Combinatorics of words

Christian Choffrut[1] and Juhani Karhumäki[2]

[1] Université Paris VII, LITP, 2, place Jussieu, 75251 Paris Cedex 05, France
email: cc@litp.ibp.fr

[2] Department of Mathematics, University of Turku, FIN-20014 Turku, Finland
email: karhumak@cs.utu.fi

1. Introduction

The basic object of this chapter is a *word*, that is a sequence – finite or infinite – of elements from a finite set. The very definition of a word immediately imposes two characteristic features on mathematical research of words, namely the *discreteness* and the *noncommutativity*. Therefore the combinatorial theory of words is a part of noncommutative discrete mathematics, which moreover often emphasizes the *algorithmic* nature of problems.

It is worth recalling that in general noncommutative mathematical theories are much less developped than commutative ones. This explains, at least partly, why many simply formulated problems of words are very difficult to attack, or to put this more positively, mathematically challenging.

The theory of words is profoundly connected to numerous different fields of mathematics and its applications. A natural environment of a word is a finitely generated free monoid, therefore connections to algebra are extensive and diversified. Combinatorics, of course, is a fundamental part of the theory of words. Less evident but fruitful connections are those to probability theory or even to topology via dynamical systems. Last but not least we mention the close interrelation of the theory of words and the theory of automata, or more generally theoretical computer science.

This last relation has without any doubt emphasized the algorithmic nature of problems on words, but even more importantly has played a major role in the process of making the theory of words to a mature scientific topic of its own. Indeed, while important results on words were til 1970's only scattered samples in the literature, during the last quarter of the century the research on words has been systematic, extensive, and we believe, also successful.

Actually, it was already at the beginning of this century when A. Thue initiated a systematic study on words, cf. [Be6] for a survey of Thue's work. However, his fundamental results, cf. [T1], [T2] and also [Be8], remained quite unnoticed for decades, mainly due to the unknown journals he used. Later many of his results were discovered several times in different connections.

The modern systematic research on words, in particular words as elements of free monoids, was initiated by M.P. Schützenberger in the sixties. Two influencial papers of that time are [LySc] and [LeSc]. This research created also the first monograph on words, namely [Len], which, however, never became widely used.

Year 1983 was important to the theory: the first book "Combinatorics on Words" [Lo] covering major parts on combinatorial problems of words appeared. Even today it is the most comprehensive presentation of the topic.

The goals of this presentation is to consider combinatorial properties of words from the point of view of formal languages. We do not intend to be exhaustive. Indeed, several important topics such as theory of codes, several problems on morphisms of free monoids, as well as unavoidable regularities like Shirshov's Theorem, are not considered in this chapter, but are discussed in other chapters of the Handbook. Neither the representations of the topics chosen are supposed to be encyclopedic.

On the other hand, the criteria we have had in our minds when choosing the material to this chapter can be summarized as follows. In addition to their relevance to formal languages we have paid a special attention to select topics which are not yet considered in textbooks, or at least to have a fresh presentation of older topics. We do not prove many of the results mentioned. However, we do prove several results either as examples of proof techniques used, or especially if we can give a proof which has not yet appeared in textbooks. We have made special efforts to fix the terminology.

The contents of our chapter is now summarized.

In Section 2 we fix our terminology. In doing so we already present some basic facts to motivate the notions. Section 3 deals with three selected problems. These problems – mappings between word monoids, binary equality languages and a separation of words by a finite automaton – are selected to illustrate different typical problems on words.

Section 4 deals with the well-known defect effect: if n words satisfy a nontrivial relation, then they can be expressed as products of at most $n-1$ words. We discuss different variations of this result some of which emphasizing more combinatorial and some more algebraic aspects. We point out differences of these results, including the computational ones, as well as consider the defect effect caused by several nontrivial relations.

In Section 5 we consider equations over words and their use in defining properties of words, including several basic ones such as the conjugacy. We also show how to encode any Boolean combination of properties, each of which expressable by an equation, into a single equation. Finally, a survey of decidable and undecidable logical theories of equations are presented.

Section 6 is devoted to a fundamental property of periodicity. We present a proof of the Theorem of Fife and Wilf which allows to analyse its optimality. We also give an elegant proof of the Critical Factorization Theorem from [CP], and finally discuss about an interesting recent characterization of ultimately periodic words due to [MRS].

In Section 7 we consider partial orderings of words and finite sets of words. As we note there normally such orderings are not finitary either in the sense that all antichains or in the sense that all chains would be finite. There are two remarkable exceptions. Higman's Theorem restricted to words states that the subword ordering, i.e., the ordering by the property being a (sparse) subword, allows only finite antichains, and is thus a well-ordering. We also consider several extensions of this ordering defined using special properties of words.

The other finiteness condition is obtained as a consequence of the validity of the Ehrenfeucht Compactness Property for words, which itself states that each system of equations with a finite number of unknowns is equivalent to one of its finite subsystems. As an application of this compactness property we can define a natural partial ordering on finite sets of words, such that it does not allow infinite chains. This, in turn, motivates us to state and solve some problems on subsemigroups of a free semigroup.

Section 8 is related to the now famous work of Thue. We give a survey on results which repetitions or abelian repetitions are avoidable in alphabets of different sizes. We also estimate the number of finite and infinite cube-free and overlap-free words over a binary alphabet, as well as square-free words over a ternary alphabet. We present, as an elegant application of automata theory to combinatorics of words, an automata-theoretic presentation due to [Be7] of Fife's Theorem, cf.[F], characterizing one-way infinite overlap-free (or 2^+-free) words over a binary alphabet. Finally, we recall the complete characterization of binary patterns which can be avoided in infinite binary words.

In Section 9, last of this chapter, we consider the complexity of an infinite word defined as the function associating to n the number of factors of length n in the considered word. Besides examples, we present a complete classification, due to [Pan2], of the complexities of words obtained as fixed points of iterated morphisms.

Finally, as a technical matter of our presentation we note that results are divided into two categories: Theorems and Propositions. The division is based on the fact whether the proofs are presented here or not. Theorems are either proved in details or outlined in the extend that an experienced reader can recover those, while Propositions are stated with only proper references to the literature.

2. Preliminaries

In this section we recall basic notions of words and sets of words, or languages, used in this chapter. The basic reference on combinatorics of words is [Lo], see also [La] or [Shy]. The notions of automata theory are not defined here, but can be found in any textbook of the area, cf. e.g. [Be1], [Harr], [HU] or [Sal1], or in appropriate chapters of this Handbook.

2.1 Words

Let Σ be a finite *alphabet*. Elements of Σ are called *letters*, and sequences of letters are called *words*, in particular, the *empty word*, which is denoted by 1, is the sequence of length zero. The set of all words (all nonempty words, resp.) is denoted by Σ^* (Σ^+, resp.). It is a monoid (semigroup, resp.) under the operation of *concatenation* or *product* of words. Moreover, obviously each word has the unique representation as products of letters, so that Σ^* and Σ^+ are *free*, referred to as the *free monoid* and *semigroup generated by* Σ.

Although we may assume for our purposes that Σ is finite we sometimes consider infinite words as well as finite ones: a *one-way infinite* word, or briefly an infinite word, can be identified with a mapping $\mathbb{N} \to \Sigma$, and is normally represented as $w = a_0 a_1 \dots$ with $a_i \in \Sigma$. Accordingly, *two-way infinite*, or *bi-infinite*, words over Σ are mappings $\mathbb{Z} \to \Sigma$. We denote the sets of all such words by Σ^ω and ${}^\omega\Sigma^\omega$, respectively, and set $\Sigma^\infty = \Sigma^* \cup \Sigma^\omega$. The notions $\mathbb{Z}$ and $\mathbb{N}$ are used to denote the sets of integers and nonnegative integers, respectively.

Let u be a word in Σ^*, say $u = a_1 \dots a_n$ with $a_i \in \Sigma$. We define $u(i)$ to denote the ith letter of u, i.e., $u(i) = a_i$. We say that n is the *length* of u, in symbols $|u|$, and note that it can be computed by the morphism $|\ | : \Sigma^* \to \mathbb{N}$ defined as $|a| = 1 \in \mathbb{N}$, for $a \in \Sigma$. The sets of all words over Σ of length k, or at most k are denoted by Σ^k and $\Sigma^{\leq k}$, respectively. By $|u|_a$, for $a \in \Sigma$, we denote the total number of the letter a in u. The *commutative image* $\pi(u)$ of a word u, often referred to as its *Parikh image*, is given by the formula $\pi(u) = (|u|_{a_1}, \dots, |u|_{a_{\|\Sigma\|}})$, where $\|\Sigma\|$ denotes the cardinality of Σ and Σ is assumed to be ordered. The *reverse* of u is the word $u^R = a_n \dots a_1$, and u is called a *palindrome* if it coincides with its reverse. For the empty word 1 we pose $1^R = 1$. By $\mathrm{alph}(w)$ we mean the minimal alphabet where w is defined.

Finally a *factorization* of u is any sequence $u_1, \dots, u_t$ of words such that $u = u_1 \dots u_t$. If words u_i are taken from a set X, then the above sequence is called an *X-factorization* of u. A related notion of an *X-interpretation* of u is any sequence of words $u_1, \dots, u_t$ from X satisfying $\alpha u \beta = u_1 \dots u_t$ for some words α and β, with $|\alpha| < |u_1|$ and $|\beta| < |u_t|$. These notions can be illustrated as in Figure 2.1.

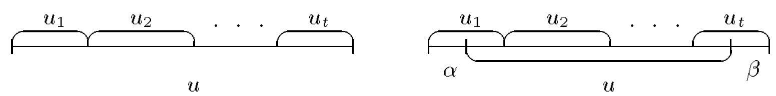


Figure 2.1. An X-factorization and an X-interpretation of u

For a pair (u, v) of words we define four relations:

u is a *prefix* of v, if there exists a word z such that $v = uz$;

u is a *suffix* of v, if there exists a word z such that $v = zu$;
u is a *factor* of v, if there exist words z and z' such that $v = zuz'$;
u is a *subword* of v, if v as a sequence of letters contains u as a subsequence, i.e., there exist words $z_1, \ldots, z_t$ and $y_0, \ldots, y_t$ such that $u = z_1 \ldots z_t$ and $v = y_0 z_1 y_1 \ldots z_t y_t$.

Sometimes factors are called *subwords*, and then subwords are called *sparse subwords*. We, however, prefer the above terminology. Each of the above relations holds if $u = 1$ or $u = v$. When these trivial cases are excluded the relations are called *proper*. A factor v of a word u can occur in u in different *positions* each of those being uniquely determined by the length of the prefix of u preceding v. For example, ab occurs in $abbaabab$ in positions 0, 4 and 6.

If $v = uz$ we write $u = vz^{-1}$ or $z = u^{-1}v$, and say that u is the *right quotient of v by z*, and that z is the *left quotient of v by u*. Consequently, the operations of right and left quotients define partial mappings $\Sigma^* \times \Sigma^* \to \Sigma^*$. Note that the above terminology is motivated by the fact that the free monoid Σ^* is naturally embedded into the free group generated by Σ. We also write $u \leq v$ ($u < v$, resp.) meaning that u is a prefix (a proper prefix, resp.) of v. Further by $\mathrm{pref}_k(v)$ and $\mathrm{suf}_k(v)$, for $k \in \mathbb{N}$, we denote the prefix and the suffix of v of length k. Finally, we denote by $\mathrm{pref}(x)$, $\mathrm{suf}(x)$, $F(x)$ and $SW(x)$ the sets of all prefixes, suffixes, factors and subwords of x, respectively.

It follows immediately that Σ^* satisfies, for all words $u, v, x, y \in \Sigma^*$ the condition

$$(1) \quad uv = xy \Rightarrow \exists t \in \Sigma^* : \ u = xt \ \text{ and } \ tv = y, \ \text{ or } \ x = ut \ \text{ and } \ v = ty.$$

Similarly, as we already noted, the length function of Σ^* is a morphism into the additive monoid $\mathbb{N}$:

$$(2) \qquad h : \Sigma^* \to \mathbb{N} \quad \text{with} \quad h^{-1}(0) = 1.$$

Conditions (1) and (2) are used to characterize the freeness of a monoid, cf. [Lev]. Consequently, Σ^* is indeed free as a monoid.

For two words u and v neither of these needs to be a prefix of another. However, they always have a unique *maximal common prefix* denoted by $u \wedge v$. Similarly, they always have among their common factors longest ones. Let us denote their lengths by $l(u, v)$. These notions allow us to define a *metric* on the sets Σ^* and Σ^ω. For example, by defining *distance functions* as

$$d(u, v) = |uv| - 2l(u, v) \quad \text{for} \quad u, v \in \Sigma^*,$$

and

$$d_\infty(u, v) = 2^{-|u \wedge v|} \quad \text{for} \quad u, v \in \Sigma^\omega,$$

(Σ^*, d) and $(\Sigma^\omega, d_\infty)$ become metric spaces.

As we shall see later the above four relations on words are *partial orderings*. The most natural *total orderings* of Σ^* are the *lexicographic* and *alphabetic orderings*, in symbols $\prec_l$ and $\prec_a$, defined as follows. Assume that the alphabet Σ is totally ordered by the ordering $\prec$. This is extended to Σ^* in the following ways:

$$u \prec_l v \text{ iff } u^{-1}v \in \Sigma^+ \text{ or } \mathrm{pref}_1((u \wedge v)^{-1}u) \prec \mathrm{pref}_1((u \wedge v)^{-1}v)$$

and

$$u \prec_a v \text{ iff } |u| < |v| \text{ or } |u| = |v| \text{ and } u \prec_l v.$$

Consequently, u is lexicographically smaller than v if, and only if, either u is a proper prefix of v, or the first symbol after the maximal common prefix $u \wedge v$ is smaller in u than in v. It follows that the orderings $\prec_a$ and $\prec_l$ coincide on words of equal length. In some respects they, however, behave quite differently: each word u is preceded only by finitely many words in $\prec_a$, while for $\prec_l$ this holds only for words composed on the smallest letter of Σ.

It follows directly from the definition that the alphabetic ordering $\prec_a$ is *compatible* with the product on two sides: for all words $u, v, z, z' \in \Sigma^*$ we have

$$u \prec_a v \text{ iff } zuz' \prec_a zvz'.$$

For the lexicographic ordering $\prec_l$ the situation is slightly more complicated. As is straightforward to see we have for all $u, v, z, z' \in \Sigma^*$,

$$u \prec_l v \text{ iff } zu \prec_l zv,$$

and

$$u \prec_l v \text{ and } u \notin \mathrm{pref}(v) \text{ implies that } uz \prec_l vz'.$$

2.2 Periods in words

We continue by defining some further notions of words, in particular those connected to periodicity.

We say that words u and v are *conjugates*, if they are obtainable from each other by the cyclic permutation $c : \Sigma^* \to \Sigma^*$ defined as

$$\begin{aligned} c(1) &= 1, \\ c(u) &= \mathrm{pref}_1(u)^{-1} u \,\mathrm{pref}_1(u) \text{ for } u \in \Sigma^+. \end{aligned}$$

Consequently, u and v are conjugates if, and only if, there exists a k such that $v = c^k(u)$. It follows that the conjugacy is an equivalence relation, each class consisting of words of the same length. It also follows that the equivalence class $[u]$ is included in $F(uu)$, or even in $F(\mathrm{pref}_1(u)^{-1}uu)$.

Next we associate periods to each word $u \in \Sigma^+$. Let $u = a_1 \ldots a_n$ with $a_i \in \Sigma$. *A period* of u is an integer p such that

$$(1) \qquad a_{p+i} = a_i \quad \text{for } i = 1, \ldots, n - p.$$

The smallest p satisfying (1) is called *the period* of u, and it is denoted by $p(u)$. It follows that any $q \geq |u|$ is a period of u, and that

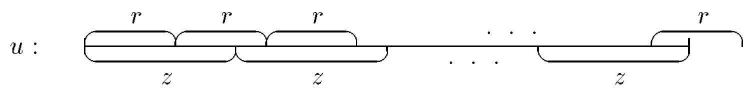
$$u \in \text{pref}(\text{pref}_{p(u)}(u))^{\omega} \text{ and } u \notin F(v^{\omega}) \text{ for any } v \in \Sigma^{\leq p(u)-1}.$$

It also follows that the conjugates have the same periods. The words in the conjugacy class $[\text{pref}_{p(u)}(u)]$ are called *cyclic roots* of u. Note that not all cyclic roots of u need to be factors of u, but at least one, namely the prefix of u of length $p(u)$, is so.

We say that a word $u \in \Sigma^+$ is *primitive*, if it is not a proper integer power of any of its cyclic roots. We claim that this is equivalent to the following condition (often used as the definition of the primitiveness):

(2) $$\forall z \in \Sigma^* : u = z^n \text{ implies } n = 1 \text{ (and hence } u = z).$$

Clearly, (2) implies the primitiveness. To see the reverse we assume that u is primitive and $u = z^n$ with $n \geq 2$. Then denoting $r = \text{pref}_{p(u)}(u)$ we have the situation depicted as



Since $|r|$ is the period necessarily $|z| \geq |r|$. Moreover, by the primitiveness $z \notin r^*$. Consequently, comparing the prefixes of length $|r|$ in the first two occurrences of z we can write

(3) $$r = ps = sp \quad \text{with} \quad p, s \neq 1.$$

The identity (3) is the most basic on combinatorics of words, and implies – after a few line proof, cf. Corollary 4.1 – that p and s are powers of a nonempty word. Therefore u would have a smaller period than $|r|$, a contradiction.

We derive directly from the above argumentation the following representation result of words.

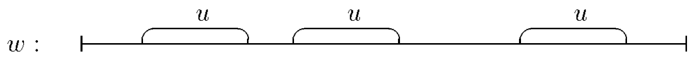
Theorem 2.1. *Each word $u \in \Sigma^+$ can be uniquely represented in the form $u = \rho(u)^n$, with $n \geq 1$ and $\rho(u)$ primitive.* □

The word $\rho(u)$ in Theorem 2.1 is called the *primitive root* of the word u.

There exist two particularly interesting subcases of primitive words: unbordered and Lyndon words. A word $u \in \Sigma^+$ is said to be *unbordered*, if none of its proper prefix is one of its suffixes. In terms of the period $p(u)$ this can be stated as

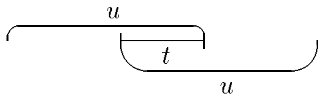
$$u \in \Sigma^+ \text{ is unbordered if, and only if, } p(u) = |u|.$$

It follows that unbordered words are primitive. Moreover, unbordered words have the following important property: different occurrences of an unbordered factor u in a word w never *overlap*, i.e., they are separate:



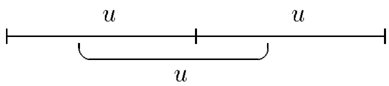
On the other hand, if $u \in \Sigma^+$ is not unbordered, i.e., is *bordered*, then it contains an overlap:

(4)



Consequently, bordered words are sometimes called *overlapping*.

As we noted the situation depicted in (4) is impossible for unbordered words. If u is only primitive, then a variant of (4) is as follows: no primitive word u can be an *inside* factor of uu, i.e., whenever $uu = pus$, then necessarily $p = 1$ or $s = 1$. Being an inside factor can, of course, be illustrated as



This, indeed, is impossible for primitive words by the argument used in (3).

