SUBSTITUTIONS, CODING PRESCRIPTIONS AND INTEGER REPRESENTATION

PAUL SURER

Abstract. Coding prescriptions are combinatorial objects linked to a substitution, that is a morphism of the free monoid. Originally they have been introduced in order to code the induced symbolic dynamical systems. In the present article we are interested in coding prescriptions of compositions and powers of substitutions. This will provide a a very general framework for representing integers. We will study their properties and find several relations with well-known systems of integer numeration.

1. Introduction

We are concerned with combinatorial aspects of morphisms of the free monoid, so-called substitutions. More precisely, we study coding prescriptions. Quickly explained, for a substitution σ over an alphabet $\mathcal{A}$, a coding prescription with respect to σ is given by a complete residue system modulo the length of $\sigma(x)$ for each letter x contained in the alphabet $\mathcal{A}$. These combinatorial objects have been introduced in [8] in order to relate the dynamical system induced by a primitive substitution with shifts of finite type.

In the actual research we associate a finite directed graph with a coding prescription. This will allow us to define a binary operation for coding prescriptions that is compatible with the composition of substitutions. It will turn out that the finite paths of a given length n of this graph correspond to coding prescriptions with respect to the nth powers of the substitution. In the principal result of the article we use the theory to represent integers by finite paths on the associated graph generalising the result of Dumont and Thomas in [1].

The main ingredient of our paper is combinatorics on words. Contrarily to related articles as [1] (that focus on the free monoid over $\mathcal{A}$) we consider the free group. The inverse letters that appear in this context correspond to negative integers. In this way we achieve a much more general result. In particular the set of representable integers usually contains negative as well as positive integers. The exact shape strongly depends on the choice of the coding prescription.

In the easiest case, when the coding prescription is given by sets of consecutive integers (we call such coding prescriptions continuous), we either obtain the representation of Dumont-Thomas or we can represent the entire set $\mathbb{Z}$. In this case we will also see that the canonical order of the integers corresponds to an ordering on the paths. For other coding prescriptions this equivalence of orderings does not hold in general, in fact, for such non-continuous coding prescriptions the representations may have quite unusual properties.

Following [1, 2] the integer representation presented by Dumont and Thomas covers several generalised integer numeration systems as Zeckendorf's Fibonacci expansion [9] or Rauzy's Tribonacci expansion [7], which are actually special cases of digit representations with respect to linear recurrences (also known as G-ary representations, see [3, 5]). Our result provides possibilities to treat these issues in a more general way. In particular, it will turn out that our setting does not only cover the cited numeration systems but also the *negaFibonacci expansion* presented recently by Knuth in part four of its famous monograph "The art of computer programming" [4]. Furthermore, we will see by some examples how to generalise the concept of G-ary expansions.

2010 *Mathematics Subject Classification.* 11A63, 68R15, 11B39.

Key words and phrases. Sustitutions, integer representation, combinatorics on words.

The research was supported by the Austrian Research Foundation (FWF), Project P28991-N35.

The article is organised as follows. In Section 2 we carefully introduce all necessary formalisms and state the main results. In Section 3 we are concerned with combinatorial aspects of coding prescriptions. We are interested in how to obtain coding prescriptions with respect to composition and powers of substitutions and study their behaviour. In Section 4 we prove the results concerning integers representation. This also involves the investigation of the structure of the set of representable integers. In Section 5 we discuss through several examples the relation of our research with existing results, as the integer numeration of Dumont-Thomas [1], the *negaFibonacci expansion* and representations with respect to linear recurrences.

2. Definitions, notations and main results

2.1. **Free monoids and free groups.** Throughout the paper we let $\mathcal{A} = \{1, \ldots, m\}$ denote a finite set (alphabet) and call its elements the letters of $\mathcal{A}$. The free monoid generated by $\mathcal{A}$ is the set of finite words $\mathcal{A}^*$ together with the concatenation of words. The neutral element is the empty word ε. Furthermore we define the set of non-empty words $\mathcal{A}^+ := \mathcal{A}^* \setminus \{\varepsilon\}$.

For our considerations we extend our alphabet by the set of "inverse letters" $\overline{\mathcal{A}} := \{\overline{x} : x \in \mathcal{A}\}$. Analogously, let $\overline{\mathcal{A}}^*$ ($\overline{\mathcal{A}}^+$, respectively) denote the set of finite (non-empty, respectively) words over $\overline{\mathcal{A}}$, and $(\mathcal{A} \cup \overline{\mathcal{A}})^*$ is the set of finite words over the extended alphabet $\mathcal{A} \cup \overline{\mathcal{A}}$. Define by $\sim$ the equivalence relation on $(\mathcal{A} \cup \overline{\mathcal{A}})^*$ induced by the cancellation of letters, that is $x\overline{x} \sim \varepsilon \sim \overline{x}x$ for each $x \in \mathcal{A}$. Then $(\mathcal{A} \cup \overline{\mathcal{A}})^* / \sim$ is the free group generated by $\mathcal{A}$.

For a word $X \in (\mathcal{A} \cup \overline{\mathcal{A}})^*$ and a set $\mathfrak{X} \subset (\mathcal{A} \cup \overline{\mathcal{A}})^*$ we write $X \in_\sim \mathfrak{X}$ if X is contained modulo $\sim$ in $\mathfrak{X}$. The monoids $\mathcal{A}^*$ as well as $\overline{\mathcal{A}^*}$ are embedded in a natural way in it the free group $(\mathcal{A} \cup \overline{\mathcal{A}})^* / \sim$ and $\mathcal{A}^* \cap \overline{\mathcal{A}}^* = \{\varepsilon\}$. Therefore, $X \in_\sim \mathcal{A}^* \cap \overline{\mathcal{A}}^*$ if and only if $X \sim \varepsilon$. Actually, all the words that appear in this article will eventually turn out to be contained in $\mathcal{A}^* \cup \overline{\mathcal{A}}^*$ (modulo $\sim$).

For a word $X = x_1 \cdots x_n \in (\mathcal{A} \cup \overline{\mathcal{A}})^*$ its inverse modulo $\sim$ is given by

$$\overline{X} := \overline{x_n}\, \overline{x_{n-1}} \cdots \overline{x_1}$$

where $\overline{\overline{x}} = x$ for all $x \in \mathcal{A}$. If $X \sim PS$ (with $P, S \in (\mathcal{A} \cup \overline{\mathcal{A}})^*$) then we obviously have $\overline{X} \sim \overline{PS} \sim \overline{S}\,\overline{P}$. For each word $X \in (\mathcal{A} \cup \overline{\mathcal{A}})^*$ with $X \in_\sim \mathcal{A}^*$ we clearly have $\overline{X} \in_\sim \overline{\mathcal{A}}^*$ and vice versa.

We adapt several known combinatorial notations to our setting. For $y \in \mathcal{A}$ and $X = x_1, \ldots, x_n \in (\mathcal{A} \cup \overline{\mathcal{A}})^*$ we define $|X|_y$ to be the difference of the number of occurrences of y and the number of occurrences of $\bar{y}$ in X, *i.e.*

$$|X|_y := \#\{j \in \{1, \ldots, n\} : x_j = y\} - \#\{j \in \{1, \ldots, n\} : x_j = \overline{y}\},$$

and

$$|X| := \sum_{y \in \mathcal{A}} |X|_y.$$

Observe that these definitions are compatible with $\sim$ and behave additively with respect to the concatenation of words, that is for all $X, X', Y \in (\mathcal{A} \cup \overline{\mathcal{A}})^*$ with $X \sim X'$ and $y \in \mathcal{A}$ we have $|XY|_y = |X|_y + |Y|_y$ and $|X|_y = |X'|_y$. This immediately implies that $|XY| = |X| + |Y|$, $|X| = |X'|$ also hold.

We define the partial ordering $\leq$ and the corresponding strict partial ordering $<$ on $(\mathcal{A} \cup \overline{\mathcal{A}})^*$ by

$$X \leq Y \iff \overline{X}Y \in_\sim \mathcal{A}^*, \quad X < Y \iff \overline{X}Y \in_\sim \mathcal{A}^+.$$

One easily verifies that the definition is compatible with $\sim$ and, hence, $\leq$ can be transferred to a partial ordering on the free group $(\mathcal{A} \cup \overline{\mathcal{A}})^* / \sim$. Observe that $|\cdot|$ is an order-preserving map from the partially ordered set $((\mathcal{A} \cup \overline{\mathcal{A}})^* / \sim, \leq)$ onto the totally ordered set $(\mathbb{Z}, \leq)$.

For two words $Y, Y' \in (\mathcal{A} \cup \overline{\mathcal{A}})^*$ the set

$$\{X \in (\mathcal{A} \cup \overline{\mathcal{A}})^* / \sim : Y \leq X \leq Y'\}$$

is a $\leq$-chain, that is a totally ordered subset, which we will refer to as the Y-Y'-chain. We can map it bijectively (and order-preservingly) via $|\cdot|$ onto the $\leq$-chain $\{|Y|, \ldots, |Y'|\}$. Such chains are of great importance in this article.

As already mentioned, the words that appear in this paper are all contained in $\mathcal{A}^* \cup \overline{\mathcal{A}}^*$ (modulo $\sim$). Note that for such words the relation $\preceq$ corresponds to the "factor"-property. Indeed, suppose that $X, Y \in \mathcal{A}^*$. Then $X \preceq Y$ means that there exists a word $S \in \mathcal{A}^*$ such that $XS \sim Y$, thus, X is a left factor of Y. The term $X \prec Y$ signifies that X is a proper left factor of Y. If $X, Y \in \overline{\mathcal{A}}^*$ then $\overline{X}, \overline{Y} \in \mathcal{A}^*$ and $X \preceq Y$ means that $\overline{X} = P\overline{Y}$ for a $P \in \mathcal{A}^*$, hence, $\overline{Y}$ is a right factor of $\overline{X}$. For $X \in_\sim \overline{\mathcal{A}}^*$ and $Y \in_\sim \mathcal{A}^*$ we always have $X \preceq \varepsilon \preceq Y$, hence, the X-Y-chain includes the empty word.

2.2. **Substitutions and coding prescriptions.** Consider an endomorphism $\sigma : \mathcal{A}^* \longrightarrow \mathcal{A}^*$ and note that σ is uniquely determined by the values $\sigma(x)$ for each $x \in \mathcal{A}$. We call σ a substitution (over $\mathcal{A}$) if

(1) $\forall x \in \mathcal{A} : \sigma(x) \nsim \varepsilon$,
(2) $\exists x \in \mathcal{A} : \lim_{n\to\infty} |\sigma^n(x)| = \infty$.

The substitution σ is *primitive* if there exists an integer $n \geq 1$ such that $|\sigma^n(y)|_x > 0$ for each two letters $x, y \in \mathcal{A}$.

A substitution σ extends in a natural way to $(\mathcal{A} \cup \overline{\mathcal{A}})^*$ by defining

$$\sigma(\bar{a}) = \overline{\sigma(a)}.$$

The setting agrees with the relation $\sim$ since for $X, Y \in (\mathcal{A} \cup \overline{\mathcal{A}})^*$ we have $\sigma(X) \sim \sigma(Y)$ whenever $X \sim Y$, making σ an endomorphism of the free group $(\mathcal{A} \cup \overline{\mathcal{A}})^*/\sim$. Observe that for words $X, Y \in_\sim \mathcal{A}^* \cup \overline{\mathcal{A}}^*$ the relation $|X| < |Y|$ implies $|\sigma(X)| < |\sigma(Y)|$. This does not hold in general for words $X, Y \in (\mathcal{A} \cup \overline{\mathcal{A}})^*$.

