) \mid a_0, a_1, a_2 \in A,\ \mathbf{u} \in \mathbb{Z}^A\}.$$

In such a set, we can introduce the relation R defined as follows: one has

$$(a_0, a_1, a_2, \mathbf{u}) R (b_0, b_1, b_2, \mathbf{u}')$$

($a_i, b_i \in A$, $i = 0, 1, 2$, $\mathbf{u}, \mathbf{u}' \in \mathbb{Z}^A$) if there exist $x_i \in A^*$ such that

$$x_i a_i \in \mathrm{Pref}(h(b_i)), \quad i = 0, 1, 2, \tag{3}$$

$$\mathbf{u}' = \tilde{h}(\mathbf{u} - \psi(x_0 x_2) + 2\psi(x_1)). \tag{4}$$

It should be clear that, for any element $\alpha \in \mathscr{L}_A$, there are finitely many $\beta \in \mathscr{L}_A$ such that $\alpha R \beta$, and they are effectively computable.

The interest in the relation R is motivated by the following technical lemma.

Lemma 1 (Carpi [3]). *If one has* $s_1, s_2, w \in A^*$, $a_0, a_1, a_2 \in A$ *and*

$$s_1 s_2 a_2 \in \mathrm{Fact}(h(w)), \quad a_0 \in \mathrm{Pref}(s_1 s_2 a_2), \quad a_1 \in \mathrm{Pref}(s_2 a_2),$$

then there exist $r_1, r_2 \in A^*, b_0, b_1, b_2 \in A$ *such that*

$$r_1 r_2 b_2 \in \mathrm{Fact}(w), \quad s_1 s_2 a_2 \in \mathrm{Fact}(h(r_1 r_2 b_2)),$$

$$b_0 \in \mathrm{Pref}(r_1 r_2 b_2), \quad b_1 \in \mathrm{Pref}(r_2 b_2),$$

$$(a_0, a_1, a_2, \psi(s_2) - \psi(s_1)) R (b_0, b_1, b_2, \psi(r_2) - \psi(r_1)).$$

If $\alpha = (a_0, a_1, a_2, \mathbf{u}) \in \mathscr{L}_A$, then we write $\|\alpha\|$ and $\ell(\alpha)$ instead of $\|\mathbf{u}\|$ and $\ell(\mathbf{u})$, respectively. Moreover, we set

$$\mathscr{L}'_A = \{(a_0, a_1, a_2, -\delta_0 \psi(a_0) + 2\delta_1 \psi(a_1) - \delta_2 \psi(a_2) \mid \delta_0, \delta_1, \delta_2 \in \{0,1\}\}.$$

Lemma 2. *Suppose* $\alpha, \beta \in \mathscr{L}_A$, $\alpha R \beta$.
- *If* $|\ell(\alpha)| \leqslant 1$ *then one has* $|\ell(\beta)| \leqslant 1$ *and either* $\|\alpha\| > \|\beta\|$ *or* $\|\alpha\| \leqslant \sqrt{2318}/6$.
- *If* $|\ell(\alpha)| \leqslant 2$ *then one has* $|\ell(\beta)| \leqslant 2$ *and either* $\|\alpha\| > \|\beta\|$ *or* $\|\alpha\| \leqslant \sqrt{202}$.

Finally, if $|\ell(\alpha)| > 2$ *then one has* $|\ell(\beta)| < |\ell(\alpha)|$.

Proof. Set $\alpha = (a_0, a_1, a_2, \mathbf{u})$, $\beta = (b_0, b_1, b_2, \mathbf{u}')$, $(a_i, b_i \in A,\ i = 0,1,2,\ \mathbf{u}, \mathbf{u}' \in \mathbb{Z}^A)$. Then, there are $x_i \in A^*$ satisfying (3) and (4). In view of (2), one has

$$|\ell(\beta)| = \frac{1}{85} |\ell(\alpha) - |x_0 x_2| + 2|x_1|| \leqslant \frac{|\ell(\alpha)| - 2}{85} + 2 \tag{5}$$

since $||x_1| - |x_0|| \leqslant 84$ and $||x_1| - |x_2|| \leqslant 84$.

First, we consider the case where $|\ell(\alpha)| \leqslant 1$. In view of (5), one has $|\ell(\beta)| \leqslant 1$ and $|x_0 x_2| - 2|x_1| \equiv \ell(\alpha) \pmod{85}$. By (1),

$$\|\alpha\| \geqslant \|\mathbf{u} - \psi(x_0 x_2) + 2\psi(x_1)\| - M \geqslant 7\|\beta\| - M,$$

where

$$M = \max_{\substack{x_i \in \mathrm{Pref}(h(A)),\ i=0,1,2 \\ |x_0 x_2| - 2|x_1| \equiv \ell(\alpha) \pmod{85}}} \| - \psi(x_0 x_2) + 2\psi(x_1)\|.$$

We conclude that either $\|\alpha\| > \|\beta\|$ or $\|\alpha\| \geqslant 7\|\beta\| - M \geqslant 7\|\alpha\| - M$, and therefore, $\|\alpha\| \leqslant M/6$. By machine computation, one gets $M = \sqrt{2318}/6$.

The case where $|\ell(\alpha)| \leqslant 2$ can be treated similarly.

Finally, if $|\ell(\alpha)| > 2$, then by (5) one gets $|\ell(\alpha)| > |\ell(\beta)|$. $\square$

Lemma 3. *For any* $\alpha \in \mathscr{L}_A$ *there exists an integer* $n \geqslant 0$ *such that whenever one has*

$$\alpha = \alpha_0 R \alpha_1 R \ldots R \alpha_n \tag{6}$$

$(\alpha_i \in \mathscr{L}_A,\ 0 \leqslant i \leqslant n)$ *there exists an index* j, $0 \leqslant j \leqslant n$ *such that* $\alpha_j \in \mathscr{L}'_A$.

Proof. First, we verify the statement in the case that $|\ell(\alpha)| \leqslant 1$. By machine computation one can verify that the statement is true when $\|\alpha\| \leqslant \sqrt{2318}/6$. Thus we suppose $\|\alpha\| > \sqrt{2318}/6$ and proceed by induction on $\|\alpha\|$. There are finitely many $\beta \in \mathscr{L}_A$ such that $\alpha R \beta$. For each one of them, one has $\|\alpha\| > \|\beta\|$ by Lemma 2, and therefore, for the sake of induction, we can suppose that there exists an integer $n_\beta \geqslant 1$ such that if one has $\beta = \alpha_1 R \alpha_2 R \ldots R \alpha_{n_\beta}$ $(\alpha_i \in \mathscr{L}_A,\ 1 \leqslant i \leqslant n)$ then $\alpha_j \in \mathscr{L}'_A$ for a suitable j, $1 \leqslant j \leqslant n_\beta$. Now, it is clear that if n is the maximal n_β and (6) is verified, then $\alpha_j \in \mathscr{L}'_A$ for a suitable j, $1 \leqslant j \leqslant n$.

In the case that $|\ell(\alpha)| = 2$, one can proceed in a similar way.

Finally, in the general case, one can achieve the proof by making induction on $|\ell(\alpha)|$. $\square$

Now, we are ready to prove the announced result concerning the distance of consecutive factors, with respect to the pseudo-metrics Δ, in words obtained by iterated application of the morphism of Keränen.

Proposition 1. *For all integer $n > 0$, there exists an integer k such that if one has $r, r' \in A^*$, $rr' \in \mathrm{Fact}(h^k(\mathscr{F}(A)))$, $|rr'| > 2 \cdot 85^k$, then $\Delta(r, r') \geqslant n$.*

For any $n > 0$, we shall denote by I_n the minimal k satisfying the previous statement.

Proof. By Lemma 3, there exists $k > 0$ such that for all $\alpha \in \mathscr{L}_A$ with $\|\alpha\| \leqslant n$, if $\alpha = \alpha_0 R \alpha_1 R \ldots R \alpha_k$, then $\alpha_j \in \mathscr{L}'_A$ for some j, $0 \leqslant j \leqslant k$.