We note that this simple lemma of primitive words is extremely useful in many concrete considerations. As a general example fast algorithms for testing the primitiveness can be based on that. Indeed, use any (linear time) pattern matching algorithm, cf. [CR], to test whether the pattern u is a factor in uu in a nontrivial way, and if "no" the primitiveness of u has been verified.

Now, we go to the second important subcase of the primitive words. A *Lyndon word* $u \in \Sigma^+$ is a word which is primitive and the smallest one in its conjugacy class $[u]$ with respect to the lexicographic ordering.

It is easy to see that a Lyndon word is unbordered. This follows since of the words vuv, vvu and uvv, with $u, v \in \Sigma^+$ and vuv primitive, the first one is never the smallest one. Indeed, by the primitiveness of vuv, we can use the argumentation of (3) to conclude that $vuv \notin \mathrm{pref}(v^\omega)$. Consequently, vuv deviates from v^ω before its end, and so uvv does it earlier and vvu later, if ever, than vuv. Therefore if $vuv \prec_l v^\omega$, then $uvv \prec_l vuv$, and otherwise $vvu \prec_l vuv$.

Let $\mathcal{L}$ denote the set of all Lyndon words. A fundamental property of these words is the following representation result:

Proposition 2.1. *Each word $u \in \Sigma^+$ admits the unique factorization as a product of nonincreasing Lyndon words, i.e., in the form*

$$u = l_1 \ldots l_n, \quad \text{with} \quad l_j \in \mathcal{L} \quad \text{and} \quad l_n \preceq_l l_{n-1} \preceq_l \ldots \preceq_l l_1.$$

The proof of Proposition 2.1 can be found in [Lo], which studies extensively Lyndon words and their applications to factorizations of free monoids. Algorithmic aspects of Lyndon words can be found in [Du2] and [BePo].

2.3 Repetitions in words

One of the most intensively studied topics of combinatorics of words is that of repetitions in words initiated already by Thue in [T1] and [T2]. This differs from the above periodicity considerations in the sense that the focus is on factors of words instead of words themselves. We state the basic definitions here to be used later in Section 8.

A nonprimitive word is a proper power of another, and hence contains a repetition of order at least 2. More generally, a word u is said to contain a *repetition of order* k, with a rational $k > 1$, if it contains a factor of the form

$$z \in \mathrm{pref}(r^{\omega}), \quad \text{with} \quad \frac{|z|}{|r|} = k.$$

In particular, if $|z| = 2|r|$ and $u = z_1 r r z_2$, with $z_1, z_2 \in \Sigma^*$, u contains a repetition of order 2, i.e., a *square* as a factor.

Special emphasis has been put to study *repetition-free* words. We define three different variants of this notion as follows. Let $k > 1$ be a *real* number. We say that $u \in \Sigma^{\infty}$ is

k-free, if it does not contain as a factor a repetition of order at least k;
k^+*-free*, if, for any $k' > k$, it is k'-free;
k^-*-free*, if it is k-free, but not k'-free for any $k' < k$.

It follows that the k^--freeness implies the k-freeness, which, in turn, implies the k^+-freeness. The reverse implications are not true in general, cf. Example 8.1 and Theorem 8.1. There exist commonly used special terms for a few most frequently studied cases: 2-free, 2^+-free and 3-free words are often called *square-free*, *overlap-free* and *cube-free* words, respectively.

In order to illustrate further the above notions we note that in the case $k = 2$, the 2-freeness means that u does not contain as a factor any square, the 2^+-freeness means that it does not contain any factor of the form $vwvwv$, with $v, w \in \Sigma^+$, and the 2^--freeness means that it does not contain any square, but does contain repetitions of order $2 - \varepsilon$, for any $\varepsilon > 0$. As an example, for the word $u = babaabaabb$ the highest order of repetitions is $2\frac{2}{3}$, since it contains the factor $(aba)^{2\frac{2}{3}} = abaabaab$. Note that although u does not contain a factor of the form $v^{2\frac{3}{5}}$ it is not $2\frac{3}{5}$-free, since it contains a repetition of order $2\frac{2}{3} > 2\frac{3}{5}$.

The above notions were generalized in [BEM], and independently in [Z], to arbitrary patterns as follows. Let Ξ be another alphabet, and P a word over Ξ, so-called *pattern*. We say that $u \in \Sigma^{\infty}$ *avoids* the pattern $P \neq 1$ in Σ, if u does not contain a factor of the form $h(p)$, where h is a morphism

$h : \Xi^* \to \Sigma^*$ with $h(x) \neq 1$ for all x in Ξ. Further a pattern P is called *avoidable* in Σ, if there exists an infinite word $u \in \Sigma^\omega$ avoiding P.

For example, the pattern xx is avoidable in Σ if there exists an infinite square-free word over Σ, and as we already indicated, the pattern $xyxyx$ is avoidable in Σ if there exists an infinite 2^+-free word over Σ. It is worth noting here that the existence of a factor of the form $vwvwv$, with $v, w \in \Sigma^+$, in u is equivalent to the existence of an overlapping factor in u, i.e., of two occurrences of a factor overlapping in u. This explains the term of overlap-free.

Natural commutative variants of the above notions can be defined, when $k \in \mathbb{N}$ and only the k-freeness is considered: we say that $u \in \Sigma^\infty$ is *abelian k-free*, if it does not contain a factor of the form $u_1 \ldots u_k$ with $\pi(u_i) = \pi(u_j)$, for $i, j = 1, \ldots, k$.

In order to motivate the use of infinite words in connection with avoidable words we note the following simple equivalence: *for each pattern P there exist infinitely many words in Σ^* avoiding P if, and only if, there exists an infinite word in Σ^ω avoiding P*. This follows directly from the finiteness of Σ. Indeed, in one direction the implication is trivial. In the other direction it follows since, by the above reason, from any infinite set L of words, each of which contains a prefix v, we can choose an infinite subset L' and a letter $a \in \Sigma$ such that each element of L' contains va, as a prefix.

We conclude this subsection by listing all the cases when the number of k-free or abelian k-free words, for an integer k, is finite. It is an exhaustive search argument which shows that this is the case for the 2-freeness in the binary alphabet, as well as the abelian 3-freeness in the binary and the abelian 2-freeness in the ternary alphabets. Figure 2.2 describes the corresponding trees $T_{2,2}$, $AT_{2,3}$ and $AT_{3,2}$, respectively. All the words of the required types (up to symmetry) are found from the paths of these trees starting at the roots.

As we shall see in Section 8, in all the other cases there exists an infinite word avoiding corresponding k-repetitions or abelian k-repetitions. By these trees all binary words of length at least 4 contain a square, and all binary words of length at least 10 contain an abelian cube. Finally, all ternary words of length at least 8 contain an abelian square.

2.4 Morphisms

As we shall see, or in fact have already seen, morphisms play an important role in combinatorics of words. *Morphisms* are mappings $h : \Sigma^* \to \Delta^*$ satisfying

$$h(uv) = h(u)h(v) \quad \text{for all} \quad u, v \in \Sigma^*.$$

In particular, necessarily $h(1) = 1$, and the morphism h is completely determined by the words $h(a)$, with $a \in \Sigma$. Therefore, as a finite set $h(\Sigma)$ of words a morphism is a very natural combinatorial object.

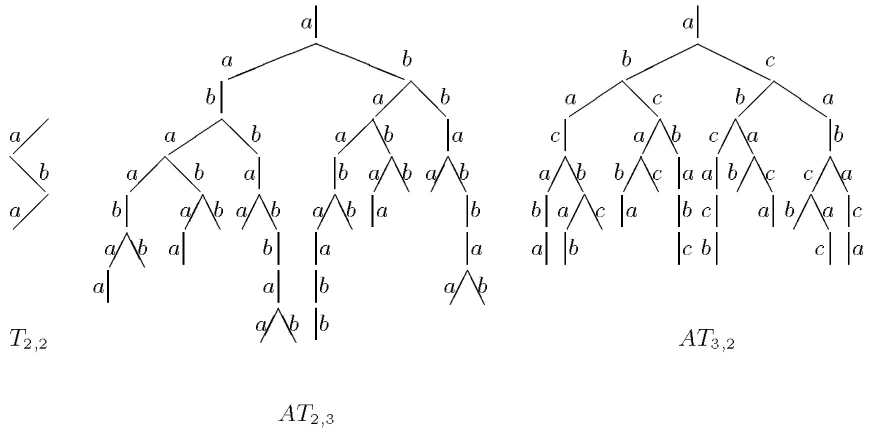


Figure 2.2. Trees $T_{2,2}$, $AT_{2,3}$ and $AT_{3,2}$

We shall need different kinds of special morphisms in our later considerations. We say that a morphism $h : \Sigma^* \to \Delta^*$ is

binary, if $\|\Sigma\| = 2$;
periodic, if there exists a z such that $h(\Sigma) \subseteq z^*$;
1-*free* or *nonerasing*, if $h(a) \neq 1$ for each $a \in \Sigma$;
uniform, if $|h(a)| = |h(b)|$ for all $a, b \in \Sigma$;
prolongable, if there exists an $a \in \Sigma$ such that $h(a) \in a\Sigma^+$;
a *prefix*, if none of the words of $h(\Sigma)$ is a prefix of another;
a *suffix*, if none of the words of $h(\Sigma)$ is a suffix of another;
a *code*, if it is injective;
of *bounded delay* p, if for each $a, b \in \Sigma$, $u, v \in \Sigma^*$ we have: $h(au) \leq h(bv)$ with $u \in \Sigma^p \Rightarrow a = b$;
simplifiable, if there exist morphisms $f : \Sigma^* \to \Gamma^*$ and $g : \Gamma^* \to \Delta^*$, with $\|\Gamma\| < \|\Sigma\|$, such that $h = g \circ f$;
elementary, if it is not simplifiable.

In many cases the alphabets Σ and Δ coincide. In the case of equations or patterns we consider morphisms $h : \Xi^* \to \Sigma^*$, i.e., the set of unknowns is denoted by Ξ. For a uniform morphism h we define its *size* as the number $|h(a)|$, with $a \in \Sigma$. Sometimes periodic morphisms are called *cyclic*. Finally, as an example of a morphism with an unbounded delay we give the morphism defined as $h(x) = a$, $h(y) = ab$ and $h(z) = bb$. Then, indeed, the word ab^ω can be factorized as $h(y)h(z)^\omega$ or $h(x)h(z)^\omega$ in $\{h(x), h(y), h(z)\}^+$.

2.5 Finite sets of words

In this last subsection we turn our attention to sets of finite words, i.e., to languages. Indeed, our main interest will be, on one hand, in words including the infinite ones, and on the other hand, on finite, or at most finitely generated, languages.

Many of the operations defined above for words extend, in a natural way, to languages. Consequently, we may talk about, for instance, a commutative image of a language, or quotients of a language by another one. As an example, let us remind that the set of all factors of words in a language X can be expressed as $F(X) = {\Sigma^*}^{-1} X {\Sigma^*}^{-1}$. We define the *size* $s(X)$ of a finite set X by the identity $s(X) = \sum_{x \in X} |x|$.

Let $X \subseteq \Sigma^*$ and $u_1, \ldots, u_t \in X$. We already said that such a sequence $u_1, \ldots, u_t$ is an X-factorization of u if $u = u_1 \ldots u_t$. Exactly as Σ was extended to Σ^* or Σ^+, we can extend the set X to a *monoid* or *semigroup it generates* by considering all X-factorizations:

$$X^* = \{u_1 \ldots u_t \mid t \geq 0,\ u_i \in X\},$$

and

$$X^+ = \{u_1 \ldots u_t \mid t \geq 1,\ u_i \in X\}.$$

Algebraically, such semigroups are subsemigroups of a finitely generated free semigroup Σ^+, and are called *F-semigroups*. Note that $1 \in X^+$ if, and only if, $1 \in X$. For convenience we concentrate to the semigroup case, and normally assume that $1 \notin X$.

Contrary to Σ^+ the semigroup X^+ need not be *free* in the sense that each $u \in X^+$ would have only one X-factorization. However, what is true is that X^+ (as a set) has the unique *minimal generating set*, namely the set Y defined by

$$Y = (X^+ - \{1\}) - (X^+ - \{1\})^2, \text{ or simply } Y = X^+ - {X^+}^2, \text{ if } 1 \notin X.$$

Indeed, any set Z generating X^+, i.e., satisfying $Z^+ = X^+$, must contain Y. On the other hand, any element of X^+, i.e., a product of elements of X, can be expressed as a product of elements of Y, so that Y generates X^+.

If each word of X^+ has exactly one Y-factorization then the semigroup X^+ is *free*, and its minimal generating set Y is a *code*, cf. [BePe].

One of our goals is to measure the complexity of a finite set $X \subseteq \Sigma^+$. A coarse classification is obtained by associating X with a number, referred to as its *combinatorial rank* or *degree*, in symbols $d(X)$, defined as

$$d(X) = \min\{\|F\| \mid X \subseteq F^*\}.$$

Consequently, $d(X)$ tells how many words are needed to build up all words of X. The simplest case corresponds to *periodic* sets, when all words of X are powers of a same word. The above notion will be compared to, but must not be confused with other notions of a rank of a set which will be called in Section 4 *algebraic ranks*, cf. [Lo].

Another way of measuring the complexity of X is to consider all relations satisfied by X. In this approach codes, i.e., those sets which satisfy only trivial relations, are the "simplest" ones. We prefer to consider these as the largest ones, since, indeed, $\|X^n\|$ assumes the maximal value namely $\|X\|^n$, for all $n \geq 1$.

To formalize the above let $X = \{u_1, \ldots, u_t\} \subseteq \Sigma^+$ be an *ordered* set of words and let $\Xi = \{x_1, \ldots, x_t\}$ be an *ordered* set of unknowns. Let $h_X : \Xi^* \to \Sigma^*$ be a morphism defined as $h_X(x_i) = u_i$. Then the set

$$R_X = \ker(h_X) \subseteq \Xi^* \times \Xi^*$$

defines all the relations in X^+. Further the subrelation

$$\min(R_X) = \{(y, z) \in R_X \mid \forall y', z' \in \Xi^+ : y' < y, z' < z \Rightarrow (y', z') \notin R_X\}$$

corresponds to *minimal* relations in X^+. Note that obviously R_X is a submonoid of the product monoid $\Xi^* \times \Xi^*$, and $\min(R_X)$ is the minimal generating set of it, i.e., $\min(R_X)$ generates R_X, and no element of $\min(R_X)$ is a nontrivial product of two elements of R_X.

Now we define a partial ordering $\preceq_r$ on the set of ordered subsets $X \subseteq \Sigma^+$ of a *fixed* cardinality as follows:

$$X \preceq_r Y \text{ if, and only if, } R_Y \subseteq R_X,$$

or equivalently if, and only if, $\min(R_Y) \subseteq \min(R_X)$. We notice that $\preceq_r$ is a partial ordering, where codes are maximal elements, i.e., for any n-element set X and any n-element code C we have $X \preceq_r C$. We also note that the equality under this ordering means the isomorphism of the corresponding F-semigroups. We call this ordering a *relation ordering*, and shall see in Section 7 that is has quite interesting properties.

We conclude this section of preliminaries with an example illustrating the above definitions.

Example 2.1. Consider the following four ordered sets

$$\begin{aligned}
X_1 &= \{a, abb, bba, baab, babb, baba\},\\
X_2 &= \{a, abb, bba, bb, babb, baba\},\\
X_3 &= \{a, abb, bba, bb, bbb, baba\},\\
X_4 &= \{a, abb, bba, bb, bbb, ba\}.
\end{aligned}$$

Using finite transducers, cf. [Be1], we can compute all words of X_1^+ having two X_1-factorizations, i.e., all nontrivial relations in X_1^+, as explicitly noticed in [Sp1]. All *minimal* such relations are computed by a transducer τ_{X_1} shown in Figure 2.3. The idea of the construction of τ_{X_1} is obvious: τ_{X_1} searches for minimal double X_1-factorizations systematically remembering at its states which of the factorizations is "ahead" and by "how much". Such a transducer contains always two isomorphic copies, so that in our illustration we can omit half of the transducer (the spotted lines in τ_{X_1}).

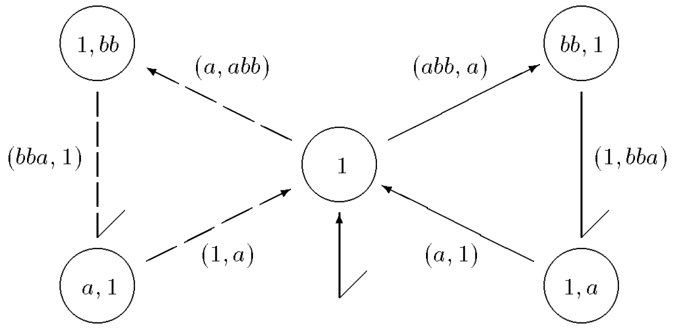


Figure 2.3. Transducer τ_{X_1}

Let us denote by id_{X_1} the identity relation of X_1^+. Then, τ_{X_1} can be transformed to compute $\min(R_{X_1})-id_{X_1}$ simply by relabelling the transitions as shown in Figure 2.4. Let us denote this transducer by τ_1. Similarly, we can compute, as shown in Figure 2.5, the transducers τ_i defining the relations $\min(R_{X_i}) - id_{X_i}$, for $i = 2, 3, 4$.

It follows that $X_4 \prec_r X_3 \prec_r X_2 \prec_r X_1 \prec_r C_6$, where C_6 is any six element code. As we shall see in Section 7, the above procedure cannot be continued for ever, i.e., each proper chain is finite. □

3. Selected examples of problems

In this section we consider three different problems which, we believe, illustrate several important aspects and techniques used in combinatorics of words. The problems deal with different possibilities of mapping Σ^* into Δ^*, a characterization of binary equality languages, and a problem of separating two words by a finite automaton.

3.1 Injective mappings between *F*-semigroups

In this subsection we consider a problem of mapping a word monoid, or more generally a finitely generated F-semigroup, into another one. Moreover, we require that such a mapping satisfy either some algebraic or automata-theoretic properties. The properties we require are that mappings are

isomorphisms;
embeddings mapping generators into generators;
general embeddings;
bijective sequential mappings.

In particular, all mappings are injective.