Definition 2.1 (Coding prescription, *cf.* [8])**.** Let σ be a substitution over the alphabet $\mathcal{A}$. A *coding prescription (with respect to σ)* is a function c with domain $\mathcal{A}^2$ that assigns to each pair of letters a finite set of integers such that

(1) $c(xx)$ is a complete residue system modulo $|\sigma(x)|$ for each $x \in \mathcal{A}$ (hence, $\#c(xx) = |\sigma(x)|$ and for all $k, k' \in c(xx)$ with $k \neq k'$ we have $k \not\equiv k' \pmod{|\sigma(x)|}$);
(2) for each $x \in \mathcal{A}$ we have $-|\sigma(x)| < k < |\sigma(x)|$ for all $k \in c(xx)$;
(3) for each $ab \in \mathcal{A}^2$ we have

$$c(ab) = \{k \in c(aa) : k < 0\} \cup \{0\} \cup \{k \in c(bb) : k > 0\}. \tag{2.1}$$

We call a coding prescription *continuous* if $c(ab)$ is a set of consecutive integers for each $ab \in \mathcal{A}^2$.

Note that $0 \in c(ab)$ for all $ab \in \mathcal{A}^2$. By the definition it is easy to see that a coding prescription is uniquely determined by defining $c(xx)$ for each $x \in \mathcal{A}$. Furthermore, c is continuous if and only if $c(xx)$ is a set of consecutive integers for all $x \in \mathcal{A}$.

We associate with a coding prescription a finite directed graph which plays an important role in the further proceeding.

Definition 2.2 (Graph associated with a coding prescription)**.** Let σ be a substitution over the alphabet $\mathcal{A}$ and c a coding prescription with respect to σ. The *graph associated with* c is the directed graph $H_{\sigma,c}$ with vertex set $\mathcal{A}^2$ and an edge from ab to a_1b_1 labelled by (D, a_1b_1) with $D \in (\mathcal{A} \cup \overline{\mathcal{A}})^*/\sim$ whenever $|D| \in c(ab)$ and

$$\sigma(\bar{a}) \preceq D\bar{a}_1 \prec D \prec Db_1 \preceq \sigma(b). \tag{2.2}$$

(hence, $D \in_\sim \mathcal{A}^* \cup \overline{\mathcal{A}}^*$).

For each vertex $ab \in \mathcal{A}$ the construction of the set of outgoing edges is based on the connection (via $|\cdot|$) between the $\sigma(\bar{a})$-$\sigma(b)$-chain and the $\leq$-chain $\{-|\sigma(a)|, \dots, |\sigma(b)|\}$. This implies that there is a one-to-one correspondence between $c(ab)$ and the set of outgoing edges $H^1_{\sigma,c}(ab)$, that is we have

$$c(ab) = \{|D| : (D, a_1b_1) \in H^1_{\sigma,c}(ab)\}.$$

The destination of an edge is determined by the predecessor and the successor of D in the chain. We see that $H_{\sigma,c}$ does not contain dead ends since each vertex ab has at least one outgoing edge that corresponds to $0 \in c(ab)$. However, in general graphs associated with coding prescriptions are not strongly connected even for primitive substitutions.

In the present article we are mainly interested in finite paths on $H = H_{\sigma,c}$ (when there is no danger of confusion we will frequently skip the indices). Therefore, we fix the following notation: for a vertex $ab \in \mathcal{A}^2$ and an integer $n > 0$ the expression $H^n(ab)$ is the set of paths of length n that start in ab. We represent the elements of $H^n(ab)$ by the corresponding label sequences.

Each vertex ab possesses (with respect to $\prec$) a minimal and a maximal outgoing edge $(D^-, a^-b^-) \in H^1(ab)$ and $(D^+, a^+b^+) \in H^1(ab)$, respectively, that satisfy $D^- \preceq D \preceq D^+$ for all $(D, a_1b_1) \in H^1(ab)$. Observe that we always have $D^- \in_\sim \overline{A}^*$ and $D^+ \in_\sim \mathcal{A}^*$. The minimal and the maximal edge may coincide, in this case we have $D^- \sim D^+ \sim \varepsilon$. Note that for each positive integer n the set $H^n(ab)$ contains a uniquely determined path that consists of minimal (maximal, respectively) edges only. We refer to this path as the minimal (maximal, respectively) path of length n.

2.3. **Main results.** In the first main result we show how to obtain a coding prescription with respect to the composition of substitutions.

Theorem 2.3. *Let σ_1 and σ_2 be two substitutions over the same alphabet $\mathcal{A}$, and c_1 and c_2 coding prescriptions with respect to σ_1 and σ_2, respectively. For each $ab \in \mathcal{A}^2$ set*

$$c_2 \odot c_1(ab) := \bigcup_{(D_1, a_1b_1) \in H^1_{\sigma_1,c_1}(ab)} \big(|\sigma_2(D_1)| + c_2(a_1b_1)\big).$$

Then $c_2 \odot c_1$ is a coding prescription with respect to $\sigma_2 \circ \sigma_1$.

The following list contains further properties of the binary operation $\odot$ and composed coding prescriptions.

- The graph associated with $c_2 \odot c_1$ is a product of the graphs associated with c_1 and c_2. In particular, for edges $(D_1, a_1b_1) \in H^1_{\sigma_1,c_1}(ab)$ and $(D_2, a_2b_2) \in H^1_{\sigma_2,c_2}(a_1b_1)$ there is an edge $(\sigma_2(D_1)D_2, a_2b_2) \in H^1_{\sigma_2\circ\sigma_1, c_2\odot c_1}(ab)$ (see Corollary 3.4).
- The binary operation $\odot$ is associative (see Corollary 3.5).
- The continuity of c_1 and c_2 is sufficient (but not necessary) for $c_2 \odot c_1$ to be continuous (see Lemma 3.8 and Corollary 3.9).

Application of Theorem 2.3 to powers of a single substitution σ yields the following result.

Theorem 2.4. *Let σ be a substitution over the alphabet $\mathcal{A}$ and c a coding prescription with respect to σ. For a positive integer n let $c^{(n)}$ denote the function with domain $\mathcal{A}^2$ defined by*

$$c^{(n)}(ab) = \left\{ \sum_{j=1}^{n} \left|\sigma^{n-j}(D_j)\right| : (D_j, a_jb_j)_{j=1}^n \in H^n_{\sigma,c}(ab) \right\}.$$

Then $c^{(n)}$ is a coding prescription with respect to σ^n. It is continuous if and only if c is continuous.

Theorem 2.4 provides a method for representing integers when the substitution σ satisfies the additional property

$$\exists ab \in \mathcal{A}^2 : (\sigma(\bar{a}) \prec \bar{a}) \wedge (b \prec \sigma(b)). \tag{2.3}$$

This requirement does not depend on the choice of a coding prescription and means that the rightmost letter of $\sigma(a)$ equals a and the leftmost letter of $\sigma(b)$ equals b and that $\sigma(a)$ as well as $\sigma(b)$ have length at least 2 (with this we want to avoid trivialities). Actually, for each substitution σ there exists a power n such that $\sigma^n(\bar{a}) \preceq \bar{a}$ and $b \preceq \sigma^n(b)$ for at least one pair $ab \in \mathcal{A}^2$ (*e.g.* [6]).

Note that if σ is a primitive substitution that satisfies $\sigma(\bar{a}) \preceq \bar{a}$ and $b \preceq \sigma(b)$ then the inequalities are automatically strict. Indeed, if we had, for instance, $b = \sigma(b)$ then $b = \sigma^n(b)$ holds for all positive integers n. In the case of a one-letter alphabet this would violate Item (2) of the definition of a substitution while for larger alphabets this would contradict the definition of primitivity.

Theorem 2.5. *Let σ be a substitution over the alphabet $\mathcal{A}$ that satisfies* (2.3) *for $ab \in \mathcal{A}^2$, and c a coding prescription with respect to σ. Define the set*

$$Z_{ab} := \bigcup_{n \geq 1} c^{(n)}(ab) \subset \mathbb{Z}.$$

Then for each element $N \in Z_{ab}$ there exists a finite path $(D_j, a_j b_j)_{j=1}^n \in H^n(ab)$ such that N can be represented as

$$N = \sum_{j=1}^{n} \left|\sigma^{n-j}(D_j)\right|. \tag{2.4}$$

Paths that represent the same integer differ by a leading sequence of (ε, ab) only, therefore, the representation is unique when we require that $D_1 \not\sim \varepsilon$ (and 0 is represented by the empty sum).

Obviously, the set Z_{ab} of representable integers always contains 0. Furthermore, if c is a continuous coding prescription then it follows from Theorem 2.4 that Z_{ab} consists of consecutive integers. Further (and less obvious) properties are collected in the next result.

Proposition 2.6. *Consider the setting of Theorem 2.5. Then the following items hold.*

1. *Z_{ab} contains positive (negative, respectively) integers if and only if $c(ab)$ contains at least one positive (negative, respectively) integer.*
2. *Z_{ab} contains all positive (negative, respectively) integers if and only if $1 \in c(ab)$ ($-1 \in c(ab)$, respectively).*
3. *If the difference between 0 and the least positive (largest negative, respectively) element of $c(ab)$ is at least 3 and σ is a primitive substitution then Z_{ab} contains gaps of arbitrary large size.*
4. *If $|\sigma(x)| \equiv 1 \pmod 2$ for all $x \in \mathcal{A}$ and c the coding prescription with*
$$c(ab) = \{-|\sigma(a)| + 1, -|\sigma(a)| + 3, \dots, -2, 0, 2, \dots, |\sigma(b)| - 1\}$$
for all $ab \in \mathcal{A}^2$ then $Z_{ab} = 2\mathbb{Z}$.

The proposition does not yield a complete description of Z_{ab} for all settings. In fact, if neither of the items can be applied then the structure of Z_{ab} is quite unpredictable and a characterisation can be a challenging task. We will analyse one such case in Example 5.5.

For some settings the canonical ordering of the integers corresponds to a lexicographical ordering of the addends of the representation (2.4). In particular:

Proposition 2.7. *Consider the setting of Theorem 2.5 and suppose that either c is a continuous coding prescription or that the conditions of Item (4) in Proposition (2.6) are satisfied. Let*

$$N := \sum_{j=1}^{n} \left|\sigma^{n-j}(D_j)\right|, \qquad N' := \sum_{j=1}^{n} \left|\sigma^{n-j}(D'_j)\right|$$

be (2.4)-representations of equal length of integers $N, N' \in Z_{ab}$. Then the following items are equivalent:

1. *$N < N'$;*
2. *$D_1, \dots, D_n \prec_{\text{lex}} D'_1, \dots, D'_n$ where $\prec_{\text{lex}}$ is the lexicographical extension of $\prec$.*
3. *$|D_1|, \dots, |D_n| <_{\text{lex}} |D'_1|, \dots, |D'_n|$.*

Remark 2.8. For comparing (2.4)-representations of different length we have to fill up the shorter sequence with a respecting number of leading ε.

3. Combinatorics of coding prescriptions

3.1. **Composing coding prescriptions.** Throughout the following considerations we let σ_1 and σ_2 denote two substitutions over the same alphabet $\mathcal{A}$, and by c_1 and c_2 coding prescriptions with respect to σ_1 and σ_2, respectively. Furthermore, we set $H_1 := H_{\sigma_1, c_1}$ and $H_2 := H_{\sigma_2, c_2}$. Our aim is to show Theorem 2.3, hence, that c_1 and c_2 induce in a natural way a coding prescription $c_2 \odot c_1$ with respect to $\sigma_2 \circ \sigma_1$ defined by

$$c_2 \odot c_1(ab) := \bigcup_{(D_1, a_1 b_1) \in H_1^1(ab)} \left(|\sigma_2(D_1)| + c_2(a_1 b_1)\right) \tag{3.1}$$

for each $ab \in \mathcal{A}^2$, and to analyse these composed coding prescriptions.

