The Degree of a Finite Set of Words

Dominique Perrin
Université Gustave Eiffel, LIGM, Marne-la-Vallée, France
dominique.perrin@esiee.fr

Andrew Ryzhikov
Université Gustave Eiffel, LIGM, Marne-la-Vallée, France
ryzhikov.andrew@gmail.com

Abstract

We generalize the notions of the degree and composition from uniquely decipherable codes to arbitrary finite sets of words. We prove that if $X = Y \circ Z$ is a composition of finite sets of words with Y complete, then $d(X) \leq d(Y) \cdot d(Z)$, where $d(T)$ is the degree of T. We also show that a finite set is synchronizing if and only if its degree equals one.

This is done by considering, for an arbitrary finite set X of words, the transition monoid of an automaton recognizing X^* with multiplicities. We prove a number of results for such monoids, which generalize corresponding results for unambiguous monoids of relations.

2012 ACM Subject Classification Theory of computation → Formal languages and automata theory

Keywords and phrases synchronizing set, degree of a set, group of a set, monoid of relations

Digital Object Identifier 10.4230/LIPIcs.FSTTCS.2020.54

Acknowledgements We thank Jean-Eric Pin and Jacques Sakarovitch for references concerning the composition of automata and transducers.

1 Introduction

Let X be a set of finite words. The set X^* of all concatenations of words in X (often called the Kleene star of X) plays an important role in formal languages theory and its applications. The set X often represents a dictionary or a code transmitted over a channel, so the case where X is finite is especially important. In general, a word in X^* can have several different factorizations over X, and it is useful to understand the relations between them. A word w is called *synchronizing* for X if for any words u, v such that $uwv \in X^*$ we have $uw, wv \in X^*$. In particular, we get that any word in X^* containing ww as a factor, that is, any word of the form $uwwv$, has a factorization where uw and wv are both in X^*, and thus can be factorized separately. A set which admits a synchronizing word is also called *synchronizing*. A set X is called *complete* if every word over the same alphabet occurs as a factor of a word in X^*.

Synchronizing words are studied a lot for uniquely decipherable codes (see e.g., Chapter 10 of [3]). A set X of words is called a *uniquely decipherable code* (often also called a *variable length code*) if every word has at most one factorization over X. Such codes play a crucial role in the theory of data compression and transmission [3].

Provided a set Z of words such that $X \subset Z^*$, one can rewrite X using Z as the alphabet, thus resulting in a new set Y. The representation $X = Y \circ Z$ is then called a decomposition of X, and the converse process of obtaining X is called composition. Decomposition of a set allows to represent it by using simpler sets as building blocks, while preserving many properties of the initial one. Conversely, compositions of codes allow to construct more complicated codes by using simple ones, so they are interesting on their own. In particular, the composition of two uniquely decipherable codes is again a uniquely decipherable code [3]. For any injective morphism $\alpha : A^* \to B^*$, $\alpha(A)$ is a code, and each code can be obtained as the image of A for some A and α [3]. Compositions of codes are then nothing more than

compositions of injective morphisms between free monoids. The notion of composition of two arbitrary finite sets of words is also natural as it corresponds to the composition of arbitrary morphisms.

Our contributions. In this paper, we transfer the notions of composition, degree and group from uniquely decipherable codes to arbitrary finite sets of words. This extends the presentation of [3] made for uniquely decipherable codes.

Provided a finite set X of words, we associate a special automaton $\mathcal{A}$ (called the flower automaton) recognizing X^* with multiplicities. Let S be the set of fixed points of an idempotent e of minimum rank in the transition relation of $\mathcal{A}$, and Γ be the set of strongly connected components of S. We consider a permutation group G_e acting on Γ. We show that all such groups are equivalent for idempotents of minimum rank (Theorem 20). Moreover, we show that for a given X all these groups are equivalent for any trim automaton recognizing X^* with multiplicities (Proposition 21). Thus this group is an invariant of a set. We introduce the degree $d(X)$ of X, which is the minimum rank of elements in the transition monoid of $\mathcal{A}$. We then show that synchronizing sets are exactly sets of degree one (Proposition 22). As our main contribution, we use the obtained results to show that for a composition $X = Y \circ Z$ of two finite sets Y, Z such that Y is complete we have $d(X) \leq d(Y) \cdot d(Z)$ (Theorem 24).

For a finite set X, all these results were previously known only for the special case of X being a uniquely decipherable code with the equality $d(X) = d(Y)d(Z)$ instead of an inequality [3]. Our generalization to the case of an arbitrary finite set requires more complicated proofs. In particular, for uniquely decipherable codes it is enough to consider a trim unambiguous automaton recognizing X^* (which is a cornerstone of the theory), while in our case we need a trim automaton recognizing X^* with multiplicities. Intuitively, such automata count the number of factorizations over X, and thus they are unambiguous when X is a uniquely decipherable code. The technical difficulties then begin with the replacement of unambiguous monoids of relations by arbitrary monoids of relations. Indeed, the multiplication of matrices there is different from the result over the Boolean semiring. In particular, the representation of maximal subgroups by permutations is still possible but more complicated.

Motivation and related results. Larger classes of codes are considered both in theory and in practice. Particular examples include multiset and set decipherable codes. A set X of words is called a *multiset* [10] (respectively, *set* [13]) *decipherable code* if every factorization of a word into codewords provides the same multiset (respectively, set) of codewords. Such codes are used if one needs to transmit only the frequencies (or the fact of occurrences) of elements, but the order of these elements does not matter. Lempel [13] reports online compilations of inventories, construction of histograms, or updating of relative frequencies as particular examples. An important property of multiset decipherable codes is that there exist examples of such codes with Kraft-McMillan sum more than one, which shows that such codes can be more efficient than uniquely decipherable codes [18]. An even wider class is that of numerically decipherable codes, which are sets with the property that every factorization of a word over such set has the same number of codewords [21]. A similar setting of multivalued encodings allows to have several different codewords for the same symbol [4]. In view of that, the transit of results from uniquely decipherable codes to arbitrary sets is interesting.

Another motivation for studying factorizations of words in X^* for an arbitrary finite set X is the area of static dictionary compression, where one looks for some specific factorization of a text over some finite dictionary [1]. The dictionary does not have to be a uniquely

decipherable code, thus a text can have several different factorizations. In this case, it is useful to know the relation between different factorizations. The parallel version of this problem is also considered [15]. In [6] a fast algorithm for checking if a given word w belongs to X^* is suggested. If the answer is positive, it also provides a factorization of w over X.

Only few results are known about decompositions and synchronization of arbitrary sets of words. The defect theorem states that every finite set of words which is not a uniquely decipherable code can be decomposed over a set of smaller size [2]. A survey of different generalizations of this theorem is presented in [11]. Synchronization in arbitrary monoids was studied in [5] and [7]. Other properties of factorizations are studied in [17, 20].

Organization of the paper. To transfer the results from uniquely decipherable code to arbitrary finite sets of words, we first set a correspondence with an adequate class of automata, namely automata recognizing with multiplicities (Sections 2 and 3). Then we introduce the notion of a composition for arbitrary finite sets of words (Section 4). We extend the theory of unambiguous monoids of relations by the theory of arbitrary monoids of relations (Section 5), and generalize the notion of the group $G(X)$ and the degree $d(X)$ of a finite set X of words (Section 6). In this way, as for codes, a set is synchronizing if and only if it is of degree 1 (Section 7). As the main result, we prove that if $X = Y \circ Z$ with Y complete, then $d(X) \leq d(Y) \cdot d(Z)$ (Section 8). In Section 9 we show that if we require Y to be complete, we do not get any new decompositions of a uniquely decipherable code other than into two uniquely decipherable codes.