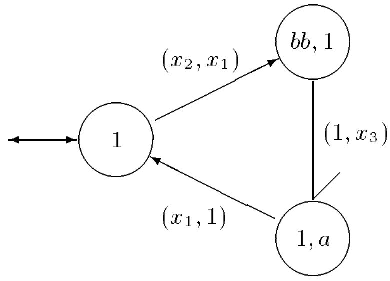


Figure 2.4. Transducer τ_1 accepting $\min(R_{X_1})$

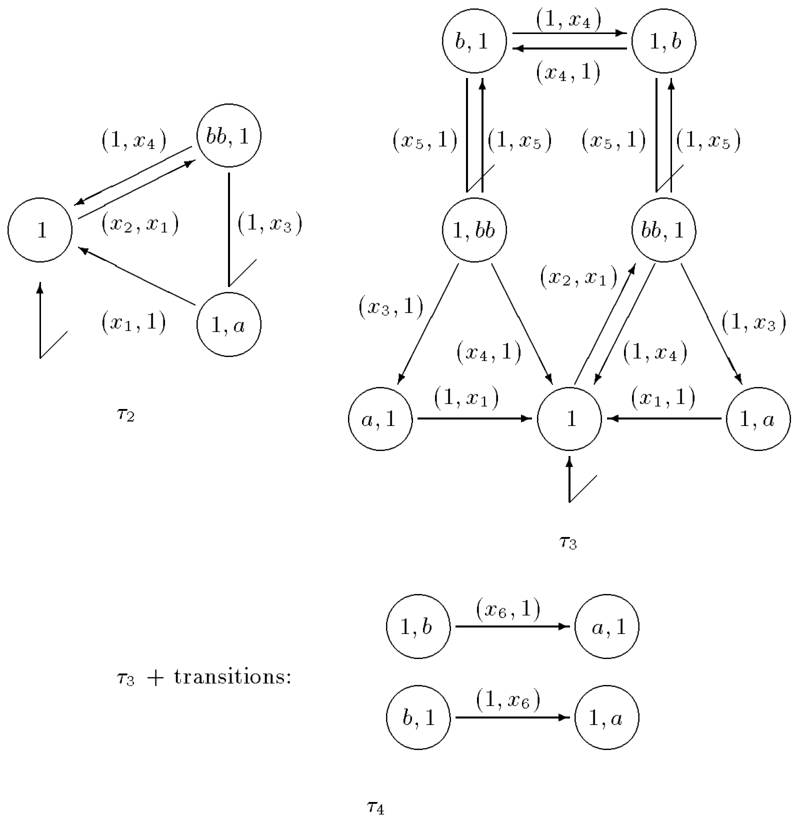


Figure 2.5. Transducers τ_2, τ_3 and τ_4

Isomorphisms. For finitely generated free semigroups a required isomorphism exists if, and only if, the minimal generating sets of the semigroups are of the same cardinality, and in such a case any bijection between those would work. Also for F-semigroups a necessary condition is that the cardinalities of the minimal generating sets are equal. Therefore, for F-semigroups the problem reduces to that of testing whether a given bijection between generating sets is an isomorphism. How this can be done is shown in Section 7.

Embeddings preserving generators. This problem can be solved by the method of the first problem: guess the bijection, and test whether it is an isomorphism.

Embeddings. Interestingly this is always possible, if only the target semigroup is not *cyclic*, i.e., a subsemigroup of u^*, for some $u \in \Sigma^+$. In order to see this we consider first free semigroups Σ_∞^+ and Σ_2^+ with countably many and two generators, respectively. Let $\Sigma_\infty = \{a_i \mid i \in \mathbb{N}\}$ and $\Sigma_2 = \{a, b\}$. Then the morphism $f : \Sigma_\infty^+ \to \Sigma_2^+$ defined as

$$f(a_i) = a^i b \quad \text{for} \quad i \in \mathbb{N},$$

gives a required embedding. This is due to the fact that f is injective, or even a prefix.

For finitely generated F-semigroups X^+ and Y^+ we proceed as follows. We allow X^+ to be countably generated, say $X = \{u_i \mid i \in \mathbb{N}\} \subseteq \Sigma^+$, and require that Y contains two noncommuting words $\alpha, \beta \in \Delta^+$. Then a required embedding $h : X^+ \to Y^+$ is obtained as the composition

$$u_i \overset{\pi}{\longmapsto} a_{i_1} \ldots a_{i_t} \overset{f}{\longmapsto} a^{i_1} b \ldots a^{i_t} b \overset{c}{\longmapsto} \alpha^{i_1} \beta \ldots \alpha^{i_t} \beta,$$

where $\pi : X^+ \to \Sigma^+$ is a natural projection, f is as above, and $c : \{a, b\}^* \to \Delta^*$ is defined by $c(a) = \alpha$ and $c(b) = \beta$. The mapping h is indeed a morphism, and moreover, injective as a composition of injective morphisms. Note that the injectivity of c follows, since α and β are assumed to be noncommuting, so that they do not satisfy any nontrivial identity, cf. Corollary 5.1.

Next we move from algebraic mappings to automata-theoretic ones.

Bijective sequential mappings. We search for a bijective sequential mapping $T : \Sigma^* \to \Delta^*$, where Σ and Δ are finite alphabets. Recall that *sequential mappings*, or sequential transductions in terms of [Be1] or deterministic generalized sequential mappings of [GR], cf. also [Sal1], are realized by deterministic finite automata over Σ, without final states and equipped with outputs in Δ^*, i.e., for each transition an output word of Δ^* is produced. Such automata are called *sequential transducers* in [Be1]. As an illustration we consider the sequential mapping $T : \{a, b, c\}^* \to \{x, y\}^*$ realized by the transducer of Figure 3.1.

The requirement that τ has to realize a bijection, implies that the underlying automaton with respect to inputs must be a *complete* deterministic automaton. Consequently, the inputs can be ignored (if only there are $\|\Sigma\|$ outgoing transitions from each state), and so we are left with the problem,

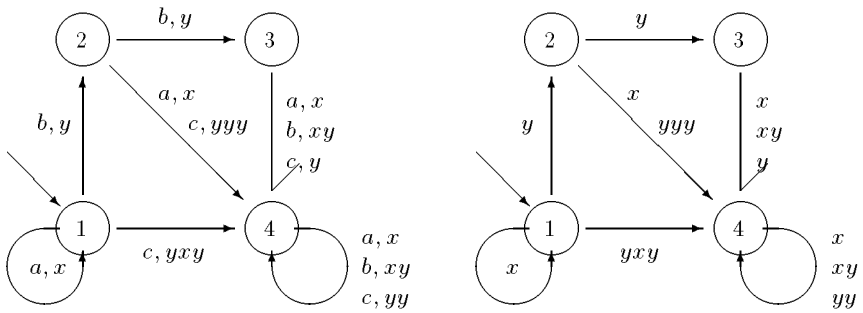


Figure 3.1. A sequential transducer τ and its underlying output automaton $\mathcal{A}$

whether the underlying output automaton of τ, say $\mathcal{A}$, accepts *unambiguously* Δ^*. Therefore we are led to the theory of finite automata with multiplicities, or in terms of [Ei] to the theory of $\mathbb{N}$-rational subsets. Now, using the Equality Theorem of [Ei] it is easy to check that the above T, constructed by Schützenberger, is actually a required bijection $\{a,b,c\}^* \to \{x,y\}^*$.

Next we introduce a systematic method from [Ch] for constructing sequential bijections $\Sigma^* \to \Delta^*$, and illustrate it in the case when $\Sigma = \{a,b,c\}$ and $\Delta = \{x,y\}$. We start from a *maximal suffix code* X over Δ, cf. [BePe]. Such sets are exactly those represented by binary trees, each node of which contains either 0 or 2 descendants. It follows that if S is the subset of all proper suffixes of words in X, then each word $u \in \Delta^*$ has the unique representation in the form $u = sx$, with $s \in S \cup \{1\}$ and $x \in X^*$. In other words, we have the following relation on $\mathbb{N}$-subsets (where we use $+$ instead of $\cup$):

$$\Delta^* = (1+S)X^*. \tag{1}$$

Now, let us return to our specific case, and fix X to be the smallest three-element maximal suffix code, i.e., $X = \{x, xy, yy\}$ (or its renaming). Consequently, $S = \{y\}$, and hence using standard properties of $\mathbb{N}$-subsets, cf. [Ei] chapter 3, we transform (1) as follows:

$$\begin{aligned}
\Delta^* &= (1+y)X^* = 1 + (X+y)X^* \\
&= 1 + (x + xy + yy + y)X^* \\
&= 1 + x(1+y)X^* + (y+yy)X^* \\
&= 1 + x\Delta^* + (y+yy)X^*.
\end{aligned}$$

This relation leads to the two state unambiguous automaton $\mathcal{A}_X$ of Figure 3.2 accepting $\{x,y\}^*$:
Here, $\mathcal{A}_X$ can be obtained, for example, by reversing the method of computing the behaviour of an $\mathbb{N}$-automaton using linear systems of equations, cf. [Ei] chapter 7.

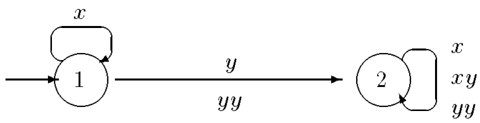
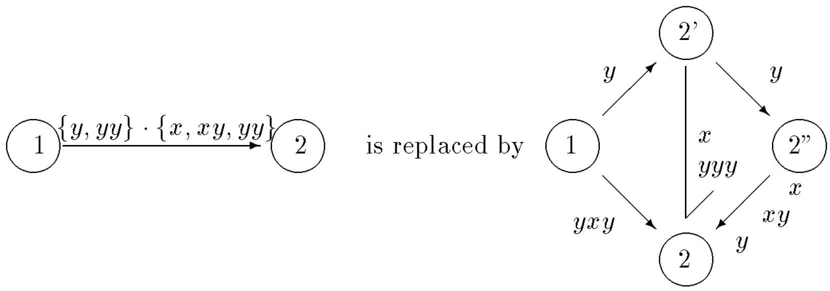


Figure 3.2. Automaton $\mathcal{A}_X$

As we mentioned a sequential bijection $\{a, b, c\}^* \to \{x, y\}^*$ is obtained from $\mathcal{A}_X$ by labelling, for each state, the outgoing transitions bijectively by $\{a, b, c\}$. We also note that the automaton $\mathcal{A}$ of Figure 3.1 can be derived from the above $\mathcal{A}_X$ by unrolling the loop of state 2 once, and redistributing the loop-free unrolled paths between two states in a suitable way:



More generally, for details cf. [Ch], if $||\Sigma|| - 1$ divides $||\Delta|| - 1$, as in our above considerations, by choosing a maximal suffix code X of the cardinality $(||\Delta|| - 1)/(||\Sigma|| - 1)$, one can construct a rational sequential bijection $\Sigma^* \to \Delta^*$ realized by a two state automaton of Figure 3.3, where x is an arbitrary letter of Δ and $S = \text{suf}(X) - (\{1\} \cup X)$.

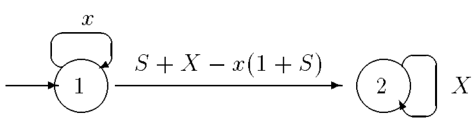


Figure 3.3. A two state automaton realizing a bijection $\Sigma^* \to \Delta^*$

An elaboration of the previous considerations, cf. [Ch], leads to the following result.

Theorem 3.1. *There exists a bijective sequential mapping $\Sigma^* \to \Delta^*$ if, and only if, $||\Sigma|| = ||\Delta||$ or $||\Sigma|| > ||\Delta|| > 1$. Moreover, if this is the case, then such a mapping is realized by a two state sequential transducer.* □

The trivial parts of Theorem 3.1 are as follows. First, if $||\Sigma|| = ||\Delta||$, then the identity mapping (or a renaming) works. Second, if $1 = ||\Delta|| < ||\Sigma||$ or $||\Sigma|| < ||\Delta||$, then simple cardinality arguments show that no required bijection exists.

Finally, we mention that a related problem searching for sequential transductions mapping a given regular set onto another regular one was considered in [MN] and [McN1].

The issues presented in this subsection deserve some comments. Due to the embedding $f : \Sigma_\infty^+ \to \Sigma_2^+$, for many problems in formal language theory, as is well-known, it is irrelevant what the cardinality of the alphabet is, as long as it is at least two. Certainly this is the case when *the property $\mathcal{P}$ to be studied is preserved under the encoding f* in the following sense. The encoded instance of a problem is still an instance of the original problem, and it has the property $\mathcal{P}$ if, and only if, the original instance has the property $\mathcal{P}$.

Let us take an example. Consider a property $\mathcal{P}$ of languages accepted by finite automata. Clearly, rational languages are closed under the encoding f, and moreover many of the natural properties, such as the finiteness, for example, holds for L if, and only if, it holds for $f(L)$. However, if we would consider $\mathcal{P}$ on languages accepted by n-state finite automata, then $\mathcal{P}$ would not be preserved under the encoding f, and hence the cardinality might matter.

There are even more natural cases when the size of the alphabet is decisive. This happens, for instance, when the problem asks something about the domain of morphisms. For example, whether for morphisms $h, g : \Sigma^* \to \Delta^*$ there exists a word $w \in \Sigma^+$ such that $h(w) = g(w)$ – this is the well-known Post Correspondence Problem for lists of length $||\Sigma||$, cf. [HK2]. On the other hand, this problem is independent of the target alphabet, as long as it contains at least two letters. Another example is the avoidability of a pattern in infinite words. Of course no embedding from an alphabet of at least three letters into a binary one preserves the square-freeness. Therefore, avoidability problems depend, in general, crucially on the size of the alphabet.

Finally, we note that normally it is enough that an encoding is injective instead of bijective. However, if bijective encodings were needed the solutions of the last problem might be useful, especially because they are defined in terms of automata theory.

3.2 Binary equality sets

As the second example we consider a simple combinatorial problem connected to the above Post Correspondence Problem. For two morphisms $h, g : \Sigma^* \to \Delta^*$ we define their *equality set* as

$$E(h, g) = \{w \in \Sigma^* \mid h(w) = g(w)\}.$$

In the next result we present a partial characterization from [EKR2] of equality sets of binary morphisms.

Theorem 3.2. *The equality set of two nonperiodic binary morphisms is always of one of the following forms*

$$\{\alpha, \beta\}^* \text{ or } (\alpha\gamma^*\beta)^* \text{ for some } \alpha, \beta, \gamma \in \Sigma^*.$$

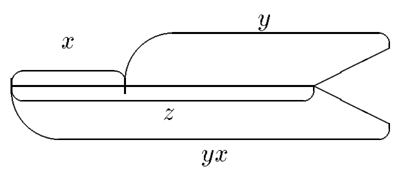
In particular, such a set is rational.

Proof. By the considerations of the previous subsection we may assume that h and g are morphisms from $\{a, b\}^*$ into itself. The proof uses essentially the following simple lemma which, we believe, is interesting on its own right.

Lemma 3.1. *Let $X = \{x, y\} \subseteq \Sigma^+$ be a nonperiodic set. Then, for each word $u, v \in X^+$, we have*

$$u \in xX^+, v \in yX^+; |u|, |v| \geq |xy \wedge yx| \Rightarrow u \wedge v = xy \wedge yx.$$

Proof of Lemma. By symmetry, we may assume that $|y| > |x|$. Let $z = xy \wedge yx$, so that, by the nonperiodicity of X, we have $|z| < |xy|$, cf. Corollary 4.1. Now, if $|z| < |x|$ we are done. In the other case we have the situation depicted as



Now, $v \in yyX^+ \cup yxX^+$ and $y \in x\Sigma^*$, so that $|v \wedge yx| > |z|$.

We shall show that also $|u \wedge xy| > |z|$, from which the claim follows, i.e., $|u \wedge v| = |z| = |xy \wedge yx|$. To see this we first note, by the identity $xy \wedge yx = z$, that z has a period $|x|$, i.e., $z \in \text{pref}(x^\omega)$. Second, by the inequality $|y| > |z| - |x|$, we conclude that y has a prefix of length $|z| - |x| + 1$ in $\text{pref}(x^\omega)$. Therefore, the words $u \in xX^+$ and xy have a common prefix of length $|z| + 1$ (in $\text{pref}(x^\omega)$). So our proof is complete. □

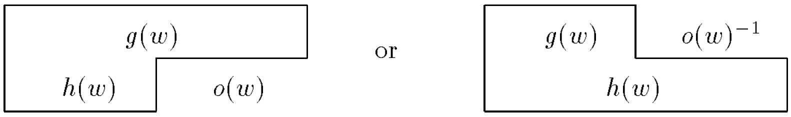
Proof of Theorem (continued). We are going to use this lemma to show that in the exhaustive search for elements in the equality set of the pair (h, g) there exists a unique situation when this search does not go deterministically. Before doing this we need some terminology.

Referring to the Post Correspondence Problem, let us call elements of $E(h, g)$ *solutions*, elements of $(E(h, g) - \{1\}) - (E(h, g) - \{1\})^2$ *minimal solutions*, and prefixes of solutions *presolutions*. Further with each presolution w we associate its *overflow* $o(w)$ as an element of the free group generated by $\{a, b\}$:

$$o(w) = h^{-1}(w)g(w).$$

Finally, we say that a presolution w (or the overflow it defines) *admits a c-continuation*, with $c \in \{a, b\}$, if wc is a presolution as well.

Now, let us consider a fixed overflow $o(w)$. Depending on whether it is an element of $\{a, b\}^*$ or not we can illustrate the situation by the following figures:



Assuming that we have the first case (the other being symmetric) we now analyse what it means that w admits both a- and b-continuations. Since $E(h, g)$ is closed under the product this can be stated that there exist words w_a and w_b, which can be chosen as long as we want, such that waw_a and wbw_b are solutions. This is illustrated in Figure 3.4.

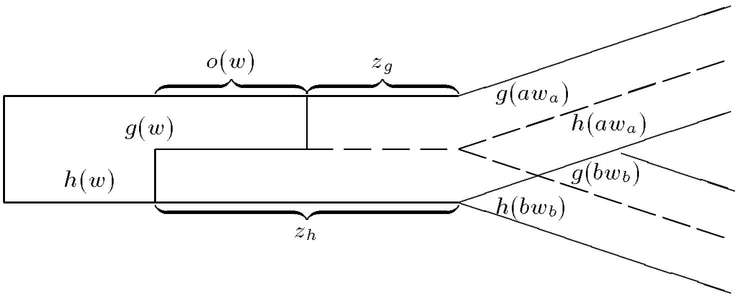


Figure 3.4. a- and b-continuations

By our choice, $g(waw_a) = h(waw_a)$ and $g(waw_b) = h(wbw_b)$. Now, by the lemma, necessarily

$$g(aw_a) \wedge g(bw_b) = g(ab) \wedge g(ba) = z_g$$

and

$$h(aw_a) \wedge h(bw_b) = h(ab) \wedge h(ba) = z_h.$$

Consequently, both waw_a and wbw_b can be solutions only if

$$o(w) = z_h z_g^{-1},$$

as already depicted in Figure 3.4. This value of $o(w)$ is the unique element of the free group depending only on the pair (h, g). In the case considered it is an element of $\{a, b\}^*$, and in the symmetric case the inverse of an element of $\{a, b\}^*$.

So we have proved that only one particular value of the overflow may allow two ways to extend presolutions into minimal solutions. Let us call such an overflow *critical*. Having this property the completion of the proof is an easy case analysis.

First, if the critical overflow does not exist. Then the presolution 1, if it is such, can be extended to a minimal solution in a unique way. Therefore $E(h,g) = \alpha^*$ for some word $\alpha \in \{a,b\}^*$. If the overflow 1 is the critical one, then the above argumentation shows that $E(h,g) = \{\alpha,\beta\}^*$ for some words $\alpha, \beta \in \{a,b\}^*$.