Lemma 3.1. *Let $(D_1, a_1 b_1) \in H_1^1(ab)$ and $D, D' \in_{\sim} \mathcal{A}^* \cup \overline{\mathcal{A}}^*$ such that $\sigma_1(\bar{a}) \preceq D \prec D_1 \prec D' \preceq \sigma_1(b)$. Then $|\sigma_2(D)| < |\sigma_2(D_1)| + k < |\sigma_2(D')|$ holds for each $k \in c_2(a_1 b_1)$.*

Proof. By assumption, D, D_1 and D' are contained in the $\sigma_1(\bar{a})$-$\sigma_1(b)$-chain. Since $D_1\overline{a_1}$ is the predecessor and $D_1 b_1$ is the successor of D_1 we clearly have $D \preceq D_1\overline{a_1}$ and $D_1 b_1 \preceq D'$. Therefore, $|\sigma_2(D)| \leq |\sigma_2(D_1)| - |\sigma_2(a_1)|$ and $|\sigma_2(D_1)| + |\sigma_2(b_1)| \leq |\sigma_2(D)|$. Now the lemma follows from the observation that for each $k \in c_2(a_1 b_1)$ we have $-|\sigma_2(a_1)| < k < |\sigma_2(b_1)|$ by definition. □

Lemma 3.2. *Let $x \in \mathcal{A}$. Then for each $k \in \{0, \ldots, |\sigma_2 \circ \sigma_1(x)| - 1\}$ we have either $k \in c_2 \odot c_1(xx)$ or $k - |\sigma_2 \circ \sigma_1(x)| \in c_2 \odot c_1(xx)$.*

Proof. Let $k \in \{0, \ldots, |\sigma_2 \circ \sigma_1(x)| - 1\}$. There exist words $U, V \in \mathcal{A}^*$ and a letter $y \in \mathcal{A}$ such that $UyV = \sigma_1(x)$ and $|\sigma_2(U)| \leq k < |\sigma_2(Uy)|$. Hence, $0 \leq k - |\sigma_2(U)| < |\sigma_2(y)|$. By definition, there exists a $k_2 \in c_2(yy)$ with $k_2 \equiv k - |\sigma_2(U)| \pmod{|\sigma_2(y)|}$. In fact, we have two possibilities: $k_2 = k - |\sigma_2(U)| \geq 0$ or $k_2 = k - |\sigma_2(U)| - |\sigma_2(y)| = k - |\sigma_2(Uy)| < 0$. We study these cases separately.

Case 1. $k_2 = k - |\sigma_2(U)|$**:** By definition the set $c_1(xx)$ contains either $|U| \geq 0$ or $|U| - |\sigma_1(x)| = \left|\overline{yV}\right| < 0$. This yields two subcases.

Case 1a. $|U| \in c_1(xx)$**:** We clearly have $\sigma_1(\bar{x}) \prec \varepsilon \preceq U \prec Uy \preceq \sigma_1(x)$. Let $a_1 \in \mathcal{A}$ such that $\sigma_1(\bar{x}) \preceq U\bar{a}_1 \prec U$. Then, by definition, $(U, a_1 y) \in H_1^1(xx)$. Since $0 \leq k_2 \in c_2(yy)$ we have by definition $k_2 \in c_2(a_1 y)$ and therefore $|\sigma_2(U)| + k_2 = |\sigma_2(U)| + k - |\sigma_2(U)| = k \in c_2 \odot c_1(xx)$.

Case 1b. $\left|\overline{yV}\right| \in c_1(xx)$**:** Here we have $\sigma_1(\bar{x}) \prec \overline{yV} \prec \overline{V}(\sim \overline{yV}y) \preceq \varepsilon \prec \sigma_1(x)$. Let $a_1 \in \mathcal{A}$ denote the letter with $\sigma_1(\bar{x}) \preceq \overline{yV}\bar{a}_1 \prec \overline{yV}$. Then, by definition, $(\overline{yV}, a_1 y) \in H_1^1(xx)$. As above we have $k_2 \in c_2(a_1 y)$ and, thus,

$$\left|\sigma_2(\overline{yV})\right| + k_2 = \left|\sigma_2(\overline{yV})\right| + k - |\sigma_2(U)| = k - |\sigma_2(UyV)| = k - |\sigma_2 \circ \sigma_1(x)| \in c_2 \odot c_1(xx).$$

Case 2. $k_2 = k - |\sigma_2(Uy)|$**:** Here we consider the element of $c_1(xx)$ that is (modulo $|\sigma_1(x)|$) equivalent to $|Uy|$. As before we have two subcases.

Case 2a. $|Uy| \in c_1(xx)$**:** We proceed similarly as in the previous cases. In particular, we have $\sigma_1(\bar{x}) \prec \varepsilon \preceq U(\sim Uy\bar{y}) \prec Uy \prec \sigma_1(x)$. Let b_1 denote the element of $\mathcal{A}$ that satisfies $Uy \prec Uyb_1 \preceq \sigma_1(x)$. Then clearly $(Uy, yb_1) \in H_1(xx)$. As $0 > k_2 \in c_2(yy)$ we have $k_2 \in c_2(yb_1)$ and, thus, $|\sigma_2(Uy)| + k_2 = |\sigma_2(Uy)| + k - |\sigma_2(Uy)| = k \in c_2 \odot c_1(xx)$.

Case 2b. $\left|\overline{V}\right| \in c_1(xx)$**:** Analogously let $b_1 \in \mathcal{A}$ such that $\sigma_1(\bar{x}) \preceq \overline{V}\bar{y} \prec \overline{V} \prec \overline{V}b_1 \preceq \sigma_1(x)$. Hence, $(\overline{V}, yb_1) \in H_1^1(xx)$, $0 > k_2 \in c_2(yb_1)$, and

$$\left|\sigma_2(\overline{V})\right| + k_2 = -|\sigma_2(V)| + k - |\sigma_2(Uy)| = k - |\sigma_2 \circ \sigma_1(x)| \in c_2 \odot c_1(xx).$$

□

Proof of Proposition 2.3. We show that all items of Definition 2.1 are satisfied.

The statement of Lemma 3.2 clearly implies that, for each $x \in \mathcal{A}$, $c_2 \odot c_1(xx)$ contains a set of representatives modulo $|\sigma_2 \circ \sigma_1(x)|$ whose elements are bounded by $|\sigma_2 \circ \sigma_1(x)|$ in modulus. From this we also immediately see that $\# c_2 \odot c_1(xx) \geq |\sigma_2 \circ \sigma_1(x)|$. On the other hand we estimate

$$\begin{aligned}
\# c_2 \odot c_1(xx) &= \# \left(\bigcup_{(D_1, a_1 b_1) \in H^1(xx)} \big(|\sigma(D_1)| + c_2(a_1 b_1) \big) \right) \\
&\leq \sum_{(D_1, a_1 b_1) \in H_1^1(xx)} \# c_2(a_1 b_1) \\
&= \sum_{(D_1, a_1 b_1) \in H_1^1(xx)} \big(\#\{k \in c_2(a_1 b_1) : k < 0\} + 1 + \#\{k \in c_2(a_1 b_1) : k > 0\} \big) \\
&= \sum_{y \in \mathcal{A}} \#\{(D_1, a_1 b_1) \in H_1^1(xx) : a_1 = y\} \cdot \#\{k \in c_2(yy) : k < 0\} + |\sigma_1(x)| \\
&\quad + \sum_{y \in \mathcal{A}} \#\{(D_1, a_1 b_1) \in H_1^1(xx) : b_1 = y\} \cdot \#\{k \in c_2(yy) : k > 0\}.
\end{aligned}$$

We claim that

$$\#\{(D_1, a_1 b_1) \in H_1^1(xx) : a_1 = y\} = \#\{(D_1, a_1 b_1) \in H_1^1(xx) : b_1 = y\} = |\sigma(x)|_y. \tag{3.2}$$

To show this we set $\sigma_1(x) = x_1x_2\cdots x_n$. Then the $\sigma_1(\bar{x})$-$\sigma_1(x)$-chain is (modulo ~) given by

$$\{\bar{x}_n\cdots\bar{x}_1, \bar{x}_n\cdots\bar{x}_2, \ldots, \bar{x}_n\bar{x}_{n-1}, \bar{x}_n, \varepsilon, x_1, x_1x_2, \ldots, x_1x_2\cdots x_{n-1}, x_1x_2\cdots x_n\}.$$

By definition $c_1(xx)$ contains 0, hence, $(\varepsilon, x_nx_1) \in H_1^1(xx)$. Furthermore, for each $k \in \{1, \ldots, n-1\}$ either $k \in c_1(xx)$ or $n - k \in c_1(xx)$. This shows that $H_1^1(xx)$ contains (modulo ~) either $(x_1\cdots x_k, x_kx_{k+1})$ or $(\bar{x}_n\cdots\bar{x}_{k+1}, x_kx_{k+1})$. Therefore, if we run through all edges $(D_1, a_1b_1) \in H_1^1(xx)$ then a_1 as well as b_1 run through all the letters of $\sigma_1(x)$ which immediately shows the claim.

Inserting this into the estimation from above yields

$$\begin{aligned}\#c_2 \odot c_1(xx) &\le \sum_{y\in\mathcal{A}} |\sigma_1(x)|_y \big(\#\{k \in c_2(yy) : k < 0\} + 1 + \#\{k \in c_2(yy) : k > 0\}\big)\\ &= \sum_{y\in\mathcal{A}} |\sigma_1(x)|_y |\sigma_2(y)| = |\sigma_2 \circ \sigma_1(x)|.\end{aligned}$$

Hence, $\#c_2 \odot c_1(xx) = |\sigma_2 \circ \sigma_1(x)|$, thus, $c_2 \odot c_1(xx)$ is a complete residue system and for each $k \in c_2 \circ c_1(x)$ we have $-|\sigma_2 \circ \sigma_1(x)| < k < |\sigma_2 \circ \sigma_1(x)|$ proving the items (1) and (2) of Definition 2.1.

To show item (3) let $ab \in \mathcal{A}^2$ and write $a^{(1)}$ and $a^{(-1)}$ ($b^{(1)}$ and $b^{(-1)}$, respectively) for the leftmost and rightmost letter of $\sigma_1(a)$ ($\sigma_1(b)$, respectively). Thus, $(\varepsilon, a^{(-1)}a^{(1)}) \in H_1^1(aa)$ and $(\varepsilon, b^{(-1)}b^{(1)}) \in H_1^1(bb)$. We clearly have

$$c_2 \odot c_1(aa) = \bigcup_{\substack{(D_1,a_1b_1)\in H_1^1(aa)\\ D_1 \in_\sim \overline{\mathcal{A}}^+}} \big(|\sigma_2(D_1)| + c_2(a_1b_1)\big) \cup c_2(a^{(-1)}a^{(1)}) \cup \bigcup_{\substack{(D_1,a_1b_1)\in H_1^1(aa)\\ D_1 \in_\sim \mathcal{A}^+}} \big(|\sigma_2(D_1)| + c_2(a_1b_1)\big),$$

$$c_2 \odot c_1(bb) = \bigcup_{\substack{(D_1,a_1b_1)\in H_1^1(bb)\\ D_1 \in_\sim \overline{\mathcal{A}}^+}} \big(|\sigma_2(D_1)| + c_2(a_1b_1)\big) \cup c_2(b^{(-1)}b^{(1)}) \cup \bigcup_{\substack{(D_1,a_1b_1)\in H_1^1(bb)\\ D_1 \in_\sim \mathcal{A}^+}} \big(|\sigma_2(D_1)| + c_2(a_1b_1)\big).$$