2 Automata

We denote by A^* the free monoid on a finite alphabet A, by 1 the empty word, and by A^+ the set $A^* \setminus \{1\}$. For notions not defined in this section see [3].

Let $\mathcal{A} = (Q, i, t)$ be an automaton on the alphabet A with Q as set of states, i as initial state and t as terminal state (we will not need to have several initial or terminal states). We do not specify in the notation the set of edges, which are triples (p, a, q) with two states $p, q \in Q$ and a label $a \in A$ denoted $p \xrightarrow{a} q$. We form paths as usual by concatenating consecutive edges. An automaton is called *trim* if there exists a path from i to every state, and from every state to t.

The *language recognized* by $\mathcal{A}$, denoted $L(\mathcal{A})$, is the set of words in A^* which are labels of paths from i to t. There can be several paths from i to t for a given label, and this motivates the introduction of multiplicities.

For a semiring K, a K-subset of A^* is a map from A^* into K. The value of a K-subset X at w is called its *multiplicity* and denoted (X, w). We denote by $K\langle\langle A\rangle\rangle$ the semiring of K-subsets of A^* and by $K\langle A\rangle$ the set of corresponding polynomials, that is the K-subsets with a finite number of words with nonzero multiplicity (on these notions, see [8]).

If X, Y are K-subsets, then $X + Y$ and XY are the K-subsets defined by

$$(X + Y, w) = (X, w) + (Y, w), \quad (XY, w) = \sum_{w=uv} (X, u)(Y, v).$$

Moreover, if X does not have a constant term, that is, if $(X, 1) = 0$, then X^* is the K-subset

$$X^* = 1 + X + X^2 + \dots$$

Since X has no constant term, for every word w, the number of nonzero terms (X^n, w) in the sum above is finite and thus X^* is well-defined.

For a set $X \subset A^*$, we denote by $\underline{X}$ the characteristic series of X, considered as an $\mathbb{N}$-subset. It is easy to verify that for $X \subset A^+$, the mutiplicity of $w \in A^*$ in $\underline{X}^*$ is the number of factorizations of w in words of X.

For an automaton $\mathcal{A} = (Q, i, t)$ on the alphabet A, we denote by $|\mathcal{A}|$ its behaviour, which is an element of $\mathbb{N}\langle\langle A \rangle\rangle$. It is the $\mathbb{N}$-subset of A^* such that the multiplicity of $w \in A^*$ in $|\mathcal{A}|$ is the number of paths from i to t labeled w in $\mathcal{A}$.

We denote by $\mu_{\mathcal{A}}$ the morphism from A^* into the monoid of $Q \times Q$-matrices with integer coefficients defined for $\mu_{\mathcal{A}}(w)_{p,q}$ as the number of paths from p to q labeled by w. Thus, the multiplicity of w in $|\mathcal{A}|$ is $(|\mathcal{A}|, w) = \mu_{\mathcal{A}}(w)_{i,t}$.

Given a set $X \subset A^+$, we say that the automaton $\mathcal{A}$ *recognizes* X^* *with multiplicities* if the behaviour of $\mathcal{A}$ is the multiset assigning to x its number of distinct factorizations in X. Formally, $\mathcal{A}$ recognizes X^* with multiplicities if $|\mathcal{A}| = \underline{X}^*$.

▶ **Example 1.** Let $X = \{a, a^2\}$. The number of factorizations of a^n in words of X is the Fibonacci number F_{n+1} defined by $F_0 = 0$, $F_1 = 1$ and $F_{n+1} = F_n + F_{n-1}$ for $n \geq 1$. The automaton $\mathcal{A}$ represented in Figure 1 recognizes X^* with multiplicities, that is $|\mathcal{A}| = (a+a^2)^*$.

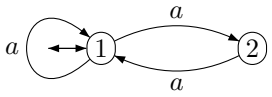


Figure 1 An automaton recognizing X^* with multiplicities.

We have indeed for every $n \geq 1$,

$$\mu_{\mathcal{A}}(a^n) = \begin{bmatrix} F_{n+1} & F_n \\ F_n & F_{n-1} \end{bmatrix}$$

For an automaton $\mathcal{A} = (Q, i, t)$ on the alphabet A, we denote by $\varphi_{\mathcal{A}}$ the morphism from A^* onto the monoid of transitions of $\mathcal{A}$. Thus, for $w \in A^*$, $\varphi_{\mathcal{A}}(w)$ is the Boolean $Q \times Q$-matrix defined by

$$\varphi_{\mathcal{A}}(w)_{p,q} = \begin{cases} 1 & \text{if } p \xrightarrow{w} q, \\ 0 & \text{otherwise} \end{cases}$$

Let $X \subset A^+$ be a finite set of words on the alphabet A. The *flower automaton* of X is the following automaton. Its set of states is the subset Q of $A^* \times A^*$ defined as

$$Q = \{(u, v) \in A^+ \times A^+ \mid uv \in X\} \cup (1, 1).$$

We often denote $\omega = (1, 1)$. There are four type of edges labeled by $a \in A$

$$\begin{aligned} (u, av) &\xrightarrow{a} (ua, v) \quad \text{for } uav \in X,\ u, v \neq 1 \\ \omega &\xrightarrow{a} (a, v) \quad \text{for } av \in X,\ v \neq 1 \\ (u, a) &\xrightarrow{a} \omega \quad \text{for } ua \in X,\ u \neq 1 \\ \omega &\xrightarrow{a} \omega \quad \text{for } a \in X. \end{aligned}$$

The state ω is both initial and terminal.

The proof of the following result is straightforward. It generalizes the fact that the flower automaton of a code recognizes X^* and is unambiguous (see Theorem 4.2.2 in [3]).

▶ **Proposition 2.** *For any finite set $X \subset A^+$, the flower automaton of X recognizes X^* with multiplicities.*

▶ **Example 3.** Let $X = \{a, ab, ba\}$. The flower automaton of X^* is represented in Figure 2. As an example, there are two paths from ω to ω labeled aba, corresponding to the two factorizations $(a)(ba) = (ab)(a)$.

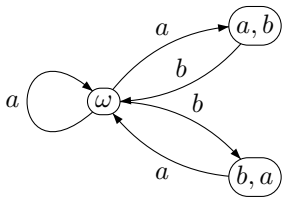


Figure 2 The flower automaton of X (Example 3).

A more compact version of the flower automaton is the *prefix automaton* $\mathcal{A} = (P, 1, 1)$ of a finite set $X \subset A^+$. Its set of states is the set P of proper prefixes of X and its edges are the $p \xrightarrow{a} pa$ for every $p \in P$ and $a \in A$ such that $pa \in P$ and the $p \xrightarrow{a} 1$ such that $pa \in X$. It also recognizes X^* with multiplicities.

▶ **Example 4.** Let $X = \{a^2, a^3\}$. The flower automaton of X is shown in Figure 3 on the left and its prefix automaton on the right.

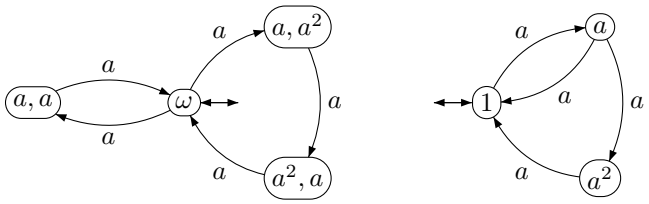


Figure 3 The flower automaton and the prefix automaton of X (Example 4).

A *reduction* from an automaton $\mathcal{A} = (P, i, t)$ onto an automaton $\mathcal{B} = (Q, j, u)$ is a surjective map $\rho : P \to Q$ such that $\rho(i) = j$, $\rho(t) = u$ and such that for every $q, q' \in Q$ and $w \in A^*$, there is a path $q \xrightarrow{w} q'$ in $\mathcal{B}$ if and only if there is a path $p \xrightarrow{w} p'$ in $\mathcal{A}$ for some $p, p' \in P$ with $\rho(p) = q$ and $\rho(p') = q'$.