Finally, if the critical overflow exists, and is different from the empty word, we proceed as follows. Let w be the prefix of a minimal solution such that $o(w)$ is critical. Clearly such a w is unique. We call a letter c *repetitive*, if there exists a word $\overline{w}_c$ such that

$$o(w) = o(wc\overline{w}_c) \text{ and } wcw' \notin E(h,g) \text{ for any } w' \in \mathrm{pref}\,\overline{w}_c.$$

Now, if neither a nor b is repetitive, then by the definition of the critical overflow, $E(h,g) = \{\alpha,\beta\}^*$ for some words $\alpha,\beta \in \{a,b\}^+$. If exactly one of the letters a and b is repetitive, then $E(h,g) = (\alpha\gamma^*\beta)^*$ for some words $\alpha,\beta,\gamma \in \{a,b\}^*$. Indeed, if a is the repetitive letter, then $\alpha = w$, $\gamma = a\overline{w}_a$, and β equals to the unique word $\hat{w}_b$ such that $wb\hat{w}_b$ is a minimal solution. Again the definition of the critical overflow guarantees the existence of $\hat{w}_b$. A similar argumentation rules out the case that both a and b are repetitive. This completes the proof of Theorem 3.2. □

Theorem 3.2 motivates a number of comments, which, we believe, illustrate nicely how intriguing simple problems of words can be.

First, the cases ruled out in Theorem 3.2, when at least one the morphisms is periodic are easy. If just one is periodic, then, by the defect theorem, cf. Theorem 5.1, the other is injective, and therefore the equality set may contain at most one minimal solution, i.e., is of the form α^* for some $\alpha \in \{a,b\}^*$. If both, in turn, are periodic, then the equality set, if not equal to $\{1\}$, consists of all words containing the letters a and b in a fixed ratio $q \in \mathbb{Q}_+ \cup \{\infty\}$. Such languages are sometimes denoted by L_q.

Second, the idea of the proof of Theorem 3.2 is not extendable into larger alphabets, since the Lemma 3.1, which is the basis of the proof, does not seem to have counterparts in larger alphabets. Note that this lemma allows to construct from a given binary nonperiodic morphism h a so-called *marked* morphism h', i.e., a morphism h' satisfying $\mathrm{pref}_1 h'(a) \neq \mathrm{pref}_1 h'(b)$, simply by applying the cyclic permutation c of Section 2.2 $|h(ab) \wedge h(ba)|$ times simultaneously to $h(a)$ and $h(b)$.

Third, and most interestingly, we compare the result of Theorem 3.2 to the problem of testing whether, for two binary morphisms h and g, there exists a word $w \neq 1$ such that $h(w) = g(w)$, i.e., to the decidability problem of PCP(2). Certainly, our proof of Theorem 3.2 is nonconstructive. As such it is, however, if not very short, at least elementary and drastically shorter than the existing decidability proofs of PCP(2), cf. [EKR1], [Pav] or also [HK2] in this handbook, which are about 20 pages long. As shown in [HKK] our existential proof of Theorem 3.2 can be made constructive, if an algorithm for PCP(2), or in fact for its slight generalization so-called GPCP(2), for definitions cf. [HK2], is known. Moreover, the arguments used in [HKK] to conclude this are short.

As a conclusion from above, we know that the equality set of two binary morphisms h and g is always of one of the three different forms, namely L_q, for some $q \in \mathbb{Q}_+ \cup \{\infty\}$, $\{\alpha, \beta\}^*$ or $(\alpha\gamma^*\beta)^*$ for some words $\alpha, \beta, \gamma \in \{a, b\}^*$. Moreover, we can effectively find it, i.e., find a k or a finite automaton accepting the equality set. Still we do not know whether the third possibility can take place!

We consider this open problem as a very nice example of challenging problems of words. Although we think this problem is difficult it is worth noting that in free groups the sets of the form $(\alpha\gamma^*\beta)^*$, with $\alpha, \beta, \gamma \in \Sigma^*$, are generated by two elements only: $\alpha\gamma^i\beta = (\alpha\gamma\beta(\alpha\beta)^{-1})^i\alpha\beta$ for $i \geq 0$.

3.3 Separating words via automata

Given two distinct words $x, y \in \Sigma^*$ we want to measure by how much they differ when processed by a finite automaton. More precisely, we want to compute the minimal size $s(x, y)$ of an automaton, i.e., the minimal cardinality of the set of states, that accepts one word and rejects the other. That this integer exists results from the fact that the free monoid is residually finite: an automaton of size $|x|$ accepting only x separates the two words.

It is easy to check that $d(x, y) = 2^{-s(x,y)}$ defines an ultrametric distance on the free monoid, once we pose $d(x, x) = 0$. Indeed, if this were not the case then for some x, y, z we would have $d(x, z) > \max\{d(x, y), d(y, z)\}$ or equivalently $s(x, z) < \min\{s(x, y), s(y, z)\}$. Then in a minimum size automaton $\mathcal{A}$ separating x and z, the words x and y are indistinguishable, i.e., they take the initial state to the same state. But this means that y and z are distinguished by $\mathcal{A}$, contradicting the minimality of $s(y, z)$.

For a fixed integer n we denote by S_n the maximum of all $s(x, y)$'s for x, y of lengths bounded by n, and we study S_n as a function of n. Here finite automaton means deterministic finite automaton, but it can be replaced by finite non-deterministic automaton, finite permutation-automaton (all letters induce a permutation of the set of states), finite monoids, finite groups etc. This question was implicitly posed in [Jo].

Surprisingly enough, it is difficult to come up with two words which would require a large automaton to be separated, say an infinite family of pairs of words for which the size of the automaton would increase as n^α for some $\alpha > 0$. Actually, using elementary number theory, it is easy to verify that two words of different lengths bounded by n can be separated by an automaton whose size is of the order of $\mathcal{O}(\log n)$. So in particular, two words of different commutative images can be separated by an automaton of size

$\mathcal{O}(\log \max\{|x|,|y|\})$. This observation can be drawn further. Indeed, assume that a factor z of length k occurs a different number of times in x and y. The above argument shows that counting the occurrences of z modulo m, for some $m = \mathcal{O}(\log n)$, discriminates x and y. As a consequence, if it is true that two different words of length n differ on the number of occurrences of some factor of length $\log n$, then these two words can be separated by a finite automaton of size $\mathcal{O}(\log^2 n)$.

The first non-trivial contribution to this problem is due to [GK], where it was proved that S_n/n tends to 0 as n tends to infinity. Approximately at the same time in [Rob1] it was proved that $S_n = \mathcal{O}(n^{2/5}\log^{3/5} n)$, and then that only a slightly worse upper bound holds when dealing with permutation automata, to with $\mathcal{O}(n^{1/2})$, see [Rob2]. We reproduce from [Rob1] a weaker result.

Theorem 3.3. *Given two words u and v of length n there exists an automaton of size $\mathcal{O}(n \log n)^{1/2}$ that accepts u if, and only if, it rejects v.*

Proof. Let us first present the proof intuitively. Let w be the shortest prefix of u that is not a prefix of v. The discriminating automaton aims at recognizing some suffix z (as an occurrence) of w by counting its position in u modulo some integer. Clearly, z may not be too large since the automaton performs a string-matching based on z. But it may not be too small either, else it might have many occurrences and the modulo counting that identifies unambiguously this occurrence might envolve a large integer. Furthermore, the length of z does not by itself guarantee a small number of occurrences. It's its period that counts, so z has to be chosen with a long period compared to its length. The proof consists in solving this trade-off.

Let $\pi(n)$ be the number of primes that are less than or equal to n. The prime number theorem asserts that there exists a constant c for which $\pi(n) > c\frac{n}{\log n}$ holds, see e.g. [HW], Theorem 6. The first claim is of pure number-theoretic nature.

Claim 1. For sufficiently large n there exists a prime number $p \le \frac{3}{c}(n \log n)^{1/2}$ such that the following holds. Let $I \subseteq [1,n]$ be a subset of less than $n^{1/2}\log^{-1/2} n$ elements and let $i \in I$ be a fixed element. Then we have

$$i \not\equiv j \bmod p, \text{ for all } i \ne j \text{ and } j \in I.$$

Proof of Claim 1. We first observe that the number of primes greater than $n^{1/2}\log^{-1/2} n$ dividing $j - i$, for some $j \in I$, is less than $2n^{1/2}\log^{-1/2} n$. This follows from the facts that $|j - i|$ is less than n and that I contains at most $n^{1/2}\log^{-1/2} n$ elements. Now the prime number theorem implies

$$\pi(\frac{3}{c}n^{1/2}\log^{1/2} n) > 3\frac{n^{1/2}\log^{1/2} n}{\frac{1}{2}\log n + \log\frac{3}{c} + \frac{1}{2}\log\log n}.$$

Here for sufficiently large n the numerator is smaller than $\log n$, i.e.,

$$\pi(\frac{3}{c}n^{1/2}\log^{1/2}n) > 3n^{1/2}\log^{-1/2}n.$$

Clearly, among these primes there is one that is greater than $n^{1/2}\log^{-1/2}n$ and that divides no $j-i$. □

The second claim concerns the period $p(w)$ of a word w, cf. Section 2.2.

Claim 2. For all $w \in \Sigma^*$, $\max\{p(wa), p(wb)\} > \frac{|w|}{2}$ holds.

Proof of Claim 2. Assume to the contrary that $p(wa), p(wb) \le \frac{|w|}{2}$. Clearly, $p(wa) \ne p(wb)$. Now wa and wb have a common prefix w of length greater than or equal to $p(wa)+p(wb)$. By the Theorem of Fine and Wilf, cf. Theorem 6.1, this contradicts the minimality of $p(wa)$ and $p(wb)$. □

The last claim gives an estimate on the size of an automaton that carries out a string-matching algorithm, see, Chapter on string-matching in this handbook.

Claim 3. Let $0 \le i < p$, be two integers and let $w \in \Sigma^*$ be a word of length $k < p$. Then there exists an automaton of size less than $2p$ that recognizes the set of words ending in w, having no other occurrence of w and for which this occurrence starts in position i modulo p.

Proof of Claim 3. We let $w = w_1 \dots w_k$ and $[p-1] = \{0, \dots, p-1\}$. The set of states of the automaton equals $[p-1] \cup \{w_1 \dots w_j | 1 \le j \le k\}$, the initial state is 0 and the final state is w. The transition function satisfies

$$w_1 \dots w_j.c = \begin{cases} w_1 \dots w_{j+1}, & \text{if} \quad c = w_{j+1}, \\ i+j+1 \bmod p, & \text{otherwise}, \end{cases}$$

and

$$\alpha.c = \alpha + 1 \bmod p,$$

if $\alpha \in [p-1] - \{i\}$ and $c \in \Sigma$ or if $\alpha = i$ and $c \ne w_1$. □

Proof of Theorem (continued). Now we contruct an automaton that separates u and v. We denote by w their maximal common prefix: $u = wau_1$ and $v = wbv_1$ for some $u_1, v_1 \in \Sigma^*$ and $a, b \in \Sigma$ with $a \ne b$.

We first rule out an easy case where $|w| < 2(n\log n)^{1/2}$. It suffices to consider the automaton accepting all words having wa as a prefix. It recognizes u if, and only if, it rejects v.

Thus we may assume that $|w| \ge 2(n\log n)^{1/2}$, and consider the suffix z of w of length $2(n\log n)^{1/2} - 1$. We have $u = w_1zau_1$, $v = w_1zbv_1$ and $w = w_1z$ for some $w_1 \in \Sigma^*$. By Claim 2, we may assume without loss of generality that $p(za) > \frac{|za|}{2}$. In particular this means that two occurrences of za are at least $\frac{|za|}{2}$ apart and therefore that there are less than $\frac{2n}{2(n\log n)^{1/2}} = (n\log n)^{1/2}$ occurrences of za in u.

If v has no occurrence of za then it suffices to construct the automaton that performs the string-matching with za as the string to be matched (see, Chapter on string-matching). We know that this can be achieved with an automaton of size $|za| = 2(n \log n)^{1/2}$.

We are left with the case where za has also an occurrence in v, i.e., $v = w_2 zav_2$ where $|v_2| < |v_1|$. Let I be the set of positions in u where the occurrences of za end. Let p be as in Claim 1 and let i be the position modulo p of the first occurrence of za in u. Then the automaton $\mathcal{A}$ accepting u and rejecting v consists of two subautomaton $\mathcal{A}_1$ and $\mathcal{A}_2$. Automaton $\mathcal{A}_1$ perfoms as prescribed by Claim 3. When the first occurrence of za is spotted then $\mathcal{A}_1$ switches to $\mathcal{A}_2$, which separates the suffixes u_1 and v_2. We know that $\mathcal{A}_2$ has size bounded by $\log n$. Thus, the automaton $\mathcal{A}$ has size $||\mathcal{A}_1|| + ||\mathcal{A}_2|| < 4(n \log n)^{1/2} + \lambda \log n$, where λ is some constant independent of n. □

4. Defect effect

The defect theorem is one of the important results on words. It is often considered to be a folklore knowledge in mathematics. This may be, at least partially, due to the fact that there does not exist just one result, but, as we shall see, rather many different results which formalize the same *defect effect* of words: *if a set X of n words over a finite alphabet satisfies a nontrivial relation E, then these words can be expressed simultaneously as products of at most $n-1$ words.* One of the older papers where this is proved is [SkSe]. It was also known in [Len].

The defect effect can be considered from different perspectives. One may concentrate on a set X satisfying one (or several) equation(s), or one may concentrate on an equation E (or a set of equations), and try to associate the notion of a "rank" with the objects studied. Our emphasis is in combinatorial aspects of words, so we concentrate on the first approach.

It follows already from the above informal formulation of a defect theorem, that it can be seen as a dimension property of words: if n words are "dependent" they belong to a "subspace of dimension" at most $n-1$. However, as we shall see in Section 4.4, words possess only very restricted dimension properties in this sense.

4.1 Basic definitions

Assume that $X \subseteq \Sigma^+$ is a finite set of words having the defect effect. This means that X is of a "smaller" size than $||X||$, but "how much smaller" depends on what properties are required from words used to build up the words of X. This is what leads to different formulations of the defect theorem.

A combinatorial formulation is based on the notion of the *combinatorial rank* or *degree* of $X \subseteq \Sigma^+$, which we already defined in Section 2.5 by the condition

(1) $$d(X) = \min\{\|F\| \mid X \subseteq F^+\}.$$

It follows immediately that $d(X) \leq \min(\|X\|, \|\Sigma\|)$, so that the finiteness of X is irrelevant. Note also that the degree of a set is not preserved under injective encodings – emphasizing the combinatorial nature of the notion.

In order to give more algebraic formulations we consider the following three conditions. Let $X \subseteq \Sigma^+$ be a finite set and S an F-semigroup. We define three properties of S as follows:

$$\begin{array}{ll} (p) & \forall p, w \in \Sigma^+ \ : \ p, pw \in S \Rightarrow w \in S; \\ (f) & \forall p, q, w \in \Sigma^+ \ : \ p, q, wp, qw \in S \Rightarrow w \in S; \\ (u) & \forall p, q, w \in \Sigma^+ \ : \ pwq \in X^+, pw, wq \in S \Rightarrow w \in S. \end{array}$$

Conditions (p) and (f) are very basic in the theory of codes, cf. [BePe]. The first one characterizes those F-semigroups having a prefix code as the minimal generating set. Such semigroups are often called *right unitary*. The second condition, which is often referred to as the *stability condition*, characterizes those F-semigroups which are *free*, i.e., have a code as the minimal generating set, cf. [LeSc] or [BePe]. The third condition, which differs from the others in the sense that *it depends also on* X, is introduced here mainly to stress the diversified nature of the defect theorem. As shown in [HK1] it characterizes those F-semigroups, where X^+ factorizes uniquely.

For the sake of completeness we prove the following simple

Lemma 4.1. *An F-semigroup S is right unitary if, and only if, it satisfies* (p).

Proof. Assume first that S is right unitary. This means that the minimal generating set, say P, of S is a prefix code. Let $p = u_1 \dots u_n$ and $pw = v_1 \dots v_m$, with $u_i, v_j \in P$, be elements of S. Now, since P is a prefix code we have $u_i = v_i$, for $i = 1, \dots, n$, and therefore $v_{n+1} \dots v_m \in P^+ = S$.

Conversely, assume that the F-semigroup S satisfies (p). Let v and q be in the minimal generating set of S. If $v < q$, then we can write $q = vt$ with $t \in \Sigma^+$. Hence, by (p), t is in S, and q is a product of two nonempty words, a contradiction with the fact that q is in the minimal generating set of S. □

For each $i = p, f, u$, F-semigroups satisfying (i) are trivially closed under arbitrary intersections. Therefore the semigroups

$$\hat{X}(i) = \bigcap_{\substack{X \subseteq S \\ S \text{ sat. } (i)}} S$$

are well-defined, and by the definition, the smallest F-semigroups of type (i) containing X. The semigroups $\hat{X}(i)$, for $i = p, f, u$, are referred to as *free hull*, *prefix hull* and *unique factorization hull* of X. Denoting by $X(i)$ the minimal generating set of $\hat{X}(i)$ we now define three different notions of an *algebraic rank* of a finite set $X \subseteq \Sigma^+$:

$$p(X) = \|X(p)\|, \quad r(X) = \|X(f)\| \quad \text{and} \quad u(X) = \|X(u)\|.$$

These numbers are called *prefix rank* or *p-rank*, *rank* or *f-rank* and *unique factorization rank* or *u-rank* of X, respectively.

The most commonly used notion of a rank of X in the literature is that of our f-rank, cf. [BPPR], or [Lo]. From the purely combinatorial point of view p-rank is often more natural. The reason we introduced all these variants, which by no means are all the possibilities, cf. [Sp2], is that they can be used to illustrate the subtle nature of the phenomenon called the defect effect.

Our next example modified from [HK1] shows that all the four different notions of a rank may lead to a different quantity.

Example 4.1. Consider the set

$$X = \{aa, aaaaba, aababac, baccd, cddaa, daa, baa\}.$$

The only minimal nontrivial relation satisfied by X is

$$aa.aababac.cddaa = aaaaba.baccd.daa.$$

Now, applying (u) we see that $aaba, bac, cd \in \hat{X}(u)$. Replacing the words *aababac*, *cddaa*, *aaaaba* and *baccd* of X by the above three words we obtain a set, where X^+ factorizes uniquely, i.e.,

$$X(u) = \{aa, aaba, bac, cd, daa, baa\}.$$

However, $X(u)^+$ is not free, since we have

$$aa.bac.daa = aaba.cd.aa, \tag{2}$$

which actually is the only nontrivial minimal relation in $X(u)^+$. It follows that $\hat{X}(u)$ is a proper subset of $\hat{X}(f)$. Applying now condition (f) to (2) we conclude that $\hat{X}(f)$ contains the words ba, c and d. But now the set

$$X(f) = \{aa, ba, c, d, baa\}$$

is a code, so that $X(f)$ is this set, as already denoted. Finally, $X(f)$ is not a prefix code, so that applying (p) to $X(f)$, or alternatively the procedure described in a moment to the original X, we obtain that

$$X(p) = \{a, ba, c, d\}.$$

Consequently, we have concluded that $p(X) < r(X) < u(X) < \|X\|$. In this example, the degree of X equals to $p(X)$. However, if we replace X by $X' = e(X)$, where $e : \{a, b, c, d\}^* \to \{a, b, c\}^*$ is a morphism defined as $e(d) = bb$ and $e(x) = x$, for $x \in \{a, b, c\}$, then the degree decreases to 3, while all the algebraic ranks remain unchanged, as is easy to conclude. Therefore we have an example of a set X' satisfying

$$3 = d(X') < p(X') < r(X') < u(X') < \|X'\| = 7. \qquad \square$$

Although we called our three notions of the rank algebraic, they do not have all desirable algebraic properties like being invariant under an isomorphism. Indeed, our next example shows that free hulls (or prefix hulls) of two finite sets generating isomorphic F-semigroups need not be isomorphic, i.e., need not have the same number of minimal generators. On the other hand, as a consequence of results in the next subsection, one can conclude that all the algebraic ranks, we defined, are closed under the encodings which are prefix codes.

