9. Infinite words generated by invertible substitutions

9.1 Introduction and preliminary

Combinatorial properties of finite and infinite words are of increasing importance in various fields of science (for a general reference, the reader should consult [228, 229] and the references therein). The combinatorial properties of the Fibonacci infinite word have been studied extensively by many authors (see for examples [53, 108, 149, 373]). This chapter looks at infinite words generated by invertible substitutions. As we shall see, words of this family generalize the Fibonacci infinite word.

The combinatorial properties of those infinite words are of great interest in various fields of mathematics and physics, such as number theory [7, 81, 137, 194], dynamical systems [277, 285, 292, 335], fractal geometry[37, 87, 186, 231], quasicrystals [70, 84, 215, 216, 231, 377, 375, 376, 380], tiling theory [37, 334], formal language and computational complexity [53, 55, 195, 248, 320, 321, 370], etc.

The purpose of this chapter is to study the properties of factors of infinite words generated by invertible substitutions over a two-letter alphabet. It is organized as follows:

Section 9.1 is devoted to the introduction and preliminaries. Section 9.2 is concerned with the structure of the semigroup of invertible substitutions for which we give a set of generators. Section 9.3 lists properties of factors of the Fibonacci word. After recalling some basic facts, we introduce what we call the singular words of the Fibonacci word and give their properties. We then establish two decompositions of the Fibonacci word involving the singular words. Some applications are derived from those decompositions, describing combinatorial properties of the Fibonacci word related to powers of its factors, for example. We also prove local isomorphisms properties of the Fibonacci word and the overlap properties of the factors.

In Sec. 9.4, we consider more general invertible substitutions. We give applications of decompositions associated with singular words, after proving some properties of factors associated with them. As we shall see, the general framework we introduce is at a much higher level of complexity than with the Fibonacci word. We conclude that section by giving a sufficient and necessary condition for the infinite words generated by two invertible substitutions to be locally isomorphic.

Let us now introduce some definitions and notation (see also Chap. 1).

Let $\mathcal{A}_n = \{a_1, a_2, \cdots, a_n\}$ be an alphabet with n letters. Let $\mathcal{A}_n^*$ and F_n respectively denote the free monoid with empty word ε as neutral element and the free group generated by $\mathcal{A}_n$.

Let us recall that a *substitution* (or a *morphism*) over $\mathcal{A}_n$ is a map $\phi : \mathcal{A}_n^* \to \mathcal{A}_n^*$, such that for any two words U and V, one has $\phi(UV) = \phi(U)\phi(V)$. The substitution ϕ is uniquely determined by the image of the elements of $\mathcal{A}_n$. We denote by $\phi = (u_1, u_2, \cdots, u_n)$ the substitution defined by $\phi(a_i) = u_i,\ 1 \le i \le n$.

A substitution over $\mathcal{A}_n^*$ can be extended to a map over $\mathcal{A}_n^{\mathbb{N}}$ by continuity, the set $\mathcal{A}_n^{\mathbb{N}}$ being equipped with the topology of simple convergence. A sequence $\mathbf{y} \in \mathcal{A}_n^{\mathbb{N}}$ is called a *fixed point* of the substitution ϕ if $\phi(\mathbf{y}) = \mathbf{y}$. If for some letter $a \in \mathcal{A}_n$, the word $\phi(a)$ begins in a and has length at least 2, then the sequence of words $\phi^n(a)$ converges to a fixed point $\phi^\omega(a) \in \mathcal{A}_n^{\mathbb{N}}$. In this chapter, we shall only consider fixed point of a substitution that can be obtained by iterating a morphism in the above manner.

Let $\phi : \mathcal{A}_n^* \to \mathcal{A}_n^*$ be a substitution over $\mathcal{A}_n^*$. Then ϕ can be naturally extended to F_n by defining $\phi(s^{-1}) = (\phi(s))^{-1}$. Denote by $\mathrm{Aut}(F_n)$ the group of automorphisms over F_n. Suppose ϕ is a substitution which is also an element of $\mathrm{Aut}(F_n)$, i.e., there exists a map $\eta : F_n \to F_n$ such that $\phi\eta = \eta\phi = Id$ on F_n, then ϕ is called an *invertible substitution* (where Id is the substitution given by $Id = (a_1, a_2, \cdots, a_n)$). The set of invertible substitutions over $\mathcal{A}_n^*$ is denoted by $\mathrm{Aut}(\mathcal{A}_n^*)$.

Set $\mathcal{A}_2 = \{a, b\}$ for convenience. We claim that the *Fibonacci substitution* $\sigma = (ab, a)$ is an invertible substitution over $\mathcal{A}_2$. To see this, choose $\eta = (b, b^{-1}a)$, then

$$\begin{aligned}
\sigma\eta(a) &= \sigma(b) = a \\
\sigma\eta(b) &= \sigma(b^{-1}a) = \sigma(b^{-1})\sigma(a) = a^{-1}ab = b \\
\eta\sigma(a) &= \eta(ab) = \eta(a)\eta(b) = bb^{-1}a = a \\
\eta\sigma(b) &= = \eta(a) = b.
\end{aligned}$$

The fixed point generated by the Fibonacci substitution $\sigma = (ab, a)$ (iterated over a) is called the *infinite Fibonacci word* and is denoted by u_σ. We shall sometimes simply call it the *Fibonacci word* when no confusion occurs.

The following theorem characterizes the structure of $\mathrm{Aut}(F_n)$.

Theorem 9.1.1. *[263, 277] The number of generators of the group* $\mathrm{Aut}(F_n)$ *is* $3n - 3$ *(these generators can be determined explicitly).*

In particular if $n = 2$, *then*

$$\mathrm{Aut}(F_2) = \langle \sigma, \quad \pi, \quad \psi \rangle,$$

where $\sigma = (ab, a)$, $\pi = (b, a)$ *and* $\psi = (a, b^{-1})$ *(where, as usual,* $\langle S \rangle$ *denotes the group generated by a set* S *of generators).*

Let us establish some terminology.

• Let $W \in \mathcal{A}_2^*$ be a word. We denote by $|W|_a$ (resp. $|W|_b$) the number of letters a (resp. b) in W. Denote also by $L(w)$ the vector $(|W|_a, |W|_b)$.

• We say that a word V is a *factor* of a word W, and we write $V \prec W$, if there exist words $U, U' \in \mathcal{A}_2^*$, such that $W = UVU'$. We say that V is a *left* (respectively *right*) *factor* of a word W, and we write $V \triangleleft W$ (respectively $V \triangleright W$), if there exists a word $U' \in \mathcal{A}_2^*$ such that $W = VU'$ (resp. $W = U'V$). One can also define factors and left factors of an infinite word (by allowing the word U' to be infinite).

• Let $W = x_1x_2 \dots x_n \in \mathcal{A}_2^*$. The *reversed word* (or *mirror image*) of W, denoted $\overline{W}$, is defined as $\overline{W} = x_n \dots x_2x_1$. A word W is called a *palindrome* if $W = \overline{W}$. We denote by $\mathcal{P}$ the set of palindromic words.

• Let $W = x_1x_2 \dots x_n$. For $1 \leq k \leq n$, we define the k-th *conjugation* of W as $C_k(W) = x_{k+1} \dots x_n x_1 \dots x_k$, and we denote by $C(W) = \{C_k(W);\ 1 \leq k \leq |W|\}$. By convention, $C_{-k}(W) = C_{|W|-k}(W)$.

• A word $W \in \mathcal{A}_2^*$ is called *primitive* if $W = U^p,\ U \in \mathcal{A}_2^*,\ p > 0$, implies $W = U$ (and thus $p = 1$). That is, W cannot be expressed as a proper power of any other word. Notice that C_k is an action from $\mathcal{A}_2^*$ to $\mathcal{A}_2^*$ preserving primitivity. That is, the word $C_k(W)$ is primitive if and only if W is itself a primitive word.

• Let $\tau : \mathcal{A}_2^* \to \mathcal{A}_2^*$ be a substitution.

1) The n-th iteration of τ is defined by $\tau^n = \tau(\tau^{n-1}),\ n \geq 2$.

2) We denote by u_τ any one of the fixed points of τ (if it exists).

3) The *incidence matrix* of τ is defined by $M_\tau = (L(\tau(a))^t, L(\tau(b))^t)$, where $L(\tau(a))^t$ is the transpose of $L(\tau(a))$. A substitution is called primitive if its incidence matrix is primitive (a nonnegative matrix is called primitive if it has a power with positive coefficients). In this chapter, unless stated otherwise, all substitutions we consider are primitive.

• Let w be a word (finite or infinite) with $|w| \geq n$. We denote by $\mathcal{L}_n(w)$ the set of factors of w of length n. We shall also write $\mathcal{L}_n$, if no confusion happens. Finally, we denote by $\mathcal{L}(w)$ the set of all factors of w. The set $\mathcal{L}(w)$ is also called the *language* of w.

• By a natural embedding, we can regard the set $\mathcal{A}_2^*$ as a subset of F_2, and if we say that w is in F_2, this means that it is reduced (see [277] and the references therein). Let $W = x_1x_2 \cdots x_n \in \mathcal{A}_2^*$. We denote by W^{-1} the *inverse word* of W, that is $W^{-1} = x_n^{-1} \cdots x_2^{-1}x_1^{-1}$. By convention, we have $WV^{-1} = U$, when $W = UV$.

• Let $W \in \mathcal{A}_2^*$. We denote by $i_W : F_2 \to F_2$ the inner automorphism defined by $U \mapsto WUW^{-1}$, for any $U \in F_2$. For any substitution τ, the composition $i_W \circ \tau$ results in $U \mapsto W\tau(U)W^{-1}$. We will say that two substitutions

ϕ, τ are *conjugate* if there exists a word $W \in \mathcal{A}_2^*$ such that $\phi = i_W \circ \tau$, which will be denoted as $\phi \sim \tau$. The set of all inner automorphisms of F_2 will denoted as $\mathrm{Inn}(F_2)$.

• From now on, the monoid generated by the set S will be denote by $\langle S \rangle$, unless stated otherwise.

• Let u and v be two infinite words over $\mathcal{A}_2$. We say that u and v are *locally isomorphic* if any factor of u (or its mirror image) is also a factor of v and vice versa. If u and v are locally isomorphic, we shall write $u \simeq v$.

9.2 Structure of invertible substitutions

This section describes a set of generators for the semigroup $\mathrm{Aut}(\mathcal{A}_2^*)$, thus revealing part of its structure.

We will prove the following conclusion: *any invertible substitution can be written as a finite product of three special invertible substitutions.* We shall see how this first result can be used as a starting point for studying invertible substitutions, since many basic properties will be derived from it.

The proof we present here is borrowed from [371]. A geometrical proof of the same result is given in [134].

Theorem 9.2.1 ([371]). *We have*

$$\mathrm{Aut}(\mathcal{A}_2^*) = \langle \pi, \varrho, \sigma \rangle = \langle \pi, \gamma, \delta \rangle,$$

where $\pi = (b, a)$, $\sigma = (ab, a)$, $\varrho = (ba, a)$, $\gamma = (ab, b)$, *and* $\delta = (ba, b)$.

The proof follows from the following lemmas.