The reduction is *sharp* if $\rho^{-1}(j) = \{i\}$ and $\rho^{-1}(u) = \{t\}$.

▶ **Proposition 5.** *Let ρ be a reduction from $\mathcal{A} = (P, i, t)$ onto $\mathcal{B} = (Q, j, u)$. Then $L(\mathcal{A}) \subset L(\mathcal{B})$, with equality if ρ is sharp.*

The term reduction is the one used in [3] and it is not standard but captures the general idea of a covering. The term conformal morphism is the one used in [19]. The following statement replaces [3, Proposition 4.2.5].

▶ **Proposition 6.** *Let $X \subset A^+$ be a finite set which is the minimal generating set of X^*. For each trim automaton $\mathcal{B} = (Q, i, i)$ recognizing X^* with multiplicities, there is a sharp reduction from the flower automaton of X onto $\mathcal{B}$.*

Proof. Let $\mathcal{A} = (P, \omega, \omega)$ be the flower automaton of X. We define a map $\rho : P \to Q$ as follows. We set first $\rho(\omega) = i$. Next, if $(u, v) \in P$ with $(u, v) \neq \omega$, then $uv \in X$. Since X is

the minimal generating set of X^*, there is only one factorization of uv into elements of X. Since $\mathcal{B}$ recognizes X with multiplicities, there is only one path $i \xrightarrow{u} q \xrightarrow{v} i$ in $\mathcal{B}$. We define $\rho\big((u,v)\big) = q$.

It is straightforward to verify that ρ is a reduction. Assume first that $q \xrightarrow{w} q'$ in $\mathcal{B}$. Let $i \xrightarrow{u} q$ and $q' \xrightarrow{v'} i$ be simple paths, that is not passing by i except at the origin or the end. Then $i \xrightarrow{uwv'} i$ and thus $uwv' = x_1x_2\cdots x_n$ with $x_i \in X$, u a proper prefix of $x_1 = uv$ and v' a proper suffix of $x_n = u'v'$. Thus $\rho\big((u,v)\big) = q$ and $\rho\big((u',v')\big) = q'$. Since $w = vx_2\cdots x_{n-1}u'$, we have in $\mathcal{A}$ a path $(u,v) \xrightarrow{w} (u',v')$. Conversely, consider a path $(u,v) \xrightarrow{w} (u',v')$ in $\mathcal{A}$. If the path does not pass by ω, then $u' = uw$, $v = wv'$ and we have a path $q \xrightarrow{w} q'$ in $\mathcal{B}$ with $\rho\big((u,v)\big) = q$ and $\rho\big((u',v')\big) = q'$. Otherwise, the path decomposes in $(u,v) \xrightarrow{v} \omega \xrightarrow{x} \omega \xrightarrow{v'} (u',v')$ with $x \in X^*$. Since $\mathcal{B}$ recognizes X^*, we have a path $i \xrightarrow{x} i$ in $\mathcal{B}$ and thus also a path $q \xrightarrow{w} q'$ with $q = \rho\big((u,v)\big)$ and $q' = \rho\big((u',v')\big)$. ◀

The statement above is false if X is not the minimal generating set of X^*, as shown by the following example.

▶ **Example 7.** Let $X = \{a, a^2\}$. There is no sharp reduction from the automaton of Figure 1 onto the one-state automaton recognizing $X^* = \{a\}^*$.

The statement is also false if the automaton $\mathcal{B}$ recognizes X^*, but does not recognize X^* with multiplicities, as shown by the following example.

▶ **Example 8.** Let $X = \{a^2\}$. The flower automaton of X is represented in Figure 4 on the left. There is no reduction onto the automaton represented on the right which also recognizes X^* (but not with multiplicities).

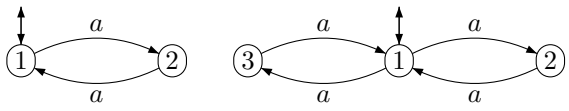


Figure 4 Two automata recognizing X^*.

3 Transducers

A literal *transducer* $\mathcal{T} = (Q, i, t)$ on a set of states Q with an *input alphabet* A and an *output alphabet* B is defined by a set of edges E which are of the form $p \xrightarrow{(a,v)} q$ with $p, q \in Q$, $a \in A$ and $v \in B \cup \{1\}$. The *input automaton* associated with a transducer is the automaton with the same set of states and edges but with the output labels removed.

The relation *realized* by the transducer $\mathcal{T}$ is the set of pairs $(u,v) \in A^* \times B^*$ such that there is a path from i to t labeled (u,v). We denote by $\varphi_{\mathcal{T}}$ the morphism from A^* to the monoid of $Q \times Q$-matrices with elements in $\mathbb{N}\langle B\rangle$ defined for $u \in A^*$ and $p, q \in Q$ by $\varphi_{\mathcal{T}}(u)_{p,q} = \sum_{p \xrightarrow{u|v} q} v$.

Let $X \subset A^+$ be a finite set. Let $\beta : B^* \to A^*$ be a *coding morphism* for X, that is, a morphism whose restriction to B is a bijection onto X. The *decoding relation* for X is the relation $\gamma = \{(u,v) \in A^* \times B^* \mid u = \beta(v)\}$. A *decoder* for X is a literal transducer which realizes the decoding relation. The *flower transducer* associated to β is the literal tranducer built on the flower automaton of X by adding an output label 1 to each edge $\omega \xrightarrow{a} (a,v)$ or $(u,av) \xrightarrow{a} (ua,v)$ and an output label b to each edge $\omega \xrightarrow{a} \omega$ such that $a \in X$ with $\beta(b) = a$ or $(u,a) \xrightarrow{a} \omega$ such that $ua = x \in X$ with $\beta(b) = x$.

▶ **Proposition 9.** *For every finite set $X \subset A^+$ with a coding morphism β, the flower transducer associated to β is a decoder for X.*

▶ **Example 10.** Let $X = \{a, ab, ba\}$ and let $\beta : u \to a, v \to ab, w \to ba$. The flower transducer associated to β is represented in Figure 5. One has

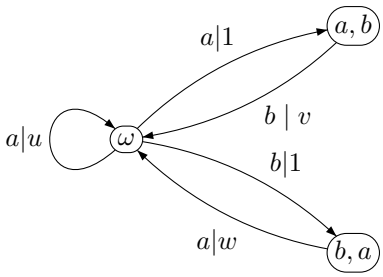


Figure 5 The flower transducer associated to β.

$$\varphi_{\mathcal{T}}(a) = \begin{bmatrix} u & 1 & 0 \\ 0 & 0 & 0 \\ w & 0 & 0 \end{bmatrix} \text{ and } \varphi_{\mathcal{T}}(b) = \begin{bmatrix} 0 & 0 & 1 \\ v & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The *prefix transducer* $\mathcal{T} = (P, 1, 1)$ is the same modification of the prefix automaton. Its states are the proper prefixes of the elements of X. There is an edge $p \xrightarrow{a|1} pa$ for every prefix p and every letter a such that $pa \in P$, and an edge $p \xrightarrow{a|b} 1$ for every prefix p and letter a such that $pa = \beta(b) \in X$. Thus the input automaton of the prefix transducer of X is the prefix automaton of X.