Example 4.2. Consider the sets

$$X = \{a, ab, babbb, abbb\} \quad \text{and} \quad Y = \{a, abb, bbba, ba\}.$$

Then X^+ and Y^+ are isomorphic, since both of these semigroups satisfy only one minimal relation, which, moreover, is the same one under a suitable orderings of sets X and Y:

$$\begin{aligned} X &: \quad a.babbb = ab.abbb \\ Y &: \quad a.bbba = abb.ba. \end{aligned}$$

From these relations we conclude, either by definitions or methods of the next subsection, that $X(p) = X(f) = \{a, b\}$, while $Y(p) = Y(f) = \{a, bb, ba\}$. □

4.2 Defect Theorems

In this subsection we show that each of our notions of a rank of a finite set X can be used to formalize the defect effect. In our algebraic cases the words from which the elements of X are built up are, by definitions, unique, while in the case of the degree the minimal F of (1) in Section 4.1 need not be unique.

Let $X \subseteq \Sigma^+$ be finite. We introduce the following procedure using simple transformations to compute the prefix hull of X. Such transformations were used already in [Ni] in connection with free groups.

Procedure P**.** Given a finite $X \subseteq \Sigma^+$, considered as an unambiguous multiset.

1. Find two words $p, q \in X$ such that $p < q$. If such words do not exist go to 4;
2. Set $X := X \cup \{p^{-1}q\} - \{q\}$ as a multiset;
3. If X is ambiguous identify the equal elements, and go to 1;
4. Output $X(p) := X$.

We obtain the following formulation of the defect theorem.

Theorem 4.1. *For each finite $X \subseteq \Sigma^+$, the minimal generating set $X(p)$ of the prefix hull of X satisfies $\|X(p)\| \leq \|X\|$, and moreover $\|X(p)\| < \|X\|$, if X satisfies a nontrivial relation.*

Proof. First of all, by the definition of the prefix hull and Lemma 4.1, the Procedure P computes $X(p)$ correctly. Hence, $\|X(p)\| \leq \|X\|$ always.

The second sentence of the theorem is seen as follows. Whenever an identification is done in step 3 a required decrease in the size of $\|X\|$ is achieved. Such an identification, in turn, is unavoidable since, if it would not occur, steps 2 and 3 would lead from a set X satisfying a nontrivial relation to a new set of strictly smaller size still satisfying a nontrivial relation. Indeed, the new nontrivial relation is obtained from the old one by substituting $q = pt$, with $t = p^{-1}q$, and by cancelling one p from the left in the old relation. Clearly, such a new relation is still nontrivial. □

Theorem 4.1 motivates a few comments. By the definition of the prefix hull as an intersection of certain free semigroups, it is not obvious that $\|X(p)\| \leq \|X\|$. Indeed, the intersection of two finitely generated free semigroups, need not be even finitely generated, cf. [Ka2]. On the other hand, the finiteness of $\|X(p)\|$ is obvious, since it must consist of factors of X.

As the second remark we note that although the proof of Theorem 4.1 is very simple, it has a number of interesting corollaries.

Corollary 4.1. *Two words $u, v \in \Sigma^+$ are powers of a word if, and only if, they commute if, and only if, they satisfy a nontrivial relation.* □

Note that the argumentation of the proof of Theorem 4.1, gives a few line proof for this basic result.

Procedure P can be applied to derive the following representation result for 1-free morphisms $h : \Sigma^* \to \Delta^*$. In order to state it let us call a morphism $e : \Sigma^* \to \Sigma^*$ *basic*, if there are two letters $a, b \in \Sigma$ such that $e(a) = ba$ and $e(c) = c$ for $c \in \Sigma - \{a\}$. Then when applied P to $h(\Sigma)$ in such a way that the identifications are done only at the end we obtain

Corollary 4.2. *Each 1-free morphism $h : \Sigma^* \to \Delta^*$ has a representation*

$$h = p \circ c \circ \pi,$$

where $p : \Delta^* \to \Delta^*$ is a prefix, $c : \Sigma^* \to \Delta^*$ is length preserving and $\pi : \Sigma^* \to \Sigma^*$ is a composition of basic morphisms. □

Obviously, Corollary 4.2 has also a two-sided variant, where p is a biprefix, and in the definition of the basic morphism the condition $e(a) = ba$ is replaced by $e(a) = ba \vee ab$.

Corollary 4.3. *The prefix hull of a finite set $X \subseteq \Sigma^+$ can be computed in polynomial time in the size $s(X)$ of X.*

Proof. Even by a naive algorithm each step in Procedure P can be done in time $\mathcal{O}(s(X)^3)$. So the result follows since the number of rounds in P is surely $\mathcal{O}(s(X))$. □

As a final corollary we note a strengthening of Theorem 4.1.

Corollary 4.4. *If a finite set* $X \subseteq \Sigma^+$ *satisfies a nontrivial* 1*-way infinite relation, i.e.,* X *does not have a bounded delay (from left to right), then* $\|X(p)\| < \|X\|$.

Proof. Indeed, it is not the property of being a noncode, but the property of not having a bounded delay (from left to right), which forces that at least one identification of words takes place in step 3 of Procedure P. □

It is interesting to note that Corollary 4.4 does not extend to 2-way infinite words.

Example 4.3. The set $X = \{abc, bca, c\}$ satisfies a nontrivial 2-way infinite relation ${}^{\omega}(abc)^{\omega} = {}^{\omega}(bca)^{\omega}$, but still even $d(X) = 3 = \|X\|$. □

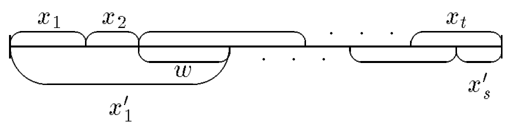
Next we turn from a prefix hull of a finite set X to two other types of hulls defined at the beginning of this subsection. Actually, from the algebraic point of view the free hull $X(f)^+$ is the most natural one, and is considered in details in [BPPR] and [Lo]. It yields the following variant of the defect theorem.

Theorem 4.2. *For each finite* $X \subseteq \Sigma^+$*, which satisfies a nontrivial relation, we have*

$$\|X(f)\| < \|X\|.$$

Without giving a detailed proof of this result we only mention the basic ideas, from which the reader can reconstruct the whole proof, cf. [Sp2]. Actually, the proof given in [BPPR] and [Lo] is even sharper defining precisely the set $X(f)$.

We start from a double X-factorization depicted as



,

where $x_i, x'_j \in X$ and $w \in \Sigma^+$. Then, by the property (f) and the definition of the free hull, w is in the free hull, i.e., in the construction of $X(f)$ we can replace x'_1 of X by w. Now, the new set obtained may be ambiguous yielding a required defect effect, or it is not a code. However, in any case it is of a smaller size than the old one guaranteeing that the procedure terminates.

Note that we already used these ideas in Examples 4.1 and 4.2.

It follows immediately that Corollary 4.3 extends to free hulls, while Corollary 4.4, of course, does not have a counterpart here. Moreover, the free hull of X satisfies the following important property, cf. [BPPR].

Proposition 4.1. *Let $X \subseteq \Sigma^+$ be finite and $X(f)$ the minimal generating set of its free hull. Then, for each $x \in X(f)$, we have $xX(f)^* \cap X \neq \emptyset$.*

The above result states that each word of $X(f)$ occurs as the first one in some $X(f)$-factorization of a word of X, a property which is, by Procedure P, also true for the prefix hull, i.e., for $X(p)$.

What we said above about free hulls extends to unique factorization hulls. The details can be found in [HK1].

Now, we are in the position to summarize our considerations of this subsection. For a finite $X \subseteq \Sigma^+$ we have

$$d(X) \leq p(x) \leq r(x) \leq u(X) \leq \|X\|, \tag{1}$$

where, moreover, the last inequality is proper if X is not a code. Here the first inequality is trivial, the second follows, by the definitions, from the fact that $X(f) \subseteq X(p)^+$, and the third similarly from the fact that $X(u) \subseteq X(f)^+$. As we saw in Example 4.1, each of the inequalities in (1) can be proper simultaneously. They, of course, can be equalities as well.

Example 4.4. Let $X = \{a, ab, cc, bccdd, dda\}$. Then the only nontrivial minimal relation is

$$a.bccdd.a = ab.cc.dda$$

from which we conclude that $X(u) = \{a, b, cc, dd\}$. But this is already a prefix code so that $X(p) = X(f) = X(u)$. Finally, the exhaustive search shows that $d(X) = 4$. Therefore we have an example for which $d(X) = p(X) = r(X) = u(X) = \|X\| - 1$. □

Although we formulated everything in this subsection for sets X not containing the empty word 1, i.e., for free semigroups, the results hold for free monoids, as well. This is because, if $1 \in X$, then trivially any rank of X is strictly smaller that $\|X\| - 1$.

4.3 Defect effect of several relations

In this subsection we consider possibilities of generalizing the above defect theorems to the case of several nontrivial relations. A natural question is: if a set of n words satisfies two "different" nontrivial relations, can these words be expressed as products of $n - 2$ words? Unfortunately, the answer to this question is negative, as we shall see in a moment.

We formalize the term "different" as follows. Let $X \subseteq \Sigma^+$ be a finite set. relations in X^+ are considered as equations with Ξ as the set of unknowns and having X as a solution, cf. Section 2.5. This requires to consider X as an ordered set, and that $\|\Xi\| = \|X\|$. This allows to state the set of all relations of X^+ as a set of equations over Ξ having X as a solution. In Section 2.5 this was referred to as R_X. Here we consider its subset consisting only of so-called *reduced* equations, i.e., equations $(u, v) \in \Xi^+ \times \Xi^+$ satisfying

$\text{pref}_1(u) \neq \text{pref}_1(v)$ and $\text{suf}_1(u) \neq \text{suf}_1(v)$. For simplicity, we prefer to denote the set of all reduced equations of X by $E(X)$.

We say that a system E of equations over the set Ξ of unknowns is *independent* in Σ^+, if no proper subset E' of E is *equivalent* to E, i.e., has exactly the same solutions as E has. Now, identities of X^+ are "different" if their corresponding equations form an independent system of equations.

Example 4.5. The pair

$$xzy = yzx \quad \text{and} \quad xzzy = yzzx$$

of equations is independent, since the former has a solution

$$x = aba,\ y = a \quad \text{and} \quad z = b,$$

which is not a solution of the latter, while the latter has a solution

$$x = abba,\ y = a \quad \text{and} \quad z = b,$$

which is not a solution of the former. However, they have a common solution of degree two, namely $x = y = a$ and $z = b$. □

Despite of Example 4.5 there are some nontrivial conditions which force sets satisfying these conditions to be of at most certain degree. Particularly useful such results are, if they guarantee that the sets are periodic.

In our subsequent considerations, unlike in those of the previous subsection, it is important that *equations are over free semigroups and not over free monoids.*

Let $\{u_1, \ldots, u_n\} = X \subseteq \Sigma^+$ be finite and $E(X) \subseteq \Xi^+ \times \Xi^+$ the set of all (reduced) equations satisfied by X. This means that $X = h(\Xi)$ for some morphism $h : \Xi^+ \to \Sigma^+$ satisfying $h(\alpha) = h(\beta)$ for all (α, β) in $E(X)$. With each equation in $E(X)$, say

$$e : x\alpha = y\beta \quad \text{with} \quad x \neq y,\ x, y \in \Xi,\ \alpha, \beta \in \Xi^*$$

we associate $\pi(e) = \{h(x), h(y)\}$, and with the system $E(X)$ we associate the following graph $G_{E(X)}$:

the set of nodes of $G_{E(X)}$ is X; and
the edges of $G_{E(X)}$ are defined by the condition: (u, v) is an edge in $G_{E(X)} \Leftrightarrow \exists e \in E(X) : \pi(e) = \{u, v\}$.

It follows that $G_{E(X)}$ defines via its compoments an equivalence relation on X. Now, the degree of X is bounded by the number of *connected components* of $G_{E(X)}$, which we denote by $c(G_{E(X)})$, cf. [HK1]. Note that in above X maybe a multiset, and this indeed is needed in the next proof.

Theorem 4.3. *For each finite $X \subseteq \Sigma^+$, we have*

$$d(X) \leq p(X) \leq c(G_{E(X)}).$$

Proof. We already know that the first inequality holds. To prove the second we proceed as in Procedure P of subsection 4.2.

Let $u - v$ be an edge in $G_{E(X)}$. Then assuming, by symmetry, that $u \leq v$ we have two possibilities:

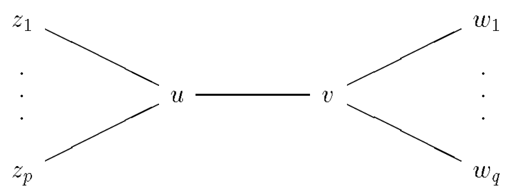
(i) if $u = v$ we identify the nodes u and v;
(ii) if $v = ut$ with $t \in \Sigma^+$, we replace X by $(X \cup \{t\}) - \{v\}$.

Let $X' \subseteq \Sigma^+$ be a multiset obtained from X by applying either (i) or (ii) once. Note that due to (ii) X' indeed can be a multiset although X would be unambiguous. Our claim is that

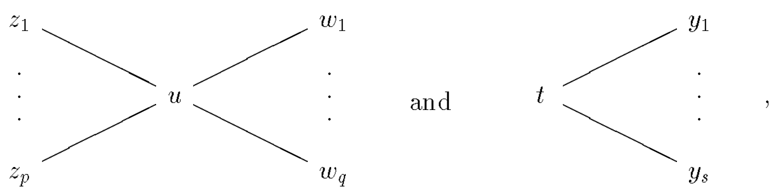
$$c(G_{E(X')}) \leq c(G_{E(X)}). \tag{2}$$

If the operation performed is (i) there is nothing to be proved. So we have to analyse what happens to the graph $G_{E(X)}$ when (ii) is performed. In particular, we have to consider what happens to a subgraph of it of the form:

(3)



Clearly, the connections $z_i - u$ remain, and connections $v - w_j$ are replaced by $u - w_j$. Moreover, v disappears, and the new node t will be connected in $G_{E(X')}$ to all y_k's in X such that $uy_k\alpha = v\beta$, with $\alpha, \beta \in X^*$, are in $E(X)$. In addition, the introduction of the new t may create some completely new edges to $G_{E(X')}$. But what is important is that, if $G_{E(X)}$ contains the subgraph (3), then $G_{E(X')}$ contains the following subgraphs



where, moreover, the nodes y_k are nodes of $E_{G(X)}$, i.e., belong to some of the components of $E_{G(X)}$. Therefore, the replacement of v by t does not increase the number of the components, so that we have proved (2).

By the construction $s(X') < s(X)$, and therefore an iterative application of the rules (i) and (ii) leads finally to the discrete graph, the edges of which are labelled by a set $\hat{X}$. It follows from the arguments of the proof of Theorem

4.1, that $\hat{X}$ is contained in the minimal generating set of the prefix hull of X. Therefore, by Theorem 4.1, $\|X(p)\| \le \|\hat{X}\|$. But, by the discreteness of $G_{E(\hat{X})}$, we have

$$\|\hat{X}\| = c(G_{E(\hat{X})}) \le c(G_{E(X)}),$$

and hence our proof is complete. □

Theorem 4.3 has a number of interesting consequences. First, we have a counterpart of Corollary 4.4: if in (1) equations are replaced by ω-equations, i.e., one-way infinite equations, but otherwise the graph – let us denote it now by $G_{E_\omega(X)}$ – is defined as $G_{E(X)}$ we obtain

Corollary 4.5. *For each finite set $X \subseteq \Sigma^+$ we have*

$$d(X) \le p(X) \le c(G_{E_\omega(X)}).$$ □

More concrete and useful corollaries are obtained, when the graph $G_{E(X)}$ is connected:

Corollary 4.6. *Let $X \subseteq \Sigma^+$ be finite. If $G_{E(X)}$ is connected, then X is periodic.* □

Corollary 4.7. *If a three element set $X = \{u, v, w\} \subseteq \Sigma^+$ satisfies the relations $ux = vy$ and $uz = wt$, with $x, y, z, t \in X^\infty$, then u, v and w are powers of a same word.* □

Corollary 4.7 should be compared to Example 4.5. It also has to be noticed that in our above considerations it is essential that X consists of only nonempty words. Indeed, the graph of the equations

$$x = yx \quad \text{and} \quad z = yz$$

is connected, but it possesses a solution of degree 2, namely $x = a$, $y = 1$ and $z = b$.

As the final application of Theorem 4.3 we give an example from [HK1], which shows that also the inside occurrence of the equations may cause some defect effect.

Example 4.6. Assume that words of X satisfy the reduced equations

$$\alpha u \gamma \;=\; \beta v \delta \quad \text{and} \quad \alpha w \varepsilon = \beta z \rho,$$

where $u, v, w, z \in X$ and $\alpha, \beta, \gamma, \delta, \varepsilon, \rho \in X^+$ and $\{\mathrm{pref}_1(\alpha), \mathrm{pref}_1(\beta)\} \neq \{v, z\}, \{u, w\}$. We claim that $d(X) \le \|X\| - 2$, which cannot be concluded simply by considering the first occurrences of the unknowns in these equations.

There are two cases to be considered. First, if $\alpha = \beta$ (in Σ^+), then u and v, as well as w and z, are in the same component proving the claim. Otherwise assuming, by symmetry, that $\alpha = \beta t$, with $t \in \Sigma^+$, and denoting $X' = X \cup \{t\}$, we see that $G_{E(X')}$ contain the edges $t - z$ and $t - v$, and still one more different from $v - z$, due to the relation $\alpha = \beta t$. Therefore $d(X) \le d(X') \le c(G_{E(X')}) \le \|X'\| - 3 = \|X\| - 2$ as claimed. □

4.4 Relations without the defect effect

This subsection is in a sense dual to the previous one, where we looked for conditions which would enforce an as large as possible defect effect. Here, motivated by Example 4.5, we try to construct as large as possible independent systems of equations having only a defect effect of a certain size, i.e., still a solution of certain degree d. Two extreme cases, namely those where $d = \|X\| - 1$ or $d = 2$, are of a particular interest. The former asks, what is the maximal number of independent equations forcing only the minimal defect effect, while the latter poses the question how many, if any, would allow only periodic solutions.

The first observation here is that there does not exist any infinite independent system of equations (with a finite number of unknowns). This is due to the validity of the Ehrenfeucht compactness property for free semigroups, considered in Section 7. Whether there can be unboundedly large such systems is an open problem.

Nontrivial bounds for the numbers of independent equations in our above problems are given in the following two examples from [KaPl2].

Example 4.7. Let $\Xi = \{x, y\} \cup \{u_i, v_i, w_i | i = 1, \dots, n\}$ be the set of unknowns and S the following system of equations over Ξ

$$S: \ xu_jw_kv_jy = yu_jw_kv_jx \quad \text{for} \quad j, k = 1, \dots, n.$$

Then clearly $\|S\| = n^2$ and $\|\Xi\| = 3n + 2$. We claim that

(i) S has a solution of degree $3n + 1$; and
(ii) S is independent.