Lemma 9.2.1. *A substitution* $\tau \in \mathrm{Aut}(F_2)$ *is an inner automorphism if and only if its incidence matrix* M_τ *is the identity matrix.*

Lemma 9.2.2 ([371]). *Let*

$$\mathcal{M} = \left\{ \begin{pmatrix} p & r \\ q & s \end{pmatrix}; \quad p, q, r, s \in \mathbb{N}, \quad \det \begin{pmatrix} p & r \\ q & s \end{pmatrix} = \pm 1 \right\},$$

then we have

$$\mathcal{M} = \langle A,\ B \rangle,$$

where

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

Proof.
Let

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad D = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad E = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}.$$

Let also $M = \begin{pmatrix} p & r \\ q & s \end{pmatrix}$ be an element of $\mathcal{M}$ with at least one entry ≥ 1. Letting the matrices A and B act on M a finite number of times, an easy induction on $\max(p, q, r, s)$ shows that it can be transformed by performing Euclidean divisions into one of the matrices I, A, B, C, D and E. Moreover, one can easily check the following relations:

$$I = A^2, \quad C = BA, \quad D = AB, \quad E = ABA.$$

This ends the proof. ∎

Observe that $\pi^{-1} = (b, a)$, $\gamma^{-1} = (ab^{-1}, b)$, $\delta^{-1} = (b^{-1}a, b)$. Also, using those relations, we get:

Lemma 9.2.3. *(1)* $\pi = \pi^{-1}$, $\pi^2 = \mathrm{id}$,
(2) $\gamma\delta = \delta\gamma$, $\delta^{-1} = \gamma\delta^{-1}$,
(3) $\delta^{-1}\pi\gamma\pi\delta = \gamma^{-1}\pi\delta\pi\gamma$,
(4) $\gamma = i_a\delta = \delta i_b$,
(5) $\pi i_a = i_b\pi$,
(6) $i_a\pi\delta = \pi\gamma$, $i_a\gamma\pi\delta = \delta\pi\gamma$, $i_b\pi\gamma\pi\delta = \pi\delta\pi\gamma$, $i_{b^{-1}}\pi\delta\pi\gamma = \pi\gamma\pi\delta$,
(7) $\pi(aW) = b\pi(W)$, $\pi(bW) = a\pi(W)$, $\gamma(aW) = ab\gamma(W)$, $\gamma(bW) = b\gamma(W)$, $\delta(aW) = ba\delta(W)$, $\delta(bW) = b\delta(W)$, *where* $W \in F_2$,
(8) $\pi(Wa) = \pi(W)b$, $\pi(Wb) = \pi(W)a$, $\gamma(Wa) = \gamma(W)ab$, $\gamma(Wb) = \gamma(W)b$, $\delta(Wa) = \delta(W)ba$, $\delta(Wb) = \delta(W)b$.

The following lemma is easily proved by induction.

Lemma 9.2.4. *Let s and $e_1, e_2 \ldots e_{2s-1}$ be positive integers. We have*

$$\delta^{-1}\pi\gamma^{e_1}\pi\gamma^{e_2}\cdots\pi\gamma^{e_{2s-2}}\pi\gamma^{e_{2s-1}}\pi\delta = \gamma^{-1}\pi\delta^{e_1}\pi\gamma^{e_2}\cdots\pi\gamma^{e_{2s-2}}\pi\delta^{e_{2s-1}}\pi\gamma.$$

Proof of Theorem 9.2.1. Let $\tau \in \mathrm{Aut}(\mathcal{A}_2^*)$. Then $M_\tau \in \mathcal{M}$. By Lemma 9.2.2, there exist $M_1, M_2, \ldots M_k \in \{A, D\}$ such that $M_\tau = M_1 \cdots M_k$. Note that $M_\pi = A$ and $M_\gamma = D$. Take $\phi = \phi_1 \cdots \phi_k$, with $\phi_i = \pi$ or $\phi_i = \gamma$, according to $M_i = A$ or D. Then, we have $M_\tau = M_\phi$. Therefore Lemma 9.2.1 implies $\tau \circ \phi^{-1} \in \mathrm{Inn}(F_2)$. This means that there exists $W \in F_2$ such that $\tau = i_W \circ \phi$, with $\phi \in \langle \pi,\ \gamma,\ \delta \rangle$. Since $\tau(a)$ and $\phi(a)$ belong to $\mathcal{A}_2^*$, we have $W \in \mathcal{A}_2^*$ or $W^{-1} \in \mathcal{A}_2^*$.

Thus we only need to prove the following claim: if $\phi \in \langle \pi,\ \gamma,\ \delta \rangle$ is such that $\phi(a) = WU$ and $\phi(b) = WV$ (with $U, V, W \in \mathcal{A}_2^*$), then $i_W \circ \phi \in \langle \pi,\ \gamma,\ \delta \rangle$.

Let $W = x_1x_2 \dots x_n$, then $i_W = i_1i_2 \dots i_n$, where $i_k = i_{x_k}$. Hence the proposition follows if we can prove that $z \in \{a, b, a^{-1}, b^{-1}\}$ implies $\mathbf{i}_z \circ \phi \in \langle \pi,\ \gamma,\ \delta \rangle$.

First consider $z = b$, and let us introduce $\tau = i_b \circ \phi$, where $\phi = \phi_1\phi_2 \dots \phi_t$ ($\phi_j \in \langle \pi,\ \gamma,\ \delta \rangle$). We shall prove that, if $\tau(a)$ and $\tau(b)$ are in $\mathcal{A}_2^*$, then $\tau \in \langle \pi,\ \gamma,\ \delta \rangle$. Lemma 9.2.3-(7) shows that γ leaves the first letter invariant, whereas π exchanges letters a and b. Therefore there exists at least one substitution equal to δ among $\phi_1, \phi_2, \dots, \phi_t$, otherwise the first letters of $\phi(a)$ and $\phi(b)$ would be a and b (or b and a), which would contradict $\tau(a), \tau(b) \in \mathcal{A}_2^*$. Let i be the smallest integer i such that $\delta = \phi_i$. The proposition will follow if we can prove $i_b\phi_1\phi_2 \dots \phi_i \in \langle \pi,\ \gamma,\ \delta \rangle$.

We may suppose $\phi_{i-1} = \pi$, since $\gamma\delta = \delta\gamma$. Furthermore, we may also suppose $\phi_{i-2} = \gamma$, since $\pi^2 = \mathrm{id}$. Therefore we have the following alternatives:

(i) either $\pi\gamma^{e_1}\pi\gamma^{e_2} \dots \pi\gamma^{e_n}\pi$,

(ii) or $\phi_1\phi_2 \dots \phi_{i-1} = \gamma^{e_1}\pi\gamma^{e_2} \dots \pi\gamma^{e_n}\pi$,

where $e_i \geq 1$.

Since the words $\phi_1\phi_2 \dots \phi_{i-1}(a)$ and $\phi_1\phi_2 \dots \phi_{i-1}(b)$ both start with the letter b, the argument above implies that n is even in case (i), and odd in case (ii).

In case (ii), Lemma 9.2.3-(4) and Lemma 9.2.4 imply

$$\begin{aligned} i_b\phi_1\phi_2 \dots \phi_{i-1}\delta &= i_b\pi\gamma^{e_1} \dots \pi\gamma^{e_{2s-1}}\pi\delta \\ &= (i_b\delta)(\delta^{-1}\pi\gamma^{e_1} \dots \pi\gamma^{e_{2s-1}}\pi\delta) \\ &= \gamma(\gamma^{-1}\pi\delta^{e_1} \dots \pi\delta^{e_{2s-1}}\pi\gamma \\ &= \pi\delta^{e_1} \dots \delta^{e_{2s-1}}\pi\gamma \in \langle \pi,\ \gamma,\ \delta \rangle. \end{aligned}$$

In case (i), Lemma 9.2.3-(2 and 4) and Lemma 9.2.4 imply

$$\begin{aligned} i_b\phi_1\phi_2 \dots \phi_{i-1}\delta &= i_b\pi\gamma^{e_1} \dots \pi\gamma^{e_{2s}}\pi\delta \\ &= (i_b\delta\gamma^{e_1})(\delta^{-1}\pi\gamma^{e_2} \dots \pi\gamma^{e_{2s}}\pi\delta) \\ &= \gamma^{e_1}\pi\gamma^{e_2}\pi \dots \pi\gamma^{e_{2s-1}}\delta^{e_{2s}}\pi\gamma \in \langle \pi,\ \gamma,\ \delta \rangle. \end{aligned}$$

Now, consider the case $z = a$, i.e., the substitution $i_a \circ \phi$. By Lemma 9.2.3-(5), we have $\tau = i_a \circ \phi = \pi\mathbf{i}_b\pi\phi$. Again, the arguments developed earlier imply $i_b\pi\phi \in \langle \pi,\ \gamma,\ \delta \rangle$, so we have $\tau \in \langle \pi,\ \gamma,\ \delta \rangle$.

The cases $i_{a^{-1}} \circ \phi$ and $i_{b^{-1}} \circ \phi$ can be handled similarly using Lemma 9.2.3-(8). ■

Remark. • Observe that $\gamma \notin \langle \pi,\ \delta \rangle$ and $\delta \notin \langle \pi,\ \gamma \rangle$. Hence the minimal number of generators of $\mathrm{Aut}(\mathcal{A}_2^*)$ is exactly 3.

• Brown [81], Kósa [212], Séébold [321], Mignosi and Séébold [249] have considered invertible substitutions in $\langle \pi, \pi_1, \pi_2 \rangle$, with $\pi_1 = (a, ab)$, $\pi_2 = (a, ba)$, and where $\langle \pi, \pi_1, \pi_2 \rangle$ denotes the semigroup generated by π, π_1, π_2.

They studied combinatorial and arithmetic properties of the infinite words generated by those invertible substitutions.

• Let us recall that a *Sturmian sequence* is an infinite sequence $x_{\alpha,\beta} = (x_n)_{n\in\mathbb{N}}$ over the alphabet $\mathcal{A}_2$ such that we either have $\forall n \in \mathbb{N}$, $x_n = a$ if $[\alpha(n+1)+\beta] = [\alpha+\beta]$, $x_n = b$ otherwise, or $\forall n \in \mathbb{N}$, $x_n = a$ if $\lceil \alpha(n+1)+\beta \rfloor = \lfloor \alpha+\beta \rfloor$, $x_n = b$ otherwise, where α, β are real numbers with $0 \leq \alpha \leq 1$ irrational (for more details, see Chap. 6). Those sequences are also sometimes called Beatty sequence, billiard trajectories, cutting sequence, etc. Equivalently, a Sturmian sequence has precisely $n+1$ distinct factors of length n for each positive integer n. Sturmian sequences possess remarkable combinatorial and arithmetical properties and, correspondingly, a long history. The interested reader is referred to Berstel [55], Berthé [60] and de Luca [110].

Mignosi and Séébold [249] have considered Sturmian substitutions, that is, substitutions preserving all Sturmian sequences. They proved that a fixed point of a substitution over $\mathcal{A}_2$ is Sturmian if and only if the substitution is invertible ([249, 371]). This motivates us to first study the set of invertible substitutions before looking at general Sturmian substitutions.

• Let $n \geq 2$ and let $\mathcal{A}_n = \{a_1, \cdots, a_n\}$ be a n-letter alphabet. Invertible substitutions over $\mathcal{A}_n$ are defined in complete analogy with invertible substitutions over $\mathcal{A}_2$. The *Rauzy substitution*, $\mathcal{R}_n : \mathcal{A}_n^* \to \mathcal{A}_n^*$, is a typical invertible substitution and is defined as follows:

$$\mathcal{R}_n(a_1) = a_1a_2,\ \mathcal{R}_n(a_2) = a_2a_3, \cdots, \mathcal{R}_n(a_{n-1}) = a_{n-1}a_n,\ \mathcal{R}_n(a_n) = a_1.$$

From its definition, we see that the Rauzy substitution is a generalization of the Fibonacci substitution. This substitution has many remarkable properties and has some interesting relations with other domains; for more details, see Chap. 7. Other interesting invertible substitutions over $\mathcal{A}_n$ are described in [137].

• Theorems 9.1.1 and 9.2.1, bring us to naturally conjecture that the number of generators of $\mathrm{Aut}(\mathcal{A}_n^*)$ is $3n-3$. However, authors in [374], have proved that the set of invertible substitutions over an alphabet of more than three letters is not finitely generated (their proof uses the notion of "prime substitution"). Also, they gave examples showing that, in that case, the structure of $\mathrm{Aut}(\mathcal{A}_n^*)$ is much more complex.

9.3 Singular words of the Fibonacci word and applications

As we have seen in the last section, the Fibonacci substitution plays an important role in the study of the structure of the group of invertible substitutions. Our investigation of the properties of infinite words generated by invertible substitutions will be guided by the study of the factors of the

Fibonacci word. Those factors have been studied in relation with mathematics and physics in theories such as number theory, fractal geometry, formal languages, computational complexity, quasicrystals etc; see for examples, [7, 70, 194, 285, 320, 370]. Moreover, the properties of the factors of the Fibonacci word have been studied extensively by many authors [53, 108, 195, 248, 320, 321, 373]. We give new properties of the factors of the Fibonacci word based on a new family of factors we call singular words. As we shall see, the most striking of those properties is that two distinct occurrences of a singular word never overlap.