Let $\mathcal{B} = (Q, j, j)$ be an automaton on the alphabet B and let $\mathcal{T} = (P, i, i)$ be a literal transducer with the input alphabet A and the output alphabet B. We build an automaton $\mathcal{A} = \mathcal{B} \circ \mathcal{T}$ on the alphabet A as follows. Its set of states is $Q \times P$ and for every $a \in A$, the matrix $\varphi_{\mathcal{A}}(a)$ is obtained by replacing in $\varphi_{\mathcal{T}}(a)$ the word $w = \varphi_{\mathcal{T}}(a)_{p,q}$ by the matrix $\varphi_{\mathcal{B}}(w)$. The initial and terminal state is (j, i). The automaton $\mathcal{A}$ is also called the *wreath product* of $\mathcal{B}$ and $\mathcal{T}$ (see [9]). The word 1 is replaced by the identity matrix, and 0 is replaced by the zero matrix of appropriate size. An example of $\mathcal{A} = \mathcal{B} \circ \mathcal{T}$ is provided in Example 13.

4 Composition

Let $Y \subset B^+$ and $Z \subset A^+$ be finite sets of words such that there exists a bijection $\beta : B \to Z$. Two such sets are called *composable*. Then $X = \beta(Y)$ is called the *composition* of Y and Z through β, where $\beta(Y) = \{\beta(y) \mid y \in Y\}$ with β naturally extended to the mapping $B^* \to Z^*$. We denote $X = Y \circ_\beta Z$. We also denote $X = Y \circ Z$ when β is clear. We say that $X = Y \circ Z$ is a *decomposition* of X.

▶ **Example 11.** Let $Y = \{u, uw, vu\}$ and $Z = \{a, ab, ba\}$ with $\beta : u \to a, v \to ab, w \to ba$. Then $X = Y \circ_\beta Z = \{a, aba\}$.

A decomposition $X = Y \circ_\beta Z$ of a finite set X is *trim* if every letter of B appears in a word of Y and every word in X is obtained in a unique way from words in Y, that is, if the restriction of β to Y is injective. For any decomposition $X = Y \circ Z$, there are $Y' \subset Y$ and $Z' \subset Z$ such that $X = Y' \circ Z'$ is trim. Indeed, if $x \in X$ has two decompositions in words of Z as $x = z_1 z_2 \cdots z_n = z'_1 z'_2 \cdots z'_{n'}$, we may remove $\beta^{-1}(z'_1 z'_2 \cdots z'_{n'})$ from Y without

changing X. A finite number of these removals gives a trim decomposition. The set Z' is obtained by removing all words in Z which correspond to the letters no longer occurring in words in Y' (we also remove such letters from B). The decomposition in Example 11 is not trim, since $aba = \beta(uw) = \beta(vu)$, but it can be made trim by taking $X = Y' \circ Z'$ with $Y' = \{u, uw\}$ and $Z' = \{a, ba\}$. In this case, $Y' \subset \{u, w\}^+$.

A set $X \subset A^*$ is *complete* if any word in A^* is a factor of a word in X^*.

▶ **Proposition 12.** *Let $Y \subset B^+$ and $Z \subset A^+$ be two composable finite sets and let $X = Y \circ_\beta Z$ be a trim decomposition. Let $\mathcal{B} = (Q, 1, 1)$ be the prefix automaton of Y and let $\mathcal{T} = (P, 1, 1)$ be the prefix transducer of Z. The automaton $\mathcal{A} = \mathcal{B} \circ \mathcal{T}$ recognizes X^* with multiplicities.*

If Y is complete, there is a reduction ρ from $\mathcal{A}$ onto the prefix automaton of Z. Moreover, the automaton $\mathcal{B}$ can be identified through β with the restriction of $\mathcal{A}$ to $\rho^{-1}(1)$.

Proof. The simple paths in $\mathcal{A}$ have the form $(1,1) \xrightarrow{z_1} (b_1, 1) \xrightarrow{z_2} (b_1b_2, 1) \cdots \xrightarrow{z_n} (1,1)$ for $x = z_1 \cdots z_n = \beta(b_1 \cdots b_n)$ in X and $z_i \in Z$. Since the decomposition is trim, there is exactly one such path for every $x \in X$ and thus $\mathcal{A}$ recognizes X^* with multiplicities.

Let us show that, if Y is complete, the map $\rho : (q, p) \to p$ is a reduction from $\mathcal{A}$ onto the prefix automaton of Z. We have to show that one has $p \xrightarrow{w} p'$ in the prefix automaton $\mathcal{C}$ of Z if and only if there exist $q, q' \in Q$ such that $(q, p) \xrightarrow{w} (q', p')$. Assume that $p \xrightarrow{w} p'$ in $\mathcal{C}$. Then we have $p \xrightarrow{w|u} p'$ in the prefix transducer $\mathcal{T}$ for some $u \in B^*$. Since Y is complete, there are some $q, q' \in Q$ such that $q \xrightarrow{u} q'$ in $\mathcal{B}$. Then $(q, p) \xrightarrow{w} (q', p')$ in $\mathcal{A}$. The converse is obvious.

Finally, the edges of the restriction of $\mathcal{A}$ to $\rho^{-1}(1)$ are the simple paths $(q, 1) \xrightarrow{z} (q', 1)$ for $z = \beta(b) \in Z$ and $q \xrightarrow{b} q'$ an edge of $\mathcal{B}$. This proves the last statement. ◀

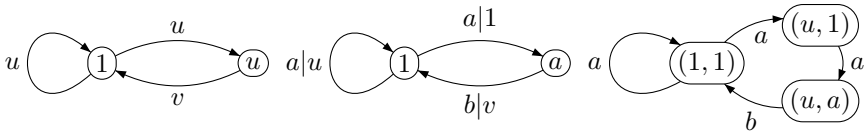


Figure 6 The prefix automaton of Y, the prefix transducer $\mathcal{T}$ of Z and the trim part of $\mathcal{A}$.

▶ **Example 13.** Let $Y = \{u, uv\}$ and $Z = \{a, ab\}$ with $\beta : u \to a, v \to ab$. We have, in view of Figure 6,

$$\varphi_{\mathcal{A}}(a) = \begin{bmatrix} \varphi_{\mathcal{B}}(u) & I \\ 0 & 0 \end{bmatrix} \text{ and } \varphi_{\mathcal{A}}(b) = \begin{bmatrix} 0 & 0 \\ \varphi_{\mathcal{B}}(v) & 0 \end{bmatrix}.$$

5 Monoids of relations

We consider monoids of binary relations and prove some results on idempotents and groups in such monoids. Few authors have considered monoids of binary relations. In [16], the Green's relations in the monoid $\mathcal{B}_Q$ of all binary relations on a set Q are considered. It is shown in [14] that any finite group appears as a maximal subgroup of $\mathcal{B}_Q$ (in contrast with the monoid of all partial maps in which all maximal subgroups are symmetric groups).

We write indifferently relations on a set Q as subsets of $Q \times Q$, as boolean $Q \times Q$-matrices or as directed graphs on a set Q of vertices.

The *rank* of a relation m on Q is the minimal cardinality of a set R such that $m = uv$ with u a $Q \times R$ relation and v an $R \times Q$ relation. Equivalently, the rank of m is the minimal number of row (resp. column) vectors (which are possibly not rows or columns of m) which generate over $\{0, 1\}$ the set of rows (resp. columns) of m.

For example, the full relation $m = Q \times Q$ has rank 1. In terms of matrices

$$m = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}$$

More generally, the rank of an equivalence relation is equal to the number of its classes.

A *fixed point* of a relation m on Q is an element $q \in Q$ such that $q \xrightarrow{m} q$. The following result appears in [20] (see also [12]).

▶ **Proposition 14.** *Let e be an idempotent relation on a finite set Q, let S be the set of fixed points of e and let Γ be the set of strongly connected components of the restriction of e to S.*

1. *For all $p, q \in Q$ we have $p \xrightarrow{e} q$ if and only if there exists an $s \in S$ such that $p \xrightarrow{e} s$ and $s \xrightarrow{e} q$.*
2. *We have*

$$e = \ell r \tag{1}$$

where $\ell = \{(p, \sigma) \in Q \times \Gamma \mid p \xrightarrow{e} s \text{ for some } s \in \sigma\}$ and $r = \{(\sigma, q) \in \Gamma \times Q \mid s \xrightarrow{e} q \text{ for some } s \in \sigma\}$.

