Return words and palindromes in specular sets*

Francesco Dolce
Université Paris-Est, France

Abstract

In this contribution we introduce specular sets, a subclass of tree sets having particular symmetries. We give some cardinality results about different types of return words in these sets. We also give some results concerning palindromes, namely we prove that tree sets of characteristic one closed under reversal are rich and that an important class of specular sets verifies G-richness. This is a joint work with V. Berthé, C. De Felice, V. Delecroix, J. Leroy, D. Perrin, C. Reutenauer and G. Rindone.

1 Introduction

In [2] the authors started a series of papers ([5, 6, 7, 8, 9]) studing the links between uniformly recurrent languages, subgroups of free groups and bifix codes. In this paper, we continue this investigation in a situation which involves groups, named specular, which are free products of a free group and of a finite number of cyclic groups of order two (see also [4]). A specular set is a subset of such a group. It is a set of words stable by taking the inverse and defined in terms of restrictions on the extensions of its elements.

The paper is organized as follows. In Section 2, we recall some notions concerning words, extension graphs and bifix codes. We also consider tree sets of arbitrary characteristic (see [9]). In Section 3, we introduce specular groups and specular sets. In Section 4 we give a construction which allows to build specular sets from tree sets of characteristic 1 (Theorem 6). In Section 5, we introduce several notions of return words: complete, right, left and mixed return words. For each of them, we prove a cardinality theorem (Theorems 9, 11 and 13). Finally, in Section 6, we make a connection with the notion of G-rich words introduced in [14] and related to the palindromic complexity of [10].

This is a joint work with Valérie Berthé, Clelia De Felice, Vincent Delecroix, Julien Leroy, Dominique Perrin, Christophe Reutenauer and Giuseppina Rindone.

2 Preliminaries

Let A be a finite alphabet. We denote by A^* the free monoid on A. We denote by ε the empty word. The *reversal* of a word $w = a_1a_2 \cdots a_n$ with $a_i \in A$ is the word $\tilde{w} = a_n \cdots a_2a_1$. A word w is said to be a *palindrome* if $w = \tilde{w}$. A set S of words is *closed under reversal* if $w \in S$ implies $\tilde{w} \in S$ for every $w \in S$.

*This work was supported by grants from Région Île-de-France.

A *factor* of a word x is a word v such that $x = uvw$ with u, w nonempty. If both u and w are nonempty, v is called an *internal factor*. A set of words on the alphabet A is said to be *factorial* if it contains the alphabet A and all the factors of its elements.

Given a set S and a word $w \in S$, we define $L(w) = \{a \in A \mid aw \in S\}, R(w) = \{a \in A \mid wa \in S\}, E(w) = \{(a,b) \in A \times A \mid awb \in S\}$. We denote also $\ell(w) = \mathrm{Card}(L(w)), r(w) = \mathrm{Card}(R(w))$ and $e(w) = \mathrm{Card}(E(w))$. A word w is *right-extendable* if $r(w) > 0$, *left-extendable* if $\ell(w) > 0$ and *biextendable* if $e(w) > 0$. A factorial set S is called *right-extendable* (resp. *left-extendable*, resp. *biextendable*) if every word in S is right-extendable (resp. left-extendable, resp. biextendable). A word w is called *right-special* if $r(w) \geq 2$. It is called *left-special* if $\ell(w) \geq 2$. It is called *bispecial* if it is both left-special and right-special. The *factor complexity* of a factorial set S of words on an alphabet A is the sequence $p_n = \mathrm{Card}(S \cap A^n)$.

Let S be a biextendable set of words. For $w \in S$, we consider the set $E(w)$ as an undirected graph on the set of vertices which is the disjoint union of $L(w)$ and $R(w)$ with edges the pairs $(a,b) \in E(w)$. This graph is called the *extension graph* of w. If the extension graph $E(w)$ is acyclic, then $\ell(w) - r(w) + e(w)$ is the number of connected components of the graph $E(w)$. A biextendable set S is called a *tree set* of *characteristic* $\chi(S)$ if for any nonempty $w \in S$, the graph $E(w)$ is a tree (acyclic and connected) and if $E(\varepsilon)$ is a union of $\chi(S)$ trees.