Consider an edge $(D_1, a_1b_1) \in H_1^1(aa)$ with $D_1 \in_\sim \overline{\mathcal{A}}^+$. Then $D_1 \prec \varepsilon$ and, thus, by Lemma 3.1, $|D_1| + c_2(a_1b_1)$ contains only negative elements. Similarly, for $D_1 \in_\sim \mathcal{A}^+$ all elements of $|D_1| + c_2(a_1b_1)$ are strictly positive. Thus,

$$\{k \in c_2 \odot c_1(aa) : k < 0\} = \bigcup_{\substack{(D_1,a_1b_1)\in H_1^1(aa)\\ D_1 \in_\sim \overline{\mathcal{A}}^+}} \big(|\sigma_2(D_1)| + c_2(a_1b_1)\big) \cup \{k \in c_2(a^{(-1)}a^{(-1)}) : k < 0\},$$

where we applied (2.1) on $c_2(a^{(-1)}a^{(1)})$. Analogously, we obtain

$$\{k \in c_2 \odot c_1(bb) : k > 0\} = \{k \in c_2(b^{(1)}b^{(1)}) : k > 0\} \cup \bigcup_{\substack{(D_1,a_1b_1)\in H_1^1(bb)\\ D_1 \in_\sim \mathcal{A}^+}} \big(|\sigma_2(D_1)| + c_2(a_1b_1)\big).$$

Now we are interested in $c_2 \odot c_1(ab)$. At first observe that the $\sigma_1(\bar{a})$-$\sigma_1(b)$-chain consists (modulo ~) of the elements of the $\sigma(\bar{a})$-$\sigma(a)$-chain that are contained in $\overline{\mathcal{A}}^+$, the elements of the $\sigma(\bar{b})$-$\sigma(b)$-chain that are contained in $\mathcal{A}^+$, and ε. Therefore, we have

$$H_1^1(ab) = \{(D_1, a_1b_1) \in H_1^1(aa) : D_1 \in_\sim \overline{\mathcal{A}}^+\} \cup \{(\varepsilon, a^{(-1)}b^{(1)})\} \cup \{(D_1, a_1b_1) \in H_1^1(bb) : D_1 \in_\sim \mathcal{A}^+\}.$$

Hence,

$$\begin{aligned}c_2 \odot c_1(ab) = &\bigcup_{\substack{(D_1,a_1b_1)\in H_1^1(aa)\\ D_1 \in_\sim \overline{\mathcal{A}}^+}} \big(|\sigma_2(D_1)| + c_2(a_1b_1)\big)\\ &\cup c_2(a^{(-1)}b^{(1)}) \cup \bigcup_{\substack{(D_1,a_1b_1)\in H_1^1(bb)\\ D_1 \in_\sim \mathcal{A}^+}} \big(|\sigma_2(D_1)| + c_2(a_1b_1)\big).\end{aligned}$$

By applying (2.1) on $c_2(a^{(-1)}b^{(1)})$ we immediately see that

$$c_2 \odot c_1(ab) = \{k \in c_2 \odot c_1(aa) : k < 0\} \cup \{0\} \cup \{k \in c_2 \odot c_1(bb) : k > 0\}.$$

□

Remark 3.3. Note that the proof immediately implies the disjointness of the union (3.1).

The next result shows that the graph induced by $c_2 \odot c_1$ corresponds to a product graph of H_1 and H_2.

Corollary 3.4. *Let $H := H_{\sigma_2\circ\sigma_1, c_2\odot c_1}$. Then for each $ab \in \mathcal{A}^2$ we have*

$$H^1(ab) = \bigcup_{(D_1,a_1b_1)\in H_1^1(ab)} \{(\sigma(D_1)D_2, a_2b_2) : (D_2, a_2b_2) \in H_1^1(a_1b_1)\}.$$

Proof. Let $ab \in \mathcal{A}^2$. Theorem 2.3 shows that $k \in c_2\odot c_1(ab)$ if and only if there are edges $(D_1, a_1b_1) \in H_1^1(ab)$ and $(D_2, a_2b_2) \in H_1^1(a_1b_1)$ such that $|\sigma(D_1)D_2| = k$. Thus, it suffices to show that (2.2) holds.

Observe that $\prec$ is invariant with respect to the concatenation of words from the left and with respect to the application of σ_2. Thus, from $\sigma_2(\bar{a}_1) \preceq D_2\bar{a}_2 \prec D_2 \prec D_2b_2 \preceq \sigma_2(b_1)$ it follows that

$$\sigma_2(D_1\bar{a}_1) \preceq \sigma_2(D_1)D_2\bar{a}_2 \prec \sigma_2(D_1)D_2 \prec \sigma_2(D_1)D_2b_2 \preceq \sigma_2(D_1b_1)$$

while $\sigma_1(\bar{a}) \preceq D_1\bar{a}_1 \prec D_1 \prec D_1b_1 \preceq \sigma_1(b)$ implies that

$$\sigma_2 \circ \sigma_1(\bar{a}) \preceq \sigma_2(D_1\bar{a}_1) \prec \sigma_2(D_1) \prec \sigma_2(D_1b_1) \preceq \sigma_2 \circ \sigma_1(b).$$

Proper combination of these two expressions yields

$$\sigma_2 \circ \sigma_1(\bar{a}) \preceq \sigma_2(D_1)D_2\bar{a}_2 \prec \sigma_2(D_1)D_2 \prec \sigma_2(D_1)D_2b_2 \preceq \sigma_2 \circ \sigma_1(b).$$

□

We will apply Theorem 2.3 and Corollary 3.4 in Example 5.6.

Corollary 3.5. *The binary operation $\odot$ is associative.*

Proof. For each $j \in \{1,2,3\}$ let σ_j denote a substitution over the alphabet $\mathcal{A}$ and by c_j a coding prescription with respect to σ_j. Let $a_0b_0 \in \mathcal{A}^2$. We prove that $(c_3 \odot c_2) \odot c_1(a_0b_0) = c_3 \odot (c_2 \odot c_1)(a_0b_0)$. For this reason let $k \in (c_3 \odot c_2) \odot c_1(a_0b_0)$. Since both $(c_3 \odot c_2) \odot c_1$ and $c_3 \odot (c_2 \odot c_1)$ are coding prescriptions, it suffices to show that k is also contained in $c_3 \odot (c_2 \odot c_1)(a_0b_0)$. By considering Theorem 3.4 there exists for each $j \in \{1,2,3\}$ an edge $(D_j, a_jb_j) \in H^1_{\sigma_j,c_j}(a_{j-1}b_{j-1})$ such that $k = |\sigma_3 \circ \sigma_2(D_1)\, \sigma_3(D_2)D_3| = |\sigma_3(\sigma_2(D_1)D_2)D_3|$. Therefore, $k \in c_3 \odot (c_2 \odot c_1)(a_0b_0)$. □

3.2. **Properties of composed coding prescriptions.**

Lemma 3.6. *Let $(D^-, a^-b^-), (D^+, a^+b^+) \in H_1^1(ab)$ the minimal and maximal edge and $k_2^- := \min c_2(a^-b^-)$, $k_2^+ := \max c_2(a^+b^+)$. Then for each $k \in c_2 \odot c_1(ab)$ we have $|\sigma_2(D^-)| + k_2^- \leq k \leq |\sigma_2(D^+)| + k_2^+$.*

Proof. Each $k \in c_2 \odot c_1(ab)$ is uniquely determined by an edge $(D_1, a_1b_1) \in H_1^1(ab)$ and a $k_2 \in c(a_1b_1)$ such that $k \in |\sigma_2(D_1)|+k_2$. If $D_1 \sim D^-$ then by the minimality of k_2^- we have $|\sigma_2(D^-)|+k_2^- \leq k$. If $D^- \prec D_1$ then observe that $k_2^- \leq 0$. Thus, $|\sigma_2(D^-)| + k_2^- \leq |\sigma_2(D^-)| \leq k$ by Lemma 3.1. The right hand inequality $k \leq |\sigma_2(D^+)| + k_2^+$ can be shown analogously. □

When we consider the union (3.1) then the last lemma stated that the minimal (maximal, respectively) element of $c_2 \odot c_1(ab)$ is contained in the subset induced by the minimal (maximal, respectively) edge of $H_1^1(ab)$. The next lemma we provides stronger result. If c_2 is continuous we will see that all subsets of the union (3.1) comply with the ordering $\prec$ on $H_1^1(ab)$. Recall that a coding prescription c is called continuous if $c(ab)$ is a set of consecutive integers for all $ab \in \mathcal{A}^2$. By construction this is equivalent to the fact that $c(xx)$ is a set of consecutive integers for each $x \in \mathcal{A}$.

Lemma 3.7. *Let $(D_1, a_1b_1), (D_1', a_1'b_1') \in H_1^1(ab)$ such that $D_1 \prec D_1'$. If c_2 is continuous then $|\sigma_2(D_1)| + k_2 < |\sigma_2(D_1')| + k_2'$ holds for all $k_2 \in c_2(a_1b_1)$ and $k_2' \in c_2(a_1'b_1')$.*

Proof. By definition we have $D_1 \prec D_1b_1$ and $D_1'\bar{a}_1' \prec D_1'$. Thus, there are two possibilities: either $D_1b_1 \preceq D_1'\bar{a}_1'$ or $D_1b_1 \sim D_1'$ (and $D_1 \sim D_1'\bar{a}_1'$).

In the first case we immediately see that

$$|\sigma_2(D_1)| + k_2 < |\sigma_2(D_1b_1)| \le |\sigma_2(D_1'\bar{a}_1')| < |\sigma_2(D_1')| + k_2'$$

holds for all $k_2 \in c_2(a_1b_1)$ and $k_2' \in c_2(a_1'b_1')$.

Now we concentrate on the case $D_1b_1 \sim D_1'$, hence, D_1' is the successor of D_1 in the $\sigma_1(\bar{a})$-$\sigma_1(b)$-chain. Since $D_1 \sim D_1'\bar{b}_1$ we have $b_1 = a_1'$. Thus, for each $k_2 \in c_2(a_1b_1)$ we have $k_2 \le \max c_2(b_1b_1)$ and for each $k_2' \in c_2(a_1'b_1')$ we have $k_2 \ge \min c_2(a_1'a_1') = \min c_2(b_1b_1)$. We obtain

$$\begin{aligned}(|\sigma_2(D_1')| + k_2') - (|\sigma_2(D_1)| + k_2) \ge& |\sigma_2(D_1b_1)| + \min c_2(b_1b_1) - |\sigma_2(D_1)| - \max c_2(b_1b_1)\\ =& |\sigma_2(b_1)| - \max c_2(b_1b_1) + \min c_2(b_1b_1).\end{aligned}$$

As c_2 is continuous, $c_2(b_1b_1)$ is a set of $|\sigma_2(b_2)|$ consecutive integers, and we have $\max c_2(b_1b_1) - \min c_2(b_1b_1) = |\sigma_2(b_1)| - 1$. Therefore the difference above is strictly positive. □

The next lemma yields sufficient and necessary conditions for the composition $c_2 \odot c_1$ to be continuous.

Lemma 3.8. *Let $x \in \mathcal{A}$. Then $c_2 \odot c_1(xx)$ is a set of consecutive integers if and only if $c_1(xx)$ and $c_2(a^-a^-)$ are sets of consecutive integers where (D^-, a^-b^-) is the minimal edge of $H_1^1(xx)$. (Note that the continuity of c_1 implies that $a^- = b^+$, where (D^+, a^+b^+) is the maximal edge of $H_1^1(xx)$).*

Proof. Set $k_1^- := \min c_1(xx) \le 0$, $k_1^+ := \max c_1(xx) \ge 0$. Let $\sigma_1(x) = x_1x_2\cdots x_n$. Then we have $D^- \sim \bar{x}_n\cdots\bar{x}_{n+k_1^-+1}$ and $D^+ \sim x_1\cdots x_{k_1^+}$ (for $k_1^- = 0$ this means $D^- \sim \varepsilon$, analogously for $k_1^+ = 0$). Furthermore, we have $a^- = x_{n+k_1^-}$ and $b^+ = x_{k_1^+ + 1}$.

