Complete codes in a sofic shift

Marie-Pierre Béal and Dominique Perrin

Institut Gaspard-Monge,
University of de Marne-la-Vallée, France
{beal,perrin}@univ-mlv.fr

Abstract. We define a code in a sofic shift as a set of blocks of symbols of the shift such that any block of the shift has at most one decomposition in code words. It is maximal if it is not strictly included in another one. Such a code is complete in the sofic shift if any block of the shift occurs within some concatenation of code words. We prove that a maximal code in an irreducible sofic shift is complete in this shift. We give an explicit construction of a regular completion of a regular code in a sofic shift. This extends the well known result of Ehrenfeucht and Rozenberg to the case of codes in sofic systems. We also give a combinatorial proof of a result concerning the polynomial of a code in a sofic shift.

1 Introduction

In this paper, we continue the study of codes in sofic shifts initiated in [1]. This generalization of the theory of (variable length) codes extends previous works of Reutenauer [2], Restivo [3] and Ashley *et al.* [4]. The main result of this paper is an extension of a classical result of Schützenberger [5] relating the notions of completeness and maximality of codes.

Let S be a sofic shift, *i.e.* the set of bi-infinite sequences of symbols labelling paths in a finite automaton. The set of factors of S, denoted by $\mathrm{Fact}(S)$, is the set of blocks appearing in the elements of S. We call S-code a set of elements of $\mathrm{Fact}(S)$ such that any element of $\mathrm{Fact}(S)$ has at most one decomposition in code words. A set of words X is S-complete if any element of $\mathrm{Fact}(S)$ occurs within some concatenation of elements of X. An S-code is maximal if it is maximal for inclusion.

We prove that, for any irreducible sofic shift S, any maximal S-code is S-complete. Moreover, we give an effective embedding of a regular S-code into an S-complete one. This extends the well known theorem of Ehrenfeucht and Rozenberg [6] to codes in a sofic shift.

Our definition of S-codes generalizes the notion introduced by Restivo [3] and Ashley *et al.* [4]. In the first place, they consider subshifts of finite type instead of the more general notion of sofic shifts. Although shifts of finite type can also be described by a finite automaton, there is a real gap between the two classes. Indeed, representations of shifts of finite type have nice strong properties of synchronization that do not have general sofic shifts. These properties are used to complete the codes. In the second place, they consider codes such that all concatenations of code words are in $\mathrm{Fact}(S)$, a condition that we do not impose. Our definition here is also slightly more general than the one used in our previous paper [1]. Indeed, we only require the unique factorization for the words of $\mathrm{Fact}(S)$ and not for all products of code words. We think that this definition is more natural. The results of [1] all extend straightforwardly to this new class.

In the last section, we give a combinatorial proof of the main result of our previous paper [1] concerning the polynomial of a finite code. This proof is interesting because it is simpler and also because it relates our result to ones due to S. Williams [7] and M. Nasu [8].

The paper is organized as follows. We first recall some basic definitions from the area of symbolic dynamics and from the theory of codes. We introduce the notions of S-code, maximal S-code, and S-complete code when S denotes a sofic shift. In Section 3, we prove that any maximal S-code is S-complete. A combinatorial proof of the result of [1] is given in the last section.

2 Codes and Sofic Shifts

2.1 Sofic Shifts

Let A be a finite alphabet. We denote by A^* the set of finite words, by A^+ the set of nonempty finite words, and by $A^{\mathbb{Z}}$ the set of bi-infinite words on A. A *subshift* is a closed subset S of $A^{\mathbb{Z}}$ which is invariant by the shift transformation σ (*i.e.* $\sigma(S) = S$) defined by $\sigma((a_i)_{i\in\mathbb{Z}}) = (a_{i+1})_{i\in\mathbb{Z}}$.

A finite *automaton* is a finite multigraph labeled on a finite alphabet A. It is denoted $\mathcal{A} = (Q, E)$, where Q is a finite set of states, and E a finite set of edges labeled by A. All states of these automata can be considered as both initial and final states.