A set of words $S \neq \{\varepsilon\}$ is *recurrent* if it is factorial and if for any $u, w \in S$, there is a $v \in S$ such that $uvw \in S$. An infinite factorial set is said to be *uniformly recurrent* if for any word $u \in S$ there is an integer $n \geq 1$ such that u is a factor of any word of S of length n. A uniformly recurrent set is recurrent. The converse is true for true for tree sets (see [9]).

An infinite word is *strict episturmian* if the set of its factors is closed under reversal and if it contains for each n exactly one right-special word u such that $r(u) = \mathrm{Card}(A)$ (see [2]). An *Arnoux-Rauzy set* is the set of factors of a strict episturmian word. Any Arnoux-Rauzy set is a recurrent tree set of characteristic 1 (see [5]).

Example 1. The *Fibonacci set* is the set of factors of the fixed-point $f^\omega(a)$ of the morphism $f : \{a,b\}^* \to \{a,b\}^*$ defined by $f(a) = ab$ and $f(b) = a$. It is an Arnoux-Rauzy set (see [13]), and thus a tree set of characteristic 1.

A *bifix code* is a set of nonempty words wich does not contain any proper prefix of proper suffix of its elements (see [3]). The *kernel* of a bifix code X is the set of words of X which are internal factors of X.

3 Specular sets

We consider an alphabet A with an involution $\theta : A \to A$, possibly with some fixed points. We also consider the group G_θ generated by A with the relations $a\theta(a) = 1$ for every $a \in A$. Thus $\theta(a) = a^{-1}$ for $a \in A$. When θ has no fixed point, we can set $A = B \cup B^{-1}$ by choosing a set of representatives of the orbits of θ for the set B. The group G_θ is then the free group on B. In general, we have $G_\theta = \mathbb{Z}^{*i} * (\mathbb{Z}/2\mathbb{Z})^{*j}$ where i is the number of orbits of θ with two elements and j the number of its fixed points. Such a group will be called a *specular group* of type (i,j). These groups are very close to free groups (see [4]).

Example 2. Let $A = \{a, b, c, d\}$ and let θ be the involution which exchanges b, d and fixes a, c. Then $G_\theta = \mathbb{Z} * (\mathbb{Z}/2\mathbb{Z})^2$ is a specular group of type $(1, 2)$.

A word on the alphabet A is *θ-reduced* if it has no factor of the form $a\theta(a)$ for $a \in A$. It is clear that any element of a specular group is represented by a unique θ-reduced word. A *specular set* on A is a biextendable set of θ-reduced words on A, closed under inverses, which is a tree set of characteristic 2. Thus, in a specular set, the extension graph of every nonempty word is a tree and the extension graph of the empty word is a union of two disjoint trees. Note that in a specular set the two trees forming $E(\varepsilon)$ are isomorphic ([4, Proposition 4.1]).

Example 3. Let $A = \{a, b\}$ and let θ be the identity on A. Then the set of factors of $(ab)^\omega$ is a specular set.

The following is a particular case of [9, Proposition 2.4].

Proposition 4. *The factor complexity of a specular set on the alphabet A is given by $p_0 = 1$ and $p_n = n(\mathrm{Card}(A) - 2) + 2$ for $n \geq 1$.*

Since a specular set is biextendable, any letter $a \in A$ occurs exactly twice as a vertex of $E(\varepsilon)$, one as an element of $L(\varepsilon)$ and one as an element of $R(\varepsilon)$. A letter $a \in A$ is said to be *even* if its two occurrences appear in the same tree. Otherwise, it is said to be *odd*. Observe that if a specular set is recurrent, there is at least one odd letter. A word $w \in S$ is said to be *even* if it has an even number of odd letters. Otherwise it is said to be *odd*.

Example 5. Let S be the set of Example 3. Both letters a and b are odd. Thus even words are exactly the words of even length.

4 Doubling maps

We now introduce a construction which allows one to build specular sets. This is a particular case of the multiplying maps introduced in [9].

A *doubling transducer* is a transducer $\mathcal{A}$ with set of states $Q = \{0, 1\}$ on the input alphabet Σ and the output alphabet A and such that:

1. the input automaton is a *group automaton*, that is, every letter of Σ acts on Q as a permutation

2. the output labels of the edges are all distinct.