This section is organized as follows. After making preliminary remarks on the Fibonacci word, we introduce the singular words and give elementary properties they satisfy. Then, we establish two decompositions of the Fibonacci word in terms of the singular words (Theorems 9.3.1 and 9.3.2), and describe some of their applications.

Those results (and, in particular, the positively separated property of the singular words) bring us in Sec. 9.3.2 to other combinatorial properties of Fibonacci infinite word, such as local isomorphism, the overlap properties of the factors, right special factors and the powers of factors of the Fibonacci word.

9.3.1 Singular words and their properties

As earlier, let $\sigma = (ab, a)$ be the Fibonacci substitution and u_σ denote the Fibonacci infinite word.

One of the motivations we have is the following. The Fibonacci word is closely related to the Fibonacci numbers defined by the recurrence formula $f_{n+2} = f_{n+1} + f_n$, with the initial conditions $f_{-1} = f_0 = 1$, as shown in Chap. 2. Consider the following decomposition of the Fibonacci word

$$\underline{a}\ \underline{b}\ \underline{aa}\ \underline{bab}\ \underline{aabaa}\ \underline{babaabab}\ \underline{aabaababaabaa}\ \underline{babaababaabaababaabab} \cdots$$

where the n-th block in the decomposition is of length f_n, $n \geq -1$. The question of whether those blocks share special properties naturally comes to mind. As we shall see, Theorem 9.3.1 will answer this question completely.

We first need to state basic facts about the Fibonacci word, for which the reader is referred to [53, 108, 285, 320]; see also Chap. 6.

Proposition 9.3.1. *Let $F_n = \sigma^n(a)$, then*

1) $|\sigma^n(a)| = Card\ C(F_n) = f_n$. That is, all conjugate of the word F_n are distinct. In particular, for any $W \in C(F_n)$, we have

$$L(W) = L(F_n) = (f_{n-1}, f_{n-2}).$$

Moreover,

$$C(F_n) = \{\overline{W};\ W \in C(F_n)\}.$$

2) $F_{n+1} = F_n F_{n-1}$.
3) For any $k \geq 1$, $\sigma^k(u_\sigma) = u_\sigma$. *That is,*

$$u_\sigma = F_k F_{k-1} F_k F_k F_{k-1} \cdots .$$

4) The word ab *is a suffix of* F_n *for odd positive* n*, and the word* ba *is a suffix of* F_n *for even positive* n.
5) $b^2 \not\prec u_\sigma$, $a^3 \not\prec u_\sigma$.
6) Any factor of u_σ *appear infinitely many times in* u_σ.
7) $W \prec u_\sigma$ *if and only if* $\overline{W} \prec u_\sigma$.

Remark. No other results than those cited in Proposition 9.3.1 will be used.

Assume $\alpha, \beta \in \mathcal{A}_2$. Then $\alpha\beta \triangleright F_n$ implies $\alpha \neq \beta$, by virtue of Proposition 9.3.1.4).

An induction using Proposition 9.3.1.1), 9.3.1.3) and 9.3.1.4) proves the following lemma.

Lemma 9.3.1. *Let* $n \geq 2$*, and assume that* $\alpha\beta$ *is a suffix of* F_n. *Then*

$$F_n = F_{n-2} F_{n-1} \alpha^{-1} \beta^{-1} \alpha\beta,$$

$$F_{n-2} F_{n-1} = F_n \beta^{-1} \alpha^{-1} \beta\alpha.$$

Remark. Observe that $F_n = F_{n-1}F_{n-2}$. Using the preceding lemma, we see that the words $F_{n-1}F_{n-2}$ and $F_{n-2}F_{n-1}$ coincide except on the two last letters which are distinct. This remarkable property will be useful when studying the factors of Fibonacci word.

Let $\alpha\beta$ be a suffix of F_n and write $W_n = \alpha F_n \beta^{-1}$.

Lemma 9.3.2. *We have:*

1) $W_n \notin C(F_n)$;

2) $\mathcal{L}_{f_n}(F_{n-1}F_n) = W_n \bigcup \{C_k(F_n);\ 0 \leq k \leq f_{n-2} - 2\}$. *In particular, as a factor,* W_n *appears only once in* $F_{n-1}F_n$.

Proof. Statement 1) follows from $\alpha \neq \beta$, $L(W_n) \neq L(F_n)$.
Lemma 9.3.1 implies that $f_{n-1}F_n = F_n F_{n-1} \alpha^{-1}\beta^{-1}\alpha\beta$ when $\alpha\beta \triangleright F_n$. Since $F_{n-1} \triangleleft F_n$, the first f_{n-1} factors of length f_n of the word $F_{n-1}F_n$ are exactly $C_k(F_n)$, $1 \leq k \leq f_{n-1}-2$, and the last factor is $F_n = C_{f_n}(F_n)$. The $(f_{n-1}-1)$-th factor is $\alpha F_n \beta^{-1} = W_n$, which ends the proof of 2). ∎

As a consequence, for any $n \geq 1$, the set $\mathcal{L}_{f_n}$ splits into two parts. The first part coincides with the conjugacy class of the word F_n, and the second part consists solely of the word W_n. We shall now describe some interesting properties of the word W_n which play an important role in the study of the factors of u_σ.

Definition 9.3.1. *The word* W_n *is called the* n-th singular word *of the Fibonacci word* u_σ. *For convenience, we define* $W_{-2} = \varepsilon$, $W_{-1} = a$ *and* $W_0 = b$. *Denote by* $\mathcal{S}$ *the set of singular words of* u_σ.

We now describe properties of the singular words:

Proposition 9.3.2. *Let* $\{W_n\}_{n \geq -1}$ *be the singular words, then*
1) If $n \geq 1$, *then*

$$L(W_n) = \begin{cases} (f_{n-1}+1, f_{n-2}-1) & \text{if } n \text{ is odd;} \\ (f_{n-1}-1, f_{n-2}+1) & \text{if } n \text{ is even.} \end{cases}$$

2) $W_n \not\prec W_{n+1}$.
3) If $\alpha \triangleright W_{n+1}$, *then* $W_{n+2} = W_n W_{n+1} \alpha^{-1} \beta$.
4) $W_n = W_{n-2} W_{n-3} W_{n-2}$, $n \geq 1$.
5) For $n \geq 1$,

$$\begin{aligned} C_{f_{n-1}-1}(F_n) &= W_{n-2} W_{n-1}, \\ C_{f_n-1}(F_n) &= W_{n-1} W_{n-2}. \end{aligned}$$

In particular,

$$\begin{aligned} &W_{n-2} \prec C_k(F_n) \text{ if and only if } 0 \leq k \leq f_{n-1}-1; \\ &W_{n-1} \prec C_k(F_n) \text{ if and only if } f_{n-1}-1 \leq k \leq f_n - 1. \end{aligned}$$

6) For $n \geq 1$, $W_{2n-1} = aaUaa$, $W_{2n} = bVb$, *where* $U, V \in \mathcal{A}_2^*$.
7) For $n \geq 2$, $1 < k < f_n$, *no proper conjugate of* W_n *is a subword of* F_∞.
8) For $n \geq 0$, $W_n^2 \not\prec F_\infty$.
9) For $n \geq -1$, W_n *is a palindrome.*
10) W_n *is not the product of two palindromes for* $n \geq 2$.
11) If $n \geq 2$, *then* W_n *is primitive.*
12) For $n \geq 1$, *we have*

$$W_n = W_n^* \left(\prod_{j=-1}^{n-2} W_j \right) = \left(\prod_{j=-1}^{n-2} W_{n-j-3} \right) W_n^*$$

$$\text{where } W_n^* = \begin{cases} a & \text{if } n \text{ is odd,} \\ b & \text{if } n \text{ is even.} \end{cases}$$

13) $W_n \not\prec \prod_{j=-1}^{n-1} W_j$.

14) Let $k \geq -1$ *and* $p \geq 1$, *and let* $U = \prod_{j=k}^{k+p} W_j$. *Then* $U \notin \mathcal{S}$.

The proofs are left to the readers as exercises.

The following theorem answers the question raised in the introductory part of this section.

Theorem 9.3.1. *We have* $u_\sigma = \prod_{j=-1}^{\infty} W_j$.

Proof. One easily checks that $F_n = abF_0F_1 \cdots F_{n-3}F_{n-2}$. By definition of the singular words W_n, we have $F_k = \alpha^{-1}W_k\beta$. Then by Proposition 9.3.1.4), we get

$$F_n = W_{-1}W_0 \cdots W_{n-2}\gamma,$$

where $\gamma = a$ if n is even, and b is n is odd. Letting $n \to \infty$ we obtain the conclusion of the theorem. ■

Some additional work [373] is needed to obtain an even more striking decomposition into singular words. In particular, this decomposition shows the non overlapping property of the singular words and plays an important role in the study of the factors of the Fibonacci word.

Definition 9.3.2. *For each* $n \geq 1$*, we consider the set* $\Sigma_n = \{W_{n-1}, W_{n+1}\}$ *as a new alphabet of two letters, and we define a morphism* $\phi_n : \mathcal{A}_2 \to \Sigma_n$ *by setting* $\phi_n(a) = W_{n+1}, \phi_n(b) = W_{n-1}$*. We call the sequence* $z = \phi_n(u_\sigma)$ *the* Fibonacci word over Σ_n.

By Proposition 9.3.1.6), the word W_n will appear in u_σ infinitely many times. We index the successive occurrences of those words as a sequence $\{W_{n,k}\}_{k\geq 1}$ and we call $W_{n,k}$ the *k-th singular word of order n.*

We can now state the main result of this section:

Theorem 9.3.2. *For any* $n \geq 0$*, we have*

$$u_\sigma = \left(\prod_{j=-1}^{n-1} W_j\right) W_{n,1}z_1W_{n,2}z_2 \cdots W_{n,k}z_kW_{n,k+1} \cdots$$

where $z = z_1z_2 \cdots z_n \cdots$ *is the Fibonacci word over* Σ_n.

For example, when $n = 2$, $W_2 = bab$ and the decomposition predicted by Theorem 9.3.2 is:

$$abaa(bab)\underline{aabaa}(bab)\underline{aa}(bab)\underline{aabaa}(bab)\underline{aabaa}(bab)\underline{aa}(bab)\underline{aabaa}(bab)\underline{aa}\cdots$$

Definition 9.3.3. *Let* $\mathbf{v} = v_1v_2 \cdots v_n \cdots$ *be an infinite word over* $\mathcal{A}_2$*. Let* $X = v_kv_{k+1} \cdots v_{k+p}$ *and* $Y = v_lv_{l+1} \cdots v_{l+m}$ *(with* $l \geq k$*) be two factors of* $\mathbf{v}$*, then the distance between the factors* X *and* Y *is defined by*

$$d(X,Y) = \begin{cases} l-k-p & \text{if } l > k+p; \\ 0 & \text{otherwise.} \end{cases}$$

If $d(X,Y) > 0$*, we say that the words* X *and* Y *are* positively separated.

Let $W = v_kv_{k+1} \cdots v_{k+p}$ $(k,p \geq 1)$ *be a factor of* $\mathbf{v}$*. If there is an integer* l*,* $1 \leq l \leq p$*, such that* $W = v_{k+l}v_{k+l+1} \cdots v_{k+l+p}$*, then we say that* W overlaps itself over $p-l$ letters.

The following statement can be seen as equivalent to the above definition. Let $U \prec u_\sigma$ be a factor of the Fibonacci word. Then, the word U overlaps itself if and only if there exist words X, Y and Z such that $U = XY = YZ$ and such that the word $UZ = XYZ$ is a factor of u_σ.

The next corollaries are consequences of Theorem 9.3.2.

Corollary 9.3.1. *Two adjacent occurrences of a singular word are positively separated. More precisely, for any n and k, we have*

$$d(W_{n,k}, W_{n,k+1}) \in \{f_{n+1}, f_{n-1}\}.$$

Moreover, one of $d(W_{n,k}, W_{n,k+1})$ and $d(W_{n,k+1}, W_{n,k+2})$ is f_{n+1}.