Set $k_2^- := \min c_2(a^-b^-) \le 0$ and $k_2^+ := \max c_2(a^+b^+) \ge 0$. By Lemma 3.6 we have $k^- := |\sigma_2(D^-)| + k_2^- = \min c_2 \odot c_1(xx)$ and $k_+ := |\sigma_2(D^+)| + k_2^+ = \max c_2 \odot c_1(xx)$.

Observe that $c_1(xx)$ is a set of consecutive integers if and only if $k_1^+ - k_1^- = |\sigma(x)| - 1 = n - 1$. At first we show that this condition is necessary. Indeed, suppose that $c(xx)$ were not a set of consecutive integers. Then $k_1^+ - k_1^- > n$ (equality cannot hold since k_1^- and k_1^+ are not equivalent modulo n). Furthermore, we must have $k_1^+, k_1^- \ne 0$. We get

$$\begin{aligned}k^+ - k^- &= |\sigma_2(D_1^+)| + k_2^+ - |\sigma_2(D_1^-)| - k_2^- \ge |\sigma_2(D_1^+)| - |\sigma_2(D_1^-)| = \left|\sigma_2(x_1\cdots x_{k_1^+})\right| - \left|\sigma_2(\bar{x}_n\cdots\bar{x}_{n+k_1^-+1})\right|\\ &= \left|\sigma_2(x_1\cdots x_{k_1^+})\right| + \left|\sigma_2(x_{n+k_1^-+1}\cdots x_{k_1^+})\right| + \left|\sigma_2(x_{k_1^++1}\cdots x_n)\right| > |\sigma_2(x_1\cdots x_n)| = \sigma_2 \circ \sigma_1(x)\end{aligned}$$

which shows that $c_2 \odot c_1(xx)$ cannot be a set of $|\sigma_2 \circ \sigma_1(x)|$ consecutive integers.

Thus, suppose that $c_1(xx)$ is a set of consecutive integers. From $k_1^+ - k_1^- = n - 1$ we conclude that $a^- = x_{n+k_1^-} = x_{n+k_1^+-n+1} = x_{k_1^++1} = b^+$. This yields $\sigma(x) \sim D^+a^-\overline{D^-}$. Therefore,

$$|\sigma_2 \circ \sigma_1(x)| - k^+ + k^- = |\sigma_2 \circ \sigma_1(x)| - |\sigma_2(D_1^+)| + |\sigma_2(D_1^-)| - k_2^+ + k_2^- = |\sigma_2(a_1^-)| - k_2^+ + k_2^-.$$

Since $b^+ = a^-$ we have $k_2^+, k_2^- \in c_2(a^-a^-)$. Thus, the last expression is positive if and only if $c_2(a^-a^-)$ is a set of consecutive integers. □

Corollary 3.9. *The continuity of c_1 and c_2 is sufficient for the continuity of $c_2 \odot c_1$.*

Now Theorem 2.4 follows as a corollary from the results of this section. Note that we explicitly apply it in Example 5.3.

Proof of Theorem 2.4. From Corollary 3.4 we immediately see that $c^{(2)} = c \odot c$ and, more generally, $c^{(n)} = c^{(n-1)} \odot c$. Hence, $c^{(n)}$ is a coding prescription with respect to σ^n. The statement concerning the continuity follows directly from Lemma 3.8. □

Observe that the disjointness of the union in Theorem 2.3 (*cf.* Remark 3.3) immediately implies the disjointness of the union in Theorem 2.4.

In the last statement of this section we want to present a special setting that shows that a statement analogue to Lemma 3.7 also may hold for non-continuous coding prescriptions. We will refer to this result in the proof of Proposition 2.7 in the next section.

Lemma 3.10. *Let σ be a substitution over an alphabet $\mathcal{A}$ such that $|\sigma(x)| \equiv 1 \pmod 2$ for all $x \in \mathcal{A}$. The map c that assigns to each pair $ab \in \mathcal{A}^2$ the set*

$$\{-|\sigma(a)|+1, -|\sigma(a)|+3, \dots, -2, 0, 2, \dots, |\sigma(b)|-1\}$$

is a coding prescription and for all $n \in \mathbb{N}$ and $ab \in \mathcal{A}^2$ the set $c^{(n)}(ab)$ consists of even integers only.

Furthermore, for all $ab \in \mathcal{A}^2$, $n \geq 1$, $(D_j, a_j b_j), (D'_j, a'_j b'_j) \in H^n_{\sigma,c}(ab)$ we have

$$\sum_{j=1}^{n} \left|\sigma^{n-j}(D_j)\right| < \sum_{j=1}^{n} \left|\sigma^{n-j}(D'_j)\right| \Longleftrightarrow (D_j)_{j=1}^n \prec_{\text{lex}} (D'_j)_{j=1}^n.$$

Proof. It is easy to see that c is a coding prescription with respect to σ. Now observe that from the condition it follows that $|\sigma^n(x)| \equiv 1 \pmod 2$ for each $x \in \mathcal{A} \cup \overline{\mathcal{A}}$ and each $n \in \mathbb{N}$. Thus, $\left|\sigma^k(X)\right| \equiv |X| \pmod 2$ for each $X \in (\mathcal{A} \cup \overline{\mathcal{A}})^*$ and each $n \in \mathbb{N}$.

Now let $ab \in \mathcal{A}^2$, $n \in \mathbb{N}$ and $k \in c^{(n)}(ab)$. Then there exists a path $(D_j, a_j b_j)_{j=1}^n \in H^n(ab)$ with

$$k = \sum_{j=1}^{n} \left|\sigma^{n-j}(D_j)\right|.$$

Since $|D_j| \equiv 0 \pmod 2$ for all $j \in \{1, \dots, n\}$ we clearly have $k \equiv 0 \pmod 2$.

Consider another path $(D'_j, a'_j b'_j)_{j=1}^n \in H^n(ab)$ and $k' := \sum_{j=1}^n \left|\sigma^{n-j}(D'_j)\right|$. Suppose that $k \neq k'$. By the disjointness of the union in Theorem 2.4 this immediately shows that the corresponding paths are not identical. If $D_j \prec D'_j$ then, as $|D'j| - |D_j| \geq 2$, there exists a $D \in \mathcal{A}^* \cup \overline{\mathcal{A}}^*$ with $D_j \prec D \prec D'_j$. Since $k, k' \in c^{(n)}(ab) = c^{(n-1)} \odot c(ab)$ we see by Lemma 3.1 that $k < k'$. On the other hand, if $D'_j \prec D_j$ then we immediately see that $k' < k$. If $D_1 = D'_1$ then $a_1 b_1 = a'_1 b'_1$, hence, $(D_j, a_j b_j)_{j=2}^n, (D'_j, a'_j b'_j)_{j=2}^n \in H^{n-1}_{\sigma,c}(a_1 b_1)$ and we can compare D_2 and D'_2 in the same way. This immediately yields that $k < k'$ if and only if $(D_j)_{j=1}^n \prec_{\text{lex}} (D'_j)_{j=1}^n$. □

4. Representing integers

We now use the results obtained so far to represent integers. Thus, in the present section we concentrate on substitutions over $\mathcal{A}$ that satisfy (2.3) for a pair of letters $ab \in \mathcal{A}^2$. Let c denote a coding prescription with respect to σ and, for convenience, $H := H_{\sigma,c}$. Recall that we defined

$$Z_{ab} = \bigcup_{n \geq 1} c^{(n)}(ab).$$

Lemma 4.1. *For each $n \in \mathbb{N}$ we have $c^{(n)}(ab) \subset c^{(n+1)}(ab)$.*

Proof. Since $0 \in c(ab)$ and $\sigma(\bar{a}) \prec \bar{a} \prec \varepsilon \prec b \prec \sigma(b)$ we have $(\varepsilon, ab) \in H^1(ab)$. Thus, since $c^{(n+1)} = c^{(n)} \odot c$ for each $n \in \mathbb{N}$, we immediately see that $\sigma(\varepsilon) + c^{(n)}(ab) \subset c^{(n+1)}(ab)$. □

Proof of Theorem 2.5. Let $N \in Z_{ab}$. For convenience set $c^{(0)}(ab) := \{0\}$ and $H^0(ab) := \varnothing$. Let $n_0 := \min\{n \geq 0 : N \in c^{(n)}(ab)\}$ and $(D_j, a_j b_j)_{j=1}^{n_0} \in H^{n_0}(ab)$ such that

$$N = \sum_{j=1}^{n_0} \left|\sigma^{n_0-j}(D_j)\right|.$$

Then this is obviously a representation (2.4) of N. At first note that $(D_j, a_j b_j)_{1=j}^{n_0}$ is the only path of length n_0 with this property. Furthermore observe that, by Lemma 4.1, $D_1 \nsim \varepsilon$, otherwise $N := \sum_{j=1}^{n_0-1} \left|\sigma^{n_0-1-j}(D_{j+1})\right| \in c^{(n_0-1)}(ab)$, which contradicts the minimality of n_0.

Now, let $(D'_j, a'_j b'_j)_{j=1}^n \in H^n(ab)$ such that $N = \sum_{j=1}^n \left|\sigma^{n-k}(D'_j)\right|$ with $n > n_0$. Then clearly $N \in c^{(n)}(ab)$ and, hence, we necessarily have that (modulo $\sim$) $(D'_j, a'_j b'_j) = (\varepsilon, ab)$ for $j \in \{1, \dots n - n_0\}$ and $(D'_j, a'_j b'_j) = (D_{j-n+n_0}, a_{j-n+n_0} b_{j-n+n_0})$ for $j \in \{n - n_0 + 1, \dots n\}$ since each element of $c^{(n)}(ab)$ corresponds to a unique path in $H^n(ab)$. This shows that $D'_1 \sim \varepsilon$.

Thus, the representation (2.4) is unique if we require that $D_1 \nsim \varepsilon$. □

We now turn to the structure of Z_{ab} and Proposition 2.6. We will divide the proof into several parts.

Lemma 4.2. *The set Z_{ab} contains positive (negative, respectively) integers if and only if $c(ab)$ contains a least one positive (negative, respectively) integer.*

Proof. If $c(ab)$ contains one positive integer then Z_{ab} also contains this integer (in fact, one easily verifies that in this case Z_{ab} contains infinitely many positive integers). On the other hand, suppose that $c(ab)$ does not contain any positive integer. Then $(\varepsilon, ab) \in H^1(ab)$ is the maximal edge and, hence, $(\varepsilon, ab)^n \in H^n(ab)$ is the maximal path of length n for all $n \in \mathbb{N}$. By Lemma 3.6 we see that $k \le 0$ for each $k \in c^{(n)}(ab)$ and, hence, $N \le 0$ for all $N \in Z_{ab}$. The negative analogue of the lemma can be shown in a similar way. □

Lemma 4.3. *The set Z_{ab} contains all positive (negative, respectively) integers if and only if $1 \in c(ab)$ ($-1 \in c(ab)$, respectively).*

Proof. At first observe that $1 \in c(ab)$ means that $(b, bb_1) \in H^1(ab)$ (for some $b_1 \in \mathcal{A}$). We show by induction that for each $n \ge 1$ we have

$$\left\{0, \dots, \left|\sigma^{n-1}(b)\right|\right\} \subset c^{(n)}(ab).$$

The case $n = 1$ holds by assumption. Now suppose that we have already proved the assertion for $n-1$. By the definition of $\odot$ and Lemma 4.1 we have

$$\begin{aligned} c^{(n)}(ab) = c^{(n-1)} \odot c(ab) \supset & \{k : k \in c^{(n-1)}(ab), k \ge 0\} \cup \sigma^{n-1}(b) + \{k : k \in c^{(n-1)}(bb_1), k \le 0\} \\ = & \{k : k \in c^{(n-1)}(bb), k \ge 0\} \cup \{k + \sigma^{n-1}(b) : k \in c^{(n-1)}(bb), k \le 0\} \\ = & \{0, \dots, |\sigma^n(b)|\}. \end{aligned}$$

The requirement $b \prec \sigma(b)$ together with the fact that σ is non-erasing yields that $\lim_{n\to\infty} |\sigma^n(b)| = \infty$, thus, Z_{ab} contains all positive integers.