By contradiction, suppose $r, r' \in A^*$, $rr' \in \mathrm{Fact}(h^k(\mathscr{F}(A)))$, $|rr'| > 2 \cdot 85^k$, $\Delta(r, r') < n$. Set

$$r_{1,0} = r, \quad r' = r_{2,0} a_{2,0}, \quad a_{2,0} \in A \quad \text{if } r' \neq \varepsilon,$$

$$r = r_{1,0} a_{2,0}, \quad r_{2,0} = r', \quad a_{2,0} \in A \quad \text{if } r' = \varepsilon,$$

and denote by $a_{0,0}, a_{1,0}$ the first letter of $r_{1,0} r_{2,0} a_{2,0}$ and $r_{2,0} a_{2,0}$, respectively. By iterated application of Lemma 3, we can find $r_{1,i}, r_{2,i} \in A^*$, $a_{0,i}, a_{1,i}, a_{2,i} \in A$ $(1 \leqslant i \leqslant k)$ such that

$$(a_{0,i-1}, a_{1,i-1}, a_{2,i-1}, \psi(r_{2,i-1}) - \psi(r_{1,i-1})) R (a_{0,i}, a_{1,i}, a_{2,i}, \psi(r_{2,i}) - \psi(r_{1,i})),$$

$$a_{0,i} \in \mathrm{Pref}(r_{1,i} r_{2,i} a_{2,i}), \quad a_{1,i} \in \mathrm{Pref}(r_{2,i} a_{2,i}),$$

$$r_{1,i} r_{2,i} a_{2,i} \in \mathrm{Fact}(h^{k-i}(\mathscr{F}(A))), \quad rr' \in \mathrm{Fact}(h^i(r_{1,i} r_{2,i} a_{2,i})),$$

for $1 \leqslant i \leqslant k$. Since $\Delta(r_{1,0}, r_{2,0}) \leqslant \Delta(r, r') + 1 \leqslant n$, one has $(a_{0,j}, a_{1,j}, a_{2,j}, \psi(r_{2,j}) - \psi(r_{1,j})) \in \mathscr{L}'_A$ for a suitable j, $0 \leqslant j \leqslant k$, that is, $\psi(r_{2,j}) - \psi(r_{1,j}) = -\delta_2 \psi(a_{2,j}) + 2\delta_1 \psi(a_{1,j}) - \delta_0 \psi(a_{0,j})$ for some $\delta_0, \delta_1, \delta_2 \in \{0, 1\}$ or, equivalently, $r_{1,j} a_{1,j}{}^{2\delta_1} \sim r_{2,j} a_{0,j}{}^{\delta_0} a_{2,j}{}^{\delta_2}$.

If $|r_{1,j}| < \delta_0$, then one has $r_{1,j} = \varepsilon$, $\delta_0 = 1$, $|r_{2,j}| = 2\delta_1 - \delta_0 - \delta_2 \leqslant 1$ and therefore $|r_{1,j} r_{2,j} a_{2,j}| \leqslant 2$; if $|r_{2,j}| < \delta_1$, then one has $r_{2,j} = \varepsilon$, $\delta_1 = 1$, $|r_{1,j}| = \delta_0 + \delta_2 - 2\delta_1 \leqslant 0$ and therefore, again, $|r_{1,j} r_{2,j} a_{2,j}| \leqslant 2$; finally, if $|r_{1,j}| \geqslant \delta_0$ and $|r_{2,j}| \geqslant \delta_1$ then one has

$r_{1,j} = a_{0,j}{}^{\delta_0} s$, $r_{2,j} = a_{1,j}{}^{\delta_1} s'$, for suitable $s, s' \in A^*$, $sa_{1,j}{}^{\delta_1} \sim a_{1,j}{}^{1-\delta_1} s' a_{2,j}{}^{\delta_2}$, $r_{1,j} r_{2,j} a_{2,j} = a_{0,j}{}^{\delta_0} s a_{1,j}{}^{\delta_1} a_{1,j}{}^{1-\delta_1} s' a_{2,j}{}^{\delta_2} a_{2,j}{}^{1-\delta_2} = a_{0,j}{}^{\delta_0} a_{2,j}{}^{1-\delta_2}$, by the Abelian square-freeness of this word, and therefore, again, $|r_{1,j} r_{2,j} a_{2,j}| \leqslant 2$. This yields a contradiction, since $rr' \in \mathrm{Fact}(h^j(r_{1,j} r_{2,j} a_{2,j}))$ and $|rr'| \leqslant 2 \cdot 85^j$. $\square$

The elements of the set $K = \bigcup_{k=1}^{\infty} \mathrm{Fact}(h^k(0))$ are the *factors of Keränen sequence*. By Proposition 1, one derives immediately the following corollary, which solves a question raised in [2].

Corollary 1. *One has*

$$\lim_{n \to +\infty} \min_{\substack{r, r' \in A^n \\ rr' \in K}} \Delta(r, r') = +\infty.$$

The following proposition provides useful information on the structure of the words of $h^2(\mathscr{F}(A))$.

Proposition 2 (Carpi [3]). *Given $a, b \in A$, there exist $\xi, \eta \in A^*$ such that $h^2(b) = \xi a \eta$ and, for all $u, v \in A^*$ such that $ubv \in \mathscr{F}(A)$, one has $h^2(u) \xi \eta h^2(v) \in \mathscr{F}(A)$.*

Roughly speaking, Proposition 2 ensures that if $w \in h^2(\mathscr{F}(A))$ and $a \in A$, then we can delete one occurrence of the letter a in w preserving its Abelian square-freeness. Moreover such a deletion can be operated in any prefixed segment of w belonging to $h^2(A)$.

This result is very powerful if combined with Proposition 1, which ensures that if $w \in h^k(\mathscr{F}(A))$ with a large k, then Abelian square-freeness is preserved by several deletions of letters, very far one from another, provided each deletion, alone, preserves Abelian square-freeness.

In the sequel of this section, we shall use this technique to produce various Abelian square-free words with some interesting properties.

Before, we need a technical lemma.

Lemma 4. *If $\varphi : A^* \to A^*$ is an Abelian square-free morphism, then no proper factor of the word $\varphi(012012)$ is an Abelian square.*

Proof. By contradiction, suppose

$$\varphi(012012) = \xi r r' \eta, \quad r \sim r' \neq \varepsilon, \quad \xi\eta \neq \varepsilon,$$

$\xi, r, r', \eta \in A^*$. Then one has $|\xi| < |\varphi(0)|$, $|\eta| < |\varphi(2)|$, otherwise the Abelian square rr' would occur in $\varphi(12012)$ or $\varphi(01201)$. ξ and η are, respectively, a prefix of $\varphi(0)$ and a suffix of $\varphi(2)$. Thus, $\varphi(012) = \xi v \eta$, for some $v \in A^*$, and $rr' = v\eta\xi v$. One has $|r| = |r'| = |v\eta\xi v|/2 > |v|$ and, therefore, $r = vs$, $r' = s'v$, $\eta\xi = ss'$, for some $s, s' \in A^*$. Moreover, $s \sim s'$ since these words are obtained by deleting v in the commutatively equivalent words r and r'. This yields a contradiction, since $ss' = \eta\xi$ is a factor of $\varphi(20)$. $\square$

For any alphabet A we define the subsets $\mathscr{F}_1(\mathscr{L}_A), \mathscr{F}_2(\mathscr{L}_A)$ of $\mathscr{L}_A$ as follows. One has $(a_0, a_1, a_2, \mathbf{u}) \in \mathscr{F}_1(\mathscr{L}_A)$ if there exist $u_1, u_2 \in A^*$ such that

$$\mathbf{u} = \psi(a_0 u_1) - \psi(a_1 u_2), \quad a_0 u_1 a_1 u_2 a_2 \in \mathscr{F}(A) \tag{7}$$

and $(a_0, a_1, a_2, \mathbf{u}) \in \mathscr{F}_2(\mathscr{L}_A)$ if there exist infinitely many pairs $(u_1, u_2) \in A^* \times A^*$ such that (7) is verified. Obviously, $\mathscr{F}_2(\mathscr{L}_A) \subseteq \mathscr{F}_1(\mathscr{L}_A)$.