Proof. 1. Choose $n > \mathrm{Card}(Q)$. Since $p \xrightarrow{e^n} q$, there is some $s \in Q$ such that $p \xrightarrow{e^i} s \xrightarrow{e^j} s \xrightarrow{e^k} q$ with $i + j + k = n$. Then $p \xrightarrow{e} s \xrightarrow{e} s \xrightarrow{e} q$ and the statement is proved. The other direction is obvious.

2. If $p \xrightarrow{e} q$, let $s \in S$ be such that $p \xrightarrow{e} s \xrightarrow{e} q$ and let σ be the strongly connected component of s. Then $p \xrightarrow{\ell} \sigma \xrightarrow{r} q$. Thus $e \leq \ell r$, which means that each element of e is not larger than the corresponding element of ℓr when these relations are considered as binary matrices. Conversely, if $p \xrightarrow{\ell} \sigma \xrightarrow{r} q$ there are $s, s' \in \sigma$ such that $p \xrightarrow{e} s$ and $s' \xrightarrow{s'} q$. Since s, s' are in the same stongly connected component, we have $s \xrightarrow{e} s'$ and we obtain $p \xrightarrow{e} s \xrightarrow{e} s' \xrightarrow{e} q$, whence $p \xrightarrow{e} q$. ◀

The decomposition of $e = lr$ given by Equation (1) is called the *column-row decomposition* of e. Note that Proposition 14 is false without the finiteness hypothesis on Q. Indeed, the relation $e = \{(x, y) \in \mathbb{R}^2 \mid x < y\}$ is idempotent, but has no fixed points.

▶ **Example 15.** The matrix

$$m = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

is an idempotent of rank 1.

For an element m of a monoid M, we denote by $H(m)$ the $\mathcal{H}$-class of m, where $\mathcal{H}$ is the Green relation $\mathcal{H} = \mathcal{R} \cap \mathcal{L}$ (see [3] for the definitions). It is a group if and only if it contains an idempotent e (see [3]). In this case, every $m \in H(e)$ has a unique inverse m^{-1} in the group $H(e)$.

The following result is the transposition of Proposition 9.1.7 in [3] to arbitrary monoids of relations. However, the result is restricted to a statement on the group $H(e)$ instead of the monoid eMe.

▶ **Proposition 16.** *Let M be a monoid of relations on a finite set Q, let $e \in M$ be idempotent and let Γ be the set of strongly connected components of the fixed points of e. For $m \in H(e)$, let $\gamma_e(m)$ be the relation on Γ defined by*

$$\gamma_e(m) = \{(\rho, \sigma) \in \Gamma \times \Gamma \mid r \xrightarrow{m} s \xrightarrow{m^{-1}} r \text{ for some } r \in \rho \text{ and } s \in \sigma\}$$

Then $m \mapsto \gamma_e(m)$ is an isomorphism from $H(e)$ onto a group of permutations on Γ.

Proof. First, $\gamma_e(m)$ is a map. Indeed, let $s \xrightarrow{m} t \xrightarrow{m^{-1}} s$ and $s' \xrightarrow{m} t' \xrightarrow{m^{-1}} s'$. If $s \xrightarrow{e} s'$, we have $t \xrightarrow{m^{-1}} s \xrightarrow{e} s' \xrightarrow{m} t'$ and thus $t \xrightarrow{e} t'$. By a symmetrical proof, we obtain that $\gamma_e(m)$ is a permutation.

Next, it is easy to verify that γ_e is a morphism.

Finally, γ_e is injective. Indeed, assume that for $m, m' \in H(e)$ we have $\gamma_e(m) = \gamma_e(m')$. Suppose that $p \xrightarrow{m} q$.

Assume first that p is a fixed point of e. Let r, r' be such that $p \xrightarrow{m} r \xrightarrow{m^{-1}} p$ and $p \xrightarrow{m'} r' \xrightarrow{m'^{-1}} p$. Since $\gamma_e(m) = \gamma_e(m')$, we obtain that r, r' are in the same element of Γ. We conclude that $p \xrightarrow{m'} r' \xrightarrow{e} r \xrightarrow{m^{-1}} p \xrightarrow{m} q$ which implies that $p \xrightarrow{m'} q$.

Now if p is not a fixed point of e, since $em = m$, there is an r such that $p \xrightarrow{e} r \xrightarrow{m} q$. By Proposition 14, there is a fixed point r' of e such that $p \xrightarrow{e} r' \xrightarrow{e} r \xrightarrow{m} q$. Then $r' \xrightarrow{m} q$ implies $r' \xrightarrow{m'} q$ by the preceding argument, and finally $p \xrightarrow{m'} q$. ◀

We denote $G_e = \gamma_e(H(e))$. The definition of γ_e can be formulated differently.

▶ **Proposition 17.** *Let M be a monoid of relations on a finite set Q and let $e \in M$ be an idempotent. Let σ, τ be two distinct connected components of fixed points of e and let $s \in \sigma, t \in \tau$. If $e_{s,t} = 1$, then $m_{t,s} = 0$ for every $m \in H(e)$ and thus $(\sigma, \tau) \notin \gamma_e(m)$. If $e_{s,t} = e_{t,s} = 0$ then $s \xrightarrow{m} t$ implies $(\sigma, \tau) \in \gamma_e(m)$.*

Proof. Assume first that $e_{s,t} = 1$ so that the restriction of e to $\{s, t\}$ is the matrix $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. If $m_{t,s} = 1$, then the restriction of m to $\{s, t\}$ is the matrix with all ones, which is impossible since no power of m can be equal to e. If the restriction of e to $\{s, t\}$ is the identity, then the restriction of $m \in H(e)$ is a permutation. Thus $(\sigma, \tau) \in \gamma_e(m)$ if and only if $s \xrightarrow{m} t$. ◀

The following extends Proposition 9.1.9 in [3]. It uses the Green relation $\mathcal{D} = \mathcal{LR} = \mathcal{RL}$. Two permutation groups G over Q and G' over Q' are called *equivalent* if there exists a bijection $\alpha : Q \to Q'$ and an isomorphism $\psi : G \to G'$ such that for all $q \in Q$ and $g \in G$ we have $\alpha(q.g) = \alpha(q).\psi(g)$, where $q.g$ is the action of g on q (see Section 1.13 of [3]). In a more standard terminology, two permutation groups are equivalent if and only if their group actions are isomorphic, though we use the terminology of [3] to simplify the comparison with the results described there.

▶ **Proposition 18.** *Let M be a monoid of relations on a finite set Q and let $e, e' \in M$ be $\mathcal{D}$-equivalent idempotents. Then the groups G_e and $G_{e'}$ are equivalent permutation groups.*

Proof. Let (a, a', b, b') be a passing system from e to e', that is such that

$$eaa' = e, \quad bb'e' = e', \quad ea = b'e'.$$

We will verify that there is a commutative diagram of isomorphisms shown in Figure 7.

$$\begin{array}{ccc} H(e) & \xrightarrow{\tau} & H(e') \\ \downarrow{\scriptstyle \gamma_e} & & \downarrow{\scriptstyle \gamma_{e'}} \\ G_e & \xrightarrow{\tau'} & G_{e'} \end{array}$$

Figure 7 Commutative diagram of isomorphisms.

We define the map τ by $\tau(m) = bma$. Then it is easy to verify that τ is a morphism and that $m' \mapsto b'm'a'$ is its inverse. Thus τ is an isomorphism.