A *sofic shift* is the set of labels of all bi-infinite paths in a finite automaton. A sofic shift is *irreducible* if there is such a finite automaton with a strongly connected graph. In this case the automaton also is said to be irreducible. An automaton $\mathcal{A} = (Q, E)$ is deterministic if, for any state $p \in Q$ and any word u, there is at most one path labeled u and going out of p. When it exists, the target state of this path is denoted by $p \cdot u$. An automaton is *unambiguous* if there is at most one path labeled by u going from a state p to a state q for any given triple p, u, q. Irreducible sofic shifts have a unique (up to isomorphisms of automata) *minimal deterministic automaton*, that is a deterministic automaton having the fewest states among all deterministic automata representing the shift. This automaton is called the *Fischer cover* of the shift. A *subshift of finite type* is defined as the bi-infinite words on a finite alphabet avoiding a finite set of finite words. It is a sofic shift. The *full shift* on the finite alphabet A is the set of all bi-infinites sequences on A, *i.e.* the set $A^{\mathbb{Z}}$.

The *entropy* of a sofic shift S is defined as

$$h(S) = \lim_{n\to\infty} \frac{1}{n} \log_2 s_n,$$

where s_n is the number of words of length n of $\mathrm{Fact}(S)$. The Fischer cover of a transitive sofic shift of null entropy is made of one cycle.

Example 1. Let S be the irreducible sofic subshift on $A = \{a, b\}$ defined by the automaton on the left of Figure 1. This automaton is the Fischer cover of S. This shift is the so-called *even system* since its bi-infinite sequences are those having an even number of b's between two a's. It is not a shift of finite type.

Let T be the irreducible shift on $A = \{a, b\}$ defined by the forbidden block bb. It is a shift of finite type. Its Fischer cover is given on the right of Figure 1. This shift is the so-called *golden mean system*.

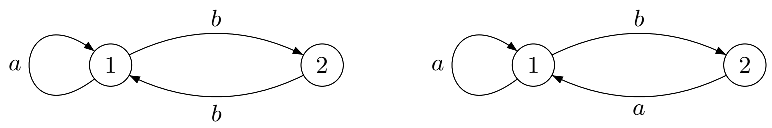

Fig. 1. The Fischer covers of the even system S on the left, and of the golden mean system T on the right.

Let S be a subshift on the alphabet A. We denote by $\mathrm{Fact}(S)$ the set of finite factors (or blocks) of elements of S. Each element of $\mathrm{Fact}(S)$ is the label of a finite path of the Fischer cover of S.

Let $\mathcal{A}$ be a finite automaton. A word w is said to be a *synchronizing* word of $\mathcal{A}$ if and only if any path in $\mathcal{A}$ labeled by w ends in a same state depending only on w. If p denotes this states, one says that w *synchronizes* to p. For instance the words a, bab are synchronizing words of the Fischer cover of the even system. In the golden mean shift, which is a shift of finite type, each word of length 1, *i.e.* a or b, is a synchronizing word. For any Fischer cover of a shift of finite type, there is a positive integer k such that any word of length k is synchronizing.

Let L be a language of finite words. A word w is a *synchronizing* word of L if and only if whenever u, v are words such that uw and wv belong to L, one has uwv belongs to L. Note that if w is a synchronizing word of an automaton $\mathcal{A}$ recognizing a sofic shift S, it is a synchronizing word of the language $\mathrm{Fact}(S)$.

It is known that the Fischer cover of an irreducible sofic shift S has a synchronizing word [9, Proposition 3.3.16]. If w is one of them, for any words u, v such that $uwv \in \mathrm{Fact}(S)$, uwv is a synchronizing word also.

2.2 Codes

Let S be a sofic shift. A set of finite words $X \subset \mathrm{Fact}(S)$ on an alphabet A is an *S-code* if and only if whenever $w = x_1x_2 \ldots x_n = y_1y_2 \ldots y_m$, where $x_i, y_j \in X$, n, m are positive integers, and $w \in \mathrm{Fact}(S)$, one has $n = m$ and $x_i = y_i$ for $1 \leq i \leq n$. Thus the classical definition of a code corresponds to the case where S is the full shift. Any code is an S-code but the converse is false as shown with the following example.

Example 2. The set $\{a, ab, ba\}$ is not a code but it is not difficult to see that it is an S-code in the even system. Indeed any word with two factorizations contains the block aba.

Let S be a sofic shift. A set X on the alphabet A is said to be *complete* in S, or *S-complete*, if X is an S-code and any word in $\mathrm{Fact}(S)$ is a factor of a word in X^*. For instance the code $X = \{a, bb\}$ is complete in the even system.