Corollary 9.3.2. *The words of length f_{n-2k} adjacent to the word W_{n+1} on the left and on the right are precisely W_{n-2k}.*

Corollary 9.3.3. *For $n \geq 1$, W_n does not overlap itself.*

Corollary 9.3.4. *Let $U \prec u_\sigma$ and assume $f_n < |U| \leq f_{n+1}$. Let W be a singular word of maximal length contained in U. Then W appears only once in U. Moreover, W must be one of the three following singular words: W_{n-1}, W_n or W_{n+1}.*

9.3.2 Some applications of singular word

The results we will now describe follow from properties of singular words, and more particularly from the fact that two occurences of a singular word are positively separated. Some of the results we give here are known (Example 1 and Example 3), but the proofs we provide are simpler.

Example 1. Power of factors. [53, 195, 248, 320]

Theorem 9.3.3. *We have:*
1) for any n, $W_n^2 \not\prec u_\sigma$;
2) for $0 \leq k \leq f_n - 1$, $\big(C_k(F_n)\big)^2 \prec u_\sigma$;
3) if $U \prec u_\sigma$, with $f_{n-1} < |U| < f_n$, then $U^2 \not\prec u_\sigma$;
4) if $0 \leq k \leq f_{n-1} - 2$, then $\big(C_k(F_n)\big)^3 \prec u_\sigma$;
5) if $f_{n-1} - 2 < k < f_n$, then $\big(C_k(F_n)\big)^3 \not\prec u_\sigma$;
6) for any $U \prec u_\sigma$, $U^4 \not\prec u_\sigma$.

Proof. 1) The statement follows from Proposition 9.3.1.5) and 9.3.2.6).

2) Let $C_k(F_n) = UV$ with $F_n = VU$. Then $U \triangleright F_n$ and $V \triangleleft F_n$. The result follows from $F_n^3 \prec u_\sigma$, since

$$\big(C_k(F_n)\big)^2 = UVUV = UF_nV \prec (F_n)^3.$$

3) Suppose that W_k is the largest singular word contained in U (as in Corollary 9.3.4), and let $U = V_1W_kV_2$. Assume that $U^2 = V_1W_kV_2V_1W_kV_2 \prec u_\sigma$. Then $W_k \not\prec V_2V_1$, otherwise Theorem 9.3.2 implies either that $W_{k+1} \prec V_1$, or $W_{k+1} \prec V_2$, which would contradict the hypothesis on W_k. Thus two singular words of the order k above are adjacent, so by Theorem 9.3.2 again, V_2V_1 must be either W_{k+1} or W_{k-1}. By Proposition 9.3.2.5), U is conjugated either to F_{k+2}, or to F_{k+1}. But these two cases are impossible, since they violate the hypothesis made on U.

4) We have that $F_nF_nF_{n-1}F_n \prec f_\infty$ since $aaba \prec f_\infty$. Let $\alpha\beta \triangleright F_{n-1}$. Then by Lemma 9.3.1, we have

$$F_n^2F_{n-1}F_n = F_n^2F_{n-1}F_{n-2}F_{n-1}\alpha^{-1}\beta^{-1}\alpha\beta = F_n^3F_{n-1}\alpha^{-1}\beta^{-1}\alpha\beta \prec u_\sigma.$$

Note that $F_{n-1} \triangleleft F_n$. Hence, if $0 \le k \le f_{n-1} - 2$, then

$$\big(C_k(F_n)\big)^3 \prec F_n^3F_{n-1}\alpha^{-1}\beta^{-1} \prec u_\sigma.$$

5) Now suppose that $f_{n-1} - 1 < k < f_n$. Then Proposition 9.3.2.5) implies $W_{n-1} \prec C_k(F_n)$. Let $C_k(F_n) = UW_{n-1}V$, so that $VU = W_{n-2}$. Thus

$$\big(C_k(F_n)\big)^3 = UW_{n-1}W_{n-2}W_{n-1}W_{n-2}W_{n-1}V.$$

Now, if $\big(C_k(F_n)\big)^3 \prec u_\sigma$, then the word $W_{n-1}W_{n-2}W_{n-1} = W_{n+1}$ would overlap itself which is impossible, by Corollary 9.3.4.

6) The statement is proved using an argument similar to that developed in 5). ■

Remark. From Theorem 9.3.3.2), we see that, any two occurences of a conjugate of F_n, $n \ge 0$, are not positively separated. This is a major difference between the conjugates of F_n and W_n.

Example 2. Local isomorphism

Denote by $T^k = T(T^{k-1})$ the kth iteration of the shift T. Properties of the singular words of the Fibonacci word can be used to obtain the following results on the local isomorphism properties of the Fibonacci word.

Theorem 9.3.4. *1) An infinite word u'_σ obtained by changing a finite number of letters in u_σ is not locally isomorphic to u_σ.*
2) Let $U \in \mathcal{A}_2^*$. Then $u_\sigma \simeq Uu_\sigma$ if and only if there exists an integer $m > -1$ such that $U \triangleright W_mW_m^*$, where W_m^* is defined as in Proposition 9.3.2.12).
3) For any $k \ge 1$, $T^k(u_\sigma) \simeq u_\sigma$.

Proof. 1) Let $u_\sigma = \prod_{j=-1}^{\infty} W_j$ as in Theorem 9.3.1. Because only a finite number of letters are changed in u_σ, we can find an integer m and words $U, V \in \mathcal{A}_2^*$, such that

$$F'_\sigma = UV \prod_{j=m}^{\sigma} W_j,$$

where $|V| = f_{m-1}$, $V \neq W_{m-1}$. Therefore by Corollary 9.3.2, $VW_m \not\prec u_\sigma$, that is, $u_\sigma \not\simeq F'_\sigma$.

2) From Theorem 9.3.1 and Proposition 9.3.2.12), for any $k > 0$ and $m \geq 0$,

$$\begin{aligned} W_{2m}au_\sigma &= W_{2m}a \left(\prod_{j=-1}^{2m+2k-1} W_j \right) \left(\prod_{j=2m+2k}^{\infty} W_j \right) \\ &= W_{2m}W_{2m+2k+1} \left(\prod_{j=2m+2k}^{\infty} W_j \right). \end{aligned}$$

Then, by Corollary 9.3.2, $W_{2m}W_{2m+2k+1} \prec u_\sigma$. That is, for any $V \prec W_{2m}au_\sigma$, we can find an integer k, such that $V \prec W_{2m}W_{2m+2k+1}$, so $V \prec u_\sigma$. The case dealing with $W_{2m+1}b$ can be proved similarly. That is, if $U \triangleright W_m W_m^*$ for some m, then $u_\sigma \simeq Uu_\sigma$. The case where U is not a right factor of any $W_m W_m^*$, follows from an argument similar to that developed in the proof of statement 1).

3) The proof follows from Proposition 9.3.1.6). ∎

Example 3. Study of right special factors of u_σ

Let us recall that a right special factor of u_σ is defined as follows: if $Ua, Ub \prec u_\sigma$, then the word U is called a *right special factor* of u_σ. The following theorem, of which we shall give a proof using singular words, is due to Berstel [53].

Theorem 9.3.5. *A word $W \prec u_\sigma$ is a right special word if and only if $W \triangleright \overline{F_n}$, for some $n \geq 0$.*

Proof. One easily checks that $\overline{F_n}$ is a right special word, for any $n \geq 0$. Therefore we only need to prove that $|\mathcal{L}_n| = n + 1$, for any $n \geq 0$.

Let $U \prec u_\sigma$ and assume $f_k < |U| \leq f_{k+1}$. An argument analogous to the one developed in Lemma 9.3.2 shows that the word U has one of the three following forms:

- $U = SW_nT$, $|ST| \leq f_{n-1}$;

- $U = SF_nT,\ S,T \neq \varepsilon,\ |ST| \leq f_{n-1},\ S \triangleright F_n,\ T \triangleleft F_n$;
- $U = ST,\ S \triangleright F_n,\ T \triangleleft F_n$.

In the first case, by Corollary 9.3.2, the factor $S \triangleright W_{n-1}$ (resp. T) is uniquely determined. Moreover, if $S \neq S'$, then $SW_nT \neq S'W_nT'$, since W_n has no overlap. Hence there are exactly $|U| - f_k + 1$ distinct words SW_nT which correspond to $|S| = 0, 1, \ldots, n - f_k$.

In the two later cases, Proposition 9.3.1.1) readily proves that there are exactly f_k different factors of length $|U|$ of F_k^3.

Summarizing the discussions above, we get $|\mathcal{L}_{|U|}| = f_k + (|U| - f_k + 1) = |U| + 1$. ∎

Example 4. Overlap of the factors of the Fibonacci word

In this example, we shall determine the factors which overlap themselves.

First recall the notations we introduced earlier. We say that the word $U \prec u_\sigma$ overlap itself if there exist words X, Y and Z such that $U = XY = YZ$ and $XYZ \prec u_\sigma$. Write $\hat{U}(Y) = UY = XYZ$. We shall say that the word U overlaps itself with overlap factor Y (or overlap length $|Y|$). The word $\hat{U}(Y)$ is called the *overlap*p of U with overlap factor Y. We denote by $\mathcal{O}(\sqcap_\sigma) = \mathcal{O}$ the set of factors overlapping themselves.

Obviously, if $U \in \mathcal{O}$, then

$$|U| + 1 \leq |\hat{U}(Y)| \leq 2|U| - 1, \qquad (*)$$

where Y is any overlap factor of U.

Lemma 9.3.3. *Let $f_n < |U| \leq f_{n+1}$, and $U \neq W_{n+1}$. Then $U \in \mathcal{O}$ if and only if $W_n \not\prec U$.*

Proof. Let $W_n \prec U$ and write $U = SW_nT$. Note that $W_n \notin \mathcal{O}$, if $U \in \mathcal{O}$. Thus any overlap of U must be of the form SW_nVW_nT. By Corollary 9.3.3,

$$|SW_nvW_nT| \geq |S| + |T| + 2f_n + f_{n-1} = |U| + f_{n+1} \geq 2|U|,$$

which is in contradiction with inequality $(*)$.

Now suppose that $W_n \not\prec U$. As in Theorem 9.3.5, we have either

- $U = SF_nT$, where $S, T \neq \varepsilon$, $|S| + |T| \leq f_{n-1}$, $S \triangleright F_n$, $T \triangleleft F_n$;
- or $U \prec F_n^2$.

In the first case, if $|T| = f_n - 1$, then $U = W_{n+1} \notin \mathcal{O}$. Now assume $|T| < f_n - 1$. We can write $F_n = TXS$, because $|S| + |T| \leq f_{n-1}$, with $S \triangleright F_n$, $T \triangleleft F_n$. Since $|T| < f_n - 1$, Theorem 9.3.3.4) implies

$$\left(C_{|T|}(F_n)\right)^3 = (XST)^3 = XSTXSTXST \prec u_\sigma.$$

That is, $U = SF_nT = STXST$ overlaps itself with factor ST.

In the second case, observe that $U \prec F_n^2$ and $|U| > f_n$. Writing $U = ST$, with $|T| = f_n$, we get $T = C_k(F_n)$ for some k, and $S \triangleright T$, so $U = SXS$. On the other hand, since $U = SC_k(F_n) \prec F_n^2$, we have $SXSXS = S\big(C_k(F_n)\big)^2 \prec F_n^3 \prec u_\sigma$. That is, $U = SXS$ overlaps itself with overlap factor S. ∎

Lemma 9.3.4. *If $U \in \mathcal{O}$, then U has a unique overlap factor.*

Proof. Let $f_n < |U| \le f_{n+1}$, and let W be a singular word contained in U with maximal length. By Corollary 9.3.5, W is one of W_{n-1}, W_n and W_{n+1}. By virtue of Lemma 9.3.3, W must be W_{n-1} since $U \in \mathcal{O}$. So we can write $U = SW_{n-1}T$. Now suppose that U admits two different overlap factors. Then W_{n-1} will appear three times in one of these two overlap factors. Since $W_{n-1} \notin \mathcal{O}$, this overlap factor must be of the form $SW_{n-1}V_1W_{n-1}V_2W_{n-1}T$. An analogous argument to the one developed in Lemma 9.3.3 brings us to a contradiction with $(*)$. ∎

The following corollary is immediate from Lemma 9.3.4 and the proof of Lemma 9.3.3.

Corollary 9.3.5. *Let $f_n < |U| \le f_{n+1}$, and assume $U \in \mathcal{O}$. Then $U = VV'V$, where $|V|$ is the overlap length.*

The following theorem summarizes the results above.