We define τ' as follows. Let $\Gamma_e, \Gamma_{e'}$ be the sets of strongly connected components of fixed points of e and e' respectively. Let θ be the relation between Γ_e and $\Gamma_{e'}$ defined by $(\sigma, \sigma') \in \theta$ if for some $s \in \sigma$ and $s' \in \sigma'$, we have $s \xrightarrow{eae'} s'$. One may verify that θ is a bijection between Γ_e and $\Gamma_{e'}$. Its inverse is the map on classes induced by $e'be = e'a'e$. Then $\tau'(n) = \theta^t n \theta$.

We verify that the diagram is commutative. Suppose that for some $m \in H(e)$ $(\sigma'_1, \sigma'_1) \in \tau'(\gamma_e(m))$. By definition of τ' there exist $\sigma_1, \sigma_2 \in \Gamma_e$ such that

$$(\sigma'_1, \sigma_1) \in \theta^t, \quad (\sigma_1, \sigma_2) \in \gamma_e(m) \text{ and } (\sigma_2, \sigma'_2) \in \theta.$$

Then for $s_1 \in \sigma_1$, $s'_1 \in \sigma'_1$, $s'_2 \in \sigma'_2$ and $s_2 \in \sigma_2$, we have

$$s'_1 \xrightarrow{e'be} s_1, \quad s_1 \xrightarrow{m} s_2 \xrightarrow{m^{-1}} s_1. \quad s_2 \xrightarrow{eae'} s'_2.$$

Then $s'_1 \xrightarrow{bma} s'_2 \xrightarrow{bm^{-1}a} s'_1$ showing that $(\sigma'_1, \sigma'_1) \in \gamma_{e'}(\tau(m))$. ◀

Note that, contrary to the case of a monoid of unambiguous relations, two $\mathcal{D}$-equivalent idempotents need not have the same number of fixed points, as shown by the following example.

▶ **Example 19.** Let M be the monoid of all relations on $Q = \{1, 2\}$. The two idempotents

$$e = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad e' = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

are $\mathcal{D}$-equivalent although the first has one fixed point and the second has two.

Let M be a monoid of relations on a finite set Q. The *minimal rank* of M, denoted $r(M)$ is the minimum of the ranks of the elements of M other than 0. The following statement generalizes Theorem 9.3.10 in [3] from unambiguous to arbitrary transitive monoids of relations. A $\mathcal{D}$-class is *regular* if it contains an idempotent. A monoid of relations on Q is *transitive* if for every $p, q \in Q$, there is an $m \in M$ such that $p \xrightarrow{m} q$.

▶ **Theorem 20.** *Let M be a transitive monoid of relations on a finite set Q. The set K of elements of rank $r(M)$ is a regular $\mathcal{D}$-class. The groups G_e for e idempotent in K are equivalent transitive permutation groups. Moreover, for a fixed point i of e, the minimal rank $r(M)$ is the index of the subgroup $\{m \in H(e) \mid i \xrightarrow{m} i\}$ in $H(e)$.*

Proof. The proof is the same as for the case of an unambiguous monoid of relations except for the last statement. Let σ, τ be two distinct strongly connected components of fixed points of e and let $s \in \sigma, t \in \tau$. Since M is transitive there is an $m \in M$ such that $s \xrightarrow{m} t$. Then eme is not 0 and thus $eme \in H(e)$. Similarly, if $n \in M$ is such that $t \xrightarrow{n} s$, then $ene \in H(e)$. This implies by Proposition 17 that the restriction of e to $\{s, t\}$ is the identity and that $(\sigma, \tau) \in \gamma_e(eme)$. Thus G_e is transitive. The last statement follows from the fact that for any transitive permutation group on a set S, the number of elements of S is equal to the index of the subgroup fixing one of the points of S (Proposition 1.13.2 of [3]). ◀

The *Suschkevitch group* of M is one of the equivalent groups G_e for e of rank $r(M)$.

6 Group and degree of a set

Let $\mathcal{A} = (P, i, i)$ and $\mathcal{B} = (Q, j, j)$ be automata and let $\rho : P \to Q$ be a reduction. For $m = \varphi_{\mathcal{A}}(w)$, the relation $n = \varphi_{\mathcal{B}}(w)$ is well defined. We denote it by $n = \hat{\rho}(m)$. Then $\hat{\rho}$ is a morphism from $\varphi_{\mathcal{A}}(A^*)$ onto $\varphi_{\mathcal{B}}(A^*)$ called the *morphism associated* with ρ. The following result extends Proposition 9.5.1 in [3] to arbitrary finite sets of words.

▶ **Proposition 21.** *Let $X \subset A^+$ be finite. Let $\mathcal{A} = (P, i, i)$ and $\mathcal{B} = (Q, j, j)$ be trim automata recognizing X^* with multiplicities. Let $M = \varphi_{\mathcal{A}}(A^*)$ and $N = \varphi_{\mathcal{B}}(A^*)$. Let E be the set of idempotents in M and F the set of idempotents in N.*

Let ρ be a sharp reduction of $\mathcal{A}$ onto $\mathcal{B}$ and let $\hat{\rho} : M \to N$ be the morphism associated with ρ. Then

1. *$\hat{\rho}(E) = F$.*
2. *Let $e \in E$ and $f = \hat{\rho}(e)$. The restriction of ρ to the set S of fixed points of e is a bijection on the set of fixed points of f, and the groups H_e and H_f are equivalent.*

Proof. 1. Let $e \in E$. Then $\hat{\rho}(e)$ is idempotent since $\hat{\rho}$ is a morphism. Thus $\hat{\rho}(E) \subset F$. Conversely, if $f \in F$, let $w \in A^*$ be such that $\varphi_{\mathcal{B}}(w) = f$. Let $n \geq 1$ be such that $e = \varphi_{\mathcal{A}}(w)^n$ is idempotent. Then $\hat{\rho}(e) = f$ since $\hat{\rho} \circ \varphi_{\mathcal{A}} = \varphi_{\mathcal{B}}$.

2. Let S be the set of fixed points of e and T the set of fixed points of f. Consider $s \in S$ and let $t = \rho(s)$. From $s \xrightarrow{e} s$, we obtain $t \xrightarrow{f} t$ and thus $\rho(S) \subset T$. Conversely, let $t \in T$. The restriction of e to the set $R = \rho^{-1}(t)$ is a non zero idempotent. Thus there is some $s \in R$ which is a fixed point of this idempotent, ans thus of e. Thus $t \in \rho(S)$.

Since $\hat{\rho}$ is a morphism from M onto N, we have $\hat{\rho}(H(e)) = H(f)$. It is clear that ρ maps a strongly connected component of e on a strongly connected component of f. To show that this map is a bijection, consider $s, s' \in S$ such that $\rho(s), \rho(s')$ belong to the same connected component. We may assume that e is not the equality relation. Let $w \in A^+$ be such that $\varphi_{\mathcal{A}}(w) = e$. Since X is finite, there are factorizations $w = uv = u'v'$ such that $s \xrightarrow{u} i \xrightarrow{v} s$ and $s' \xrightarrow{u'} i \xrightarrow{v'} s'$. Then we have $j \xrightarrow{v} \rho(s) \xrightarrow{w} \rho(s') \xrightarrow{u'} j$. Since ρ is sharp, this implies $i \xrightarrow{vwu'} i$ and finally $s \xrightarrow{uvwu'v'} s'$. This shows that $s \xrightarrow{e} s'$. A similar proof shows that $s' \xrightarrow{e} s$. Thus, s, s' belong to the same connected component of e.