The condition (i) is easy to fulfill: choose $x = y$, whence all the equations become trivial, so that a required solution can be found in a free semigroup of $3n + 1$ generators.

That the set S is independent is more difficult to see. We have to show that, for each pair (j, k), there exists a solution of the system

$$S(j, k) = S - \{xu_jw_kv_jy = yu_jw_kv_jx\},$$

which is not a solution of S. To find out such a solution is not obvious, however, here is such a solution:

$$(1) \qquad \left\{ \begin{array}{lcl} x & = & b^2ab, \\ y & = & b, \\ u_t & = & \left\{ \begin{array}{ll} ba & \text{if } t = j, \\ bab & \text{otherwise}, \end{array} \right. \\ w_{t'} & = & \left\{ \begin{array}{ll} bab^2 & \text{if } t' = k, \\ b & \text{otherwise}, \end{array} \right. \\ v_t & = & \left\{ \begin{array}{ll} ba & \text{if } t = j, \\ a & \text{otherwise}. \end{array} \right. \end{array} \right.$$

Then if $t = j$ and $t' = k$, we compute

$$xu_jw_kv_jy = b^2ab.ba\ldots \neq b.ba.bab^2\ldots = yu_jw_kv_jx$$

to note that (1) is not a solution of S. The verification that it is a solution of $S - S(j,k)$ is a matter of simple calculations:

$$\begin{array}{lcl} t \neq j \wedge t' \neq k & : & b^2ab.bab.b.a.b = b.bab.b.a.b^2ab, \\ t \neq j \wedge t' = k & : & b^2ab.bab.bab^2.a.b = b.bab.bab^2.a.b^2ab, \\ t = j \wedge t' \neq k & : & b^2ab.ba.b.ba.b = b.ba.b.ba.b^2ab. \end{array}$$

□

Example 4.8. Let $\Xi = \{x_i, y_i, u_i, w_i, v_i | i = 1,\ldots,n\}$ be a set of $5n$ unknowns, and S' the following system of n^3 equations

$$S' \;:\; x_iu_jw_kv_jy_i = y_iu_jw_kv_jx_i \quad \text{for} \quad i,j,k = 1,\ldots,n.$$

Hence S' is obtained from the system S of Example 4.7 by introducing index i for x and y, and by allowing it to range from $1,\ldots,n$. The solution (1) of S is extended by setting

$$(2) \qquad \left\{ \begin{array}{lcl} x_{t''} & = & \left\{ \begin{array}{ll} b^2ab & \text{if } t'' = i, \\ a & \text{otherwise,} \end{array} \right. \\ y_{t''} & = & \left\{ \begin{array}{ll} b & \text{if } t'' = i, \\ a & \text{otherwise.} \end{array} \right. \end{array} \right.$$

It follows directly from the computations of Example 4.7, that the solution described by (1) and (2) satisfies all the equations of S' except one, namely $x_iu_jw_kv_jy_i = y_iu_jw_kv_jx_i$.

Note that S' has a nonperiodic solution in Σ^+. □

The message of Example 4.7 is that in a free semigroup there can be $\Omega(n^2)$ independent equations in n unknowns without forcing larger than the minimal defect effect, i.e., has still a solution of degree $n-1$. Similarly, Example 4.8 shows that there can be $\Omega(n^3)$ independent equations in n unknowns having a nonperiodic solution.

Examples 4.7 and 4.8 motivate several comments and open problems. First, one may think that in Example 4.7 the requirement that Σ contains at least $3n+1$ generators makes the whole example artificial. However, if instead of the degree, i.e., the combinatorial rank, for example the prefix rank would be considered, then the example can be encoded into the binary alphabet. Indeed, encoding the alphabet of $3n+1$ letters into the binary one by a prefix encoding, we can find for S a solution over the binary alphabet having the p-rank equal to $3n+1$.

Second, if the systems of equations are solved in a free monoid, instead of a free semigroup, then the bounds of Examples 4.7 and 4.8 can be improved to $\Omega(n^3)$ and $\Omega(n^4)$, respectively, cf. [KaPl2].

Third, we state two open problems.

Problem 4.1. Improve the bounds $\Omega(n^2)$ and $\Omega(n^3)$ in Examples 4.7 and 4.8. In particular, can they be exponential?

Problem 4.2. Does there exist an independent system of three equations with three unknowns having a nonperiodic solution in Σ^+ ?

Problem 4.2 is connected to Example 4.5, as well as to Corollary 4.7. Our guess is that the answer to this problem is "no". However, the problem does not seem to be easy.

4.5 The Defect Theorem for equations

In this subsection we turn our focus explicitly from sets to equations, i.e., from solutions of equations to equations itself. The *rank* of an equation $u = v$, with the unknowns Ξ is defined as the maximal rank of its solutions $h : \Xi^+ \to \Sigma^+$ over all free semigroups Σ^+. Consequently, different notions of the rank of a finite set seem to lead to different notions of the rank of an equation. Fortunately, this is not true, at least as long as the rank of a set is defined in one of the four ways we did. To establish this is the goal of this subsection.

We start by comparing the combinatorial rank d and the prefix rank p. This is done in two lemmas, the first one being obvious from the definitions.

Lemma 4.2. *Each solution $h : \Xi^+ \to \Sigma^+$ of an equation over Ξ satisfies $d(h(\Xi)) \leq p(h(\Xi))$.*

The second lemma is less obvious, and shows that, with each solution h, we can associate so-called *principal* solution of [Len].

Lemma 4.3. *Let $u = v$ be an equation over Ξ. For each solution $h : \Xi^+ \to \Sigma^+$ of the equation $u = v$, there exists another solution, $h' : \Xi^+ \to \Sigma'^+$ such that $d(h'(\Xi)) = p(h(\Xi))$.*

Proof. Let the minimal generating set of the prefix hull of $h(\Xi)$ be $U = \{u_1, \ldots, u_d\}$. Consequently, for each $x \in \Xi$, $h(x)$ has a U-factorization, say

$$h(x) = u_{i_1} \ldots u_{i_t}. \tag{1}$$

Let $\theta : \Sigma' \leftrightarrow U$ be a one-to-one mapping, where Σ' is an new alphabet and denote by $c_i \in \Sigma'$ the element corresponding to $u_i \in U$ in this mapping. Next we define a morphism $h' : \Xi^+ \to \Sigma'^+$ by setting, for each $x \in \Xi$,

$$h'(x) = c_{i_1} \ldots c_{i_t} \Leftrightarrow h(x) = u_{i_1} \ldots u_{i_t} \quad \text{with} \quad u_{i_j} \in U.$$

By construction $\theta(h'(x)) = h(x)$ holds for all $x \in \Xi$ and since θ is injective, we have $h'(x) = h'(v)$ showing that h' is a solution and by its definition, the minimal generating set of the prefix hull of $h'(\Xi)$ is Σ'^+. Consequently, $d(h'(\Xi)) \leq d = p(h(\Xi))$.

If $d(h'(\Xi)) < d$, there would be at most $d-1$ words of Σ'^+, such that each word $h'(a)$ could be expressed as a product of these words. Therefore also words in (1) could be expressed as products of at most $d-1$ words of U^+. This, however, contradicts with the fact that each u_i must be the last factor in at least one of the factorizations (1), cf. Proposition 4.1. Hence, necessarily $d(h'(\Xi)) = p(h(\Xi))$, as required. □

Both of the Lemmas 4.2 and 4.3 can be extended to the other algebraic ranks. The detailed proofs, using Proposition 4.1 and its counterpart for the u-rank, are left to the reader.

Now we are ready to formulate our main result of this section.

Theorem 4.4. *Let $u = v$ be an equation over Ξ. The rank of the equation $u = v$, defined as the maximal rank of its solutions, is independent of which of our four ranks is used to define the rank of a solution.* □

Theorem 4.4 allows to denote the *rank* of an equation simply by $r(E)$, as well as restate the defect theorem for equations.

Theorem 4.5. *For each nontrivial equation E over the unknowns Ξ, the rank $r(E)$ of E satisfies $r(E) < \|\Xi\|$.* □

Note that, as shown by the proof of Theorem 4.4, for all algebraic ranks the rank of an equation can be defined over a fixed free semigroup Σ^+ containing at least two generators, but the combinatorial rank requires it to be defined over all free semigroups Σ^+.

We already noted that the p-rank and the f-rank of a finite set of words can be computed in a polynomial time. The same holds for the u-rank, but as we shall see in the next subsection, is known not to hold for the combinatorial rank. Computing the rank of an equation is essentially more complicated. However, as shown in the next section, this can be achieved by applying Makanin's algorithm.

4.6 Properties of the combinatorial rank

We conclude Section 4 by pointing out some further differences between the combinatorial rank and the algebraic ranks.

First, however, we emphasize the usefulness of the notion of the combinatorial rank, or of the degree. The most important cases are the both extremes, namely when a degree of a finite set $X \subseteq \Sigma^+$ equals 1 or $\|X\|$. The former corresponds to periodic sets, and the usefulness of the notion of the degree in connection with periodic sets was already seen, for instance, in Theorem 4.3. In the other extreme we call a finite $X \subseteq \Sigma^+$ *elementary*, if $d(X) = \|X\|$, and *simplifiable* otherwise. Note that this definition is consistent with that of an elementary morphism defined in Section 2.4.

A striking example of the usefulness of the above notions is an elegant proof of the D0L equivalence problem in [ER1], cf. also [RoSa1]. A crucial step in this proof was the following result.

Theorem 4.6. *An elementary morphism has a bounded delay.*

Proof. Follows directly from Corollary 4.4. Indeed, a morphism $h : \Sigma^+ \to \Delta^+$ having an unbounded delay satisfies $d(h(\Sigma)) < \|\Sigma\|$ so that it is not elementary. □

Our next example shows that the elementary sets are not closed under composition of sets in the sense of codes, cf. [BePe].

Example 4.9. Let $X = \{b, cab, cabca\}$. Then its composition with itself is

$$X \circ X = \{cab, cabcabcab, cabcabcabcabcab\} \subseteq (cab)^+.$$

Consequently, $d(X \circ X) = 1$, while $d(X) = 2$. □

As shown in [Ne2] it is not difficult to modify Example 4.9 to show that, for each $n \in \mathbb{N}$, there exists a set $X_n \subseteq \Sigma^+$ such that $d(X_n \circ X_n) - d(X_n) \geq n$. In [Ne2] it is also considered how the degree of a set behaves with respect to certain operations, in particular with respect to rational operations.

Finally, we deal with the problem of computing the degree of a given set. This seems to be computationally very difficult, as a contrast to Corollary 4.3 (or its variants to the other algebraic ranks), which shows that the algebraic ranks are computable in polynomial time. This also explains why we didn't give any procedure to compute a set F in the definition of the degree: no fast method for that is known, or even likely to be discovered, as we now demonstrate.

The complexity results for the degree, due to [Ne1], are as follows:

Theorem 4.7. *(i) The problem of deciding, for a given finite set $X \subseteq \Sigma^+$ and for a given number k, whether $d(X) \leq k$ is NP-complete.*

(ii) The problem of deciding whether a given finite set is simplifiable is NP-complete.

Actually, the problem of (i) remains NP-complete even if k is fixed to be any number larger than 2. The choice $k = 2$ makes the problem computationally easy: as shown in [Ne3] it can be solved in time $\mathcal{O}(n \log^2 m)$, where $n = s(X)$ and $m = \max\{|x| \mid x \in X\}$. Note also that (ii) is equivalent to saying that the elementariness problem is in the class of co-NP-complete problems, cf. [GJ]. In particular, it is not likely that a polynomial time algorithm will be found for it.

We do not present the proof of Theorem 4.7 here, but in order to give an intuition why the result holds, we show, in the next example, that a related problem is NP-complete. Actually, the NP-completeness of this is the first step in the proof of Theorem 4.7 in [Ne1].

Example 4.10. (Strong Factorizability Problem.) The problem asks to decide, for a finite set $X \subseteq \Sigma^+$ and for a number k, whether there exists a set $Y \subseteq \Sigma^+$ such that

$$X \subseteq Y^+, \ \|Y\| \leq k \text{ and } X \cap Y = \emptyset.$$

If such a Y exists, we say that X is *strongly k-factorizable*, and we refer this problem to as the *SF-problem*. Note that if we drop from the SF-problem the requirement $X \cap Y = \emptyset$, we obtain the problem (i) of Theorem 4.7.

Obviously the SF-problem is in NP. So to prove its NP-completeness we have to reduce it to some known NP-complete problem, which will be the following variant of the *vertex cover problem*, referred to as the *special* vertex cover problem, or the SVC-problem for short. For the NP-completeness of this, which is a straightforward modification of the NP-completeness of the ordinary vertex cover problem, we refer to [Ne2] or [GJ].

The *SVC-problem* asks to decide for a given graph $G = (V, E)$, with $\|V\| = \|E\|$ and having no isolated points, and for a given natural number k, whether there exists a subset V' of V, with $\|V'\| \leq k$, such that the set of edges connected to V' equals that of all edges of G. In other words, the SVC-problem asks whether a graph of the required type has a *vertex cover* of size at most k. Now let

$$((V, E), k) \ \text{ with } \ \|V\| = \|E\| = n \ \text{ and } \ 1 \leq k \leq n - 1$$

be an instance of the SVC-problem. We associate it with an instance

$$(X, k + n)$$

of the SF-problem by defining a subset $X \subseteq VTV$, where T is a renaming of E under the mapping $c : E \rightarrow T$, as follows

$$(1) \qquad \alpha a \beta \in X \Leftrightarrow \alpha, \beta \in E \ \text{ and } \ a = c(\alpha, \beta).$$

We have to show that

$G = (V, E)$ has a vertex cover V' with $\|V'\| \leq k$ if, and only if,
X is $(k + n)$-strongly factorizable.

First, assume that G has a vertex cover of size at most k. Let $\alpha a \beta$ be a word in X. It is factorized as $\alpha.a\beta$, if $\alpha \in V'$, and $\alpha a.\beta$, if $\alpha \in V'$. Now let B be the set of all words of length 2 in these factorizations. Then, by (1), $\|B\| = n$, so that $\|V' \cup B\| = \|V'\| + n \leq k + n$. Therefore, X is $(k+n)$-strongly factorizable.

Second, assume that X is $(k + n)$-strongly factorizable via Y. We define a partition of X

$$X = X_1 \cup X_2 \ \text{ with } \ X_1 \cap X_2 = \emptyset$$

as follows. The word $\alpha a \beta \in X_1$ if, and only if, it is factorized in Y as $\alpha a.\beta$ or $\alpha.a\beta$, and therefore $\alpha a \beta \in X_2$ if, and only if, it is factorized in Y as $\alpha.a.\beta$.

Let V_i, for $i = 1, 2$, consists of those letters of V which occur in the above factorizations of words of X_i. Similarly, let $T_i \subseteq T \cup TV \cup VT$, for $i = 1, 2$, consists of those words in $Y - V$ which occur in these factorizations of words of X_i. Finally, for each $w \in X_2$, i.e., w being factorized as $\alpha.a.\beta$, we pick up either α or β from V_2, and denote by V_2' the set of all letters picked up when w ranges over X_2. Now, we set

$$K = V_1 \cup V_2'.$$

Then, by the construction, K is a vertex cover. It also follows that the sets T_1, T_2 and $V_1 \cup V_2$ are pairwise disjoint, and moreover, by (1), we have $\|T_i\| = \|X_i\|$, for $i = 1, 2$. Consequently, we obtain the following relation

$$\begin{aligned} \|Y\| &= \|V_1 \cup V_2 \cup T_1 \cup T_2\| = \|V_1 \cup V_2\| + \|T_1\| + \|T_2\| \\ &= \|V_1 \cup V_2\| + \|X_1\| + \|X_2\| = \|V_1 \cup V_2\| + \|X\| \end{aligned}$$

implying, since $\|Y\| \leq k + n = k + \|X\|$, that

$$\|K\| = \|V_1 \cup V_1'\| \leq \|V_1 \cup V_2\| \leq k.$$

Therefore, the graph (V, E) has a vertex cover of size at most k, completing our proof. □

5. Equations as properties of words

Two elements x and y of a group are said conjugate if there exists an element z such that equation $x = zyz^{-1}$ holds. In order to extend this definition to monoids, one has to eliminate the inverses which can be easily achieved by multiplying two handsides by the element z to the right yielding equation

$$xz = zy$$

The purpose of this section is to discuss the connection between equations in words and some properties of words. We think that little is known so far and that much remains to be done.

5.1 Makanin's result

We already noted that the p-rank and the f-rank of a finite set of words can be computed in a polynomial time. The same holds for the u-rank, but as we have seen in Section 4.6, it does not hold for the combinatorial rank. Computing the rank of an equation is essentially more complicated since we aim at computing the maximal rank over a (usually) infinite set of solutions. However, this can be achieved by applying Makanin's algorithm which is one of the major advances in combinatorial free monoid theory.

We recall that given an alphabet Ξ of unknowns and an alphabet Σ of constants, Ξ and Σ being disjoint, an *equation with constants* is a pair $(u, v) \in (\Xi \cup \Sigma)^* \times (\Xi \cup \Sigma)^*$, also written $u = v$. A *solution* is a morphism

$h : (\Xi \cup \Sigma)^* \rightarrow \Sigma^*$ leaving Σ invariant, i.e., satisfying $h(a) = a$ for all $a \in \Sigma$, for which the following holds

$$h(u) = h(v).$$

For example, the equation $ax = xb$ with $a \neq b \in \Sigma$ and $x \in \Xi$ has no solution since the left handside has one more occurrence of a than the right handside, and the equation $ax = xa$ has the solution $x = a$.

We have the famous result of Makanin, cf. [Mak].

Proposition 5.1. *There exists an algorithm for solving an equation with constants.*

The exact complexity of the problem is unknown but several authors have contributed to lower the complexity of the original algorithm which was an exponential function of height 5. Actually, this complexity depends on the complexity of computing the minimal solutions of diophantine equations. We refer the interested reader to [Ab1], [Do] and [KoPa] for the latest results on this topic. Several sofware packages have been produced which work relatively well up to length, see e.g., [Ab2].

5.2 The rank of an equation

One of the most direct consequences of Makanin's result is the fact that the rank of an equation may be effectively computed, cf. [Pec].

Theorem 5.1. *Given an equation without constants $u = v$, its rank can be effectively computed.*

Proof. The idea of the proof is as follows. Let Ξ be the set of unknowns and denote by ι some mapping of Ξ onto some disjoint subset Σ with $||\Sigma|| < ||\Xi||$. Consider the morphism $\theta : \Xi^* \rightarrow (\Xi \cup \Sigma)^*$ defined for all $x \in \Xi$ by $\theta(x) = \iota(x)x$. Then the rank of $u = v$ is the maximum cardinality of $||\iota(\Xi)||$ for which the equation with unknowns $\theta(u) = \theta(v)$ has a solution. For example, starting with the equation $xyz = zyx$ we would be led to define the 4 equations $axayaz = azayax$, $axaybz = bzayax$, $axbyaz = azbyax$, $axbybz = bzbyax$ and to apply Makanin's result to each of these equation.