Theorem 9.3.6. *Let U be a word such that $f_n < |U| \le f_{n+1}$. Moreover assume that $U \ne W_{n+1}$ and $U \prec u_\sigma$. Then $U \in \mathcal{O}$ if and only if $W_n \not\prec U$. If $U \in \mathcal{O}$, then U has a unique overlap factor V satisfying $U = VV'V$ and $|V| = |U| - f_n$. In particular, $C_k(F_n) \in \mathcal{O}$ if and only if $0 \le k \le f_n - 2$.*

Remark. Note that:
i) $f_{n+1} < 2f_n < f_{n+2} < 3f_n < f_{n+3}$;
ii) for any k, $W_{n+1} \not\prec \big(C_k(F_n)\big)^2$;
iii) for any k, $W_{n+2} \not\prec \big(C_k(F_n)\big)^3$.

The following corollary is an immediate consequence of Theorem 9.3.6.

Corollary 9.3.6. *For any k, $\big(C_k(F_n)\big)^2 \in \mathcal{O}$, $\big(C_k(F_n)\big)^3 \in \mathcal{O}$.*

Remark. If $W^2 \not\prec u_\sigma$ and W has no overlap factor, then two consecutive occurences of w are positively separated. Moreover, we can give for those words a decomposition similar to the one we gave for singular words.

For example, let $W = abab$, by Theorem 9.3.3.3) and Theorem 9.3.6, $W^2 \not\prec u_\sigma$ and $W \notin \mathcal{O}$, so W is positively separated. The following decomposition illustrates the remark above:

$$aba(abab)\underline{aaba}(abab)\underline{a}(abab)\underline{aaba}(abab)\underline{aaba}(abab)\underline{a}(abab)\underline{aaba}(abab)\cdots$$

We totally order $\mathcal{A}_2$ by setting $a < b$ and we extend this order to the set $\mathcal{A}_2^*$ of all words lexicographically. A *Lyndon word* is a word strictly smaller than its proper right factors. Melançon [241] established the factorization of the Fibonacci word into a nonincreasing product of Lyndon words. He also described the intimate links between the Lyndon factorization and the factorization into singular words, and also gave a self-similarity property of the Lyndon factorization. Those results enabled him to give new proofs for Theorems 9.3.1 and 9.3.2. He also introduced generalizations of singular words to the Sturmian words. Using another approach, Cao and Wen [88] generalized singular words and the corresponding factorizations to general Sturmian words and gave some applications to the studies of the factor properties of Sturmian words.

9.4 Properties of factors of the invertible infinite words

In Sec. 9.3 we discussed some factor properties of the Fibonacci word. It is natural to seek for more general invertible substitutions based on the knowledge we developed on the Fibonacci substitution.

Section 9.4.1 is concerned with elementary factor properties. Section 9.4.2 takes a close look at decompositions of the invertible infinite words according to singular words, from which we obtain some applications. Finally, Sec. 9.4.3 is devoted to local isomorphisms.

9.4.1 Elementary properties of factors

From Theorem 9.2.1, we know that any invertible substitution may be written as

$$T := \tau^{n_k} \circ \pi \circ \tau^{n_{k-1}} \circ \pi \circ \cdots \circ \tau^{n_2} \circ \pi \circ \tau^{n_1},$$

where $\tau \in \{\sigma, \varrho\}$, $n_k, n_1 \geq 0$, $n_j \geq 1, 2 \leq j \leq k-1$.

Set

$$\Sigma := \sigma^{n_k} \circ \pi \circ \sigma^{n_{k-1}} \circ \pi \circ \cdots \circ \sigma^{n_2} \circ \pi \circ \sigma^{n_1},$$

where $n_k, n_1 \geq 0$, $n_j \geq 1, 2 \leq j \leq k-1$.

Lemma 9.4.1.2) below gives $M_\Sigma = M_T$, since $M_\sigma = M_\varrho$. Thus by Theorem 9.4.4, the infinite words u_Σ and u_T are locally isomorphic. Thus we do not need to distinguish u_Σ and u_T if we focus on factor properties (see Sec. 9.4.3 below). Consequently, when studying properties of factors of an infinite word generated by an invertible substitution, it is enough to consider factors of Σ.

As a consequence, we only to consider the subclass of Σ with $n_j \geq 1, 1 \leq j \leq k$, we denote by $\mathcal{G}_\sigma$ (the other three cases $n_k \geq 1$, $n_1 = 0$; $n_k = 0$, $n_1 \geq 1$ and $n_k = n_1 = 0$ can be treated in a similar way).

This subsection is structured as follows. After establishing auxiliary lemmas (Lemmas 9.4.1-9.4.3), we discuss the prefix and suffix of $\tau(\alpha)$ ($\tau \in$

$\mathcal{G}_\sigma, \alpha \in \mathcal{A}_2$), the powers of a and b that can appear in $\tau(\alpha)$ (Lemma 9.4.5), the length and primitivity of $\tau(\alpha)$ ((Lemmas 9.4.6 and Lemma 9.4.7). Lemma 9.4.9 establishes relationships between $\tau(a)$ and $\tau(b)$ which play a role analog to Lemma 9.3.1 for the Fibonacci word. The remaining part of this subsection is devoted to the palindrome properties of τ.

The following lemmas can be proved easily, using the definition of the incidence matrix and Lemma 9.2.2.

Lemma 9.4.1. *Let $\tau, \tau_1, \cdots, \tau_n$ be substitutions. Then*

1) $L(\tau(w)) = L(w)M_\tau^t$;

2) let $\tau = \tau_1\tau_2\cdots\tau_n$; then, $M_\tau = M_{\tau_1}M_{\tau_2}\cdots M_{\tau_n}$, where M_τ and M_{τ_j} are the substitutional matrices associated with τ and τ_j, respectively.

3) For any $M \in \mathcal{M}$, there exists $\tau \in \langle \pi, \varrho, \sigma \rangle =$ such that $M_\tau = M$.

Lemma 9.4.2. *For any $W \in \mathcal{A}_2^*$, we have*

1) $(\sigma\pi)^n(a) = a$, $(\sigma\pi)^n(b) = a^n b$; $(\pi\sigma)^n(a) = b^n a$,$(\pi\sigma)^n(b) = b$;

2) $L(\sigma(W)) = (|W|_a + |W|_b, |W|_a)$, $L(\pi(W)) = (|W|_b, |W|_a)$;

3) $L((\sigma\pi)^n(W)) = (|W|_a + n|W|_b, |W|_b)$, $L((\pi\sigma)^n(W)) = (|W|_a, n|W|_a + |W|_b)$.

Let $\alpha, \gamma \in \mathcal{A}_2$. If $\alpha\gamma$ appears in the context, we mean that $\alpha \neq \gamma$, moreover if we write $\alpha\gamma \triangleright \tau(b)$, we assume $|\tau(b)| \geq 2$.

The following lemmas can easily be proved.

Lemma 9.4.3. *Let $\tau \in \mathcal{G}_\sigma$, we have:*

1) If $\displaystyle\sum_{j=1}^{k} n_j + k - 1 \in 2\mathbb{N}+1$, then $ab \triangleright \tau(a)$. If $\displaystyle\sum_{j=1}^{k} n_j + k - 1 \in 2\mathbb{N}$, then $ba \triangleright \tau(a)$;

2) $\tau(b) \triangleleft \tau(a)$, $\tau(a) \triangleleft \tau^n(a)$, $\tau(b) \triangleleft \tau^n(b)$;

3) $$\tau = \begin{cases} (\sigma\pi)^n\sigma & \text{for some } n \text{ if } |\tau(b)| = 1; \\ \sigma(\sigma\pi)^n\sigma & \text{for some } n \text{ if } |\tau(b)| = 2; \\ \sigma & \text{if } |\tau(a)| = 2. \end{cases}$$

4) If $|\tau(b)| \geq 2$ and $\alpha\gamma \triangleright \tau(a)$, then $\beta\alpha \triangleright \tau(b)$.

Lemma 9.4.4. *Let $\tau = (\sigma\pi)^n\sigma^2$ and let $W \in \mathcal{A}_2^*$. Then*

1) $a^{n+1}b \triangleleft \tau(W)$;

2) $a \triangleright W \Rightarrow ba \triangleright \tau(W)$, $b \triangleright W \Rightarrow a^{n+1}b \triangleright \tau(W)$, $ab \triangleright W \Rightarrow a^{n+2}b \triangleright \tau(W)$;

3) The highest power of a appearing in $\tau(W)$ is a^{n+2} if $aab \prec \sigma^2(W)$ and a^{n+1} otherwise; 4) The factor of $\tau(W)$ located between two consecutive occurences of b is either a^{n+1} or a^{n+2}.

Proof. Let $\sigma^2(W) = x_1x_2\cdots x_{m-1}x_m$. Since $\sigma^2(a) = aba, \sigma^2(b) = ab$, it is easy to see that:

(i) $x_1x_2 = ab$;

(ii) $x_{m-1}x_m = ab$ or ba according to $b \triangleright W$ or $a \triangleright W$;

(iii) $a^3 \not\prec \sigma^2(W), b^2 \not\prec \sigma^2(W)$;

(iv) the factor of $\sigma^2(W)$ located between two consecutive occurences of b is either a^2 or a.

Observe that

$$\tau(W) = (\sigma\pi)^n\sigma^2(W) = (\sigma\pi)^n(x_1)(\sigma\pi)^n(x_2)\cdots(\sigma\pi)^n(x_{m-1})(\sigma\pi)^n(x_m).$$

Hence, the conclusions of the lemma follow from the discussion above and Lemma 9.4.3). ∎

Lemma 9.4.5. *Let $\tau \in \mathcal{G}_\sigma$, and let $m(\tau(a))$ (resp. $m(\tau(b))$) be the integer such that $a^{m(\tau(a))}$ (resp. $b^{m(\tau(b))}$) is the highest power of a (resp. b) appearing in $\tau(a)$ (resp. $\tau(b)$). Let $p(\tau(a))$ be $m(\tau(a))$ or $m(\tau(a)) - 1$. Then*

1) $m(\tau(b)) \leq m(\tau(a))$;

2) $a^{p(\tau(a))}b \triangleleft \tau(a)$;

3) either $a^{p(\tau(a))}b \triangleright \tau(a)$, or $ba \triangleright \tau(a)$;

4) $a^{p(\tau(a))+1} \prec F_\tau$, $a^{p(\tau(a))+2} \not\prec F_\tau$, $b^2 \not\prec F_\tau$;

5) any factor of F_τ located between two consecutive occurences of b is either $a^{p(\tau(a))+1}$ or $a^{p(\tau(a))}$.

Proof. Let $\tau = \sigma^{n_k}\pi\sigma^{n_{k-1}}\pi\cdots\sigma^{n_2}\pi\sigma^{n_1}$.

If $n_j = 1$ for all j, $1 \leq j \leq k$, then $\tau = (\sigma\pi)^k\sigma$. This case of the lemma can be proved easily.

Now suppose that there exists at least one index j, such that $n_j > 1$. Let j_0 be the largest integer satisfying that condition. Then

$$\tau = (\sigma\pi)^{k-j_0}\sigma^2(\sigma^{n_{j_0}-2}\cdots\sigma^{n_2}\pi\sigma^{n_1}). \tag{2}$$

Hence the proof of the lemma follows from Lemma 9.4.4 by treating the different cases of (2) (the proof is technical, but not difficult, and we omit the details). ∎

We now determine the length of $\tau(a)$ and $\tau(b)$. Let $\mu, \nu \in \mathbb{N}$. Denote by $F_n(\mu, \nu)$ the integer sequence satisfying the same recurrence as the Fibonacci numbers, but with initial condition $F_{-1}(\mu, \nu) = \nu, F_0(\mu, \nu) = \mu$. Thus for $W \in \mathcal{A}_2^*$, it is readily seen that

$$L(\sigma^n(W)) = (F_n(|W|_a, |W|_b), F_{n-1}(|W|_a, |W|_b)),$$
$$L(\pi(W)) = (|W|_b, |W|_a).$$

As a consequence, it is not difficult to prove the following lemma.