On the other hand suppose that $1 \notin c(ab)$. By the definition of a coding prescription this is equivalent to $1 \notin c(bb)$ which is in turn equivalent to $-|\sigma(b)| + 1 \in c(bb)$ (note that $-|\sigma(b)| + 1$ is the minimum of $c(bb)$). The corresponding edge in $H^1(bb)$ is given by (D, bb_1) with $D \sim \sigma(\bar{b})b$. Since $b \prec \sigma(b)$ we have $D \in_\sim \overline{\mathcal{A}}^+$ and, hence, $(D, bb_1) \in H^1(bb_1)$. Therefore, for each $n \ge 2$ the path $(D, bb_1)^n$ is contained in $H^n_{\sigma,c}(bb)$ and, thus, $c^{(n)}(bb)$ contains

$$\sum_{j=0}^{n-1} |\sigma^n(D)| = \sum_{j=0}^{n-1} \left(\left|\sigma^{j+1}(\bar{b})\right| + \left|\sigma^j(b)\right|\right) = -|\sigma^n(b)| + 1.$$

Since $c^{(n)}$ is a coding prescription with respect to σ^n we conclude that $c^{(n)}(bb)$ and $c^{(n)}(ab)$ do not contain 1 for each $n \ge 1$. This implies that Z_{ab} does not contain 1 and, hence, Z_{ab} does not contain all positive integers.

The negative analogue of the lemma can be shown similarly. □

Lemma 4.4. *Let σ be primitive and c not continuous. If $c(ab)$ contains negative integers such that $\max\{k \in c(ab) : k < 0\} \le -3$ then $Z_{ab} \cap (-\infty, 0)$ contains gaps of arbitrary large size. Analogously, if $c(ab)$ contains positive integers such that $\min\{k \in c(ab) : k > 0\} \ge 3$ then $Z_{ab} \cap (0, \infty)$ contains gaps of arbitrary large size.*

Proof. Let $(D_1, a_1b_1) \in H^1(ab)$ the edge with $|D_1| = \max\{k \in c(ab) : k < 0\} \le -3$. Set $D := D_1b_1$ and observe that $D_1 \prec D \prec \overline{a} \prec \varepsilon$.

We claim that for each $n \ge 1$ and for all $j \ge 1$ we have

$$\{\left|\sigma^j(D)\right|, \dots, \left|\sigma^j(\overline{a})\right|\} \cap c^{(n)}(ab) = \varnothing. \tag{4.1}$$

We will show this by induction on n. For $n = 1$ we have by assumption that $\{|D|, \dots, |\overline{a}|\} \cap c(ab) = \varnothing$ which shows the case $j = 1$. By the definition of coding prescriptions we have $k > -|\sigma(a)|$ for all $k \in c(ab)$. Thus, (4.1) also holds for $j > 1$.

Now we want to show (4.1) for arbitrary n where we may suppose that our claim holds for $n-1$. For $j \ge n$ this is clear since $c^{(n)}$ is a coding prescription with respect to σ^n and, thus, all elements of $c^{(n)}(ab)$ are strictly larger than $|\sigma^n(\overline{a})|$. Thus, we concentrate on the case $j \le n-1$. We have $c^{(n)} = c^{(n-1)} \odot c$. Therefore, each element $k \in c^{(n)}(ab)$ is given by an edge $(D_1', a_1'b_1') \in H^1(ab)$

and a $k' \in c^{(n-1)}(a_1'b_1')$ such that $k = \left|\sigma^{n-1}(D_1')\right| + k'$. If $D_1' \in_\sim \mathcal{A}^+$ then $k > 0$ by Lemma 3.1. If $D_1 \sim \varepsilon$ then $a_1b_1 = ab$ and, hence, $k = k' \in c^{n-1}(ab)$. By the assumption on the induction we have $k \notin \{\left|\sigma^j(D)\right|, \dots, \left|\sigma^j(\overline{a})\right|\}$ for any $j \in \mathbb{N}$. If $D_1 \in_\sim \overline{\mathcal{A}}^+$ then we necessarily have $D_1' \preceq D_1 \prec D$. Thus, by Lemma 3.1, $k = \left|\sigma^{n-1}(D_1')\right| + k' < \left|\sigma^{n-1}(D)\right| \leq \left|\sigma^j(D)\right|$ for all $j \leq n-1$.

Finally, we observe that the primitivity of σ ensures that $\left|\sigma^j(\overline{a})\right| - \left|\sigma^j(D)\right|$ becomes arbitrary large since $D \prec \overline{a}$ and, hence, $\overline{D\overline{a}} \in_\sim \mathcal{A}^+$.

The positive analogue of the lemma can be shown in the same way. ☐

Proof of Proposition 2.6. The items (1) and (2) correspond to Lemma 4.2 and Lemma 4.3, respectively. Lemma 4.4 immediately implies Item 3. Finally, Item 4 follows from Lemma 3.10. ☐

Proof of Proposition 2.7. The equivalence of (2) and (3) follows from the fact that $|\cdot|$ is an order preserving isomorphism between the $\sigma(\bar{a})$-$\sigma(b)$-chain and the $\leq$-chain $\{-|\sigma(a)|, \dots, |\sigma(b)|\}$.

For σ and c as in (4) of Proposition 2.6 we showed the equivalence of (1) and (2) in Lemma 3.10. For continuous coding prescription this can be shown analogously by using Lemma 3.7. ☐

5. Examples, applications and open problems

5.1. **Relation with the results of J.M. Dumont and A. Thomas.** In [1] we find the following approach for representing non-negative integers: for a substitution σ over the alphabet $\mathcal{A}$ we define the *prefix graph* $\tilde{H} = \tilde{H}_\sigma$ to be the directed graph with vertex set $\mathcal{A}$ and an edge from x to x_1 labelled by $(P, x_1) \in \mathcal{A}^* \times \mathcal{A}$ if $Px_1 \preceq \sigma(x)$.

Theorem 5.1 (*cf* [1, 1.5. Théorème])**.** *Let σ be a substitution over the alphabet $\mathcal{A}$ (which is supposed to be primitive) and $x \in \mathcal{A}$ such that $x \prec \sigma(x)$. For each integer $N > 1$ there exists a unique path $(P_j, x_j)_{j=1}^n \in \tilde{H}_\sigma^n(x)$ with $P_1 \not\sim \varepsilon$ such that*

$$N = \sum_{j=1}^{n} \left|\sigma^{n-j}(P_j)\right|. \tag{5.1}$$

We clearly see the parallels to our results and, in fact, the above statement corresponds to Theorem 2.5 for a special coding prescription.

Lemma 5.2. *Let c be the coding prescription that assigns to each pair $ab \in \mathcal{A}^2$ a set of non-negative integers and $H := H_{\sigma,c}$. Then the following items hold:*

- *Let $ab \in \mathcal{A}^2$ and $(D, a_1b_1) \in H^1(ab)$. Then $(D, b_1) \in \tilde{H}^1(b)$ (modulo $\sim$).*
- *Let $x \in \mathcal{A}$ and $(P, x_1) \in \tilde{H}^1(x)$. Then for each a we have $(P, a_1x_1) \in H^1(ax)$ where $a_1 \in \mathcal{A}$ such that $\sigma(\bar{a}) \preceq P\bar{a}_1$ (thus, if $P \not\sim \varepsilon$ then a_1 does not depend on a).*

Proof. The proof is straightforward and is left to the reader. ☐

We clearly see that if the substitution σ satisfies (2.3) then Theorem 5.1 and Theorem 2.5 (for the special coding prescription) are equivalent, that is representations (5.1) and (2.4) of equal length are identical.

On the other hand, the conditions of Theorem 5.1 are weaker than (2.3). To recover the representation (5.1) from Theorem 2.5 we have to consider higher powers of σ since there exists an integer $n > 1$ and a letter $a \in \mathcal{A}$ such that $\sigma^n(\bar{a}) \prec \bar{a}$ (see Example 5.3).

5.2. **Examples concerning Fibonacci numbers.** Consider the sequence $(F_j)_{j\geq 1}$ of Fibonacci numbers defined by $F_0 = 0$, $F_1 = 1$, and $F_j = F_{j-1} + F_{j-2}$ for $j > 1$. Each positive integer can be uniquely represented by the sum of pairwise not consecutive Fibonacci numbers of index at least 2 (Zeckendorf representation, *cf.* [9]). The Fibonacci expansion is the encoding of the Zeckendorf representation as a 0-1 sequence where the rightmost digit corresponds to F_2. Clearly, the Fibonacci expansion of any integer does not contain consecutive occurrences of 1.

Now let σ be the Fibonacci-substitution defined by $\sigma : 1 \mapsto 12, 2 \mapsto 1$ over the alphabet $\mathcal{A} = \{1, 2\}$. It is well known (see the examples in [1]) that for this substitution (5.1) corresponds exactly to the Zeckendorf representation. This is easy to see by considering the prefix-graph $\tilde{H}_\sigma$ (see Figure 1) and the observation $\left|\sigma^i(1)\right| = F_{i+2}$ and $\left|\sigma^i(2)\right| = F_{i+1}$ for each $i \in \mathbb{N}$.

In the following examples we want to study integer numeration with respect to the Fibonacci numbers from the point of view of Theorem 2.5. Observe that we cannot apply it directly to the Fibonacci substitution since neither $\sigma(\bar{1}) \prec \bar{1}$ nor $\sigma(\bar{2}) \prec \bar{2}$ holds. We rather consider $\sigma^2 : 1 \mapsto 121, 2 \mapsto 12$ since then (2.3) is satisfied for $ab = 11$ as well as $ab = 12$.

Example 5.3. The Fibonacci substitution possesses only two different coding prescriptions, namely c_1, that assigns to each pair $ab \in \mathcal{A}^2$ a set of non-negative integers, and c_2 that assigns to each pair $ab \in \mathcal{A}^2$ a set of non-positive integers. In the present example we are interested in the first one, that is $c_1(11) = c_1(21) = \{0,1\}$ and $c_1(22) = c_1(12) = \{0\}$. The associated graph H_{σ,c_1} is depicted in Figure 1 (centre). We see that the prefix graph left of it can be obtained by projecting each pair $ab \in \mathcal{A}^2$ onto its right component b. Note that in Example 5.6 we will use c_2.

The coding prescription $c_1^{(2)}$ with respect to σ^2 can be easily obtained by considering the paths of length 2 in H_{σ,c_1} (see Theorem 2.4). Less surprisingly, $c_1^{(2)}$ also assigns to each pair of letters a set of non-negative integers. We find the associated graph $H := H_{\sigma^2,c_1^{(2)}}$ on the right in Figure 1.

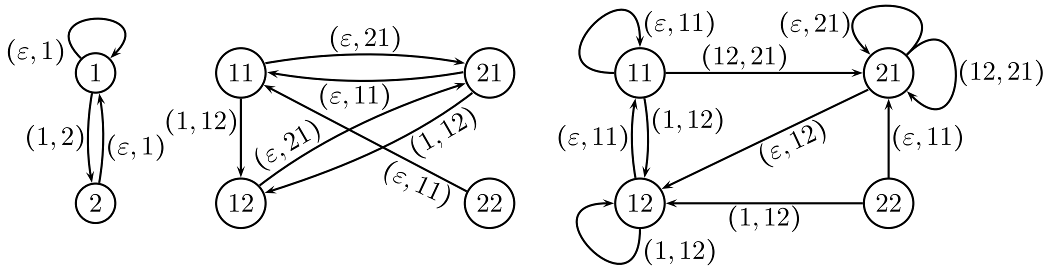


FIGURE 1. Graphs related with the Fibonacci substitution $\sigma : 1 \mapsto 12, 2 \mapsto 1$. Left: the prefix graph $\tilde{H}_\sigma$; centre: the graph H_{σ,c_1} for c_1 the coding prescription that assigns to each pair a set of non-negative integers; right: the graph $H_{\sigma^2,c_1^{(2)}}$.