Lemma 5. *Let $\alpha = (a_0, a_1, a_2, \mathbf{u}) \in \mathscr{L}_A - \mathscr{F}_2(\mathscr{L}_A)$. Then there exist arbitrarily large n such that one can factorize $h^n(0) = \xi_2 a_2 \xi_2'$, $h^n(2) = \xi_0 a_0 \xi_0'$, $h^n(20) = \xi_1 a_1 \xi_1'$, with $\xi_i, \xi_i' \in A^*$, $i = 0, 1, 2$, and*

$$\|\psi(\xi_1' \xi_0 a_0) - \psi(\xi_2' \xi_1 a_1) - \mathbf{u}\| \leqslant 53.$$

Proof. Let v_0, v_1 be the shortest prefixes of $h(0)$ which do not contain a_0, a_1, respectively, and v_2 the shortest suffix of $h(2)$ which do not contain a_2. Then fix $b_i, c_j \in A$ $(1 \leqslant i \leqslant p,\ 1 \leqslant j \leqslant q)$, such that

$$\psi(b_1 b_2 \ldots b_p) - \psi(c_1 c_2 \ldots c_q) = \mathbf{u} + \psi(v_0) - \psi(v_1^2 v_2 a_2).$$

By Proposition 2, for a sufficiently large n, one can factorize

$$h^n(1) = t_0 b_1 t_1 b_2 t_2 \ldots b_p t_p = s_0 c_1 s_1 c_2 s_2 \ldots c_q s_q$$

with $t_i, s_j \in A^*$, $|t_i|, |s_j| \geqslant 2 \cdot 85^{l_{p+q}}$ and

$$s_{j-1} s_j, t_{i-1} t_i, h^n(0) s_0, h^n(0) t_0, s_p h^n(2), t_q h^n(2) \in \mathscr{F}(A). \tag{8}$$

Set

$$s = s_0 s_1 \ldots s_q, \quad t = t_0 t_1 \ldots t_p,$$

$$h^n(0) = v_0 a_0 \eta_0 = v_1 a_1 \eta_1, \quad h^n(2) = \eta_2 a_2 v_2,$$

$$u_1 = \eta_0 s h^n(2) v_1, \quad u_2 = \eta_1 t \eta_2.$$

One has

$$\begin{aligned} \psi(a_0 u_1) - \psi(a_1 u_2) &= \psi(a_0 \eta_0 s h^n(2) v_1) - \psi(a_1 \eta_1 t \eta_2) \\ &= \psi(s v_1^2 a_2 v_2) - \psi(t v_0) \\ &= \mathbf{u} \end{aligned}$$

and

$$h^n(0) s h^n(20) t h^n(2) = v_0 a_0 \eta_0 s h^n(2) v_1 a_1 \eta_1 t \eta_2 a_2 v_2 = v_0 a_0 u_1 a_1 u_2 a_2 v_2.$$

One cannot have $h^n(0) s h^n(20) t h^n(2) \in \mathscr{F}(A)$, for arbitrarily large n, otherwise, by the previous equation, one would derive $\alpha = (a_0, a_1, a_2, \mathbf{u}) \in \mathscr{F}_2(\mathscr{L}_A)$. Thus, we suppose that, for arbitrarily large n, $h^n(0)sh^n(20)th^n(2)$ contains an Abelian square rr' ($r,r' \in A^*$, $r \sim r' \neq \varepsilon$). Such an rr' cannot be a factor of $h^n(0)sh^n(20)t$ or $sh^n(20)th^n(2)$, otherwise respectively $h^n(01201)$ or $h^n(12012)$ should contain a factor $\bar{r}\bar{r}'$, where $\bar{r},\bar{r}'$ are obtained by suitably inserting letters b_i, c_j in r, r' respectively. One would derive $\Delta(\bar{r},\bar{r}') \leqslant p + q + \Delta(r,r') = p+q$, since $\bar{r},\bar{r}'$ are obtained by, at most, $p+q$ insertions of letters in r,r', and therefore by Proposition 1, $|rr'| \leqslant |\bar{r}\bar{r}'| \leqslant 2 \cdot 85^{I_{p+q}}$. Thus rr' should be a factor of one of the words (8), yielding a contradiction.

Thus, we can suppose

$$h^n(0) = \zeta_2\zeta_2', \quad h^n(2) = \zeta_0\zeta_0', \quad \zeta_2' sh^n(20)t\zeta_0 = rr'$$

($\zeta_2, \zeta_2', \zeta_0, \zeta_0' \in A^*$). Moreover, $|r| = |rr'|/2 \geqslant |\zeta_2' s|$, $|r'| = |rr'|/2 \geqslant |t\zeta_0|$, and therefore

$$h^n(20) = \zeta_1\zeta_1', \quad r = \zeta_2' s\zeta_1, \quad r' = \zeta_1' t\zeta_0$$

($\zeta_1, \zeta_1' \in A^*$). One derives

$$\begin{aligned}\psi(\zeta_1'\zeta_0) - \psi(\zeta_2'\zeta_1) &= \psi(s) - \psi(t)\\ &= \mathbf{u} + \psi(v_2) - \psi(v_1^2 v_0 a_0)\end{aligned}$$

and therefore

$$\|\psi(\zeta_1'\zeta_0) - \psi(\zeta_2'\zeta_1) - \mathbf{u}\| = \|\psi(v_2) - \psi({v_1}^2 v_0 a_0)\| \leqslant 29 \tag{9}$$

since $|v_i| \leqslant 7$, $i = 0,1,2$.

Define $\xi_i, \xi_i' \in A^*$, $i = 0,1,2$, by setting $\xi_i = \zeta_i v_i'$, $\zeta_i' = v_i' a_i \xi_i'$, where v_i' is the shortest word satisfying these relations or, if such a v_i does not exist, by setting $\xi_i a_i v_i'' = \zeta_i$, $v_i''\zeta_i' = \xi_i'$, with $|v_i''|$ minimal. One easily checks that

$$h^n(0) = \xi_2 a_2 \xi_2', \quad h^n(2) = \xi_0 a_0 \xi_0', \quad h^n(20) = \xi_1 a_1 \xi_1',$$

$$\Delta(\xi_1', \zeta_1') \leqslant 8, \quad \Delta(\xi_0 a_0, \zeta_0) \leqslant 8, \quad \Delta(\xi_2', \zeta_2') \leqslant 8, \quad \Delta(\xi_1 a_1, \zeta_1) \leqslant 8.$$

One derives, in view of (9),

$$\|\psi(\xi_1'\xi_0 a_0) - \psi(\xi_2'\xi_1 a_1) - \mathbf{u}\| \leqslant \|\psi(\zeta_1'\zeta_0) - \psi(\zeta_2'\zeta_1) - \mathbf{u}\| + 24 \leqslant 53. \qquad \square$$

The next proposition shows that, under suitable conditions, it is possible to find an Abelian square-free word of the form $a_0 u_1 a_1 u_2 a_2$ with assigned a_0, a_1, a_2 and $\psi(u_1) - \psi(u_2)$.

Proposition 3. *Set*

$$\mathscr{L}_A'' = \{(a_0, a_1, a_2, \mathbf{u}) \in \mathscr{L}_A \mid \mathbf{u} < -2\psi(a_1) \text{ or } \mathbf{u} > \psi(a_0 a_2)\}$$

Then $\mathscr{L}_A - (\mathscr{L}_A'' \cup \mathscr{F}_2(\mathscr{L}_A))$ is a finite set.