Lemma 9.4.6. *Let $\tau = \sigma^{n_k}\pi\cdots\sigma^{n_2}\pi\sigma^{n_1} \in \mathcal{G}_\sigma$. Then*

$$\begin{aligned} |\tau(a)| &= F_{n_k+1}(\mu_{k-1}, \nu_{k-1}); \\ |\tau(b)| &= F_{n_k}(\mu_{k-1}, \nu_{k-1}); \\ (L(\tau(a))^t, L(\tau(b))^t) &= \begin{pmatrix} F_{n_k}(\mu_{k-1}, \nu_{k-1}) & F_{n_k-1}(\mu_{k-1}, \nu_{k-1}) \\ F_{n_k-1}(\mu_{k-1}, \nu_{k-1}) & F_{n_k-2}(\mu_{k-1}, \nu_{k-1}) \end{pmatrix} = M_\tau, \end{aligned}$$

where $\mu_0 = 1, \nu_0 = 0$, $\mu_{k-1} = F_{n_{k-1}}(\mu_{k-2}, \nu_{k-2})$, $\nu_{k-1} = F_{n_k}(\mu_{k-2}, \nu_{k-2})$.

Lemma 9.4.7. *Let τ be an invertible substitution. For all $\alpha \in \mathcal{A}_2$, the word $\tau(\alpha)$ is primitive.*

Proof. Suppose that $\tau(\alpha)$ is not primitive. Then there exist a word $W \in \mathcal{A}_2^*$ and an integer $p \geq 2$, such that $\tau(\alpha) = W^p$. Hence $L(\tau(\alpha))) = (p|W|_a, p|W|_b)$. Thus $\det M_\tau$ will be divisible by p, which is in contradiction with the fact $\det M_\tau = \pm 1$. ∎

We are now going to establish an important factorization of $\tau \in \mathcal{G}_\sigma$ which will be used in the sequel. The next result is obvious.

Lemma 9.4.8. *If $\tau \in \mathcal{G}_\sigma$, then $\tau(b) \triangleleft \tau(a)$.*

We can write $\tau(a) = \tau(b)U^*$, by virtue of the preceding lemma. Let $U \triangleleft \tau(a)$ with $|U| = |U^*|$. We can also write $\tau(a) = UV$ with $|V| = |\tau(b)|$.

Lemma 9.4.9. *Let $\tau \in \mathcal{G}_\sigma$ with $|\tau(b)| = 2$ and let $\tau(a) = \tau(b)u^* = UV$ as above. If $\alpha\beta \triangleright \tau(a)$, then*
1) $U = U^*$;
2) $\tau(b)\alpha^{-1}\beta^{-1} = V\beta^{-1}\alpha^{-1}$;
3) $\tau(a) = \tau(b)U = U\tau(b)\alpha^{-1}\beta^{-1}\alpha\beta$.

Proof. Since $\alpha\beta \triangleright \tau(a)$, we have $\beta\alpha \triangleright \tau(b)$.

We prove the lemma by induction with respect to the length of τ in terms of $\mathcal{A}_\sigma := \{\sigma, \pi\}$.

1) The inequality $|\tau(b)| = 2$ holds only if τ has length at least 2 and $\tau = \sigma^2$. In this case, $\tau(a) = \sigma^2(a) = aba = \tau(b)a$, $U^* = U = a$, and the statement is true.

Suppose that for $\tau = \tau_n \cdots \tau_1$, we have

$$\tau_n \cdots \tau_1(a) = \tau_n \cdots \tau_1(b)U_n^* = U_n V_n, \tag{3}$$

with $U_n^* = U_n$ (so $|V_n| = |\tau_n \cdots \tau_1(b)|$).

Let $\tau = \tau_{n+1}\tau_n \cdots \tau_1$. Write

$$\tau(a) = \tau_{n+1}\tau_n \cdots \tau_1(a) = \tau_{n+1}\tau_n \cdots \tau_1(b)U_{n+1}^* = U_{n+1}V_{n+1},$$

with $|U_{n+1}^*| = |U_{n+1}|$. We will prove that $U_{n+1}^* = U_{n+1}$.

If $\tau_{n+1} = \pi$, the proof is trivial. Suppose now that $\tau_{n+1} = \sigma$. By virtue if (3), we have:

$$\sigma(\tau_n \cdots \tau_1(a)) = \sigma(U_n V_n) = \sigma(U_n)\sigma(V_n) = \sigma(\tau_n \cdots \tau_1(b))\sigma(U_n^*).$$

Thus $U_{n+1}^* = \sigma(U_n^*)$, $U_{n+1} = \sigma(U_n)$. So we get $U_{n+1}^* = U_{n+1}$, since $U_n^* = U_n$ by induction.

2) The case of $|\tau(b)| = 2$ can be checked directly. Let $\tau = \tau_n \cdots \tau_1$, then by 1) we have

$$\tau_{n+1}\tau_n \cdots \tau_1(a) = \tau_{n+1}\tau_n \cdots \tau_1(b)U_{n+1} = U_{n+1}V_{n+1}.$$

Let $\alpha\beta \triangleright \tau_n \cdots \tau_1(a)$ and suppose that $V_n\beta^{-1}\alpha^{-1} = \tau_n \cdots \tau_1(b)\alpha^{-1}\beta^{-1}$. We shall prove

$$V_{n+1}\alpha^{-1}\beta^{-1} = \tau_{n+1}\tau_n \cdots \tau_1(b)\beta^{-1}\alpha^{-1}.$$

As in 1), it suffices to consider $\tau_{n+1} = \sigma$. Since $\alpha\beta \triangleright \tau_{n+1}\tau_n \cdots \tau_1(b)$, $\beta\alpha \triangleright \sigma(V_{n+1})$, we have

$$\begin{aligned} V_{n+1} = \sigma(V_n) &= \sigma(\tau_n \cdots \tau_1(b)\alpha^{-1}\beta^{-1}\alpha\beta) \\ &= \sigma(\tau_n \cdots \tau_1(b))(\sigma(\beta\alpha))^{-1}\sigma(\alpha\beta). \end{aligned}$$

Notice that $\sigma(\alpha\beta) = a\beta\alpha$ for any $\alpha \neq \beta$, and therefore we obtain

$$V_{n+1}\alpha^{-1}\beta^{-1} = \tau_{n+1}\tau_n \cdots \tau_1(b)\beta^{-1}\alpha^{-1}.$$

3) The conclusion follows immediately from 1) and 2). ∎

Lemma 9.4.10. *Let $W \in \mathcal{P}$, then $\sigma(W)a$, $a^{-1}\sigma(W) \in \mathcal{P}$.*

Proof. Since $W \in \mathcal{P}$, we can write W in the following form:

$$a^{k_1}b^{l_1}a^{k_2}b^{l_2} \cdots a^{k_m}\zeta a^{k_m} \cdots b^{l_2}a^{k_2}b^{l_1}a^{k_1},$$

where $\zeta \in \{a, b, \varepsilon\}$, $k_j, l_j = 1$, if $2 \leq j \leq m-1$, and $k_1, k_m = 0$.

Thus we have

$$\begin{aligned} \sigma(W)a &= (ab)^{k_1}a^{l_1} \cdots (ab)^{k_m}\sigma(\zeta)(ab)^{k_m} \cdots a^{l_2}(ab)^{k_2}a^{l_1}(ab)^{k_1}a \\ &= (ab)^{k_1}a^{l_1} \cdots (ab)^{k_m}\sigma(\zeta)a(ba)^{k_m} \cdots a^{l_2}(ba)^{k_2}a^{l_1}(ba)^{k_1}. \end{aligned}$$

Since $\sigma(\zeta)a$ is equal to a, aba, aa respectively according to ζ being ε, a, b, $\sigma(W)a$ is a palindrome with center a, b, ε respectively by the formula above. ∎

Corollary 9.4.1. *Let $\zeta \in \mathcal{A}_2$. Suppose that $\tau \in \mathcal{G}_\sigma$ with $\beta \triangleright \tau(\zeta)$, then $\alpha\tau(\zeta)\beta^{-1} \in \mathcal{P}$.*

In particular, if $|\tau(\zeta)| \geq 3$, and $\alpha\beta \triangleright \tau(\zeta)$, then $\tau(\zeta)\beta^{-1}\alpha^{-1} \in \mathcal{P}$.

Proof. Let $\tau = \tau_n\tau_{n-1}\cdots\tau_1$.

We prove this corollary by induction as in Lemma 9.4.9. We only prove the case $\zeta = a$, the case of $\zeta = b$ can be proved in a similar way.

The case of $n = 1$ is evident.

Now suppose that $\beta \triangleright \tau_n\tau_{n-1}\cdots\tau_1(a)$ and $\alpha\tau_n\tau_{n-1}\cdots\tau_1(a)\beta^{-1} \in \mathcal{P}$.

If $\tau_{n+1} = \pi$, notice that the word $\pi(\tau_n\tau_{n-1}\cdots\tau_1(a))$ is obtained by exchanging the letters a and b in the word $\tau_n\tau_{n-1}\cdots\tau_1(a)$. So by the induction hypothesis, we get $\beta\pi(\tau_n\tau_{n-1}\cdots\tau_1(a))\alpha^{-1} \in \mathcal{P}$.

Now suppose that $\tau_{n+1} = \sigma$. Notice that $\alpha \triangleright \sigma\tau_n\tau_{n-1}\cdots\tau_1(a)$. Thus

$$\begin{aligned}\beta\sigma\tau_n\tau_{n-1}\cdots\tau_1(a)\alpha^{-1} &= \beta\sigma(\alpha^{-1}\alpha\tau_n\tau_{n-1}\cdots\tau_1(a)\beta^{-1}\beta)\alpha^{-1}\\ &= \beta\sigma(\alpha^{-1})\sigma(\alpha\tau_n\tau_{n-1}\cdots\tau_1(a)\beta^{-1})\sigma(\beta)\alpha^{-1}.\end{aligned}$$

By the induction hypothesis, $\alpha\tau_n\tau_{n-1}\cdots\tau_1(a)\beta^{-1} \in \mathcal{P}$. On the other hand, by a simple calculation, we have either $\beta\sigma(\alpha^{-1}) = a^{-1}$, $\sigma(\beta)\alpha^{-1} = \varepsilon$, or $\beta\sigma(\alpha^{-1}) = \varepsilon$, $\sigma(\beta)\alpha^{-1} = a$. Hence by Lemma 9.4.10, we get

$$\beta\sigma\tau_n\tau_{n-1}\cdots\tau_1(a)\alpha^{-1} \in \mathcal{P}.$$

■

Let $\tau(a) = \tau(b)U$ be the factorization of Lemma 9.4.9. Let $\tau = \sigma^{n_k}\pi\cdots\sigma^{n_2}\pi\sigma^{n_1}$, then $U = \sigma^{n_k}\pi\cdots\sigma^{n_2}\pi\sigma^{n_1-1}(b)$. Thus by Lemma 9.4.3.3), 9.4.9 and Corollary 9.4.1, we get:

Corollary 9.4.2. *Let $\tau \in \mathcal{G}_\sigma$ and let $\alpha\beta \triangleright \tau(a)$.*

1) If $|\tau(b)| = 1$, then $\tau(a) = a^n b$ for some n, i.e., $\tau(a)$ is the product of palindromes a^n and b;

2) If $|\tau(b)| = 2$, then $\tau(a)$ is a palindrome;

3) If $|\tau(b)| > 2$, then $\tau(a)$ is the product of palindromes $\tau(b)\alpha^{-1}\beta^{-1}$ and $\beta\alpha u$.

Lemma 9.4.11. *Suppose that $W \in \mathcal{A}_2^*$ is a primitive word and suppose that $W = U_1U_2$ is a product of two palindromes U_1 and U_2.*

1) If $|U_1|, |U_2| \in 2\mathbb{N}+1$, then for any k, $C_k(W) \notin \mathcal{P}$.

2) If $|U_1|, U_2| \in 2\mathbb{N}$, then $C_{|U_1|/2}(W) \in \mathcal{P}$, $C_{|U_1|+|U_2|/2}(W) \in \mathcal{P}$. For other k, $C_k(W) \notin \mathcal{P}$;

3) If $|U_1| \in 2\mathbb{N}$, $|U_2| \in 2\mathbb{N}+1$ (or $|U_1| \in 2\mathbb{N}+1$, $|U_2| \in 2\mathbb{N}$), then $C_{|U_1|/2}(W) \in \mathcal{P}$ (or $C_{|U_1|+|U_2|/2}(W) \in \mathcal{P}$). For other k, $C_k(W) \notin \mathcal{P}$.