More precisely, assume the rank of $u = v$ is r, i.e., there exists a morphism $h : \Xi^* \rightarrow \Sigma^*$ such that $h(u) = h(v)$ and $r(X) = r$ where $X = h(\Xi)$. Deleting, if necessary, some unknowns it is always possible to assume that the morphism is nonerasing. Furthermore, without loss of generality, we may assume that the free hull $X(f)^* = \Sigma^*$. Indeed, let $\alpha : \Sigma' \leftrightarrow h(\Xi)$ be an one-to-one mapping, where Σ' is a new alphabet. Then there exists an unique solution $h' : \Xi^+ \rightarrow \Sigma'^+$ such that $\alpha(h'(x)) = h(x)$ holds for all $x \in \Xi$. We have $X(f) = \Sigma'$ and $r(h'(\Xi) = r(h(\Xi)$. Let ι be the mapping that associates the initial letter of $h(x)$ to each x. By Proposition 4.1, we know that $\Sigma = \{\iota(x) | x \in \Xi\}$. Consider the morphism $\theta : \Xi^* \rightarrow (\Xi \cup \Sigma)^*$ satisfying

$\theta(x) = \iota(x)x$. Then the morphism $g(x) = (\iota(x))^{-1}h(x)$ satisfies the equation with constants $\theta(u) = \theta(v)$.

Example 5.1. With $\Xi = \{x, y, z\}$ and $xyz = zyx$, we have the solution $x = a, y = bab, z = aba$. Then by θ we obtain an equation with unknowns $axbyaz = azbyax$ for which $g(x) = 1, g(y) = ab, g(z) = ba$ is a solution. □

Proof of Theorem (continued). Conversely, let ι be a mapping of Ξ onto some Σ with $\Xi \cap \Sigma = \emptyset$ and $||\Sigma|| < ||\Xi||$. Consider the morphism $\theta : \Xi^* \to (\Xi \cup \Sigma)^*$ defined for all $x \in \Xi$ by $\theta(x) = \iota(x)x$, and assume that the equation with unknowns $\theta(u) = \theta(v)$ has a solution g. The morphism $h(x) = \iota(x)g(x)$ is clearly a solution of $u = v$. Now we claim that its rank is greater than or equal to $||\Sigma||$. Indeed, let $X \subseteq \Sigma^*$ be the minimal generating set of the free hull of $h(\Xi^*)$: $h(x) \in X^*$ for all $x \in \Xi$. Every element in X appears as the leftmost factor in the decomposition of some $h(x)$. If $||X|| < ||\Sigma||$, then some letter of Σ does not appear in the leftmost position contradicting the definition of h □

Actually, this result carries over to the rank of equations with constants, after a suitable extension of the notion of rank.

5.3 The existential theory of concatenation

Makanin's result can be interpreted either as a statement on systems of equations and inequations, or equivalently as a statement of formulae of the existential theory of concatenation. More precisely, it has been observed that at the cost of introducing new unknowns, negations and disjunctions can be expressed as conjunctions of equations and further that all conjunctions are equivalent to a single equation. In other words, starting from a Boolean combination of equations on the unknowns Ξ, it is possible to define a single equation on the unknowns $\Xi \cup \Xi'$ for some Ξ', whose set of solutions restricted to the unknowns Ξ equals the set of solutions of the Boolean combination.

It is worthwhile considering the power of equations in expressing properties or n-ary relations on words, for some integer n. Following the tradition, we call *diophantine* a relation on words $R(x_1, \ldots, x_n)$ that is equivalent to a formula of the form

$$(1) \quad \exists y_1, \ldots, \exists y_m \lambda(x_1, \ldots, x_n, y_1, \ldots, y_m) = \rho(x_1, \ldots, x_n, y_1, \ldots, y_m)$$

with $\lambda = \rho$ an equation. For example, "x is imprimitive" can be expressed as

$$\exists y, z : x = 1 \lor (x = yz \land yz = zy \land y \neq 1 \land z \neq 1)$$

and "x and y are conjugate" can be expressed with two extra unknowns as

$$\exists u, v : x = uv \land y = vu,$$

or with one extra unknown only as

$$(2) \qquad \exists z : xz = zy.$$

These formulae are diophantine. No characterization of diophantine relations seems to exist in the literature. There is no available tool either for showing that a given property is not diophantine, a natural candidate would be, e.g., primitivity. Neither do we know which are the properties that are diophantine and whose negation also is diophantine. Intuitively, this imposes very strong restrictions on the property, one such example being "x is a prefix of y". Yet another area of research is to study the hierarchy of diophantine formulae where the number of existential quantifiers is taken into account, i.e., the integer m of equation (1). In this vein, it was shown in [Sei] that the relation "x is a prefix of y" can not be expressed without an extra variable.

Let us now briefly show how to reduce a Boolean combination of equations to a single equation. Assuming that Σ contains two different constants a and b, the system consisting of the two equations $x = y$ and $u = v$ is equivalent to the single equation $xauxbu = yavybv$ as noticed in [Hm]. To check this, identify the unknowns with their images under the solution h and observe that xau, xbu, yav and ybv have all the same length, to wit half the common length of the left- and right-handsides. Thus $xau = yav$ and $xbu = ybv$ holds. If $x \neq y$, say $|x| < |y|$ without loss of generality, then the first equation says that there is an occurrence of a in position $|x|$ in y, while the second says that this occurrence is equal to b.

Similarly, as noted, e.g., in [CuKa2], introducing new unknowns, the inequation $x \neq y$ is equivalent to a disjunction of equations saying that x and y are prefixes of each other or that their maximum common prefix is a proper prefix of both. Hence three new unknowns are needed here in this reduction. Finally, with the help of more unknowns a disjunction of equations can be expressed as a conjunction of equations as we show in a moment. So, in terms of logics, Makanin's result implies that the existential fragment of the theory of concatenation is decidable. We formulate the above as.

Theorem 5.2. *For any Boolean combination B of equations with Ξ as the set of unknowns we can construct a single equation E with $\Xi \cup \Xi'$ as the set of unknowns such that solutions of B and those of E restricted to Ξ are exactly the same.* □

As we said, in the process of reducing a Boolean combination to a single equation new unknowns are introduced. A more precise computation of how many are needed has been studied though the issue of the exact number is not yet settled. In particular reducing a disjunction to conjunctions has received various solutions. Büchi and Senger used 4 new unknowns in [BS], Senger in his thesis needs 3, while Serge Grigorieff achieves the same result with 2. It is an open question whether or not one unknown suffices though it is suspected it does not.

We reproduce here the unpublished proof of S. Grigorieff.

Theorem 5.3. *The disjunction $x = u \vee y = v$ is equivalent to a formula of the form*

$$\exists z \exists t \lambda(x, y, u, v) = z\rho(x, y, u, v)t$$

where $\lambda(x, y, u, v)$ and $\rho(x, y, u, v)$ are words over the alphabet $\{x, y, u, v, a, b\}$ and z, t are new variables.

Proof. First by observing that $x = u \vee y = v$ is equivalent to $xv = uv \vee uy = uv$, without loss of generality we may start with a disjunction of the form $x = u \vee x = v$. By making the further observation that $x = u \vee x = v$ is equivalent to $xa = ua \vee xa = va$ we may assume that x, u, v are nonempty words.

Now we use the pairing function $< x, y >= xayxby$. We set

$$\begin{aligned} \rho(x, u, v) &=< uuu, vvv >^2 x < uuu, vvv >^3 x < uuu, vvv >^2 \\ \lambda(x, u, v) &=< uuu, vvv >^3 u < uuu, vvv >^3 u < uuu, vvv >^2 v \\ &\qquad < uuu, vvv >^3 v < uuu, vvv >^3 \end{aligned}$$

Making use of the primitivity of $< uuu, vvv >$, a case study shows that the factor ρ fits in λ in only two places, either

$$\lambda =< uuu, vvv > \rho v < uuu, vvv >^3 v < uuu, vvv >^3$$

implying that $x = u$, or

$$\lambda =< uuu, vvv >^3 u < uuu, vvv >^3 u\rho < uuu, vvv >$$

implying that $x = v$. □

Finally we note that Makanin's result is on the borderline between the decidability and the undecidability. Indeed, [Marc] established the undecidability of the fragment $\forall\exists^4$-positive of the concatenation theory, further improved to $\forall\exists^3$-positive. The previous reduction of disjunctions yields the undecidability of the theory consisting of formulae of the form

$$\forall\exists^5 \lambda(x_1, \ldots, x_6) = \rho(x_1, \ldots, x_6),$$

where $\lambda = \rho$ is an equation.

5.4 Some rules of thumb for solving equations by "hand"

There is unfortunately no method, in the practical sense of the word, for solving equations. We list here just a few simple-minded tricks that are widely used when dealing with real equations. Most of them lead to proving that the equation has only cyclic solutions by reducing the initial equation to the equations that are well-known, such as Levi's Lemma, cf. (1) in Section 2.1, the conjugacy, cf. e.g. (2), or the commutativity, cf. Corollary 4.1.

First of all, conditions on the lengths of the unknowns are expressed as linear equations over the positive integers. When some of these unknowns have length 0 then the number of unknowns reduces. An elaboration of this idea is exemplified by the following well-known fact that appears when solving the general equation $x^n y^m = z^p$ for $n, m, p \geq 2$. Let us verify that $x^2y^2 = z^2$ implies that x, y and z are powers of the same elements. Indeed, observing that xy^2x is a conjugate of z^2, there exists a conjugate z' of z such that $xy^2x = z'^2$. Since xy and yx have same length, they are both equal to z' implying that $x, y \in t^*$ for some word t and thus that $z \in t^*$ also.

Splitting represents another approach. In the easy cases, there is a prefix of the left- and right-handsides that have the same length, i.e., $zxyxzy = yxxzyz$ splits into $zxyx = yxxz$ and $zy = yz$. This ideal situation is rare, however a variant of it is not so seldom. Assume a primitive word x has an occurrence in both handsides of the equation, say $uxv = u'xxv'$ where $u, u', v, v' \in \Xi^*$ are products of unknowns. Assume further $|u'| \leq |u| \leq |u'x|$. Then the equation splits into $u = u'x$ and $v = v'$ or into $u = u'$ and $v = xv'$. Combinatorial problems on words in the theory of finite automata, rational relations,varieties etc..., usually come up as families of equations involving a parameter, e.g., $xy^n z = zy^n t$ with $x, y, z, t \in \Xi$ and $n \in \mathbb{N}$. Then the above condition on the lengths can be enforced by choosing an appropriate value of n.

Another technique proves useful in some very special cases. It was the starting point of the theory developped in [Len] and it consists, for fixed lengths of a solution, to compute the "freest" solution with these lengths. As an illustration let us consider the equation

$$xyz = zyx \tag{3}$$

and assume $|x| = 3$, $|y| = 5$, $|z| = 1$, with a total length of 9. Write

$$x = x_1x_2x_3, y = y_1y_2y_3y_4y_5, z = z_1.$$

The idea is to identify the positions which bear the same letter in both handsides, such as 3 and 9 (carrying x_3) and 5 and 3 (carrying y_2).

x_1	x_2	x_3	y_1	y_2	y_3	y_4	y_5	z_1
z_1	y_1	y_2	y_3	y_4	y_5	x_1	x_2	x_3

More precisely, define a graph whose 9 vertices are in one-one correspondence with the 9 occurrences of letters in the solution, and whose non-oriented edges are the pairs $(i, j), 0 \leq i, j \leq 9$, where the letter in position i in the left handside is equal to the letter in position j in the right handside of (3) or vice versa. Then each connected component of the graph is associated with a distinct letter in the target alphabet. In other words, the "richest" alphabet for a solution of (3) has cardinality equal to the number of connected components of the graph.

If we had chosen $|x| = 2$, $|y| = 4$ and $|z| = 1$ for a total length of 7, then we would have found one connected component, actually one Hamiltonian path $1, 5, 4, 3, 2, 6$, i.e., the richest solution would be cyclic.

Fixing the lengths may look like too strong a requirement, however, this very technique allows us to prove in the next section that the Theorem of Fine and Wilf is sharp, i.e., that on a binary alphabet there exist only two words of length p and q, p and q coprimes, whose powers have a common prefix of length exactly equal to $p+q-2$.

6. Periodicity

Periodicity is one of the fundamental properties of words. Depending on the context, and traditions, the term has had several slightly different meanings. What we mean by it in different contexts is recalled here, cf. also Section 2.2. The other goals of this section is to present three fundamental results on periodicity of words, namely the Periodicity Theorem of Fine and Wilf, the Critical Factorization Theorem, and recent characterizations of ultimately periodic 1-way infinite words and periodic 2-way infinite words.

6.1 Definitions and basic observations

We noted in Section 2.2 that each word $w \in \Sigma^+$ has the unique period $p(w)$ as the length of the minimal u such that

$$w \in F(u^\omega). \tag{1}$$

Such a $p(w)$ is called *the period* of w as a distinction of *a period* of w which is the length of any u satisfying (1). When the period refers to a word, and not to the length, then *the periods* of w are all the conjugates of the minimal u in (1), often called *cyclic roots* of w. Similarly *periods* of w are all conjugates of words u satisfying (1). Finally, we call w *periodic*, if $|w| \geq 2p(w)$, i.e., w contains at least two consecutive factors of its same cyclic root. *Local* variants of these notions are defined in Section 6.3.

In connection with infinite words *periodic* 1-way and 2-way infinite words are defined as words of the forms u^ω and ${}^\omega u^\omega$, with $u \in \Sigma^+$, respectively. By an *ultimately periodic* 1-way infinite word we mean a word of the form uv^ω, with $u \in \Sigma^*$ and $v \in \Sigma^+$. Formally, the word ${}^\omega u^\omega$, for instance, is defined by the condition

$${}^\omega u^\omega(i) = u(i \bmod |u|), \text{ for all } i \in \mathbb{Z}.$$

Finally, a language $L \subseteq \Sigma^*$ is *periodic*, if there exists a $z \in \Sigma^*$ such that $L \subseteq z^*$.

There should be no need to emphasize the importance of periodicity either in combinatorics of words or in formal language theory. Especially in the latter theory periodic objects are drastically simpler than the general ones: the fundamental difficulty of the noncommutativity is thus avoided. Therefore one tries to solve many problems of languages by reducing them to periodic languages, or at least to cases where a “part” of the language is periodic.

Based on the above it is important to search for the *periodicity forcing conditions*, i.e., conditions which forces that the words involved form a periodic language. We have already seen several such conditions, cf. Section 4:

any nontrivial relation on $\{x,y\} \subseteq \Sigma^*$;
any pair of nontrivial identities on $X = \{x,y,z\} \subseteq \Sigma^+$ of the form $x\alpha = y\beta$, $y\gamma = z\delta$ with $\alpha,\beta,\gamma,\delta \in X^*$;
any condition on $X = \{x_1,\ldots,x_n\} \subseteq \Sigma^+$ satisfying: the transitive closure of the relation ρ defined as

$$x\rho y \Leftrightarrow xX^\omega \cap yX^\omega \neq \emptyset$$

equals $X \times X$.

Another classical example of a periodicity forcing condition is the equation, cf. [LySc] or [Lo],

$$x^n y^n = z^k \quad \text{with} \quad n,m,k \geq 2.$$

As we observed in Section 5 many properties of words are expressable in terms of solutions of equations. Thus it is often of interest to know whether such languages, or more generally parts of such languages, either are always periodic or can be periodic. By considerations of Section 5, Makanin's algorithm can be used to test this. Indeed, we only have to add to the system S defining the property suitable predicates of the forms

$$xy = yx \quad \text{or} \quad xy \neq yx,$$

and transform the whole predicate into one equation.

For example, if we want to know, whether there exist words x, y, z, u and v satisfying the equation $\alpha = \beta$ in these unknowns such that x, y and z are powers of a same word, and u and v are not powers of a same word, we consider the system

$$\begin{cases} \alpha &= \beta \\ xy &= yx \\ xz &= zx \\ uv &\neq vu, \end{cases}$$

and test whether it has a solution.

6.2 The Periodicity Theorem of Fine and Wilf

Our first result of this section is the classical *periodicity theorem* of Fine and Wilf, cf. [FW]. Intuitively it determines how far two periodic events have to match in order to guarantee a common period. Interestingly, although the result is clearly a result on sequences of symbols, i.e., on words, it was first presented in connection with real functions!

Theorem 6.1. *(Periodicity Theorem). Let $u, v \in \Sigma^+$. Then the words u and v are powers of a same word if, and only if, the words u^ω and v^ω have a common prefix of length $|u| + |v| - \gcd(|u|, |v|)$.*

Proof. We first note that we can restrict to the basic case, where $\gcd(|u|, |v|) = 1$. Indeed, if this is not the case, say $|u| = dp$ and $|v| = dq$, with $\gcd(p, q) = 1$, then considering u and v as elements of $(\Sigma^d)^+$ the problem is reduced to the basic case with only a larger alphabet.

So assume that $|u| = p$, $|v| = q$ and $\gcd(p, q) = 1$. The implication in one direction is trivial. Therefore, we assume that u^ω and v^ω have a common prefix of length $p+q-1$. Assuming further, by symmetry, that $p > q$ we have the situation depicted in Figure 6.1. Here the vertical dashline denotes how far the words can be compared, the numbers tell the lengths of the words u and v, and the arrows the procedure defined below.

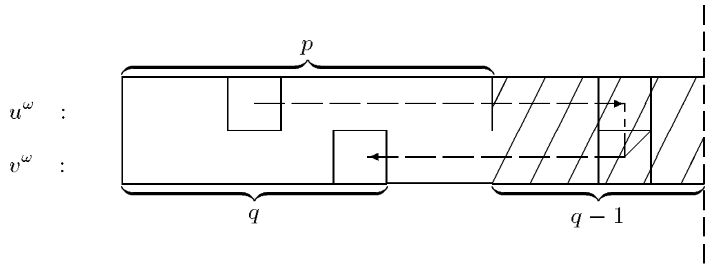


Figure 6.1. An illustration of the procedure

We denote by i, for $i = 1, \ldots, p + q - 1$, the corresponding position in the common prefix of u^ω and v^ω. Next we describe a *procedure* to fix new positions with the same value as a given initial one i_0. Let $i_0 \in [1, q-1]$. Then, by the assumption, the position obtained as follows, cf. arrows in Figure 6.1, gets the same value as i_0

$$(1) \qquad i_0 \xrightarrow{+p} i_0 + p \xrightarrow{\text{mod } q} i_1 = i_0 + p \pmod{q},$$

where i_1 is reduced to the interval $[1, q]$. Moreover, since $\gcd(p, q) = 1$, i_1 is different from i_0. If i_1 is also different from q we can repeat the procedure, and the new position obtained is different from the previous ones. If the procedure can be continued $q - 1$ steps, then all the positions in the shadowed area will be fixed, so that these together with i_0 make v unary. Hence, so is u, and we are done.

The procedure (1) can indeed be continued $q - 1$ steps if i_0 is chosen as

$$i_0 + (q - 1)p \equiv q \pmod{q}.$$

This is possible since $\gcd(p,q) = 1$. After this choice all the values $i_0 + jp \pmod q$, for $j = 0, \ldots, q-2$, are different from q, which was the assumption of the procedure (1). □

In terms of periods of a word and the distance of words, cf. Section 2.1, Theorem 6.1 can be restated in the following forms, the latter of which does not require that the comparison of words has to be started from either ends.