Proof. Suppose $(a_0, a_1, a_2, \mathbf{u}) \in \mathscr{L}_A - (\mathscr{L}_A'' \cup \mathscr{F}_2(\mathscr{L}_A))$. By Lemma 5, there are arbitrarily large n such that

$$h^n(0) = \xi_2 a_2 \xi_2', \quad h^n(2) = \xi_0 a_0 \xi_0', \quad h^n(20) = \xi_1 a_1 \xi_1',$$

$$\|\psi(\xi_1' \xi_0 a_0) - \psi(\xi_2' \xi_1 a_1) - \mathbf{u}\| \leqslant 53$$

$(\xi_i, \xi_i' \in A^*,\ i = 0, 1, 2)$. Set

$$b_1 b_2 \dots b_p = 0^{\max\{0, k_0\}} 1^{\max\{0, k_1\}} 2^{\max\{0, k_2\}} 3^{\max\{0, k_3\}},$$

$$c_1 c_2 \dots c_q = 0^{\max\{0, -k_0\}} 1^{\max\{0, -k_1\}} 2^{\max\{0, -k_2\}} 3^{\max\{0, -k_3\}}$$

$(b_i, c_j \in A,\ 1 \leqslant i \leqslant p,\ 1 \leqslant j \leqslant q)$ where $(k_0, k_1, k_2, k_3) = \psi(\xi_1' \xi_0 a_0) - \psi(\xi_2' \xi_1 a_1) - \mathbf{u}$. Then one has

$$\psi(b_1 b_2 \dots b_p) - \psi(c_1 c_2 \dots c_q) = \psi(\xi_1' \xi_0 a_0) - \psi(\xi_2' \xi_1 a_1) - \mathbf{u},$$

$$p + q = |k_0| + |k_1| + |k_2| + |k_3| \leqslant k_0^2 + k_1^2 + k_2^2 + k_3^2 \leqslant 53^2.$$

By Proposition 2, for a sufficiently large n, one can factorize

$$h^n(1) = t_0 b_1 t_1 b_2 t_2 \dots b_p t_p = s_0 c_1 s_1 c_2 s_2 \dots c_q s_q$$

with $t_i, s_j \in A^*$ satisfying (8), $|s_i|, |t_j| \geqslant 2 \cdot 85^I$, $I = I_{53^2}$. Set

$$s = s_0 s_1 \dots s_q, \quad t = t_0 t_1 \dots t_p,$$

$$w = a_2 \xi_2' s \xi_1 a_1 \xi_1' t \xi_0 a_0.$$

For a sufficiently large n, one has $w \notin \mathscr{F}(A)$. Otherwise, indeed, for $u_1 = (\xi_1' t \xi_0)^M$, $u_2 = (\xi_2' s \xi_1)^M$, one would have

$$\psi(a_0 u_1) - \psi(a_1 u_2) = \psi(\xi_1' t \xi_0 a_0) - \psi(\xi_2' s \xi_1 a_1) = \mathbf{u},$$

$$a_0 u_1 a_1 u_2 a_2 = w^M \in \mathscr{F}(A),$$

and therefore, in view of the arbitrariness of n, $(a_0, a_1, a_2, \mathbf{u}) \in \mathscr{F}_2(\mathscr{L}_A)$.

Thus one has

$$w = \zeta_0 r r' \zeta_1, \quad r \sim r' \neq \varepsilon$$

for suitable $\zeta_0, \zeta_1, r, r' \in A^*$,

$$h^n(0) s h^n(20) t h^n(2) = \xi_2 \zeta_0 r r' \zeta_1 \xi_0',$$

and

$$h^n(012012) = \bar{\xi}_2 \bar{\zeta}_0 \bar{r} \bar{r}' \bar{\zeta}_1 \bar{\xi}_0',$$

where $\bar{\xi}_2, \bar{\zeta}_0, \bar{r}, \bar{r}', \bar{\zeta}_1, \bar{\xi}_0'$ are obtained by suitably inserting the letters b_i, c_j in ξ_2, ζ_0, r, r', ζ_1, ξ_0' respectively.

By contradiction, suppose $|\bar{\xi}_2\bar{\zeta}_0| \geqslant 85^I$. Then one has

$$h^{n-I}(012012) = 0v, \quad \bar{\xi}_2\bar{\zeta}_0 = h^I(0)\xi_3, \quad h^I(v) = \xi_3\bar{r}\bar{r}'\bar{\zeta}_1\bar{\xi}_0',$$

and $v \in \mathscr{F}(A)$), by Lemma 4. One derives $\Delta(\bar{r},\bar{r}') \leqslant p+q+\Delta(r,r') \leqslant 53^2$, since $\bar{r},\bar{r}'$ are obtained by, at most, $p+q$ insertions of letters in r,r'. By Lemma 1, one derives $|rr'| \leqslant |\bar{r}\bar{r}'| \leqslant 2\cdot 85^I$, and therefore rr' occurs in one of the words (8). This yields a contradiction.

We conclude that $|\xi_2| \leqslant |\bar{\xi}_2\bar{\zeta}_0| < 85^I$ and, by a symmetrical argument, $|\xi_0'| < 85^I$.

Set $T = 8\cdot 54$ and suppose, by contradiction, $|\xi_1| < |\xi_0| - T$. Since ξ_0, $\xi_1 a_1$ are both prefixes of $h^n(20)$, one has $\xi_0 = \xi_1 a_1 v$, for some $v \in A^*$ such that $|v| \geqslant T$. Moreover, since ξ_2', ξ_1' are both suffixes of $h^n(20)$ and $|\xi_1'| = 2\cdot 85^n - |\xi_1| - 1 \geqslant 2\cdot 85^n - |\xi_0| - 1 \geqslant 85^n > |\xi_2'|$, one has $\psi(\xi_1') > \psi(\xi_2')$.

Set $\mathbf{u}' = \psi(\xi_1'\xi_0 a_0) - \psi(\xi_2'\xi_1 a_1) - \mathbf{u}$. No component of $\mathbf{u}' + \psi(a_2)$ is larger than 54, for $\|\mathbf{u}' + \psi(a_2)\| \leqslant 54$, and no component of $\psi(v)$ is smaller than 54, because any letter of A occurs at least once in any word of $\mathrm{Fact}(h(A^*))$ of length 8 and, consequently, at least 54 times in any word of $\mathrm{Fact}(h(A^*))$ of length T. One derives

$$\begin{aligned}
\mathbf{u} - \psi(a_0a_2) &= \psi(\xi_1'\xi_0a_0) - \psi(\xi_2'\xi_1a_1) - \mathbf{u}' - \psi(a_0a_2)\\
&> \psi(v) - (\mathbf{u}' + \psi(a_2))\\
&\geqslant 0.
\end{aligned}$$

This yields a contradiction, because $(a_0,a_1,a_2,\mathbf{u}) \in \mathscr{L}_A''$. Symmetrically, a contradiction is yielded if one supposes $|\xi_1'| < |\xi_2'| - T$.

We conclude that $|\xi_1| \geqslant |\xi_0| - T$, $|\xi_1'| \geqslant |\xi_2'| - T$ and, consequently,

$$\begin{aligned}
-T \leqslant |\xi_1a_1| - |\xi_0a_0| &= 85^n - |\xi_1'| + |\xi_0'|\\
&\leqslant 85^n - |\xi_2'| + T + |\xi_0'|\\
&= |\xi_2a_2| + T + |\xi_0'|\\
&\leqslant 2\cdot 85^I + T,
\end{aligned}$$

$$\begin{aligned}
-T \leqslant |\xi_1'| - |\xi_2'| &= 85^n - |\xi_1a_1| + |\xi_2a_2|\\
&\leqslant 85^n - |\xi_0a_0| + T + |\xi_2a_2|\\
&= |\xi_2a_2| + T + |\xi_0'|\\
&\leqslant 2\cdot 85^I + T.
\end{aligned}$$

Since ξ_1a_1, ξ_0a_0 are both prefixes and ξ_2', ξ_1' are both suffixes of $h^n(20)$, one derives $\Delta(\xi_1a_1, \xi_0a_0) \leqslant 2\cdot 85^I + T$, $\Delta(\xi_2', \xi_1') \leqslant 2\cdot 85^I + T$, and consequently $\Delta(\xi_1'\xi_0a_0, \xi_2'\xi_1a_1) \leqslant 4\cdot 85^I + 2T$. One derives $\|\mathbf{u}\| \leqslant 4\cdot 85^I + 2T + 53$ and this inequality proves the finiteness of $\mathscr{L}_A - (\mathscr{L}_A'' \cup \mathscr{F}_2(\mathscr{L}_A))$. □

Proposition 4. *For all $u,u' \in A^*$ at least one of the following conditions is verified:*

(i) $\psi(u') \geqslant \psi(u)$,

(ii) $\psi(u') < \psi(u)$,
(iii) *for all* $m > 0$, *there exist* $u'', w, w' \in A^*$ *such that*

$$w \sim uu'', \quad w' \sim u'u'', \quad ww' \in \mathscr{F}(A), \quad |ww'| > m.$$