An S-code X is *maximal* if it is not strictly included in another S-code.

In [2] is an example of an S-complete code which is not maximal. Indeed, let us consider the shift of finite type S defined on the alphabet $A = \{a, b\}$ and avoiding the blocks aa and bb. The S-code $X = \{ab\}$ is S-complete but not maximal since X is strictly included in the S-code $Y = \{ab, ba\}$.

There is a connection between complete S-codes and a concept which has been studied in symbolic dynamics. This explains why the results proved in Section 4 are related with the results of Williams [7] and Nasu [8]. Let X be a complete S-code. Let $\mathcal{A} = (Q, E)$ be the Fischer cover of S. We build an automaton $\mathcal{B}$ computed from X and $\mathcal{A}$ as follows. The set of states of $\mathcal{B}$ contains the set of states Q of $\mathcal{A}$. For each path in $\mathcal{A}$ labeled by a word in X going from a state p to a state q, we build a path in $\mathcal{B}$ from p to q with dummy states inbetween. Let T be the subshift of finite type made of the bi-infinite paths of the graph of $\mathcal{B}$. The labelling of the paths in the automaton $\mathcal{B}$ defines a block map ϕ from T to S. The set X is an S-code if and only if ϕ is finite-to-one. It is S-complete if and only if ϕ is onto. Thus statements on complete S-codes can be reformulated as statements on finite-to-one factor maps between irreducible sofic shifts.

3 Completion of an S-Code

The following result generalizes the theorem of Ehrenfeucht and Rozenberg [6]. As in the case of the extension to subshifts of finite type obtained in [4], the proof uses the same type of construction as the one of [6]. It requires however, as we shall see, a careful adaptation to extend to sofic shifts.

Theorem 1. *Let S be an irreducible sofic shift. If X is an S-code, there is an S-code Y such that $X \subseteq Y$ and Y is S-complete. If X is moreover regular, Y can be chosen regular and is computable in an effective way.*

A nonempty word w of A^* is called *unbordered* if no proper nonempty left factor of w is a right factor of w. In other words, w is unbordered if and only if $w \in uA^+ \cap A^+u$ implies $u = \varepsilon$, where ε denotes the empty word.

The following lemma provides the construction of an unbordered word in the set of factors of an irreducible sofic shift. It replaces the construction used in [5, Proposition 3.6] for the case of the full shift.

Lemma 1. *Let S be an irreducible sofic shift which has a positive entropy. Let z be a word in $\mathrm{Fact}(S)$. Then there is a word y in $\mathrm{Fact}(S)$ such that z is a factor of y and y is unbordered.*

Proof. Let $\mathcal{A}$ be the Fischer cover of S. Let m be the number of states of $\mathcal{A}$ and let k be the length of z. Since S has a positive entropy, there are two distinct nonempty words u, v labels of first return paths in $\mathcal{A}$ to state p. The words u and v are not two powers a same word since $\mathcal{A}$ is deterministic. Moreover $\{u, v\}^*$ is a submonoid of $\mathrm{Fact}(S)$. Let $w = u^{k+m}v^{k+m}$. Since $k + m \geq 2$, by [10, Theorem 9.2.4 pp. 166], w is a primitive word. It follows that the Lyndon word w' conjugate to w is unbordered (see for instance [10, Proposition 5.1.2 p. 65]). Since $\mathcal{A}$ is irreducible, there are two words b_1, b_2 of length at most m such that the word $y = w'b_1zb_2w' \in \mathrm{Fact}(S)$.

We claim that y is unbordered. This fact is trivial by considering the length, greater than $2k + 2m$, of w', the length $k + 2m$ of b_1zb_2 and the fact that w' is unbordered.

Proof (Sketch of proof of Theorem 1). Let S be an irreducible sofic shift. We denote by $\mathcal{A}$ the Fischer cover of S. Let X be an S-code.

Let us suppose that X is not S-complete. Consequently there is a word z in $\mathrm{Fact}(S)$ which is not in $\mathrm{Fact}(X^*)$.