Moreover, for every $m \in H(e)$, one has $s \xrightarrow{m} t \xrightarrow{m^{-1}} s$ if and only if $\rho(s) \xrightarrow{\hat{\rho}(m)} \rho(t) \xrightarrow{\hat{\rho}(m^{-1})} \rho(s)$. Thus H_e and H_f are equivalent permutation groups. ◀

Let $X \subset A^+$ be a finite set and let $\mathcal{A}$ be the flower automaton of X. The *degree* of X, denoted $d(X)$ is the minimal rank of the monoid $M = \varphi_{\mathcal{A}}(A^*)$. The *group of* X is the Suschkevitch group of M. Proposition 21 shows that the definitions of the group and of the degree do not depend on the automaton chosen to recognize X^*, provided one takes a trim automaton recognizing X^* with multiplicities.

7 Synchronization

Let $X \subset A^+$ be a finite set of words. A word $x \in A^*$ is *synchronizing* for X if for every $u, v \in A^*$, $uxv \in X^* \Rightarrow ux, xv \in X^*$. A set X is *synchronizing* if there is a synchronizing word $x \in X^*$. The next proposition generalizes Proposition 10.1.11 of [3]

▶ **Proposition 22.** *A finite set $X \subset A^+$ is synchronizing if and only if its degree $d(X)$ is 1.*

Proof. Let $\mathcal{A} = (Q, i, i)$ be a trim automaton recognizing X^*.

Assume first that $d(X) = 1$. Let $x \in X^*$ be such that $\varphi_{\mathcal{A}}(x)$ has rank 1. If $uxv \in X^*$, we have $i \xrightarrow{u} p \xrightarrow{x} q \xrightarrow{v} i$ for some $p, q \in Q$. Since $\varphi_{\mathcal{A}}(x)$ has rank 1, we deduce from $i \xrightarrow{x} i$ and $p \xrightarrow{x} q$ that we have also $i \xrightarrow{x} q$ and $p \xrightarrow{x} i$. Thus $ux, xv \in X^*$, showing that x is synchronizing.

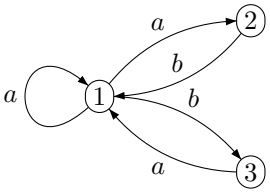
Assume conversely that X is synchronizing. Let $x \in X^*$ be a synchronizing word. Replacing x by some its power, we may assume that $\varphi_{\mathcal{A}}(x)$ is an idempotent e. Let $m \in H(e)$ and let $w \in \varphi_{\mathcal{A}}^{-1}(m)$. Since $H(e)$ is finite, there is some $n \geq 1$ such that $m^n = e$. Then $(me)^n = e$ implies that $(wx)^n \in X^*$. Since x is synchronizing, we obtain $wx \in X^*$ and since $\varphi_{\mathcal{A}}(wx) = me = m$, this implies $w \in X^*$. This shows that $H(e)$ is contained in $\varphi_{\mathcal{A}}(X^*)$ and thus that $d(X) = 1$ by Theorem 20. ◀

▶ **Example 23.** Consider again $X = \{a, ab, ba\}$ (Example 3). The flower automaton of X is represented again for convenience in Figure 8 (left).

The minimal rank of the elements of $\varphi_{\mathcal{A}}(A^*)$ is 1. Indeed, we have

$$\varphi_{\mathcal{A}}(a^2) = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}$$

Accordingly, aa is a synchronizing word.



	$1,2$	$1,3$	3
$1,3$	$*\ a^2$	$*\ a^2b$	$*\ a^2b^2$
$1,2$	$*\ ba^2$	$*\ ba^2b$	ba^2b^2
2	$*b^2a^2$	b^2a^2b	$b^2a^2b^2$

Figure 8 The flower automaton of X (left) and the set K of elements of rank 1 (right).

The set K of elements of rank 1 is represented in Figure 8 (right). For each $\mathcal{H}$-class, we indicate on its left the set of states p such that the row of index p in nonzero. Similarly, we indicate above it the set of states q such that the column of index q is nonzero. A star $*$ indicates an $\mathcal{H}$-class which is a group. Note that

$$\varphi_{\mathcal{A}}(a^2b) = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

has two fixed points but only one strongly connected class, in agreement with fact that it is of rank 1.

8 Groups and composition

Given a transitive permutation group G on a set Q, an *imprimitivity relation* of G is an equivalence on Q compatible with the group action. If θ is such an equivalence relation, we denote by G_θ the permutation group induced by the action of G on the classes of θ. The groups induced by the action on the class of an element $i \in Q$ by the action of the elements of G stabilizing the class of i are all equivalent. We denote by G^θ one of them. For two permutation groups G, H on sets P and Q respectively, we denote $G \leq H$ if there is an imprimitivity equivalence θ on Q such that $G = H_\theta$.

The next theorem generalizes Proposition 11.1.2 of [3].

▶ **Theorem 24.** *Let $X \subset A^+$ be a finite set with a trim decomposition $X = Y \circ Z$, where Y is complete. There exists an imprimitivity equivalence θ of $G = G(X)$ such that*

$$G^\theta \leq G(Y), \quad G_\theta = G(Z).$$

In particular, $d(X) \leq d(Y) \cdot d(Z)$.

Proof. Let $\mathcal{B} = (Q, i, i)$ be the flower automaton of Y and let $\mathcal{T}$ be the prefix transducer of Z. Let $\mathcal{A} = \mathcal{B} \circ \mathcal{T}$. By Proposition 12, there is a reduction ρ from $\mathcal{A} = (Q \times P, (i,1), (i,1))$ onto the prefix automaton $\mathcal{C}$ of Z.

Let e be an idempotent of minimal rank in $\varphi_{\mathcal{A}}(X^*)$. Let S be the set of fixed points of e and let Γ be the set of connected components (scc) of the elements of S. Let $\hat{S}$ be the set of fixed points of $\hat{e} = \hat{\rho}(e)$ and let $\hat{\Gamma}$ be the set of corresponding scc's. If $s, s' \in S$ are in the same scc, then $\rho(s), \rho(s')$ are in the same scc of $\hat{S}$. Thus, we have a well-defined map $\bar{\rho} : \Gamma \to \hat{\Gamma}$ such that $s \in \Gamma$ if and only if $\rho(s) \in \bar{\rho}(\Gamma)$.

We define an equivalence θ on Γ by $\sigma \equiv \sigma'$ if $\bar{\rho}(\sigma) = \bar{\rho}(\sigma')$. Let $m \in H(e)$ and suppose that $(\sigma, \tau), (\sigma', \tau') \in \gamma_e(m)$. If $\sigma \equiv \sigma' \bmod \theta$, then $\tau \equiv \tau' \bmod \theta$. Let indeed $s \in \sigma, s' \in \sigma'$ and $t \in \tau, t' \in \tau'$. We have by definition of γ_e

$$s \xrightarrow{m} t \xrightarrow{m^{-1}} s \text{ and } s' \xrightarrow{m} t' \xrightarrow{m^{-1}} s'$$

and thus

$$\rho(s) \xrightarrow{\hat{\rho}(m)} \rho(t) \xrightarrow{\hat{\rho}(m)^{-1}} \rho(s) \text{ and } \rho(s') \xrightarrow{\hat{\rho}(m)} \rho(t') \xrightarrow{\hat{\rho}(m)^{-1}} \rho(s')$$

This implies that $\rho(t) \xrightarrow{\hat{e}} \rho(t')$ and $\rho(t') \xrightarrow{\hat{e}} \rho(t)$. But since $\gamma_{\hat{e}}(\hat{m})$ is a permutation, this forces $\bar{\rho}(\tau) = \bar{\rho}(\tau')$ and finally $\tau \equiv \tau' \bmod \theta$. Since the action of $H(e)$ on the classes of θ is the same as the action of $H(\hat{e})$, we have $G(Z) = G_\theta$.