We define two maps $\delta_0, \delta_1 : \Sigma^* \to A^*$ corresponding to initial states 0 and 1 respectively. Thus $\delta_0(u) = v$ (resp. $\delta_1(u) = v$) if the path starting at state 0 (resp. 1) with input label u has output v. The pair $\delta_{\mathcal{A}} = (\delta_0, \delta_1)$ is called a *doubling map*. The *image* of a set T on the alphabet Σ by the doubling map $\delta_{\mathcal{A}}$ is the set $S = \delta_0(T) \cup \delta_1(T)$. We denote by $i \stackrel{\alpha|a}{\to} j$ the edge of $\mathcal{A}$ from the state i to the state j having input label α and output label a. We define an involution $\theta_{\mathcal{A}}$ as the map such that $i \stackrel{\alpha|a}{\to} j$ and $1-j \stackrel{\alpha|\theta(a)}{\to} 1-i$ are the two edges of $\mathcal{A}$ having α as input label.

Theorem 6. *For any tree set T of characteristic 1 closed under reversal, the image of T by a doubling map $\delta_{\mathcal{A}}$ is a specular set relative to the involution $\theta_{\mathcal{A}}$.*

Example 7. Let $\Sigma = \{\alpha, \beta\}$ and let T be the Fibonacci set. Let $\delta_{\mathcal{A}}$ be the doubling map given by the transducer of Figure 1 on the left. The letter α acts as the transposition of the two states $0, 1$, while β acts as the identity.

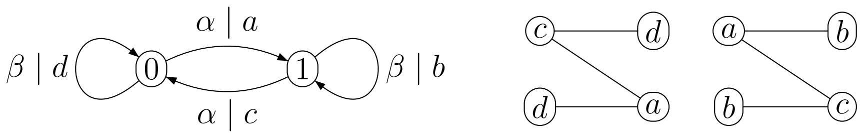

Figure 1: A doubling transducer and the extension graph $E(\varepsilon)$.

Then $\theta_{\mathcal{A}}$ is the involution θ of Example 2 and the image of T by δ is a specular set S on the alphabet $A = \{a, b, c, d\}$. The graph $E(\varepsilon)$ is represented in Figure 1 on the right. The letters a, c are odd while b, d are even.

5 Return words

Let S be a factorial set of words and let $X \subset S$ be a set of nonempty words. A *complete return word* to X is a word of S with a proper prefix in X, a proper suffix in X but no internal factor in X. We denote by $\mathcal{CR}(X)$ the set of complete return words to X. The set $\mathcal{CR}(X)$ is a bifix code. If S is uniformly recurrent, $\mathcal{CR}(X)$ is finite for any finite set X. For $w \in S$, we denote $\mathcal{CR}(w)$ instead of $\mathcal{CR}(\{w\})$. Thus $\mathcal{CR}(x)$ is the usual notion of a complete return word (see [11] for example).

Example 8. Let S be the specular set of Example 7. One can compute the sets $\mathcal{CR}(a) = \{abca, abcda, acda\}$, $\mathcal{CR}(b) = \{bcab, bcdacdab, bcdacdacdab\}$, $\mathcal{CR}(c) = \{cabc, cdabc, cdac\}$ and $\mathcal{CR}(d) = \{dabcabcabcd, dabcabcd, dacd\}$.

The following result is a direct consequence of [9, Theorem 5.2].

Theorem 9 (Cardinality Theorem for complete return words)**.** *Let S be a recurrent specular set on the alphabet A. For any finite nonempty bifix code $X \subset S$ with empty kernel, one has* $\mathrm{Card}(\mathcal{CR}(X)) = \mathrm{Card}(X) + \mathrm{Card}(A) - 2$.

Example 10. Let S be the specular set of Example 7. One has, for example, $\mathcal{CR}(\{a, c\}) = \{abc, ac, ca, cda\}$ and $\mathcal{CR}(\{b, d\}) = \{bcab, bcd, dab, dacd\}$. Both sets have four elements in agreement with Theorem 9.