We first assume that S has a null entropy. This means that the Fischer cover $\mathcal{A}$ is made of a unique cycle. One can assume that there is a state p such that p has no outgoing path in $\mathcal{A}$ labeled in X. Otherwise X is already S-complete. Since $\mathcal{A}$ is irreducible, one can assume without loss of generality that z is the label of a path in $\mathcal{A}$ going from a state p to itself, and that z is moreover a synchronizing word of $\mathcal{A}$. We set $Y = X \cup \{z\}$. Now we show that Y is an S-code. Assume the contrary and consider a relation

$$x_1x_2 \ldots x_n = y_1y_2 \ldots y_m,$$

with $x_1x_2 \ldots x_n \in \mathrm{Fact}(S)$, $x_i, y_j \in Y$, and $x_n \neq y_m$. The set X being an S-code, at least one of the words x_i, y_j must be z. Hence, for instance $x_1x_2 \ldots x_n = x_1x_2 \ldots x_r z x_{r+1} \ldots x_n$. The word $zx_{r+1} \ldots x_n$ is the label of a path in $\mathcal{A}$ going through the state p after reading the label z. Since p has no outgoing path in $\mathcal{A}$ labeled in X, it follows that $x_{r+1} \ldots x_n = z^{n-r}$. Hence there is a positive integer k such that $x_1x_2 \ldots x_n = x_1x_2 \ldots x_r z^k$ with $x_1, x_2, \ldots, x_r \neq z$. Since z is not a factor of X^*, there is also a positive integer l such that $y_1y_2 \ldots y_m = y_1y_2 \ldots y_t z^l$ with $y_1, y_2, \ldots, y_t \neq z$. The above relation becomes

$$x_1x_2 \ldots x_r z^k = y_1y_2 \ldots y_t z^l,$$

which contradicts the hypothesis that $x_n \neq y_m$ since $z \notin \mathrm{Fact}(X^*)$. It is trivial that Y is S-complete.

We may now assume that S has a positive entropy. Without loss of generality, by extending z on the right, one can moreover assume that z is a synchronizing word. By Lemma 1, we construct a word $y \in \mathrm{Fact}(S)$ which is unbordered and has z as factor. Moreover y is a synchronizing word of $\mathcal{A}$.

If L is a language of finite words, we denote by $u^{-1}L$ (resp. Lu^{-1}) the set of words z such that $uz \in L$ (resp. $zu \in L$).

We define the sets U and Y by

$$U = y^{-1}\operatorname{Fact}(S)y^{-1} - X^* - A^*yA^*, \tag{1}$$
$$Y = X \cup y(Uy)^*. \tag{2}$$

The rest of the proof consists in verifying the following three properties.

- The set Y is a subset of $\operatorname{Fact}(S)$.
- The set Y is an S-code.
- The set Y is S-complete.

It is clear from Equations (1) and (2) that Y is regular when X is regular. It can be computed in an effective way from these equations. □

Remark 1. Note that our proof shows that, if S is an irreducible sofic shift with a positive entropy, and X is a code, then X can be completed into a code Y (*i.e* a code for the full shift) which is S-complete. We do not know whether this property also holds for irreducible shifs of entropy zero.

In [11, 3] (see also [4]), it is proved that if S is an irreducible shift of finite type and X a code with $X^* \subseteq \operatorname{Fact}(S)$ which is not S-complete, X can be embedded into an S-complete set which is moreover a code (*i.e* a code for the full shift). The proof of our theorem allows us to recover this result. Indeed, when $X^* \subseteq \operatorname{Fact}(S)$, our construction build an S-code Y which is a code. Moreover, the S-complete code Y that we have built satisfies also $Y^* \subseteq \operatorname{Fact}(S)$, when $X^* \subseteq \operatorname{Fact}(S)$. This is due to the strong synchronization properties of the Fischer cover of an irreducible shift of finite type.

Example 3. We consider the even system S of Example 1 on the alphabet $A = \{a, b\}$. Let $X = \{a, ba\}$. The set X is an S-code but it is not S-complete since for instance $z = bb$ does not belong to $\operatorname{Fact}(X^*)$. The regular completion of X is obtained following the proof of Theorem 1. We replace z by bba in order to get a synchronizing word. The proof of Lemma 1 says that the word $a^2b^4bbaa^2b^4$ is an unbordered word of $\operatorname{Fact}(S)$. But a smaller y can be chosen. For instance $y = bba$ also is an unbordered word of $\operatorname{Fact}(S)$. We then define U and Y as in Equations (1) and (2). The set Y is a regular S-complete code.