Proof. Set

$$u = b_1 b_2 \dots b_p, \quad u' = c_1 c_2 \dots c_q$$

$(b_i, c_j \in A,\ 1 \leqslant i \leqslant p,\ 1 \leqslant j \leqslant q)$. By Proposition 2, for a sufficiently large n, one can factorize

$$h^n(1) = t_0 b_1 t_1 b_2 t_2 \dots b_p t_p = s_0 c_1 s_1 c_2 s_2 \dots c_q s_q$$

with $t_i, s_j \in A^*$ verifying (8), $|t_i|, |s_j| \geqslant 2 \cdot 85^{lp+q}$. Set

$$s = s_0 s_1 \dots s_q, \quad t = t_0 t_1 \dots t_p, \quad w = h^n(0) s h^n(2), \quad w' = h^n(0) t h^n(2).$$

If n has been chosen sufficiently large, one has $|ww'| > m$, $w \sim uu''$, $w' \sim u'v''$ for suitable $u'', v'' \in A^*$ and also $uu'u'' \sim h^n(0) s h^n(2) u' \sim h^n(012) \sim h^n(0) t h^n(2) u \sim uu'v''$. One derives $u'' \sim v''$. Thus, if ww' is Abelian square-free, then (iii) is verified.

If, on the contrary, $ww' = h^n(0) s h^n(2) h^n(0) t h^n(2) \notin \mathscr{F}(A)$, then proceeding similarly to the proof of Lemma 5 one finds

$$h^n(0) = \zeta_0 \zeta_0', \quad h^n(20) = \zeta_1 \zeta_1', \quad h^n(2) = \zeta_2 \zeta_2',$$

$$\psi(\zeta_1' \zeta_2) - \psi(\zeta_0' \zeta_1) = \psi(s) - \psi(t) = \psi(u) - \psi(u').$$

Set $w_1 = (\zeta_1' h^n(1) \zeta_2)^M$, $w_1' = (\zeta_0' h^n(1) \zeta_1)^M$. If $\zeta_0 \zeta_2' \neq \varepsilon$, then $(w_1 w_1')^M = \zeta_0' h^n(1201) \zeta_2$ is a proper factor of $h^n(012012)$ and therefore, by Lemma 4, it is Abelian square-free. Moreover, if n is sufficiently large, one has $|w_1 w_1'| > m$, $w_1 \sim uu_1''$, and $w_1' \sim u'u_1''$, since $\psi(w_1') - \psi(u') = \psi(\zeta_0' h^n(1) \zeta_1) - \psi(u') = \psi(\zeta_1' h^n(1) \zeta_2) - \psi(u) = \psi(w_1) - \psi(u) = \psi(u_1'')$. In this case (iii) is verified.

Finally, if $\zeta_0 \zeta_2' = \varepsilon$, then $\psi(u) - \psi(u') = \psi(\zeta_1') - \psi(h^n(0)) + \psi(h^n(2)) - \psi(\zeta_1)$ and from $h^n(20) = \zeta_1 \zeta_1'$ we derive that (i) or (ii) is verified, according to whether $|h^n(2)| \leqslant |\zeta_1|$ or not. □

For larger alphabets, we can state the following result.

Proposition 5. *Let A be an alphabet with* $\mathrm{Card}(A) \geqslant 5$, $a_0, a_1, a_2 \in A$, *and* $\mathbf{u} \in \mathbb{Z}^A$. *Then one of the following two conditions holds true*
(i) *There exist* $\delta_0, \delta_1, \delta_2 \in \{0, 1\}$ *such that* $\mathbf{u} = \delta_2 \psi(a_2) - 2\delta_1 \psi(a_1) + \delta_0 \psi(a_0)$.
(ii) *There exist* $u_1, u_2 \in A^*$, *such that* $a_0 u_1 a_1 u_2 a_2 \in \mathscr{F}(A^*)$ *and* $\mathbf{u} = \psi(a_0 u_1) - \psi(a_1 u_2)$. *Moreover, given* $\mathbf{u}' \in \mathbb{Z}^A$, u_1, u_2 *can be chosen in such a way that* $\psi(a_0 u_1) > \mathbf{u}'$, $\psi(a_1 u_2) > \mathbf{u}'$.

This proposition was proved in [2], in the case that $\mathrm{Card}(A) \geqslant 6$ (the further condition $\psi(a_0 u_1) > \mathbf{u}', \psi(a_1 u_2) > \mathbf{u}'$ in (ii) was not explicitly stated, but it is an evident consequence of the proof).

The proof was based on a property of the Abelian square-free morphism of Pleasants [8] analogous to the property of the morphism of Keränen stated in Proposition 1. To extend Proposition 5 to the case of a 5-letter alphabet, it is sufficient to replace in the proof the morphism of Pleasants with the morphism of Keränen, which is defined on a smaller alphabet.

By Proposition 5, one derives the following corollary.

Corollary 2. *Let A be an alphabet with* $\mathrm{Card}(A) \geqslant 5$. *Then*

$$\begin{aligned}\mathscr{F}_1(\mathscr{L}_A) &= \mathscr{F}_2(\mathscr{L}_A)\\ &= \mathscr{L}_A - \left\{ (a_0, a_1, a_2, \delta_0\psi(a_0) - 2\delta_1\psi(a_1) + \delta_2\psi(a_2) \,\middle|\, \begin{matrix} a_0, a_1, a_2 \in A, \\ \delta_0, \delta_1, \delta_2 \in \{0,1\} \end{matrix} \right\}. \end{aligned} \tag{10}$$

Proof. By Proposition 5, the right member of Eq. (10) is included in $\mathscr{F}_2(\mathscr{L}_A)$, which, in turn, is included in $\mathscr{F}_1(\mathscr{L}_A)$. Thus it is sufficient to show that it contains $\mathscr{F}_1(\mathscr{L}_A)$.

If $(a_0, a_1, a_2, \mathbf{u}) \in \mathscr{F}_1(\mathscr{L}_A)$, then there are $u_1, u_2 \in A^*$ such that $\mathbf{u} = \psi(a_0u_1) - \psi(a_1u_2)$ and $a_0u_1a_1u_2a_2 \in \mathscr{F}(A)$. One derives, for all $\delta_0, \delta_1, \delta_2 \in \{0,1\}$, $\psi({a_0}^{1-\delta_0}u_1{a_1}^{\delta_1}) \neq \psi({a_1}^{1-\delta_1}u_2{a_2}^{\delta_2})$, that is, $\psi(a_0u_1) - \psi(a_1u_2) \neq \delta_0\psi(a_0) - 2\delta_1\psi(a_1) + \psi(a_2)$. Thus $(a_0, a_1, a_2, \mathbf{u})$ belongs to the right member of Eq. (10). □

4. Abelian square-free substitutions

In this section, we study conditions for Abelian square-free substitutions and substitutions with bounded Abelian squares.

The first one concerns commutative injectivity.

Proposition 6. *Let A, B be two alphabets, with* $\mathrm{Card}(A \cap \mathrm{dom}(\sigma)) \geqslant 5$. *If $\sigma : A^* \to B^*$ is a substitution with bounded Abelian squares then it is commutatively injective.*

Proof. Without loss of generality, we can assume $\mathrm{dom}(\sigma) = A^*$. By contradiction, suppose σ is not commutatively injective. Then, we can find words $u, u' \in A^*$, $v \in \sigma(u)$, $v' \in \sigma(u')$ such that $v \sim v'$ but $u \not\sim u'$. Set $\mathbf{u} = \psi(u) - \psi(u')$. By Proposition 5, for any $a \in A$ such that $\mathbf{u} \notin \langle \psi(a) \rangle$ there are $u_1, u_2 \in A^*$ such that $\mathbf{u} = \psi(u_1) - \psi(u_2)$, $au_1au_2a \in \mathscr{F}(A)$. Moreover, one can assume $\psi(au_1) > \psi(u)$. One derives $au_1 \sim uu''$, for some $u'' \in A^*$ and $\psi(au_2) = \psi(au_1) - \mathbf{u} = \psi(u'u'')$.