Let S be a factorial set. For any nonempty word $w \in S$, a *right return word* to w in S is a word u such that wu is a complete return word to w. One defines symmetrically the *left return words*. We denote by $\mathcal{R}(x)$ the set of right return words to w in S and by $\mathcal{R}'(w)$ the corresponding set of left return words. Note that when S is closed under inverses, one has $\mathcal{R}(w)^{-1} = \mathcal{R}'(w^{-1})$. One can prove that in a specular set S, for every $w \in S$, all words in $\mathcal{R}(w)$ are even.

Theorem 11 (Cardinality Theorem for right return words)**.** *Let S be a recurrent specular set. For any $w \in S$, one has* $\mathrm{Card}(\mathcal{R}(w)) = \mathrm{Card}(A) - 1$.

Example 12. Let S be the specular set of Example 7. One has $\mathcal{R}(a) = \{bca, bcda, cda\}$, $\mathcal{R}(b) = \{cab, cdacdab, cdacdacdab\}$, $\mathcal{R}(c) = \{abc, dabc, dac\}$ and $\mathcal{R}(d) = \{abcabcd, abcabcabcd, acd\}$.

Two words u, v are said to *overlap* if a nonempty suffix of one of them is a prefix of the other. In particular a nonempty word overlaps with itself. We now consider the return words to $\{w, w^{-1}\}$ with w such that w and w^{-1} do not overlap. This is true for every w in a specular set S where the involution θ has no fixed point. With a complete return word u to $\{w, w^{-1}\}$, we associate a word $N(u)$ obtained as follows: if u has w as prefix, we erase it and if u has a suffix w^{-1}, we also erase it. Note that these two operations can be made in any order since w and w^{-1} cannot overlap. The *mixed return words to* w are the words $N(u)$ associated with the words $u \in \mathcal{CR}(\{w, w^{-1}\})$. We denote by $\mathcal{MR}(w)$ the set of mixed return words to w in S. The reason for this definition comes from the study of natural codings of a linear involution (see [8]). Note that $\mathcal{MR}(w)$ is closed under inverses and that $w\mathcal{MR}(w)w^{-1} = \mathcal{MR}(w^{-1})$.

Theorem 13 (Cardinality Theorem for mixed return words)**.** *Let S be a recurrent specular set on the alphabet A. For any $w \in S$ such that w, w^{-1} do not overlap, one has* $\mathrm{Card}(\mathcal{MR}(w)) = \mathrm{Card}(A)$.

Example 14. Let S be the specular set of Example 7. One has $\mathcal{MR}(b) = \{cab, c, dac, dab\}$. Since b, d do not overlap, $\mathcal{MR}(b)$ has four elements in agreement with Theorem 13.

6 Palindromes

The notion of palindromic complexity originates in [10] where it is proved that a word of length n has at most $n+1$ palindrome factors. A word of length n is *rich* (or full) if it has $n+1$ palindrome factors and a factorial set is *rich* (or full) if all its elements are rich. By a result of [12], a recurrent set closed under reversal is rich if and only if every complete return word to a palindrome in S is a palindrome. It is known that all Arnoux-Rauzy sets are rich [10] and also all natural codings of interval exchanges defined by a symmetric permutation [1]. The following proposition generalizes results of [1, 10].

Proposition 15. *Let T be a recurrent tree set of characteristic* 1 *closed under reversal. Then T is rich.*

In [14], this notion was extended to that of G-rich, where G is a finite group of morphisms and antimorphisms of A^* containing at least one antimorphism. A set S closed under G is G-rich if for every $w \in S$, every complete return word to the G-orbit of w is fixed by a nontrivial element of G ([14, Theorem 29]).

Let S be a specular set obtained as the image of a tree set of characteristic 1 by a doubling map $\delta_{\mathcal{A}}$. Let us define the antimorphism $\sigma : u \mapsto u^{-1}$ for $u \in S$. From Section 4 it follows that both edges $i \overset{\alpha|a}{\to} j$ and $1-i \overset{\alpha|\sigma(a)}{\to} 1-j$ are in $\mathcal{A}$. Let us define the morphism τ obtained by replacing each letter $a \in A$ by $\tau(a)$ if there are edges $i \overset{\alpha|a}{\to} j$ and $1-j \overset{\alpha|\tau(a)}{\to} 1-i$ in $\mathcal{A}$. We denote by $G_{\mathcal{A}}$ the group generated by the σ and τ. Actually, we have $G_{\mathcal{A}} = (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$.