We derive the following corollary which generalizes to codes in irreducible sofic shifts the fact that any maximal code is complete [5, Theorem 5.1].

Corollary 1. *Let S be an irreducible sofic shift. Any maximal S-code is S-complete.*

4 Polynomial of a Code

In the sequel, S is an irreducible sofic shift recognized by its Fischer cover $\mathcal{A} = (Q, E)$. Let $\mu_{\mathcal{A}}$ (or μ) be the morphism from A^* into $\mathbb{N}^{Q \times Q}$ defined as follows. For each word u, the matrix $\mu(u)$ is defined by

$$\mu(u)_{pq} = \begin{cases} 1 & \text{if } p \cdot u = q \\ 0 & \text{otherwise.} \end{cases}$$

The matrix $\alpha_{\mathcal{A}}(u)$ (or $\alpha(u)$) is defined by $\alpha(u) = \mu(u)u$. Thus the matrix $\alpha(u)$ is obtained from $\mu(u)$ by replacing its coefficients 1 by the word u. The coefficients of $\alpha(u)$ are either 0 or u. In this way α is a morphism from A^* into the monoid of matrices with elements in the set of subsets of A^*.

The morphism α is extended to subsets of A^* by linearity.

For a finite set X, we denote by p_X the polynomial in commuting variables:

$$p_X = \det(I - \alpha(X)).$$

The following result is proved in [1]. It is a generalization of a result of C. Reutenauer [2] who has proved it under more restrictive assumptions.

Theorem 2. *Let S be an irreducible sofic shift and let X be a finite complete S-code. The polynomial p_A divides p_X.*

Example 4. For the even shift and the set $X = \{aa, ab, ba, bb\}$, we have

$$\alpha(A) = \begin{bmatrix} a & b \\ b & 0 \end{bmatrix} \quad \text{and} \quad \alpha(X) = \begin{bmatrix} aa+bb & ab \\ ba & bb \end{bmatrix},$$

and $p_A = 1 - a - bb$, $p_X = 1 - aa - 2bb + b^4 = (1 + a - bb)(1 - a - bb)$.

We present here two combinatorial proofs of this result, which come as an alternative to the analytic proof presented in [1]. Both proofs rely on the reduction of automata with multiplicities.

The first proof goes along the same line as the proof of a result of S. Williams presented in Kitchen's book [12, p. 156], giving a necessary condition to the existence of a finite-to-one factor map between irreducible sofic shifts.

We first build as in Section 2 an automaton $\mathcal{B}$ computed from X and $\mathcal{A}$ as follows. The set of states of $\mathcal{B}$ contains the set of states Q of $\mathcal{A}$. For each path in $\mathcal{A}$ labeled by a word in X going from state p to state q, we build a path in $\mathcal{B}$ from p to q with dummy states inbetween as shown in Example 5. The automaton $\mathcal{B}$ is unambiguous if and only if the set X is an S-code. It represents S if and only if the set X is S-complete.

Example 5. Consider the code $X = \{aa, ab, ba, bb\}$ in the even system S. The automaton $\mathcal{B}$ is represented in the right part of Figure 2.

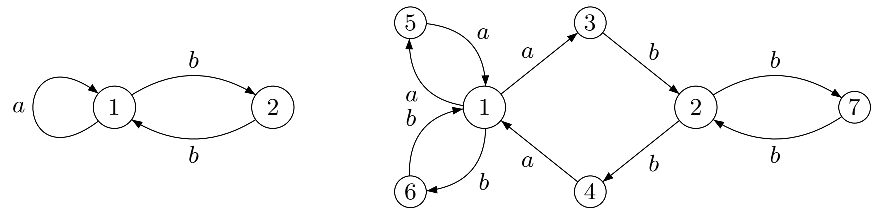

Fig. 2. The automaton $\mathcal{A}$ (on the left), and the automaton $\mathcal{B}$ computed from $\mathcal{A}$ and $X = \{aa, ab, ba, bb\}$ (on the right).

Since X is a complete S-code, $\mathcal{B}$ is unambiguous and represents S. Without loss of generality, one can assume that $\mathcal{B}$ is irreducible. Otherwise, one keeps only a strongly connected component of $\mathcal{B}$ representing S. By construction,

$$p_A = \det(I - \alpha_{\mathcal{A}}(A)) \quad \text{and} \quad p_X = \det(I - \alpha_{\mathcal{B}}(A)).$$

Hence, Theorem 2 is a consequence of the following result.