Let $v'' \in \sigma(u'')$. Then $vv'' \in \sigma(uu'')$, $v'v'' \in \sigma(u'u'')$, and therefore there are $t \in \sigma(au_1)$, $t' \in \sigma(au_2)$ such that $t \sim vv'' \sim v'v'' \sim t'$. From $tt' \in \sigma(au_1au_2)$, one derives $tt' \in \mathscr{S}(\sigma)$.

Remark that $|vv'|$, and consequently $|tt'|$ can be assumed to be arbitrarily large, since one can always replace u, u', v, v' by $uw, u'w, v\bar{w}, v'\bar{w}$ where $w \in A^*$ and $\bar{w} \in \sigma(w)$. □

To extend the previous result to substitutions defined on a 4-letter alphabets, one needs to add the hypothesis of commutative functionality.

Proposition 7. *Let A,B be two alphabets, with* $\mathrm{Card}(A)=4$ *and $\sigma: A^* \to B^*$ a commutatively functional substitution with bounded Abelian squares. Then σ is commutatively injective.*

Proof. First, let us verify that $\varepsilon \notin \sigma(A)$. Suppose the contrary. Without loss of generality, we can assume $A=\{0,1,2,3\}$, $\sigma(0)=\varepsilon$. Let h be the morphism of Keränen, considered in Section 3, and set $h^n(012)=0w_n$, $n \geqslant 1$, $w \in A^*$. Then $w_n 0 w_n \in \mathscr{F}(A)$, by Lemma 4, and therefore $\sigma(w_n 0 w_n)=\sigma(w_n w_n) \subseteq \mathscr{S}(\sigma)$ for all n.

Now, by contradiction, suppose there are $u,u' \in A^*$, $v \in \sigma(u)$, $v' \in \sigma(u')$ such that $v \sim v'$ but $u \not\sim u'$.

By Proposition 4, if neither $\psi(u') \geqslant \psi(u)$ nor $\psi(u') < \psi(u)$, then for any $n>1$, there are $w,w',u'' \in A^*$ such that $w \sim uu''$, $w' \sim u'u''$, $|u''|>n$ and $ww' \in \mathscr{F}(A)$, and one can proceed as in the case of an alphabet with 5 or more letters.

If, on the contrary, $\psi(u') \geqslant \psi(u)$, then one has $u' \sim uu''$, for some $u'' \in A$, and $v' \sim vv''$, for all $v'' \in \sigma(u'')$. One derives $v \sim vv''$, that is, $v''=\varepsilon$ and therefore $u''=\varepsilon$, $u \sim u'$, which contradicts our assumption.

Symmetrically, one has $u \sim u'$ when $\psi(u') < \psi(u)$. □

Our next goal is to find an effective characterization of commutatively functional Abelian square-free substitutions.

Let $\sigma: A^* \to B^*$ be a commutatively injective substitution. Then, we can consider the sets

$$\mathscr{S}'(\sigma)=\left\{ y_0 v_1 x_1 y_1 v_2 x_2 \;\middle|\; \begin{array}{l} a_i \in A,\ x_i, y_i \in B^*,\ x_i y_i \in \sigma(a_i),\ i=0,1,2, \\ u_j \in A^*,\ v_j \in \sigma(u_j),\ j=1,2, \\ a_0 u_1 a_1 u_2 a_2 \in \mathscr{F}(A),\ y_0 v_1 x_1 \sim y_1 v_2 x_2 \end{array} \right\},$$

$$\mathscr{L}_\sigma=\left\{ (a_0,a_1,a_2,\tilde{\sigma}(\psi(x_0 x_2)-2\psi(x_1)) \;\middle|\; \begin{array}{l} a_i \in A, \\ x_i \in \mathrm{Pref}(\sigma(a_i)), \\ i=0,1,2 \end{array} \right\}. \tag{11}$$

If σ is a finite substitution, then $\mathscr{S}(\sigma)-\mathscr{S}'(\sigma)$ is a finite, effectively computable set. Indeed, if $rr' \in \mathscr{S}(\sigma)$, $r \sim r' \neq \varepsilon$, then there are $a_i \in A$, $v_i \in \sigma(a_i)$, $1 \leqslant i \leqslant n$, $n \geqslant 1$ such that $a_1 a_2 \ldots a_n \in \mathscr{F}(A)$ and $rr' \in \mathrm{Fact}(v_1 v_2 \ldots v_n)$. If $r,r' \notin \mathrm{Fact}(\sigma(A))$, then one has $v_{i_0}=x_0 y_0$, $v_{i_1}=x_1 y_1$, $v_{i_2}=x_2 y_2$, $r=y_0 v_{i_0+1} \ldots v_{i_1-1} x_1$, $r'=y_1 v_{i_1+1} \ldots v_{i_2-1} x_2$, for suitable $x_k, y_k \in B^*$, $k=0,1,2$, $1 \leqslant i_0 \leqslant i_1 \leqslant i_2 \leqslant n$, and therefore $rr' \in \mathscr{S}'(\sigma)$. In other terms, the elements of $\mathscr{S}(\sigma)-\mathscr{S}'(\sigma)$ have the form rr' with $r \sim r'$ and either $r \in \mathrm{Fact}(\sigma(A))$ or $r' \in \mathrm{Fact}(\sigma(A))$.

Then to decide whether a finite substitution is Abelian square-free or has bounded Abelian squares, one is reduced to study $\mathscr{S}'(\sigma)$.

The next Proposition provides a tool to check the finiteness or emptiness of $\mathscr{S}'(\sigma)$.

Proposition 8. *Let A,B be two alphabets and $\sigma: A^* \to B^*$ a commutatively injective and commutatively functional substitution. The set $\mathscr{S}'(\sigma)$ is finite (respectively, empty) if and only if $\mathscr{F}_2(\mathscr{L}_A) \cap \mathscr{L}_\sigma=\emptyset$ (respectively, $\mathscr{F}_1(\mathscr{L}_A) \cap \mathscr{L}_\sigma=\emptyset$).*

Proof. We shall prove that $\mathscr{F}_2(\mathscr{L}_A)\cap\mathscr{L}_\sigma\neq\emptyset$ if and only if $\mathscr{S}'(\sigma)$ is infinite. In a similar way, one can prove that $\mathscr{F}_1(\mathscr{L}_A)\cap\mathscr{L}_\sigma\neq\emptyset$ if and only if $\mathscr{S}'(\sigma)$ is non-empty.

Suppose $\mathscr{F}_2(\mathscr{L}_A)\cap\mathscr{L}_\sigma\neq\emptyset$. Then, there are $a_i\in A$, $x_i\in\mathrm{Pref}(\sigma(a_i))$, $i=0,1,2$, and infinitely many pairs $(u_1,u_2)\in A^*\times A^*$ such that

$$a_0u_1a_1u_2a_2\in\mathscr{F}(A), \tag{12}$$

$$\psi(a_0u_1)-\psi(a_1u_2)=\tilde{\sigma}(\psi(x_0x_2)-2\psi(x_1)).$$

Fix words $v_j\in\sigma(u_j)$, $j=1,2$ and $y_i\in B^*$ such that

$$x_iy_i\in\sigma(a_i),\quad i=0,1,2. \tag{13}$$

Then one has

$$x_0y_0v_1x_1y_1v_2x_2y_2\in\sigma(a_0u_1a_1u_2a_2)\subseteq\sigma(\mathscr{F}(A)), \tag{14}$$

$$\tilde{\sigma}(\psi(x_0x_2)-2\psi(x_1))=\tilde{\sigma}(\psi(x_0y_0v_1)-\psi(x_1y_1v_2)).$$

One derives $\psi(x_0x_2)-2\psi(x_1)=\psi(x_0y_0v_1)-\psi(x_1y_1v_2)$ or, equivalently,

$$y_0v_1x_1\sim y_1v_2x_2, \tag{15}$$

and therefore, $y_0v_1x_1y_1v_2x_2\in\mathscr{S}'(\sigma)$. Letting (u_1,u_2) vary between all the possible pairs, one obtains infinitely many elements of $\mathscr{S}'(\sigma)$.