Finally, let $\sigma \in \Gamma$ be the scc of the initial state $(i,1)$ and let I be its class modθ. Thus $d(X) = \mathrm{Card}(I) d(Z)$. Let $x \in X^*$ be such that $\varphi_{\mathcal{A}}(x) = e$ and let $y = \beta^{-1}(x)$. Then $f = \varphi_{\mathcal{B}}(y)$ is an idempotent of $\varphi_{\mathcal{B}}(B^*)$ of rank $d(Y)$. Let U be the set of fixed points of f and let Φ be the set of scc of U for the action of f. Let σ be the equivalence on Φ induced by the equivalence $r \equiv s$ if $(r,1), (s,1)$ belong to the same scc for e. Then σ is an imprimitivity equivalence for $G(Y)$ such that $G(Y)_\sigma = G^\theta$. Thus $G^\theta \leq G(Y)$ and $\mathrm{Card}(I) \leq d(Y)$, which implies $d(X) \leq d(Y) \cdot d(Z)$. ◀

▶ **Example 25.** Let $Z = \{a, ab, ba, ca\}$ and $X = Z^2$. We have $X = Y \circ_\beta Z$ with $Y = \{u, v, w, x\}^2$ and $\beta : u \mapsto a, v \mapsto ab, w \mapsto bc, x \mapsto ca$. The word aa is synchronizing for Z and thus $d(Z) = 1$. In contrast, we have $d(Y) = 2$ and $G(Y) = \mathbb{Z}/2\mathbb{Z}$. It can be verified that the word ca^2b is synchronizing for X and thus $d(X) = 1$. Thus $d(X) < d(Y) \cdot d(Z) = 2 \cdot 1 = 2$. Thus the case of a strict inequality can occur. This is made possible by the fact that Z is not a code. Indeed, we have $(ab)(ca) = a(bc)a$.

9 Decompositions of codes

Finally, we use the developed techniques to show that for a uniquely decipherable code X for all the trim decompositions of the form $X = Y \circ Z$ with Y complete we have that Z (and thus Y) is a uniquely decipherable code as well. It shows that, as long as we require Y to be complete, we do not get any new trim decompositions of uniquely decipherable codes even if we decompose them as arbitrary sets of words.

▶ **Proposition 26.** *Let $X = Y \circ Z$ be a trim decomposition of a finite set X. If X is a uniquely decipherable code and if Y is complete, then Z is a uniquely decipherable code.*

Proof. Since β is trim, Y is a uniquely decipherable code. Let $\beta : B \to Z$ be the coding morphism for Z such that $X = Y \circ_\beta Z$. Assume that $z \in Z^*$ is a word with more than one factorization into words of Z. Let $u, v \in B^*$ two distinct elements in $\beta^{-1}(z)$. Let $\mathcal{A}$ be the flower automaton of Y. Let $y \in Y^*$ be such that $\varphi_{\mathcal{A}}(y)$ has minimal rank. Then yuy, yvy are not zero since Y is complete. Thus $\varphi_{\mathcal{A}}(yuy), \varphi_{\mathcal{A}}(yvy)$ belong to the $\mathcal{H}$-class of $\varphi_{\mathcal{A}}(y)$ which is a finite group. Let e be its idempotent. There are integers n, m, p such that $\varphi_{\mathcal{A}}(y)^n = \varphi_{\mathcal{A}}(yuy)^m = \varphi_{\mathcal{A}}(yvy)^p = e$. Since $y \in Y^*$, this implies that $e \in \varphi_{\mathcal{A}}(Y^*)$ and thus that $(yuy)^m, (yvy)^p$ are in Y^*. We conclude that Y is not a uniquely decipherable code, a contradiction. ◀

This is false if we do not require Y to be complete. Consider a code $X = \{ab, abaab, abbab\}$, which can be decomposed into $X = Y \circ Z$ with $Y = \{u, uvu, uwu\}$ and $Z = \{ab, a, b\}$. The decomposition is obviously trim, the set X is a uniquely decipherable code, but the set Z is not a uniquely decipherable code.

References

1 Timothy C. Bell, John G. Cleary, and Ian H. Witten. *Text compression*. Prentice-Hall, Inc., 1990.

2 Jean Berstel, Dominique Perrin, Jean-Francois Perrot, and Antonio Restivo. Sur le théorème du défaut. *Journal of Algebra*, 60(1):169–180, 1979. doi:10.1016/0021-8693(79)90113-3.

3 Jean Berstel, Dominique Perrin, and Christophe Reutenauer. *Codes and Automata*. Cambridge University Press, 2009.

4 Renato M. Capocelli, Luisa Gargano, and Ugo Vaccaro. A fast algorithm for the unique decipherability of multivalued encodings. *Theoretical Computer Science*, 134(1):63–78, 1994. doi:10.1016/0304-3975(94)90278-X.

5 Arturo Carpi and Flavio D'Alessandro. On incomplete and synchronizing finite sets. *Theor. Comput. Sci.*, 664:67–77, 2017. doi:10.1016/j.tcs.2015.08.042.

6 Julien Clément, Jean-Pierre Duval, Giovanna Guaiana, Dominique Perrin, and Giuseppina Rindone. Parsing with a finite dictionary. *Theoretical Computer Science*, 340(2):432–442, 2005. doi:10.1016/j.tcs.2005.03.030.

7 Aldo de Luca, Dominique Perrin, Antonio Restivo, and Settimo Termini. Synchronization and simplification. *Discrete Mathematics*, 27(3):297–308, 1979. doi:10.1016/0012-365X(79)90164-X.

8 Samuel Eilenberg. *Automata, Languages and Machines*, volume A. Academic Press, 1974.

9 Samuel Eilenberg. *Automata, Languages and Machines*, volume B. Academic Press, 1976.

10 Fernando Guzmán. Decipherability of codes. *Journal of Pure and Applied Algebra*, 141(1):13–35, 1999. doi:10.1016/S0022-4049(98)00019-X.

11 Tero Harju and Juhani Karhumäki. Many aspects of defect theorems. *Theoretical Computer Science*, 324(1):35–54, 2004. Words, Languages and Combinatorics. doi:10.1016/j.tcs.2004.03.051.

12 Evelyne Le Rest and Michel Le Rest. Une représentation fidèle des groupes d'un monoïde de relations binaires sur un ensemble fini. *Semigroup Forum*, 21(2-3):167–172, 1980. doi:10.1007/BF02572547.

13 Abraham Lempel. On multiset decipherable codes (corresp.). *IEEE Trans. Inf. Theor.*, 32(5):714–716, 1986. doi:10.1109/TIT.1986.1057217.

14 J.S. Montague and R.J. Plemmons. Maximal subgroups of the semigroup of relations. *Journal of Algebra*, 13(4):575–587, 1969.

15 H. Nagumo, Mi Lu, and Karan Watson. Parallel algorithms for the static dictionary compression. In *Proceedings DCC '95 Data Compression Conference*, pages 162–171, 1995.

16 R. J. Plemmons and M. T. West. On the semigroup of binary relations. *Pacific J. Math.*, 35(3):743–753, 1970.

17 Evelyne Barbin-Le Rest and Stuart W. Margolis. On the group complexity of a finite language. In *Proceedings of the 10th Colloquium on Automata, Languages and Programming*, pages 433–444, Berlin, Heidelberg, 1983. Springer-Verlag.

18 Antonio Restivo. A note on multiset decipherable codes. *IEEE Trans. Inf. Theor.*, 35(3):662–663, 1989. doi:10.1109/18.30991.

19 Jacques Sakarovitch. *Elements of Automata Theory.* Cambridge University Press, 2009.

20 Marcel-Paul Schützenberger. A property of finitely generated submonoids of free monoids. In G. Pollak, editor, *Algebraic Theory of Semigroups*, pages 545–576. North-Holland, 1979.

21 Andreas Weber and Tom Head. The finest homophonic partition and related code concepts. *IEEE Transactions on Information Theory*, 42(5):1569–1575, 1996.