Proposition 1. *Let S be an irreducible sofic shift and let $\mathcal{A}$ be its Fischer cover. If $\mathcal{B}$ is an unambiguous and irreducible automaton representing S,* $\det(I-\alpha_{\mathcal{A}}(A))$ *divides* $\det(I - \alpha_{\mathcal{B}}(A))$.

Proof (Sketch of proof). The *degree* of a word u in an automaton is defined as the number of paths labeled by u. The degree of an automaton is the minimal non-null value of the degree of words. Any unambiguous irreducible automaton of degree k has the following property: for any word u of degree k and any word w such that uwu has a non-null degree, uwu has degree k.

We first assume that the Fischer cover $\mathcal{A}$ of S is codeterministic (or left resolving): for any state $p \in Q$ and any word u, there is at most one path labeled by u and ending at p. In this case the degree of $\mathcal{A}$ is $d = 1$. Indeed, since $\mathcal{A}$ is a Fischer cover, it has a synchronizing word. Since $\mathcal{A}$ is codeterministic, each synchronizing word has degree 1.

Let v (resp. w) be a word which has a non-null and minimal degree k (resp. $d = 1$) in $\mathcal{B}$ (resp. in $\mathcal{A}$). Since $\mathcal{B}$ is irreducible, there are words z, z' such that $vzwz'v$ has a non-null degree. Hence $vzwz'v$ has degree k in $\mathcal{B}$ and degree $d = 1$ in $\mathcal{A}$. We set $u = vzwz'v$.

An $\mathbb{N}$-automaton with a set of states Q is a triple $\langle I, \mu, T\rangle$, where I and T are two vectors — respectively initial row vector and final column vector — with entries in $\mathbb{N}$, and where μ is a morphism from A^* into $\mathbb{N}^{Q\times Q}$. It is equivalently defined by the triple $\langle I, \alpha(A), T\rangle$. Two $\mathbb{N}$-automata $\langle I, \mu, T\rangle$ and $\langle J, \mu', F\rangle$ are *equivalent* if and only if, for any word $w \in A^*$, $I\mu(w)T = J\mu'(w)F$.

Let $\mathbf{1}_{\mathcal{A}}$ be the row-vector with all coefficients equal to 1 of size the number of states of $\mathcal{A}$, and $\mathbf{1}^t_{\mathcal{A}}$ its transpose. It follows from the definition of the word u that the two $\mathbb{N}$-automata $\mathcal{C} = \langle k\mathbf{1}_{\mathcal{A}}\mu_{\mathcal{A}}(u), \quad \mu_{\mathcal{A}}, \quad \mu_{\mathcal{A}}(u)\mathbf{1}^t_{\mathcal{A}}\rangle$, and $\mathcal{D} = \langle d\mathbf{1}_{\mathcal{B}}\mu_{\mathcal{B}}(u), \quad \mu_{\mathcal{B}}, \quad \mu_{\mathcal{B}}(u)\mathbf{1}^t_{\mathcal{B}}\rangle$, are equivalent.

The standard Schützenberger reductions of the $\mathbb{N}$-automata $\mathcal{C}$ and $\mathcal{D}$ over the field $\mathbb{R}$ are similar. The reduction of each $\mathbb{N}$-automaton is obtained through a left reduction followed by a right reduction (see for instance [13] or [14]).

Since u has degree 1, the initial row (resp. final column) vector of $\mathcal{C}$ has a unique non-null coefficient. Consequently, since $\mathcal{A}$ is deterministic (resp. codeterministic) and irreducible, the automaton $\mathcal{C}$ is left (resp. right) reduced. Hence $\mathcal{C}$ is already reduced.