Example 16. Let S be the recurrent specular set defined in Example 7. One has $G_{\mathcal{A}} = \{\mathrm{id}, \sigma, \tau, \sigma\tau\}$, where σ and τ are defined by $\sigma(a) = a, \sigma(b) = d, \sigma(c) = c, \sigma(d) = a$, and $\tau(a) = c, \tau(b) = d, \tau(c) = a, \tau(d) = b$. Note that $\sigma\tau = \tau\sigma$ is the antimorphism fixing b, d and exchanging a and c.

We now connect the notions of richness and G-richness, with an analogous result of Proposition 15 for specular sets.

Proposition 17. *Let T be a recurrent tree set of characteristic 1 on the alphabet Σ, closed under reversal and let S be the image of T under a doubling map $\mathcal{A}$. Then S is $G_{\mathcal{A}}$-rich.*

Example 18. Let S be the specular set of Example 7. S is $G_{\mathcal{A}}$-rich with respect to the group $G_{\mathcal{A}}$ of Example 16. The $G_{\mathcal{A}}$-orbit of a is the set $X = \{a, c\}$. The four words of $\mathcal{CR}(X)$ (see Example 10) are fixed by $\sigma\tau$.

References

[1] P. Baláži, Z. Masáková, and E. Pelantová. Factor versus palindromic complexity of uniformly recurrent infinite words. *Theoretical Computer Science*, 380(3):266–275, 2007.

[2] J. Berstel, C. De Felice, D. Perrin, C. Reutenauer, and G. Rindone. Bifix codes and Sturmian words. *Journal of Algebra*, 369:146–202, 2012.

[3] J. Berstel, D. Perrin, and C. Reutenauer. *Codes and Automata.* Cambridge University Press, 2009.

[4] V. Berthé, C. De Felice, V. Delecroix, F. Dolce, J. Leroy, D. Perrin, C. Reutenauer, and G. Rindone. Specular sets. In *Combinatorics on Words*, number 9304 in Springer LNCS, pages 210–222. 2015.

[5] V. Berthé, C. De Felice, F. Dolce, J. Leroy, D. Perrin, C. Reutenauer, and G. Rindone. Acyclic, connected and tree sets. *Monatshefte für Mathematik*, 176:521–550, 2015.

[6] V. Berthé, C. De Felice, F. Dolce, J. Leroy, D. Perrin, C. Reutenauer, and G. Rindone. The finite index basis property. *Journal of Pure and Applied Algebra*, 219:2521–2537, 2015.

[7] V. Berthé, C. De Felice, F. Dolce, J. Leroy, D. Perrin, C. Reutenauer, and G. Rindone. Maximal bifix decoding. *Dicrete Mathematics*, 338:725–742, 2015.

[8] V. Berthé, V. Delecroix, F. Dolce, D. Perrin, C. Reutenauer, and G. Rindone. Return words of linear involutions and fundamental groups. *Ergodic Theory and Dynamical Systems*, 2015. (to appear).

[9] F. Dolce and D. Perrin. Neutral and tree sets of arbitrary characteristic. *Theoretical Computer Science*, 2016. (to appear).

[10] X. Droubay, J. Justin, and G. Pirillo. Episturmian words and some constructions of de Luca and Rauzy. *Theoretical Computer Science*, 255(1-2):539–553, 2001.

[11] F. Durand. A characterization of substitutive sequences using return words. *Discrete Mathematics*, 179(1-3):89–101, 1998.

[12] A. Glen, J. Justin, S. Widmer, and L.Q. Zamboni. Palindromic richness. *European Journal of Combinatorics*, 30(2):510–531, 2009.

[13] M. Lothaire. *Algebraic Combinatorics on Words*. Cambridge University Press, 2002.

[14] E. Pelantová and Š. Starosta. Palindromic richness for languages invariant under more symmetries. *Theoretical Computer Science*, 518:42–63, 2014.