Conversely, suppose $\mathscr{S}'(\sigma)$ is infinite. Then we can find infinitely many tuples

$$(a_0,a_1,a_2,x_0,x_1,x_2,y_0,y_1,y_2,u_1,u_2,v_1,v_2)\in A^3\times(\mathrm{Pref}(\sigma(A)))^3\times(\mathrm{Suff}(\sigma(A)))^3\times(A^*)^2\times(B^*)^2,$$

with $v_j\in\sigma(u_j)$, $j=1,2$ satisfying Eqs. (12), (13) and (15). At least for one of the finitely many tuples $(a_0,a_1,a_2,x_0,x_1,x_2,y_0,y_1,y_2)$, there are infinitely many 4-tuples (u_1,u_2,v_1,v_2) satisfying such relations. Such tuples satisfy the equation

$$\begin{aligned}\tilde{\sigma}(\psi(x_0x_2)-2\psi(x_1))&=\tilde{\sigma}(\psi(x_0x_2y_0v_1x_1)-\psi(x_1x_1y_1v_2x_2))\\&=\tilde{\sigma}(\psi(x_0y_0v_1)-\psi(x_1y_1v_2))\\&=\psi(a_0u_1)-\psi(a_1u_2).\end{aligned}$$

We conclude that $(a_0,a_1,a_2,\psi(a_0u_1)-\psi(a_1u_2))\in\mathscr{F}_2(\mathscr{L}_A)\cap\mathscr{L}_\sigma$. $\square$

Remark. In proving that $\mathscr{S}'(\sigma)\neq\emptyset\Rightarrow\mathscr{F}_1(\mathscr{L}_A)\cap\mathscr{L}_\sigma\neq\emptyset$ and $\mathrm{Card}(\mathscr{S}'(\sigma))=\infty\Rightarrow\mathscr{F}_2(\mathscr{L}_A)\cap\mathscr{L}_\sigma\neq\emptyset$ one does not use the hypothesis of commutative functionality.

Now, we can characterize commutatively functional Abelian square-free substitutions.

Corollary 3. *A commutatively functional substitution $\sigma:A^*\to B^*$ is Abelian square-free if and only if it satisfies the following three conditions*:

(i) *σ is commutatively injective,*

(ii) $\mathscr{F}_1(\mathscr{L}_A) \cap \mathscr{L}_\sigma = \emptyset$,
(iii) $\mathscr{S}(\sigma) - \mathscr{S}'(\sigma) = \emptyset$.

Proof. If σ is Abelian square-free, then (i) is verified, by Proposition 6. Once (i) is verified, in view of Proposition 8, one has $\mathscr{S}(\sigma) = \emptyset$ if and only if (ii) and (iii) are satisfied. □

In a similar way, one proves the following result, concerning substitutions with bounded Abelian squares.

Corollary 4. *A commutatively functional substitution* $\sigma : A^* \to B^*$ *such that* $\mathrm{Card}(A) \geqslant 4$ *has bounded Abelian squares if and only if it satisfies the following two conditions*:
(i) σ *is commutatively injective*,
(ii) $\mathscr{F}_2(\mathscr{L}_A) \cap \mathscr{L}_\sigma = \emptyset$.

By the Remark to the proof of Proposition 8, one obtains the following corollary.

Corollary 5. *A substitution* $\sigma : A^* \to B^*$ *satisfying conditions* (i)–(iii) *of Corollary* 3 *is Abelian square-free. A finite substitution* $\sigma : A^* \to B^*$ *satisfying conditions* (i) *and* (ii) *of Corollary* 4 *has bounded Abelian squares.*

Corollaries 3 and 4 reduce the problem of deciding whether a commutatively functional substitution $\sigma : A^* \to B^*$ is Abelian square-free or with bounded Abelian squares, to the emptiness problem of the sets $\mathscr{F}_i(\mathscr{L}_A) \cap \mathscr{L}_\sigma$, $i = 1, 2$. For an alphabet with at least 5 letters, $\mathscr{F}_1(\mathscr{L}_A) = \mathscr{F}_2(\mathscr{L}_A)$ is described by Corollary 2.

Proposition 9. *There exists an algorithm to decide whether a commutatively functional finite substitution* σ *is Abelian square-free* (*respectively, with bounded Abelian squares*).

Proof. We have seen that one is reduced to decide the emptiness of the sets $\mathscr{F}_i(\mathscr{L}_A) \cap \mathscr{L}_\sigma$, $i = 1, 2$.

If $\mathrm{Card}(A) \geqslant 5$, this problem is decidable, since $\mathscr{L}_\sigma$ is finite and $\mathscr{F}_1(\mathscr{L}_A) = \mathscr{F}_2(\mathscr{L}_A)$ is co-finite by Corollary 2.

Now, let us consider the case where A is a 4-letter alphabet. We claim that $\mathscr{F}_1(\mathscr{L}_A) \cap \mathscr{L}_\sigma \subseteq \mathscr{L}''_A$, where $\mathscr{L}''_A$ is the set defined in Proposition 3.

Indeed, proceeding as in the proof of Proposition 8, one can show that if $(a_0, a_1, a_2, \mathbf{u}) \in \mathscr{F}_1(\mathscr{L}_A) \cap \mathscr{L}_\sigma$ then one has

$$\mathbf{u} = \psi(a_0 u_1) - \psi(a_1 u_2), \quad a_0 u_1 a_1 u_2 a_2 \in \mathscr{F}(A), \quad y_0 v_1 x_1 \sim y_1 v_2 x_2$$

for suitable $u_i \in A^*$, $v_i \in \sigma(u_i)$, $x_i \in \mathrm{Pref}(\sigma(a_i))$, $y_j \in \mathrm{Suff}(\sigma(a_j))$, $i = 1, 2$, $j = 0, 1$.

If, by contradiction, one has $\mathbf{u} < -2\psi(a_1)$, then one derives $\psi(a_0u_1a_1) < \psi(u_2)$, and therefore, for any $w_i \in \sigma(a_i)$, $(i=0,1)$, $\psi(y_1v_2x_2) = \psi(y_0v_1x_1) \leqslant \psi(w_0v_1w_1) < \psi(v_2) \leqslant \psi(y_1v_2x_2)$, which gives a contradiction.

Symmetrically, a contradiction is yielded if one supposes $\mathbf{u} > \psi(a_0a_2)$, and therefore $(a_0, a_1, a_2, \mathbf{u}) \in \mathscr{L}''_A$.

Thus,

$$\mathscr{F}_1(\mathscr{L}_A) \cap \mathscr{L}_\sigma = \mathscr{L}_\sigma \cap (\mathscr{F}_1(\mathscr{L}_A) \cap \mathscr{L}''_A),$$
$$\mathscr{F}_2(\mathscr{L}_A) \cap \mathscr{L}_\sigma = \mathscr{L}_\sigma \cap (\mathscr{F}_2(\mathscr{L}_A) \cap \mathscr{L}''_A),$$

and $\mathscr{F}_1(\mathscr{L}_A) \cap \mathscr{L}''_A, \mathscr{F}_2(\mathscr{L}_A) \cap \mathscr{L}''_A$ are finite sets. □

Remark. Unfortunately we do not know what are the elements of the finite sets $\mathscr{F}_1(\mathscr{L}_A) \cap \mathscr{L}''_A, \mathscr{F}_2(\mathscr{L}_A) \cap \mathscr{L}''_A$ considered in the previous proof. Thus, even if there exists an algorithm deciding whether a commutatively functional morphism, defined on 4 letters, is Abelian square-free or with bounded Abelian squares, it remains unknown. This problem does not arise when one deals with alphabets with 5 or more letters, in view of Proposition 2.

The last result concerns commutatively functional substitutions with bounded Abelian squares, defined on 5 or more letters.

Proposition 10. *Let A, B be alphabets, with* $\mathrm{Card}(A) \geqslant 5$ *and $\sigma : A^* \to B^*$ a commutatively functional substitution with bounded Abelian squares. Then $\mathscr{S}(\sigma) \subseteq \sigma(A^2)$.*

This result ensures that, under the considered hypotheses, to check the Abelian square-freeness of a word of $\sigma(\mathscr{F}(A))$, it is sufficient to verify that its factors lying in $\sigma(A^2)$ are Abelian square-free.