Finally, one can prove that the transition matrix of $\mathcal{D}$ is similar to a matrix having a principal subblock equal to the transition matrix of its left (or right) reduced form. It follows that $\det(I - \alpha_{\mathcal{A}}(A))$ divides $\det(I - \alpha_{\mathcal{B}}(A))$. The extension to sofic shifts that may not have a codeterministic Fischer cover can be obtained with a specialization argument (see [1]). □

Example 6. We continue with Example 5. The word bab has degree 2 in $\mathcal{B}$ and 1 in $\mathcal{A}$. Hence the $\mathbb{N}$-automata

$$\mathcal{C} = \langle \begin{bmatrix} 0 \ 2 \end{bmatrix},\ \mu_{\mathcal{A}}(A) = \begin{bmatrix} a & b \\ b & 0 \end{bmatrix},\ \begin{bmatrix} 0 \\ 1 \end{bmatrix} \rangle, \text{ and } \mathcal{D} = \langle \begin{bmatrix} 0\ 1\ 0\ 0\ 0\ 1\ 0 \end{bmatrix},\ \mu_{\mathcal{B}}(A), \begin{bmatrix} 0\ 1\ 0\ 0\ 0\ 1\ 0 \end{bmatrix}^t \rangle,$$

are equivalent. We obtain a right-reduction of the automaton $\mathcal{D} = \langle I, E = \alpha_{\mathcal{B}}(A), T \rangle$ by computing a basis of the vector space generated by the vectors in $\mu(A^*)T$. We can choose the basis $(T, \mu(b)T, \mu(ab)T)$ since $\mu(a)T = 0$, $\mu(bb)T = T$, $\mu(bab)T = T$ and $\mu(aab)T = \mu(ab)T$. This basis is extended to a basis of $\mathbb{R}^7$, for instance with the first 4 column vectors $e_1, \dots e_4$ of the canonical basis of $\mathbb{R}^7$.

Let F and H be the matrices

$$F = \begin{bmatrix} \begin{bmatrix} 0 & b & b \\ b & 0 & 0 \\ 0 & a & a \end{bmatrix} & \begin{bmatrix} b & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ a & 0 & 0 & 0 \end{bmatrix} \\ \\ \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} & \begin{bmatrix} -a & -b & a & 0 \\ -b & 0 & 0 & b \\ a & 0 & 0 & 0 \\ -a & 0 & 0 & 0 \end{bmatrix} \end{bmatrix}, \qquad H = \begin{bmatrix} \begin{bmatrix} 0 & b \\ b & a \end{bmatrix} & \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ \\ \begin{bmatrix} 1 & 0 \end{bmatrix} & \begin{bmatrix} 0 \end{bmatrix} \end{bmatrix}.$$

We get that E is similar to F. Let us denote by G the upper left block matrix of size 3 of F. The right-reduced automaton $\langle \left(2\ 0\ 0 \right), G = \left(\begin{smallmatrix} 0 & b & b \\ b & 0 & 0 \\ 0 & a & a \end{smallmatrix} \right), \left(\begin{smallmatrix} 1 \\ 0 \\ 0 \end{smallmatrix} \right) \rangle$ can be now reduced on the left side. We get that G is similar to H. The upper left block matrix of size 2 of G is similar to $\alpha_{\mathcal{A}}(A)$. As a consequence, $\det(I - \alpha_{\mathcal{A}}(A)) = 1 - a - bb$ divides $\det(I - H)$ which divides $\det(I - F) = \det(I - \alpha_{\mathcal{B}}(A)) = (1 - a - bb)(1 + a - bb)$.

A variant of the combinatorial proof uses an argument due to Nasu [8].

We denote by M (resp. M') the matrix $M = \sum_{a \in A} \mu_{\mathcal{A}}(a)$ and (resp. $M' = \sum_{a \in A} \mu_{\mathcal{B}}(a)$). It is known from the Perron-Frobenius theory that M and M' have the same positive spectral radius λ, the logarithm of λ being called the topogical entropy of the sofic shift S [9]. Let U, V (resp. U', V') be two real positive left and right eigenvectors of M (resp. of M') for the eigenvalue λ. One can choose these vectors such that $UV = U'V' = 1$. With these settings, the two $\mathbb{R}$-automata $\mathcal{C} = \langle U, \mu_{\mathcal{A}}, V \rangle$ and $\mathcal{D} = \langle U', \mu_{\mathcal{B}}, V' \rangle$ are equivalent.

The proof of this equivalence relies on the following arguments. One first divides $\mu_{\mathcal{A}}$ and $\mu_{\mathcal{B}}$ by λ to assume that $\lambda = 1$.