Proof. We only prove 1), the other cases can be discussed in the same way.

Notice that at least one of $|U_1|$, $|U_2|$ is larger than 2 (otherwise U_1, U_2 will be a^2 or b^2). Suppose that without loosing generality $|U_1| = 3$. Then we can write $U_1 = xU_1'x$, where $x \in \mathcal{A}_2$, $U_1' \in \mathcal{P}$. Thus $C_1(W) = C_1(U_1U_2) = (U_1')(xU_2x)$ is a product of two palindromes, and $C_1(U_1U_2) \notin \mathcal{P}$. Since $|U_1'|, |xU_2x| \in 2\mathbb{N}+1$ and $U_1', xU_2x \in \mathcal{P}$, we can repeat the discussion above which follows the proof. ■

Lemma 9.4.12. *Let $\tau \in \mathcal{G}_\sigma$. Suppose that $|\tau(a)| \in 2\mathbb{N}$. If $\tau(a) = U_1U_2$ is a product of two palindromes, then $|U_1|, |U_2| \in 2\mathbb{N}+1$.*

Proof. If $|U_1|, |U_2| \in 2\mathbb{N}$, then $|U_1|_a + |U_1|_b$, $|U_2|_a + |U_2|_b \in 2\mathbb{N}$. Thus both components of $L(\tau(a))$ are even which follows that $|\det M_\tau| \in 2\mathbb{N}$. But we know that $\det M_\tau$ must be ± 1. ■

From Corollary 9.4.2, Lemma 9.4.11, Lemma 9.4.12 and Proposition 9, 10, 11 of [109], we obtain

Proposition 9.4.1. *Let $\tau \in \mathcal{G}_\sigma$.*
1) Any conjugate of $\tau(a)$ is primitive.
2) $|C(\tau(a))| = |\tau(a)|$. That is, all conjugates of $\tau(a)$ are distinct.
3) Any conjugation of $\tau(a)$ (containing $\tau(a)$ itself) is either a palindrome, or a product of two palindromes. In the later case, the factorization is unique.
4) $C(\tau(a)) = \overline{C(\tau(a))}$, where $\overline{C(\tau(a))} = \{\overline{W};\ W \in C(\tau(a))\}$.
5) If $|\tau(a)| \in 2\mathbb{N}$, there is no palindrome in $C(\tau(a))$; if $|\tau(a)| \in 2\mathbb{N}+1$, there is only one palindrome in $C(\tau(a))$.

9.4.2 Decomposition into singular words and its application

Similarly as in the case of the Fibonacci infinite word, we will introduce singular words for the invertible substitutions in $\mathcal{G}_\sigma$ and study their properties.

Let $\tau \in \mathcal{G}_\sigma$ and let $F_\tau = \tau(F_\tau) = x_1x_2\cdots x_k\cdots$ be the fixed point of τ. Then for any $n \in \mathbb{N}$, we have

$$F_\tau = \tau^n(F_\tau) = \tau^n(x_1)\tau^n(x_2)\cdots\tau^n(x_k)\cdots \tag{4}$$

We now discuss the properties of the factors of the infinite word F_τ.

We denote by $A_n = \tau^n(a)$, $B_n = \tau^n(b)$.

Notice that if $\tau \in \mathcal{G}_\sigma$, then for any $n \geq 2$, $\tau^n \in \mathcal{G}_\sigma$. Therefore the conclusions about τ which we have obtained above hold also for τ^n.

Let W be a factor of F_τ of length $|A_n|$. By (4), W will be contained in one of the following four words: A_nA_n, A_nB_n, B_nA_n, $A_nB_nA_n$. By Lemma 9.4.3.2), $B_n \triangleleft A_n$, so $A_nB_n \prec A_nA_n$.

Let $W \prec A_nB_nA_n$, write $W = W_1B_nW_2$. Then $W_1 \triangleright A_n, W_2 \triangleleft A_n$. Let $A_n = B_nU_n$ be the factorization as in Lemma 9.4.9. Then $U_n \triangleleft A_n$ by Lemma 9.4.9.3). Since $|W| = |A_n| = |B_n| + |U_n|$, $|W_2| < |U_n|$. Consequently $W_2 \triangleleft U_n$, and hence $W = W_1B_nW_2 \prec A_nB_nU_n = A_nA_n$.

By the discussions above, any factor of length $|A_n|$ of F_τ will be contained either in A_nA_n or B_nA_n which we are going to study respectively.

Lemma 9.4.13. *Let $W \prec A_nA_n$ with $|W| = |A_n|$, then $W \in C(A_n)$.*

Proof. Since $W \prec A_nA_n$, and $|W| = |A_n|$, we can write $W = UV$ with $U \triangleright A_n, V \triangleleft A_n$, and $|U| + |V| = |A_n|$, which follows that $A_n = VU$. That is $W = C_{|V|}(A_n)$. ■

Since all conjugates of A_n are distinct by Proposition 9.4.1.2), we get

Corollary 9.4.3. *The set of the factors of length $|A_n|$ of A_n^k, $k = 2$ is exactly $C(A_n)$*

Now we study the factors of the word B_nA_n. For this aim, we introduce *singular words* of F_τ as follows:

Let $\alpha \triangleright A_n$ (so $\beta \triangleright B_n$). Then the n-th singular words with respect a or b of F_τ are defined respectively as:

$$W_{n,a} = \beta A_n \alpha^{-1}, \quad W_{n,b} = \alpha B_n \beta^{-1}.$$

From the definition of the singular words, we get immediately

Lemma 9.4.14. *With the notations above, we have*

$$\begin{pmatrix} L(W_{n,a}) \\ L(W_{n,b}) \end{pmatrix} = \begin{cases} \begin{pmatrix} |A_n|_a - 1 & |A_n|_b + 1 \\ |B_n|_a + 1 & |B_n|_b - 1 \end{pmatrix} & \textit{if } \alpha = a; \\ \begin{pmatrix} |A_n|_a + 1 & |A_n|_b - 1 \\ |B_n|_a - 1 & |B_n|_b + 1 \end{pmatrix} & \textit{if } \alpha = b \end{cases}$$

Since $L(w) = L(A_n)$ for any $W \in C(A_n)$, we get from Lemma 9.4.14,

Corollary 9.4.4. $\forall n \in \mathbb{N},\ W_{n,a} \notin C(A_n),\ W_{n,b} \notin C(B_n)$.

Let $\alpha\beta \triangleright A_n$ (so $\beta\alpha \triangleright B_n$). By Lemma 9.4.9.3),

$$B_nA_n = B_nu_nB_n\alpha^{-1}\beta^{-1}\alpha\beta = A_nB_n\alpha^{-1}\beta^{-1}\alpha\beta.$$

Since $A_nB_n\alpha^{-1}\beta^{-1} \prec A_nB_n \prec A_nA_n$, we see that the first $|B_n| - 2$ factors of length $|A_n|$ of B_nA_n are those of A_nA_n which are distinct from each other by Corollary 9.4.3, and the $(|B_n| - 1)$-th factor is exactly $W_{n,a}$ which is not in $C(A_n)$ by Corollary 9.4.4. We get from the analysis above

Lemma 9.4.15. *Let $W \prec B_nA_n$ with $|W| = |A_n|$. Then W is either in $C(A_n)$ or equal to $W_{n,a}$. In particular, as a factor, $W_{n,a}$ appears only once in B_nA_n.*

From Proposition 9.4.1.4), Lemma 9.4.13, Corollary 9.4.3, and Lemma 9.4.15, we obtain

Proposition 9.4.2. *Let $\tau \in \mathcal{G},\ n \in \mathbb{N}$. Then*

1) $\mathcal{L}_{|A_n|} = C(A_n(a)) \bigcup \{W_{n,a}\}, \quad \mathcal{L}_{|B_n|} = C(B_n(b)) \bigcup \{W_{n,b}\}$;

2) $|\mathcal{L}_{|A_n|}| = |A_n| + 1, \quad |\mathcal{L}_{|B_n|}| = |B_n| + 1$;

3) $\mathcal{L}_{|A_n|} = \overline{\mathcal{L}_{|A_n|}}, \quad \mathcal{L}_{|B_n|} = \overline{\mathcal{L}_{|B_n|}}$

Remark. Let τ_1 and τ_2 be two invertible substitutions. It results from Proposition 9.4.2-3) and the definitions of local isomorphisms that F_{τ_1} and F_{τ_2} are locally isomorphic if and only if $\mathcal{L}(F_{\tau_1}) = \mathcal{L}(F_{\tau_2})$.

We now discuss the properties of the singular words.

Proposition 9.4.3. *$\forall n, m \in \mathbb{N}$, $W_{n,a}, W_{n,b}$ and $\beta A_n^m B_n \alpha^{-1} \in \mathcal{P}$.*

Proof. The fact that $W_{n,a}, W_{n,b} \in \mathcal{P}$ follows immediately from the definition of the singular words and Proposition 9.4.1.

Let $A_n = B_n U_n$ be the factorization as in Lemma 9.4.9, then for any $m \geq 1$,

$$\beta A_n^m B_n \alpha^{-1} = ((\beta B_n \alpha^{-1})(\alpha U_n \beta^{-1}))^m (\beta B_n \alpha^{-1}),$$

by Corollary 9.4.2, $\beta B_n \alpha^{-1}, \alpha u_n \beta^{-1} \in \mathcal{P}$. Thus $\beta A_n^m B_n \alpha^{-1} \in \mathcal{P}$. ∎

Lemma 9.4.16. *$W_{n,a} \triangleright W_{n+2,a}$, $W_{n,b} \triangleright W_{n+2,b}$.*

Proof. Let $\alpha\beta \triangleright \tau^n(a)$, then $\alpha\beta \triangleright \tau^{n+2}(a)$ by Lemma 9.4.3.3), thus

$$W_{n+2,a} = \alpha \tau^{n+2}(a) \beta^{-1} = \alpha \tau^n(\tau^2(a)) \beta^{-1}.$$

By Lemma 9.4.3.1), $a \triangleleft \tau^2(a)$, so $\tau^2(a) = ua$ for some $u \in \mathcal{A}_2$. Thus $W_{n+2,a} = \alpha \tau^n(u) \tau^n(a) \beta^{-1}$ which yields the conclusion of the lemma. ∎

Lemma 9.4.17. *Either $a^{p(\tau(a))+1} \triangleright W_{n,a}$, $a^{p(\tau(a))+1} \triangleleft W_{n,a}$, or $b \triangleright W_{n,a}$, $b \triangleleft W_{n,a}$, where $p(\tau(a))$ is defined as in Lemma 9.4.5.2).*

Proof. This follows from Lemma 9.4.5.2), Proposition 9.4.2 and the definition of $W_{n,a}$. ∎

Proposition 9.4.4. *$\forall n \in \mathbb{N}$, if $W_{n,a}$ is not a power of a, then $W_{n,a}$ is primitive.*

Proof. Suppose that $W_{n,a}$ is not a power of a. If $W_{n,a} = w^p$ for some $w \in \mathcal{A}_2^*$ and $p = 2$, then $W_{n,a}$ will be equal to one of its conjugates. But by Lemma 9.4.17, we will then have either $b^2 \prec F_\tau$ or $a^{2p(\tau(a))+2} \prec F_\tau$. This is impossible because of Lemma 9.4.4.4. ∎

Let $F_\infty = x_1 \cdots x_n \cdots$ be an infinite word over $\mathcal{A}_2$. Let $W \prec F_\infty$ be a word which appears infinitely many times in F_∞. If any two adjacent $W's$ (i.e., two successive occurrences) appearing in F_∞ are separated by a factor of F_∞ (this assertion is equivalent to $\nexists U \prec F_\infty$ such that $U = XYZ$ with $W = XY = YZ$), we say that the word W possesses the *positive separation property*, and the factor is called a *separate factor* (with respect to W). If any separate factor is not equal to ε, we say that W is *strongly separated* (otherwise, we say that W is *weakly separated*).

Definition 9.4.1. *A word $W \in \mathcal{A}_2^*$ of the form $W = (UV)^k U$, where $k > 0$, UV is a primitive word, is called a* sesquipower. *The positive integer k is said to be the* order *of the sesquipower. A sesquipower of order larger than 1 is called a* strong sesquipower.