Proof. By contradiction, let $u \in \mathscr{S}(\sigma) - \mathrm{Fact}(\sigma(A^2))$. There are $r, r' \in B^*$, $a_i \in A$, $w_i \in \sigma(a_i)$, $1 \leqslant i \leqslant n$, $n > 2$ such that

$$u = rr', \quad r \sim r',$$
$$a_1a_2 \ldots a_n \in \mathscr{F}(A), \tag{16}$$
$$w_1w_2 \ldots w_n = prr'q.$$

One can factorize

$$p = w_1w_2 \ldots w_{i_0-1}x_0,$$
$$pr = w_1w_2 \ldots w_{i_1-1}x_1,$$
$$prr' = w_1w_2 \ldots w_{i_2-1}x_2,$$

with $1 \leqslant i_0 \leqslant i_1 \leqslant i_2 \leqslant n$, $x_j \in \mathrm{Pref}(w_{i_j})$, $j=0,1,2$. One derives $x_0 u = x_0 rr' = w_{i_0} w_{i_0+1} \ldots w_{i_2-1} x_2$ and consequently $i_2 - i_0 > 1$, for $rr' \notin \mathrm{Fact}(\sigma(A^2 \cap \mathscr{F}(A)))$. One easily finds

$$\psi(x_0 x_2) - 2\psi(x_1) = \psi(w_{i_0} w_{i_0+1} \ldots w_{i_1-1}) - \psi(w_{i_1} w_{i_1+1} \ldots w_{i_2-1})$$

and consequently,

$$\tilde{\sigma}(\psi(x_0 x_2) - 2\psi(x_1)) = \psi(a_{i_0} a_{i_0+1} \ldots a_{i_1-1}) - \psi(a_{i_1} a_{i_1+1} \ldots a_{i_2-1}).$$

that is,

$$(a_{i_0}, a_{i_1}, a_{i_2}, \psi(a_{i_0} a_{i_0+1} \ldots a_{i_1-1}) - \psi(a_{i_1} a_{i_1+1} \ldots a_{i_2-1})) \in \mathscr{L}_\sigma.$$

By Proposition 4, $\mathscr{F}_2(\mathscr{L}_A) \cap \mathscr{L}_\sigma = \emptyset$ and therefore, in view of Corollary 2,

$$\psi(a_{i_0} a_{i_0+1} \ldots a_{i_1-1}) - \psi(a_{i_1} a_{i_1+1} \ldots a_{i_2-1}) = \delta_0 \psi(a_0) - 2\delta_1 \psi(a_1) + \delta_2 \psi(a_2)$$

for suitable $\delta_0, \delta_1, \delta_2 \in \{0,1\}$ or, equivalently,

$$a_{i_0} a_{i_0+1} \ldots a_{i_1-1} a_{i_1}{}^{2\delta_1} \sim a_{i_1} a_{i_1+1} \ldots a_{i_2-1} a_{i_0}^{\delta_0} a_{i_2}^{\delta_2}.$$

By deleting $a_{i_0}{}^{\delta_0} a_{i_1}{}^{\delta_1}$ in both members, we get

$$a_{i_0}^{1-\delta_0} a_{i_0+1} \ldots a_{i_1-1} a_{i_1}^{\delta_1} \sim a_{i_1}^{1-\delta_1} a_{i_1+1} \ldots a_{i_2-1} a_{i_2}^{\delta_2}$$

or

$$a_{i_0}^{\delta_1-\delta_0} \sim a_{i_0}^{1-\delta_1} a_{i_0+1} \ldots a_{i_2-1} a_{i_2}^{\delta_2}$$

or

$$a_{i_0}^{1-\delta_0} a_{i_0+1} \ldots a_{i_2-1} a_{i_2}^{\delta_1} \sim a_{i_2}^{\delta_2-\delta_1},$$

according to whether $i_0 < i_1 < i_2$ or $i_0 = i_1$ or $i_1 = i_2$. In all cases, (16) is contradicted. $\square$

In order to illustrate the results presented in this Section, we consider the following Example.

Example. Let $h: A^* \to A^*$ be the morphism of Keränen, defined in Section 3. For $1 \leqslant n \leqslant 85^2$ and $a \in A$, we denote by $w_{a,n}$ the word obtained by deleting the n-th letter in $h^2(a)$. Consider the substitution $\sigma: A^* \to A^*$ defined by

$$\sigma(0) = \{w_{0,3034}\},$$
$$\sigma(1) = \{w_{1,2184}, w_{1,6774}\},$$
$$\sigma(2) = \{w_{2,3459}, w_{2,7199}\},$$
$$\sigma(3) = \{w_{3,6349}\}.$$

One has

$$L_\sigma \subseteq \{(a_0, a_1, a_2, \delta_0\psi(a_0) - 2\delta_1\psi(a_1) + \delta_2\psi(a_2)) | a_i \in A,\ \delta_i \in \{0,1\},\ i=0,1,2\}$$

Consequently, $\mathscr{F}_1(\mathscr{L}_A) \cap \mathscr{L}_\sigma = \emptyset$. Indeed, if one has

$$\psi(a_0 u_1) - \psi(a_1 u_2) = \delta_0 \psi(a_0) - 2\delta_1 \psi(a_1) + \delta_2 \psi(a_2)$$

$a_i \in A$, $\delta_i \in \{0,1\}$, $u_j \in A^*$, $i = 0,1,2$, $j = 1,2$, then ${a_0}^{1-\delta_0} u_1 {a_1}^{\delta_1} \sim a_1^{1-\delta_1} u_2 a_2^{\delta_2}$, and therefore $a_0 u_1 a_1 u_2 a_2 \notin \mathscr{F}(A)$. By Proposition 8, $\mathscr{S}'(\sigma) = \emptyset$ and therefore σ has bounded Abelian squares, which can be explicitly determined. It turns out that

$$w_{1,6774} w_{2,3459}, \; w_{1,6774} w_{2,7199}, \; w_{2,7199} w_{3,6349}, \; w_{2,7199} w_{0,3034} \tag{17}$$

contain Abelian squares, while any word of $\sigma(\mathscr{F}(A))$ avoiding these four words is Abelian square-free. One can also consider the substitution $\varphi : A^* \to A^*$ defined by

$$\varphi(a) = \sigma(h(a)) - (T_1 \cup T_2),$$

where T_1 is the set of the words containing one of (17) as a factor and T_2 is the set of the words with $w_{1,6774}$ or $w_{2,7199}$ as a suffix. Clearly, $\varphi(\mathscr{F}(A)) \subseteq \sigma(h(\mathscr{F}(A))) \subseteq \sigma(\mathscr{F}(A))$ and $\mathrm{Fact}(\varphi(\mathscr{F}(A)))$ do not contain any of (17). Thus φ is an Abelian square-free substitution.

References

[1] J. Berstel, Axel Thue's work on repetitions in words, in: P. Leroux, C. Reutenauer (Eds.), Séries formelles et combinatoire algébrique, Publications du LaCIM, Université du Québec a Montréal, 1990, pp. 65–80.
[2] A. Carpi, On abelian power-free morphisms, Int. J. Algebra Comput. 3 (1993) 151–167.
[3] A. Carpi, On the number of Abelian square-free words on four letters, Discrete Appl. Math. 81 (1998) 155–167.
[4] F.M. Dekking, Strongly nonrepetitive sequences and progression-free sets, J. Combin. Theory, Ser. A 27 (1979) 181–185.
[5] P. Erdős, Some unsolved problems, Magyar Tud. Akad. Mat. Kutató Int. Közl. 6 (1961) 221–264.
[6] A.A. Evdokimov, Strongly asymmetric sequences generated by a finite number of symbols, Sov. Math. Dokl. 9 (1968) 536–539.
[7] V. Keränen, Abelian squares are avoidable on 4 letters, in: W. Kuich (Ed.), Proc. ICALP '92, Lecture Notes in Computer Science, vol. 623, Springer, Berlin, 1992, pp. 41–52.
[8] P.A.B. Pleasants, Nonrepetitive sequences, Proc. Cambridge Phil. Soc. 68 (1970) 267–274.