For any word $x \in A^*$ and any $\mathbb{R}$-automaton $\mathcal{S} = \langle I, \mu, T \rangle$, we denote by $\pi_{\mathcal{S}}(x)$ the real coefficient $I\mu(x)T$. Hence $\mathcal{C}$ and $\mathcal{D}$ are equivalent if and only if $\pi_{\mathcal{C}}(x) = \pi_{\mathcal{D}}(x)$ for any $x \in A^*$. The functions $\pi_{\mathcal{C}}$ and $\pi_{\mathcal{D}}$ define two rational probability measures on A^* [15]. These measures satisfy the following properties.

- A right (and left) invariance property: for any $x \in A^*$, with $\mathcal{S}$ equal to $\mathcal{C}$ or $\mathcal{D}$.

$$\sum_{w \in A^k} \pi_{\mathcal{S}}(xw) = \pi_{\mathcal{S}}(x).$$

– An ergodic property: for any $x \in A^*$, with $\mathcal{S}$ equal to $\mathcal{C}$ or $\mathcal{D}$.

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \sum_{w \in A^i} \pi_{\mathcal{S}}(xwy) = \pi_{\mathcal{S}}(x)\pi_{\mathcal{S}}(y).$$

Moreover, since the automata $\mathcal{A}$ and $\mathcal{B}$ are unambiguous, one can show that there are positive real numbers ρ, ρ' such that for any $x \in A^*$, $\pi_{\mathcal{C}}(x) \leq \rho\, \pi_{\mathcal{D}}(x)$ and $\pi_{\mathcal{D}}(x) \leq \rho'\, \pi_{\mathcal{C}}(x)$. The equivalence of $\mathcal{C}$ and $\mathcal{D}$ follows from these inequalities. The reduction of the automata is used to finish the proof as before.

Acknowledgments The authors would like to thank an anonymous reviewer for detecting an error in Lemma 1 in a preliminary version of this paper.

References

1. Béal, M.P., Perrin, D.: Codes and sofic constraints. Theoret. Comput. Sci. **340**(2) (2005) 381–393
2. Reutenauer, Ch.: Ensembles libres de chemins dans un graphe. Bull. Soc. Math. France **114**(2) (1986) 135–152
3. Restivo, A.: Codes and local constraints. Theoret. Comput. Sci. **72**(1) (1990) 55–64
4. Ashley, J., Marcus, B., Perrin, D., Tuncel, S.: Surjective extensions of sliding-block codes. SIAM J. Discrete Math. **6**(4) (1993) 582–611
5. Berstel, J., Perrin, D.: Theory of codes. Volume 117 of Pure and Applied Mathematics. Academic Press Inc., Orlando, FL (1985) http://www-igm.univ-mlv.fr/~berstel/LivreCodes/Codes.html.
6. Ehrenfeucht, A., Rozenberg, G.: Each regular code is included in a maximal regular code. RAIRO Inform. Théor. Appl. **20**(1) (1986) 89–96
7. Williams, S.: Lattice invariants for sofic shifts. Ergodic Theory and Dynamical Systems **11** (1991) 787–801
8. Nasu, M.: An invariant for bounded-to-one factor maps between transitive sofic subshifts. Ergodic Theory Dynam. Systems **5**(1) (1985) 89–105
9. Lind, D., Marcus, B.: An Introduction to Symbolic Dynamics and Coding. Cambridge University Press, Cambridge (1995)
10. Lothaire, M.: Combinatorics on words. Volume 17 of Encyclopedia of Mathematics and its Applications. Addison-Wesley Publishing Co., Reading, Mass. (1983)
11. Restivo, A.: Codes with constraints. In: Mots. Lang. Raison. Calc. Hermès, Paris (1990) 358–366
12. Kitchens, B.P.: Symbolic dynamics. Universitext. Springer-Verlag, Berlin (1998) One-sided, two-sided and countable state Markov shifts.
13. Berstel, J., Reutenauer, C.: Rational series and their languages. Volume 12 of EATCS Monographs on Theoretical Computer Science. Springer-Verlag, Berlin (1988)
14. Sakarovitch, J.: Éléments de Théorie des Automates. Vuibert, Paris (2003) english translation to appear, Cambridge University Pres.
15. Hansel, G., Perrin, D.: Mesures de probabilités rationnelles. In: Mots. Lang. Raison. Calc. Hermès, Paris (1990) 335–357