From Lemma 9.4.5, Lemma 9.4.13, Lemma 9.4.15, Formula (4) and the definition of $W_{n,a}$, we obtain the following

Theorem 9.4.1. *Let $\tau \in \mathcal{G}_\sigma$ and let $\alpha\beta \triangleright A_n$. For $n \geq 1$, we have*

$$F_\tau = W_0 W_{n,a} z_1^{(n)} W_{n,a} z_2^{(n)} \cdots z_k^{(n)} W_{n,a} z_{k+1}^{(n)} \cdots ,$$

where $W_0 = A_n^{p(\tau(a))} B_n \alpha^{-1}$, z_j is either equal to $\beta A_n^{p(\tau(a))-1} B_n \alpha^{-1}$ or to $\beta A_n^{p(\tau(a))} B_n \alpha^{-1}$, and $p(\tau(a))$ is defined as in Lemma 9.4.4.2). We call above decomposition of F_τ the $n-$decomposition of F_τ,

Denote by $F_\tau(z^{(n)})$ the infinite word $z_1^{(n)} z_2^{(n)} \ldots z_k^{(n)} \ldots$.

From the discussions above and Proposition 9.4.3, we obtain

Theorem 9.4.2. *Let $\tau \in \mathcal{G}_\sigma$, then for any n,*

1) $W_{n,a}$ is strongly separated;

2) The separate factor is either $\beta A_n^{p(\tau(a))-1} B_n \alpha^{-1}$, or $\beta A_n^{p(\tau(a))} B_n \alpha^{-1}$;

3) The separate factors are sesquipower palindromes of order $p(\tau(a))$ and $p(\tau(a)) - 1$ respectively.

Similarly as in the Fibonacci word, the right special factors of an infinite word F_∞ are defined as follows: if $Ua, Ub \prec F_\infty$, then the word U is called a right special factor of F_∞. Since $\{W, W \prec F_\tau\} = \{\overline{W}, W \prec F_\tau\}$, the right special factors of F_τ are equivalent to $aU, bU \prec F_\tau$.

From Corollary 9.4.3, Lemma 9.4.15, Theorem 9.4.2, by an analogous argument as that of Example 3 of Sec. 9.3, we get

Theorem 9.4.3. *Let $\tau \in \mathcal{G}_\sigma$. Then for any $n \geq 1$, there is a unique right special factor of length n. Moreover W is a right special factor if and only if $W \triangleleft F_\tau$.*

Example. Let $\tau = (\sigma\pi)^{n-1}\sigma$, then

$$\tau = (a^n b, a), \quad p(\tau(a)) = n.$$

Consider $1-$decomposition of F_τ, we have

$$ab \triangleright A_1^n = (a^n b)^n, \quad W_0 = (A_1)^n B_1 a^{-1} = (a^n b)^n, \quad W_{1,a} = a(a^n b) b^{-1} = a^{n+1}.$$

Put

$$A = b A_1^n B_1 a^{-1} = b(a^n b)^n a a^{-1} = b(a^n b)^{n-1}, \quad B = b(a^n b)^n.$$

When $n = 1$, $\tau = \sigma$, then $\forall m \in \mathbb{N}$, $F_\sigma(z^{(m)}) = F_\sigma$ over $\{A, B\}$;
When $n = 2$, $\tau = \sigma\pi\sigma$, then

$$F_{\sigma\pi\sigma}(z^{(1)}) = ABABAABABAABABABAABABAABABABAABA\ldots$$

We see that $F_{\sigma\pi\sigma}(z) \neq F_{\sigma\pi\sigma}$ over $\{A, B\}$, A calculation on computer shows that

$$F_{\sigma\pi\sigma}(z^{(1)}) = F_\tau = F_\tau(z^{(1)})$$

where $\tau = (ab, aba) = \sigma\sigma\pi$. ■

Remark. As said in the introduction, the motivation of this section is to generalize the properties of the factors of the Fibonacci word to the case of all invertible substitutions. By comparing with the results of Sec. 9.3, we see that one main difference is the structure of the singular words and the singular decomposition of the infinite word according to the singular words. In fact, the structure of the singular words of a general invertible substitution is much more complicated than that of the Fibonacci substitution (for example, all singular words and separate words of the Fibonacci word are sesquipowers of order 1, i.e, weak sesquipowers. On the other hand, those of a general invertible substitution are strong sesquipowers). But we have seen that the singular words still play an important role in the studies of the factors, such as power of factors, overlap of factors and local isomorphism. On the other hand, singular words and their applications can be generalized to the case of Sturmian words [88].

9.4.3 Local isomorphism of invertible substitutions

By the definition of local isomorphism, we see that if two infinite sequence u and v are locally isomorphic, then u and v have the same set of factors, so if we are interested in the factor properties only, we need not to distinguish u and v.

The aim of this section is to give a sufficient and necessary condition for local isomorphism of the sequences generated by invertible substitution.

The following lemma establishes the relation between two invertible substitution and their substitutive matrices.

Lemma 9.4.18 ([277]). *Let τ_1, τ_2 be two invertible substitutions, then the following two assertions are equivalent:*
1) $M_{\tau_1} = M_{\tau_2}$;
2) there exists a word $W \in \mathcal{A}_2^*$, such that $\tau_1 = i_W\tau_2$ (or $\tau_1 = i_{W^{-1}}\tau_2$).

Proposition 9.4.5. *Let τ_1, τ_2 be two substitutions and let F_{τ_1}, F_{τ_2} be the fixed points of τ_1 and τ_2 respectively. If there exist $m, n \in \mathbb{N}$, such that $\tau_1^m \sim \tau_2^n$, then $F_{\tau_1} \simeq F_{\tau_2}$.*

Proof. It is clear that $F_{\tau^n} \simeq F_\tau$ for any substitution and $n \in \mathbb{N}$. Thus we prove only that $\tau_1 \sim \tau_2 \Longrightarrow F_{\tau_1} \simeq F_{\tau_2}$. It is easy to check that

$$\tau i_W = i_{\tau(W)}\tau, \quad W \in \mathcal{A}_2^* \tag{1}$$

If $\tau_1 = i_W\tau_2$, then from (1), by induction, we have

$$\tau_1^n = (i_W\tau_2)^n = i_W i_{\tau_2(W)} \cdots i_{\tau_2^{n-1}(W)}\tau_2^n \quad \text{for some } W \in \mathcal{A}_2^*.$$

Let $W_n = W\tau_2(W)\cdots\tau_2^{n-1}(W)$. We have $\tau_1^n = i_{W_n}\tau_2^n$, $\tau_1^n = W_n\tau_2^n W_n^{-1}$, that is $\tau_1(a)^n W_n = W_n\tau_2^n(a)$. Thus $\tau_1^n(a)$ and $\tau_2^n(a)$ are conjugate, that is, there exist U_n and V_n, such that $\tau_1^n(a) = U_nV_n$, and $\tau_2^n(a) = V_nU_n$. This means that, both U_n and V_n are factors of $\tau_1^n(a)$ and $\tau_2^n(a)$. Since $|\tau_1^n(a)|$ tends to infinity, so does $|U_n|$ (or $|V_n|$). Let U be a factor of F_{τ_1}, $U \prec U_n$ (or $W \prec V_n$) for n large enough, thus $U \prec F_{\tau_2}$. Inversely, by the same argument, if $U \prec F_{\tau_2}$, then $U \prec F_{\tau_1}$. We thus prove that $F_{\tau_1} \simeq F_{\tau_2}$.

The case of $\tau_1 = i_{W^{-1}}\tau_2$ can be treated in the same manner. ■

From this proposition and Lemma 9.4.18, we have immediately

Proposition 9.4.6. *Let τ_1, τ_2 be two invertible substitutions and let F_{τ_1}, F_{τ_2} be the fixed points of τ_1 and τ_2 respectively. If $M_{\tau_1} = M_{\tau_2}$, then $F_{\tau_1} \simeq F_{\tau_2}$.*

Now, by using Lemma 9.4.1.3), we can prove the following

Theorem 9.4.4. *Let τ_1, τ_2 be two invertible substitutions and let F_{τ_1}, F_{τ_2} be the fixed points of τ_1 and τ_2 respectively. If there exists a matrix M, such that $M_{\tau_1} = M^m$, $M_{\tau_2} = M^n$, $m, n \in \mathbb{N}$, then $F_{\tau_1} \simeq F_{\tau_2}$.*

Proof. By Lemma 9.4.1.3, there exists an invertible substitution τ, such that $M_\tau = M$. From Proposition 9.4.6, we have

$$F_{\tau_1} \simeq F_{\tau_1^n} \simeq F_{\tau^n} \simeq F_{\tau^m} \simeq F_{\tau_2^m} \simeq F_{\tau_2}. ■$$

Theorem 9.4.5. *Let τ_1, τ_2 be two invertible substitutions. The following assertions are equivalent:*

(1) $M_{\tau_1} = M_{\tau_2}$,

(2) there exists $W \in \mathcal{A}_2^*$ such that $\tau_1 = i_W \circ \tau_2$, or $\tau_2 = i_w\tau_1$,

(3) F_{τ_1} and F_{τ_2} are locally isomorphic, and $|\tau_1(a)| = |\tau_2(a)|$, $|\tau_1(b)| = |\tau_2(b)|$.

Proof. From Theorem 9.4.4, (1) implies (3).

Let us now show that assertion (3) implies the other ones. By the remark made after Proposition 9.4.2, we have $\mathcal{L}_{|\tau_1(a)|}(F_{\tau_1}) = \mathcal{L}_{|\tau_2(a)|}(F_{\tau_2})$. Hence $\tau_2(a) \in \mathcal{L}_{|\tau_1(a)|}(F_{\tau_1})$. By Proposition 9.4.2, we have either $\tau_2(a) \in C(\tau_1(a))$ or $\tau_2(a) = W_{1,a}$. Let W be one conjugate of $\tau_2(a)$ distinct from $\tau_2(a)$ (this possible by Proposition 9.4.2).

We have $L(\tau_2(a)) = L(W)$. If $\tau_2(a) = W_{1,a}$, then by Proposition 9.4.2-1), $W \in C(\tau_1(a))$. Therefore, by Proposition 9.4.2-2), $L(\tau_2(a)) \neq L(W)$, which the previous discussion has shown to be impossible.

Therefore we necessarily have $\tau_2(a) \in C(\tau_1(a))$. It follows that $L(\tau_1(a)) = L(\tau_2(a))$. In the same way, we also have $L(\tau_1(b)) = L(\tau_2(b))$. This proves $M_{\tau_1} = M_{\tau_2}$. ∎

Remark. By introducing *patterns* in high dimensional space (see for example [21, 272]), we can generalize the definition of local isomorphism to the colored lattice, that are applicable in the studies of quasicrystals [380]. But it is a difficult problem to characterize local isomorphism for the general case. In fact, in Theorem 9.4.4, we only deal with the invertible substitutions over the alphabet of two letters. We do not know any results for the general substitutions (which are non invertible) over $\mathcal{A}_2$. For the alphabet over more than two letters, the question is still open even for invertible substitutions.

We mention here two results related to Theorem 9.4.4.

- The condition of Theorem 9.4.4 can be relaxed further as follows [371]:

  Theorem 9.4.6. *Let τ_1, τ_2 be two primitive invertible substitutions over $\mathcal{A}_2$. The following assertions are equivalent:*
  (1) There exist a primitive invertible substitution τ_0 and positive integers $m, n \in \mathbb{N}$, such that
  $$M_{\tau_1} = M_{\tau_0}^m, \quad M_{\tau_2} = M_{\tau_0}^n;$$
  (2) F_{τ_1} and F_{τ_2} are locally isomorphic.

  In [130], Durand has proved the following result:
  If two primitive substitutions over $\mathcal{A}_n$ have the same nonperiodic fixed point, then they have some powers which have the same eigenvalues.
- Let τ is a substitution, we denote by Φ_τ the trace mapping induced from τ. Let $\tau_1, \tau_2 \in \mathrm{Aut}(\mathcal{A}_2^*)$, then F_{τ_1} and F_{τ_2} are locally isomorphic if and only $\Phi_{\tau_1} = \Phi_{\tau_2}$, [277]. For the definition and properties of trace mapping, see also [277] and Chap. 10.

Acknowledgements. The author would like to thank Guy Mélançon for his careful reading of the original manuscript, and for his modifications to make it suitable as a chapter of this book.