Now we apply Theorem 2.5. By comparing $H^n(11)$ and $H^n(21)$ for each $n \geq 1$ one easily verifies that the choice of the initial point is irrelevant. We choose $ab = 11$. We have $Z_{11} = \{0, 1, 2 \ldots, \}$ (since $c(11)$ consists of non-negative integers only and $1 \in c(11)$). For each $N \in Z_{11}$ there is a uniquely determined path $(D_j, a_j b_j)_{j=1}^n \in H^n(11)$ with $D_1 \not\sim \varepsilon$ that satisfies

$$N = \sum_{j=1}^{n} \left|\sigma^{2(n-j)}(D_j)\right|.$$

We clearly have that $\left|\sigma^{2(n-j)}(1)\right| = F_{2(n-j)+2}$ and $\left|\sigma^{2(n-j)}(12)\right| = \left|\sigma^{2(n-j)} \circ \sigma(1)\right| = F_{2(n-j)+3}$ for each $j \geq 1$. Therefore, the setting also yields the Zeckendorf representation. Each edge represents two digits of the Fibonacci expansion: $(D_j a_j b_j)$ corresponds to 00 if $D_j \sim \varepsilon$, $D_j \sim 1$ yields the string 01 and $D_j \sim 12$ gives 10. One easily verifies that no consecutive digits 1 can occur.

Of course, applying Theorem 5.1 on σ^2 would have led to the same result as in our previous example. Hence, the coding prescription $c^{(2)}$ is less interesting. Fortunately there are all in all 8 different coding prescriptions with respect to σ^2. For each of them Theorem 2.5 yields a different way of representing integers. We study two more settings explicitly. It will turn out that they behave completely differently.

Example 5.4. Let $c(11) = c(12) := \{-1, 0, 1\}$ and $c(21) = c(22) := \{0, 1\}$. Clearly, c is a coding prescription with respect to σ^2. Less obvious, c is actually a composition of the two coding prescriptions with respect to σ. More precisely, one readily verifies that $c = c_1 \odot c_2$ (*cf.* Example 5.3). The associated graph $H = H_{\sigma^2,c}$ is depicted in Figure 2 (left).

Due to Proposition 2.6 we have $Z_{11} = \mathbb{Z}$ and $Z_{21} = \{0, 1, 2, \ldots\}$. By comparing the finite paths with initial points 11 and 21 we see that starting in 21 does not yield alternative representations. Thus, we apply Theorem 2.5 on $ab = 11$. This allows us to represent each integer $N \in Z_{11} = \mathbb{Z}$ as

$$N = \sum_{j=1}^{n} \left|\sigma^{2(n-j)}(D_j)\right|$$

where $(D_j, a_j b_j)_{j=1}^n \in H^n(11)$. By removing leading occurrences of ε we obtain the unique representation. Since $D_j \in_{\sim} \{\overline{1}, \varepsilon, 1\}$ we have $\left|\sigma^{2(n-j)}(D_j)\right| \in \{-F_{2(n-j+1)}, 0, F_{2(n-j+1)}\}$, thus, N is represented as sum of positive or negative Fibonacci numbers of even index. We also see that the sum does not contain both $-F_{2i}$ and $-F_{2i+2}$, for any positive integer i.

Note that c is continuous. Hence, if

$$N' = \sum_{j=1}^{n} \left|\sigma^{2(n-j)}(D'_j)\right|$$

for a path $(D'_j, a'_j b'_j)_{j=1}^n \in H^n(11)$ then

$$N < N' \iff (D_j)_{j=1}^n \prec_{\text{lex}} (D'_j)_{j=1}^n.$$

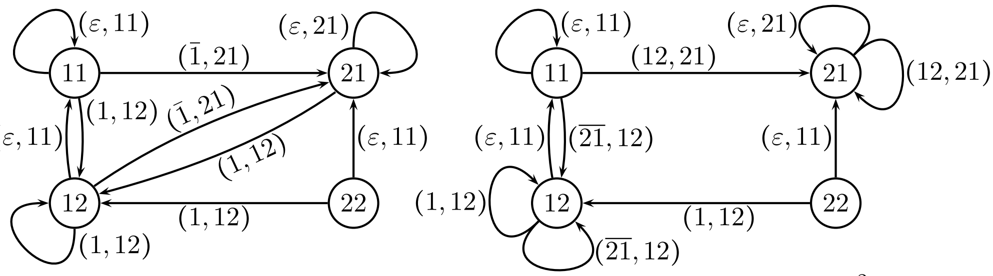


FIGURE 2. Graphs associated with coding prescriptions with respect to σ^2, the square of the Fibonacci substitution. Left: $H_{\sigma^2,c}$ for $c : 11 \mapsto \{-1, 0, 1\}, 22 \mapsto \{0, 1\}$; right: $H_{\sigma^2,c}$ for $c : 11 \mapsto \{-2, 0, 2\}, 22 \mapsto \{0, 1\}$.

Example 5.5. We consider the coding prescription c with respect to σ^2 defined by $c(11) := \{-2, 0, 2\}, c(22) := \{0, 1\}$ (thus, $c(12) = \{-2, 0, 1\}, c(21) = \{0, 2\}$). As in the previous example we can concentrate on the vertex 11. By Proposition 2.6 Z_{11} contains negative integers as well as positive ones but not all all of them.

A more exact characterisation of Z_{11} is a little tricky. We will deduce some properties by analysing $H = H_{\sigma^2,c}$. Consider a finite path $(D_j, a_j b_j)_{j=1}^n \in H^n(11)$ and

$$D := \sigma^{2n-2}(D_1)\sigma^{2n-4}(D_2)\cdots\sigma^2(D_{n-1})D_n,$$

thus $|D| \in c^{(n)}(11)$. We clearly have that the first occurrence of a non-empty word decides whether $D \in_{\sim} \mathcal{A}^+$ (and $|D| > 0$) or $D \in_{\sim} \overline{\mathcal{A}^+}$ (and $|D| < 0$). Especially, $D \in_{\sim} \mathcal{A}^+$ if and only if $a_n b_n = 21$ and by definition this is equivalent to $\varepsilon \preceq D\overline{1} \prec D \prec D2 \preceq \sigma^{2n}(1)$. Thus, each positive element of $c^{(n)}(11)$ corresponds to an occurrence of 21 in $\sigma^{2n}(1)$. Now observe that (2.3) ensures the existence of an infinite word (Fibonacci word) $(u_j)_{j\geq 1}$ over $\mathcal{A}$ that starts with $\sigma^{2n}(1)$ for any n. By the above considerations we see that for a positive integer $N \in Z_{11}$ we clearly have $u_N u_{N+1} = 21$. This shows that Z_{11} does not contain consecutive integers.

On the other hand, $c^{(n)}$ is a coding prescription for each $n \geq 1$. Therefore,

$$\{k - \left|\sigma^{2n}(1)\right| : k \in c^{(n)}(11), k > 0\} \cup \{k \in c^{(n)}(11) : k \leq 0\} = \{-\left|\sigma^{2n}(1)\right| + 1, -\left|\sigma^{2n}(1)\right| + 2, \ldots, -1, 0\}$$

(where the union is disjoint). This immediately shows that the difference between two consecutive negative elements in Z_{11} is at most 2.

In fact, $Z_{11} \cap (0, \infty)$ contains gaps of arbitrary large size. To see this we need another strategy. Consider again an arbitrary path $(D_j, a_j b_j)_{j=1}^n \in H^n(11)$. Then the path corresponds to a non-negative element of $c^{(n)}(11)$ if and only if $D_j \in_{\sim} \{\varepsilon, 12\}$ holds for all $j \in \{1, \ldots, n\}$. Therefore the largest element of $c^{(n)}(11)$ is given by the path $(12, 21)^n$ that yields (since $\sigma^i(12) = F_{i+3}$ for $i \geq 0$)

$$\max c^{(n)}(11) = \sum_{j=1}^{n} \left|\sigma^{2(n-j)}(12)\right| = \sum_{j=1}^{n} F_{2j+1} = F_{2n+2} - F_2 = F_{2n+2} - 1,$$

while the smallest positive integer in $c^{(n+1)}(11) \setminus c^{(n)}(11)$ obviously corresponds to the path $(12, 21), (\varepsilon, 21)^n$ and, thus, equals $\left|\sigma^{2n}(12)\right| = F_{2n+3} > F_{2n+2} - 1$. We see that Z_{11} does not contain integers between $F_{2n+2}-1$ and F_{2n+3} for any $n \geq 1$. As $F_{2n+3} - F_{2n+2} = F_{2n+1}$ we conclude that for growing n this gap becomes arbitrary large.

Without any problem we can define Fibonacci numbers with negative indices in the same way, thus, by setting $F_n = F_{n-1} + F_{n-2}$ for all $n \in \mathbb{Z}$ (with $F_0 = 0$, $F_1 = 1$). One easily verifies that $F_{-n} := (-1)^{n-1} F_n$ holds for all $n \in \mathbb{Z}$. The Fibonacci numbers with negative indices are known as the *negaFibonacci numbers*. Knuth states in [4] that each integer can be represented uniquely as a sum of pairwise not consecutively indexed negaFibonacci numbers. The *negaFibonacci expansion* is the corresponding $0-1$-string where the rightmost digit corresponds to $F_{-1} = 1$. In the last example concerning Fibonacci numbers we will see that Theorem 2.5 also covers the representation by negaFibonacci numbers.

Example 5.6. Let $\sigma' : 1 \mapsto 21, 2 \mapsto 1$ denote the flipped Fibonacci substitution. Note that for each integer $i > 1$ we have $|\sigma'^i(1)| = |\sigma^i(1)| = F_{i+2}$ and $|\sigma'^i(2)| = |\sigma^i(2)| = F_{i+1}$. The function c' with domain $\{1,2\}^2$ defined by $c'(11) = c'(21) := \{0,1\}$, $c'(12) = c'(22) := \{0\}$ is obviously a coding prescription with respect to σ'.

We again need the Fibonacci-substitution $\sigma : 1 \mapsto 12, 2 \mapsto 1$ and the coding prescription c_2 defined by $c_2(11) = \{-1,0\}$ and $c_2(22) = \{0\}$ (*cf.* Example 5.3). We easily see that $\sigma' \circ \sigma(1) = 211$ and $\sigma' \circ \sigma(2) = 21$. For determining the coding prescription $c' \odot c_2$ with respect to $\sigma' \circ \sigma$ we need the graph associated with c_2 (see Figure 3 left). We obtain

$$\begin{aligned}
c' \odot c_2(11) &= \sigma'(\overline{2}) + c'(12) \cup \sigma'(\varepsilon) + c'(21) = \{-1\} \cup \{0,1\} = \{-1,0,1\},\\
c' \odot c_2(12) &= \sigma'(\overline{2}) + c'(12) \cup \sigma'(\varepsilon) + c'(21) = \{-1\} \cup \{0,1\} = \{-1,0,1\},\\
c' \odot c_2(21) &= \sigma'(\varepsilon) + c'(11) = \{0,1\} = \{0,1\},\\
c' \odot c_2(22) &= \sigma'(\varepsilon) + c'(11) = \{0,1\} = \{0,1\}.
\end{aligned}$$

The right part in Figure 3 shows the associated graph $H := H_{\sigma' \circ \sigma, c' \odot c_2}$.

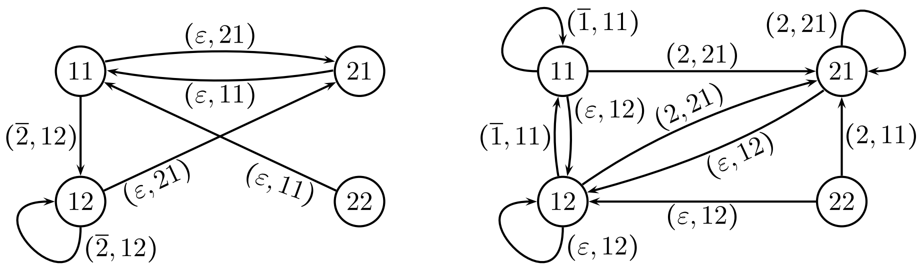


FIGURE 3. On the left the graph H_{σ, c_2} is depicted, on the right we see the graph associated with the composed coding prescription $c' \odot c_2$ with respect to $\sigma' \circ \sigma$.

Condition (2.3) holds for $ab = 12$, thus, we can apply Theorem 2.5 in order to represent integers. As $c' \odot c_2(12) = \{-1,0,1\}$ we see that $Z_{12} = \mathbb{Z}$, hence, we can represent all integers by paths that start in 12. Consider a path $(D_j, a_j b_j)_{j=1}^n \in H^n(12)$. The corresponding integer is given by

$$N := \sum_{j=1}^{n} \left|(\sigma' \circ \sigma)^{n-j}(D_j)\right|. \tag{5.2}$$

We have $D_j \in_\sim \{\overline{1}, \varepsilon, 2\}$ for all $j \in \{1, \ldots, n\}$. Now observe that

$$\begin{aligned}
\left|(\sigma' \circ \sigma)^{n-j}(2)\right| &= \left|\sigma^{2(n-j)}(2)\right| = F_{2n-2j+1} = F_{-(2n-2j+1)},\\
\left|(\sigma' \circ \sigma)^{n-j}(\overline{1})\right| &= \left|\sigma^{2(n-j)}(\overline{1})\right| = -F_{2n-2j+2} = F_{-(2n-2j+2)}.
\end{aligned}$$

We also see that if $D_j \sim 2$ then $D_{j+1} \nsim \overline{1}$, thus, two consecutively indexed negaFibonacci numbers cannot occur in the sum. Thus, if we remove the leading empty words in (5.2) we obtain the negaFibonacci representation due to uniqueness. Each addend in (5.2) represents two digits in the negaFibonacci expansion. To recover it (from left to right) we have to replace $\overline{1}$ by 10, ε by 00, and 2 by 01. Since $c' \odot c_2$ is continuous we also see by Proposition 2.7 that comparing two integers corresponds to the (lexicographical) comparison of pairs of digits in the negaFibonacci expansions (where, loosely speaking, $10 < 00 < 01$ holds).

5.3. **Expansions with respect to a linear recurrence.** Instead of the Fibonacci sequence we can also consider another linear recurrent sequence. For example, define $(G_j)_{j\geq 0}$ by $G_0 := 1, G_1 := 5$ and $G_j = 4G_{j-1} + 3G_{j-2}$ for $j > 1$. Following [5], each positive integer N can be expressed by its *G-ary representation*

$$N = \sum_{j=1}^{n} d_j G_{n-j} \qquad d_j \in \{0, \dots, 4\}$$

such that $d_j d_{j+1} \leq_{\text{lex}} 4, 2$ holds for all $j \in \{1, \dots, n-1\}$. The integers d_j are the so-called *G-ary digits*. Due to [2] we can recover the G-ary representation by applying Theorem 5.1 on the substitution $1 \mapsto 11112, 2 \mapsto 111$.

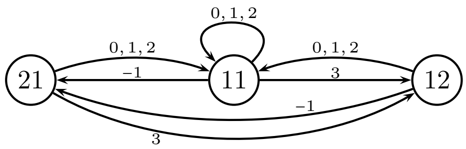
Now let $\sigma : 1 \mapsto 11121, 2 \mapsto 111$ and observe that $|\sigma^j(1)| = G_{j+1}$. Since $1 \prec \sigma(1)$ and $\sigma(\bar{1}) \prec \bar{1}$ we can apply Theorem 2.5 with the initial vertex 11. In the following examples we do this for two different coding prescriptions in order to represent integers with respect to our recurrent sequence $(G_j)_{j\geq 0}$ in an alternative way.

Example 5.7. In the first example we consider the coding prescription c defined by $c(11) = \{-1, 0, 1, 2, 3\}$ and $c(22) = \{0, 1, 2\}$. Since $-1, 1 \in c(11)$ we have $Z_{11} = \mathbb{Z}$, hence, for each integer N there exists a unique path $(D_j, a_j b_j)_{j=1}^n \in H_{\sigma,c}^n(11)$ with $D_j \not\vdash \varepsilon$ such that

$$N = \sum_{j=1}^{n} \left|\sigma^{n-j}(D_j)\right|.$$

Now observe that, by construction, $\bar{1} \preceq D_j \preceq 111$ for each $j \in \{1, \dots n\}$. Thus, $\left|\sigma^{n-j}(D_j)\right| = |D_j| G_{n-j}$. In other words, the sequence $(k_j)_{j=1}^n$ with $k_j = |D_j|$ for each $j \in \{1, \dots, n\}$ is a digit sequence, analogue to the G-ary digits, but with digits from -1 to 3.

To characterise the digit strings that may occur we consider the graph $H_{\sigma,c}$. and replace each label (D, ab) by the corresponding digit $|D|$. It is depicted below. Observe that we skipped the vertex 22 since it is not reachable from 11. Furthermore, we joined edges with the same origin end destination.

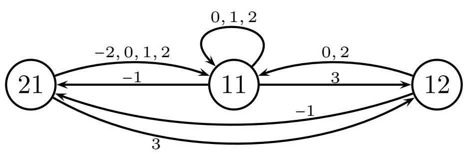


Without any difficulty we see that the digit sequences comply with the lexicographical order condition: for each $N \in \mathbb{Z}$ there is exactly one sequence $(k_j)_{j=1}^n \in \{-1, \dots, 3\}^*$ with $k_1 \neq 0$ and $-1, 0 \leq_{\text{lex}} k_j, k_{j+1} \leq_{\text{lex}} 3, 2$ for each $j \in \{1, \dots, n-1\}$ such that $N = \sum_{j=1}^n k_j G_{n-j}$. On the other hand, each digit sequence with these properties really occurs as a path that starts in 11.

Note that c is continuous, hence, the canonical ordering of the integers is reflected by the lexicographical ordering of the corresponding digit sequences (where we possibly have to fill up with leading zeros to obtain sequences of the same length).

Example 5.8. For the same substitution we now consider the coding prescription $c(11) = \{-1, 0, 1, 2, 3\}$ and $c(22) = \{-2, 0, 2\}$. There are a lot of parallels to the previous example. As before we can apply Theorem 2.5 with initial point 11 and since $c(11)$ did not change we have $Z_{11} = \mathbb{Z}$.

Each label of $H_{\sigma,c}$ has the shape (D, ab) with $\bar{1}\bar{1} \preceq D \preceq 111$, thus, $|\sigma^j(D)| = |D| G_j$ and $-2 \leq |D| \leq 3$. Again we obtain a digit representation with respect to the linear recurrent sequence $(G_j)_{j\geq 1}$. The graph for the digit sequences has a similar shape as in Example 5.7, however, we see that the digit sequences are not characterised by a lexicographical order condition.



Since c is not continuous we cannot expect that the lexicographical order of the digit sequences corresponds to the canonical order of the integers. Indeed, when we consider the digit sequences $3, 2$ and $1, -1, -2$ then we clearly have $0, 3, 2 <_{\text{lex}} 1, -1, -2$. But obviously

$$3G_1 + 2G_0 = 15 + 2 = 17 > G_3 - 1G_2 - 2G_1 = 23 - 5 - 2 = 16.$$

5.4. **Decomposition of coding prescriptions.** In this last subsection we want to state some questions that arise in context with the composition of coding prescriptions. In fact, we already touched on this problem in Example 5.4. Consider a substitution σ and two different coding prescriptions c_1, c_2 with respect to σ. Then $\odot$ yields 4 coding prescriptions with respect to σ^2, namely, $\sigma_1^{(2)}$, $\sigma_2^{(2)}$, $\sigma_1 \odot \sigma_2$ and $\sigma_2 \odot \sigma_1$. In general these coding prescriptions are pairwise different (but it is not clear whether this is always the case). However, there are other coding prescriptions with respect to σ^2 that are not given as an $\odot$-product of coding prescriptions with respect to σ.

These considerations can be extended to any arbitrary power of σ^n. Therefore, it seems to be reasonable to call a coding prescription c with respect to σ^n *reducible (over σ)* when there exist coding prescriptions c_1 and c_2 with respect to σ^{n_1} and σ^{n_2}, respectively, with $n_1 + n_2 = n$ and $c_1 \odot c_2 = c$. We call a coding prescription with respect to σ^n *irreducible* when it is not reducible. Our first question concerns the characterisation of irreducible coding prescriptions.

Problem 5.9. Give a characterisation of the irreducible coding prescriptions with respect to a substitution σ^n.

Note that the coding prescriptions that assign to each pair a set of non-negative (non-positive, respectively) integers are clearly reducible. Beside this obvious fact there does not seem to be an easy answer to this question.

The distinction between reducible and irreducible coding prescriptions allows us to represent each coding prescription c with respect to σ^n as $\odot$-product of irreducible coding prescriptions. More precisely, there exist irreducible coding prescriptions $c_1, \dots, c_j$ with respect to $\sigma^{n_1}, \dots, \sigma^{n_j}$ and positive powers $p_1, \dots, p_j$ such that $c = c_1^{(p_1)} \odot c_2^{(p_2)} \odot \cdots c_j^{(p_j)}$ (hence, $n = k_1 n_1 + \cdots k_j n_j$). Note that this product is in general not commutative. The coding prescription c is itself irreducible if and only if $j = 1$, $n_1 = n$ and $p_1 = 1$.

Problem 5.10. Is this decomposition unique?

References

[1] J.-M. Dumont and A. Thomas, *Systèmes de numération et fonctions fractales relatifs aux substitutions*, Theor. Comput. Sci., 65 (1989), pp. 153–169.
[2] ———, *Digital sum problems and substitutions on a finite alphabet*, J. Number Theory, 39 (1991), pp. 351–366.
[3] P. Grabner and R. Tichy, *Contributions to digit expansions with respect to linear recurrences*, J. Number Theory, 36 (1990), pp. 160–169.
[4] D. E. Knuth, *The Art of Computer Programming, Volume 4, Fascicle 1: Bitwise Tricks & Techniques; Binary Decision Diagrams*, Addison-Wesley Professional, 12th ed., 2009.
[5] A. Pethö and R. F. Tichy, *On digit expansions with respect to linear recurrences*, J. Number Theory, 33 (1989), pp. 243–256.
[6] M. Queffélec, *Substitution Dynamical Systems—Spectral Analysis*, vol. 1294 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, second ed., 2010.
[7] G. Rauzy, *Nombres algébriques et substitutions*, Bull. Soc. Math. France, 110 (1982), pp. 147–178.
[8] P. Surer, *Coding of substitution dynamical systems as shifts of finite type*, Ergodic Theory Dyn. Syst., 36 (2016), pp. 944–972.
[9] E. Zeckendorf, *Représentation des nombres naturels par une somme de nombres de Fibonacci ou de nombres de Lucas*, Bull. Soc. R. Sci. Liège, 41 (1972), pp. 179–182.

Institut für Mathematik, Universität für Bodenkultur (BOKU), Gregor-Mendel-Strasse 33, 1180 Wien, Austria

E-mail address: paul@surer.at

URL: http://www.palovsky.com