Corollary 6.1. *If a word $w \in \Sigma^+$ has periods p and q, and it is of the length at least $p + q - \gcd(p,q)$, then it also has a period $\gcd(p,q)$.* □

Corollary 6.2. *For any two words $u, v \in \Sigma^+$, we have*

$$l(u^\omega, v^\omega) \geq |u| + |v| - \gcd(|u|,|v|) \Rightarrow \rho(u) \text{ and } \rho(v) \text{ are conjugates.}$$ □

We tried to make the proof of Theorem 6.1 as illustrative as possible. At the same time it shows clearly, why the bound given is optimal, and even more, as we shall see in Example 6.1.

Theorem 6.1 allows, for each pair (p,q) of coprimes, the existence of a word w of length $p+q-2$ having the periods p and q. Let $W_{p,q}$ be the set of all such words, and define

$$PER = \bigcup_{\gcd(p,q)=1} W_{p,q}.$$

So, we excluded unary words from PER.

Example 6.1. We claim that, for each pair (p,q) of coprimes, $W_{p,q}$ contains exactly one word (up to a renaming), which moreover is binary. These observations follow directly from our proof of Theorem 6.1. The idea of that proof, namely filling positions in the shorter word v, can be illustrated in Figure 6.2. The nodes of this cycle correspond the positions of v, two labelled by ? are those which are missing from the shadowed area of Figure 6.1, and each arrow corresponds one application of the procedure (1). By the construction, starting from any position, and applying (1) the letter in the new position may differ from the previous one, only when to a position labelled by ? is entered. Consequently, during the cycle it may change at most twice, but, in fact, the latter change is back to the value of i_0. The fact that all positions are visited is due to the condition $\gcd(p,q) = 1$. Hence, we have proved our claim.

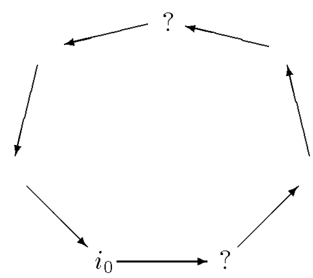


Fig. 6.2. The case PER

As a concrete example, the word of length 12 in PER starting with a and having the periods 5 and 9 is as depicted below:

a	a	a	b	a	a	a	a	b	a	a	a
2	1	5	4	3	2	1	5	4	3	2	1

Here the word, that is $(aaaba)^2aa = (aaabaaaab)^1aaa$, is described on the upper line, and the order of filling the positions, starting from the second one, on the lower line. Note that the change can take place in steps number 4 and 5, but the latter must assume the same value as the next one encountered in the procedure, which is the value of the second position. □

Example 6.2. Consider a word w of length $dp+dq-d-1$ with $\gcd(p,q)=1$, having the periods dp and dq, but not d, for some d. The argumentation of Example 6.1 shows that such a word exists, proving that in all cases the bound given in Theorem 6.1 is optimal. Moreover, for each $i=1,\ldots,d-1$, the positions $i+jd$ are filled by the same letter a_i, while in position $d+jd$ the situation is as in Example 6.1: they are uniquely filled by a word in $W_{p,q}$. □

Example 6.1 can be generalized also as follows.

Example 6.3. Let $p,q,k \in \mathbb{N}$, with $p>q$, $\gcd(p,q)=1$ and $2 \leq k \leq q$. Then there exists a unique word w_k up to a renaming such that

$$|w_k| = p+q-k, \quad \|\mathrm{alph}(w_k)\| = k \quad \text{and} \quad w_k \quad \text{has periods} \quad p \quad \text{and} \quad q.$$

Indeed, the considerations of Example 6.1 extend immediately to this case, when the number of ?'s in Figure 6.2 is k. It follows that all words of length $p+q-k$, with $\gcd(p,q)=1$, having periods p and q are morphic images of w_k under a length preserving morphism. □

We conclude this section by reminding that the set PER has remarkable combinatorial properties, cf. e.g. [dLM], [dL] and [BdL]. For example, all finite Sturmian words are characterized as factors or words in PER.

Finally we recall a result of [GO], which characterizes the set of all periods of an arbitrary word w.

6.3 The Critical Factorization Theorem

Our second fundamental periodicity result is the *Critical Factorization Theorem* discovered in [CV], and developped into its current form in [Du1], cf. also [Lo]. Our proof is from [CP]. The difference between [CV] and [Du1] was essentially, in terms of Figure 6.3 below, that [CV] considered only the case (i).

Intuitively the theorem says that the period $p(w)$ of a word $w \in \Sigma^+$ is always locally detectable in at least one position of the word. To make this precise we have to define what we mean by a local period of w at some position. We say that p is *a local period of w at the position $|u|$*, if $w = uv$, with $u, v \neq 1$, and there exists a word z, with $|z| = p$, such that one of the following conditions holds for some words u' and v':

$$(1) \qquad \begin{cases} (i) & u = u'z \text{ and } v = zv' \ ; \\ (ii) & z = u'u \text{ and } v = zv' \ ; \\ (iii) & u = u'z \text{ and } z = vv' \ ; \\ (iv) & z = u'u = vv' \ . \end{cases}$$

Further *the local period of w at the position $|u|$*, in symbols $p(w, u)$, is defined as the smallest local period of w at the position u. It follows directly from (1), cf. also Figure 6.3, that $p(w, u) \leq p(w)$.

The intuitive meaning of the local period is clear: around that position there exists a factor of w having as its minimal period this local period. The situations of (i), (ii) and (iv) in (1) can be depicted as in Figure 6.3.

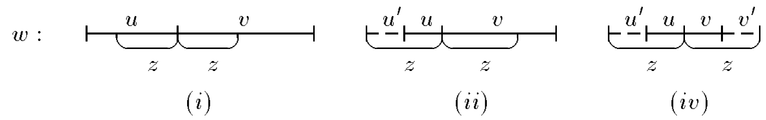


Figure 6.3. The illustration of a local period

Now, we can say precisely what the above local detectability means. It means that there exists a factorization $w = uv$, with $u, v \neq 1$, such that

$$p(w, u) = p(w),$$

i.e., the local period at position $|u|$ is that of the (global) period of w. The corresponding factorization $w = uv$ is called *critical*. The theorem claims that each word possesses at least one critical factorization (if it possesses any nontrivial factorization at all).

Example 6.4. Consider the words $w_1 = aababaaa$ and $w_2 = a^n ba^n$. The periods of these words are 6 and $n + 1$. The local periods of w_1 in positions $1, 2, \ldots, 7$ are $1, 5, 2, 6, 6, 1, 1$, respectively. For example, at position 4 we have $w = aaba.baaa$ so that $z = baaaba$ contains $baaa$ as a prefix and $aaba$ as a suffix, but no shorter z can be found to satisfy (1). The word w_1 has two critical factorizations. The critical factorizations of w_2 are $a^n b.a^n$ and $a^n.ba^n$, showing that there are none among the first $n - 1$ factorizations. □

Example 6.5. As an application of Lyndon words we show that, any word $w \in \Sigma^+$ satisfying $|w| \geq 3p(w)$, has a critical factorization. Indeed, we can write

$$w = ullv,$$

where $u, v \in \Sigma^*$ and l is the Lyndon word in the class $[\mathrm{pref}_{p(w)}(w)]$. As we noted in Section 2.2 Lyndon words are unbordered. Consequently, the factorization $w = u.llv$ is critical. Hence, in a critical factorization we can even choose $1 \leq |u| \leq p(w)$. □

To extend Example 6.5 for all words is much more difficult.

Theorem 6.2 (Critical Factorization Theorem). *Each word $w \in \Sigma^+$, with $|w| \geq 2$, possesses at least one factorization $w = uv$, with $u, v \neq 1$, which is critical, i.e., $p(w) = p(w, u)$. Moreover, u can be chosen such that $|u| < p(w)$.*

Proof. Our proof from [CP] not only shows the existence of a critical factorization, but also gives a method to define such a factorization explicitly. We may assume that w is not unary, i.e., $p(w) > 1$.

Let $\preceq_l$ be a lexicographic ordering of Σ^+, and $\preceq_r$ another lexicographic ordering obtained from $\preceq_l$ by reversing the order of letters, i.e., for $a, b \in \Sigma$, $a \preceq_l b$ if, and only if, $b \preceq_r a$. Let v and v' be the maximal suffixes of w with respect to the orderings $\preceq_l$ and $\preceq_r$, respectively. We shall show that one of the factorizations

$$w = uv \quad \text{or} \quad w = u'v'$$

is critical. More precisely, it is the factorization $w = uv$, if $|v| \leq |v'|$, and $w = u'v'$ otherwise. In addition, in both the cases

$$|u|,\ |u'| < p(w). \tag{2}$$

We need two auxiliary results. The first one holds for any lexicographic ordering $\preceq$.

Claim 1. If v is the lexicographically maximal suffix of w, then no nonempty word t is both a prefix of v and a suffix of $u = wv^{-1}$.

Proof of Claim 1. Assume that $u = xt$ and $v = ty$. Then, by the maximality of v, we have $tv \preceq v$ and $y \preceq v$. Since $v = ty$ these can be rewritten as $tty \preceq ty$ and $y \preceq ty$. Now, from the former inequality we obtain that $ty \preceq y$, which together with the latter one means that $y = ty$. Therefore, t is empty as claimed. □

The second one, which is obvious from the definitions, claims that the orderings $\preceq_l$ and $\preceq_r$ together define the prefix ordering $\leq$.

Claim 2. For any two words $x, y \in \Sigma^+$, we have

$$x \preceq_l y \ \text{ and } \ x \preceq_r y \Leftrightarrow x \leq y, \ \text{ i.e., } \ x \ \text{ is a prefix of } \ y. \qquad □$$

Proof of Theorem (continued). Assume first that $|v| \leq |v'|$. We intend to show that the factorization $w = uv$ is critical. First we show that $u \neq 1$. If this is not the case, and $w = at$, with $a \in \Sigma$, then $w = v = v'$. Therefore, by the definitions of v and v', we have both $t \preceq_l w$ and $t \preceq_r w$. So, by Claim 2, t is a prefix of $w = at$, and hence $t \in a^+$, i.e., $p(w) = 1$. This, however, was ruled out at the beginning. Hence, the word u, indeed, is nonempty.

From now on let us denote $p(w, u) = p$. By Claim 1, we cannot have $p \leq |u|$ and $p \leq |v|$ simultaneously. Hence, if $p \leq |u|$, then necessarily $p > |v|$, implying that v is a suffix of u. This, however, would contradict with the maximality of v, since $v \prec_l vv$. So we have concluded that $p > |u|$. Since p is a local period at the position $|u|$, there exists a word z such that $p = |zu|$, and the words zu and v are comparable in the prefix ordering, i.e., one of the words v and zu is a prefix of another. We consider these cases separately.

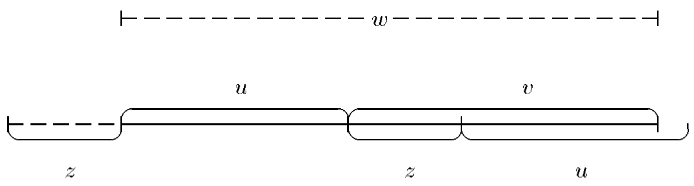


Figure 6.4. The case $p = |zu| > |v|$

Case I: $p > |v|$. Now, the situation can be depicted as in Figure 6.4. It follows that $|uz|$ is a period of $uv = w$, i.e., $p(w) \leq |uz|$. On the other hand, the period $p(w)$ is always at least as large as any of its local periods, so that $p(w) \geq p(w, u) = p = |uz|$. Therefore, $p(w) = p(w, u)$ showing that the factorization $w = uv$ is critical.

Case II: $p \leq |v|$. Now the illustration is as shown in Figure 6.5, where also the words u' and v' from the factorization $w = u'v'$ are shown.

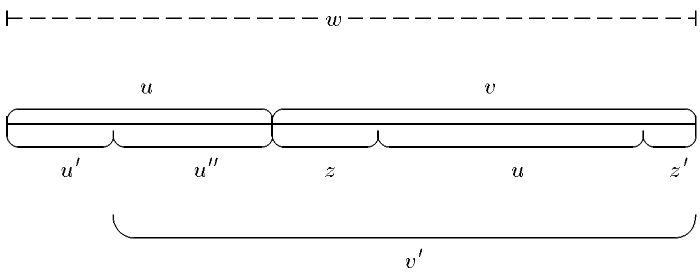


Figure 6.5. The case $p = |zu| \leq |v|$

Since $p \leq |v|$, and $|v| \leq |v'|$ we indeed have words u', u'' and z' such that $u = u'u''$ and $v = zuz'$. We have to show, as in Case I, that uv has a period $|zu|$.

By the maximality of v', the suffix $u''z'$ of v' satisfies $u''z' \preceq_r v' = u''v$, implying that $z' \preceq_r v$. On the other hand, the maximality of v yields the relation $z' \preceq_l v$. Therefore we conclude from Claim 2 that z' is a prefix of zuz'. It follows that $z' \in \text{pref}(zu)^\omega$, and hence $w = uv = uzuz' \in \text{pref}(uz)^\omega$, showing that w has a period $p = |zu|$. Consequently, also the Case II is completed.

It remains to be proved that (2) holds true, i.e., $|u| < p(w)$ and $|u'| < p(w)$. The former follows from the fact $|u| < p$, which we already proved, and the latter from our assumption $|u'| \leq |u|$.

Finally, to complete the proof of Theorem 6.2 we have to consider the case $|v'| \geq |v|$. But this reduces to the above case by interchanging the orderings $\preceq_l$ and $\preceq_r$. □

As we already noted we proved more than the existence of a critical factorization. Namely, we proved that such a factorization can be found by computing a lexicographically maximal suffix of a word, or in fact two of those with respect to two different orderings. There exist linear time algorithms for such computations, cf. [CP] or [CR]. For example, one can use the suffix tree construction of [McC].

The Critical Factorization Theorem is certainly a very fundamental result on periodicity of words. It is probably due to its subtle nature, as shown also by the above proof, that it has not been applied as much as it would have deserved.

One application of the theorem, which actually is the source of its discovery, cf. [CV], is as follows. To state it we have to recall the notion of an X-interpretation of a word defined in Section 2.1. An X-interpretation of a word $w \in \Sigma^+$ is a sequence $x, x_1, \ldots, x_n, y$ of words such that

$$xwy = x_1 \ldots x_n,$$

where $x_i \in X$, for $i = 1, \ldots, n$, x is a proper prefix of x_1 and y is a proper suffix of x_n. Two X-interpretations $x, x_1, \ldots, x_n, y$ and $x', x'_1, \ldots, x'_m, y'$ of w are *disjoint*, if for each $i \leq n$ and $j \leq m$, we have $x^{-1}x_1 \ldots x_i \neq x'^{-1}x'_1 \ldots x'_j$. Now an application of Theorem 6.2 yields, cf. [Lo]:

Proposition 6.1. *Let $w \in \Sigma^+$ and $X \subseteq \Sigma^+$ be a finite set satisfying $p(x) < p(w)$ for all $x \in X$. Then w has at most $\|X\|$ disjoint X-interpretations.*

Proposition 6.1 requires two remarks. First the disjointness is essential: if X-interpretations are required to be only different, then taking X to be a noncode the number of different X-interpretations could grow exponentially on $|w|$. In Proposition 6.1 the growth is bounded by a constant.

Second, the bound is close to the optimal one as noted in [Lo]: for each $n \geq 2$, words of the form $w \in (a^{2n-2}b)^+$ have exactly $n-1$ disjoint X-interpretations for $X = \{a^n, a^iba^i \mid i = 0, \ldots, n-1\}$.

Another elegant application of Theorem 6.2 was given in [CP], where it was used to describe an efficient pattern matching algorithm.

6.4 A characterization of ultimately periodic words

In this subsection we introduce a recent characterization of ultimately periodic words from [MRS]. The characterization is in terms of local properties of the considered word, or more precisely, in terms of repetitions at the ends of finite prefixes of the considered word. Variants for 2-way infinite words are presented, too.

Clearly, if $w = a_0a_1\ldots$, with $a_i \in \Sigma$, is ultimately periodic, then the following condition holds for any real number ρ:

(1) $\exists n = n(\rho) \in \mathbb{N} : \forall m \geq n : \mathrm{pref}_m w$ contains a repetition of order at least ρ as a suffix.

Our next simple example shows that infinite words satisfying (1) for $\rho = 2$ need not be ultimately periodic.

Example 6.6. Let $X = \{ab, aba\}$. Note that X is an ω-code, i.e., each word in X^ω has a unique X-factorization, due to the fact that any binary nonperiodic set is such, by Corollary 5.1. We consider infinite words in X^ω satisfying that in their X-factorizations

- (i) there are no two consecutive blocks of ab; and
- (ii) there are no three consecutive blocks of aba.

Let X_2 be the set of all such words. Obviously, the set X_2 is nondenumerable, and therefore contains words which are not ultimately periodic. Moreover, we claim that words in X_2 satisfy (1) for $\rho = 2$.

To see this we consider all possible sequences of ab- or aba-blocks immediately preceding ab (aba, resp.) in X-factorizations, and note that any position of ab (aba, resp.) is an endpoint of a square in these left extensions of ab (aba, resp.). Luckily there is only a finite number of cases to be considered as illustrated in Figure 6.6.

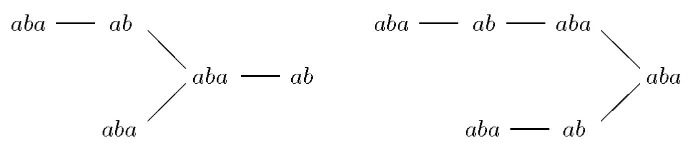


Figure 6.6. An exhaustive search for left extensions of ab and aba

A concrete example of a word which satisfies (1) for $\rho = 2$, and is not ultimately periodic is obtained by starting from $abaaba$ and extending it on the right nonperiodically by the blocks ab and aba. This particular word satisfies (1) for $\rho = 2$ with $n = 6$. □

Example 6.6 does not extend to higher integer repetitions, i.e., to the case $\rho = 3$, as we shall see in the next theorem. The proof of it is a modification due to A. Restivo from the proof of a more general result in [MRS].

Theorem 6.3. *A word $w \in \Sigma^\omega$ is ultimately periodic if, and only if, it satisfies (1) for $\rho = 3$, i.e., contains a cube as a suffix of any long enough prefix of w.*

Proof. To prove the nontrivial part we assume that w satisfies (1) with $\rho = 3$. We start with an auxiliary result.

Lemma 6.1. *Let $w = v^2$. If w has a period q satisfying $\frac{2}{3}|v| < q < |v|$, then $w = ux^3$, with $|x| = |v| - q$.*

Proof of Lemma. Denoting $w = zt$, with $|t| = q$, we can illustrate the situation in Figure 6.7.

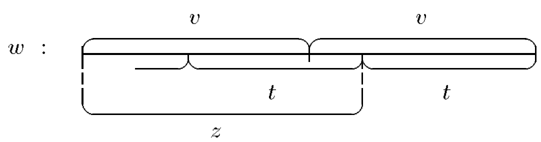


Figure 6.7. Factorizations of w with $|v| = p$ and $|t| = q$

By the Theorem of Fine and Wilf, v and t has a common period, and therefore $|v| - |t|$ is a period of z. By our assumption $\frac{2}{3}|v| < q = |t|$ implying that

$$3(|v| - |t|) < p.$$

It follows that z contains as a suffix a cube x^3, with $|x| = |v| - |t|$. Now, the lemma follows since any suffix of z is a suffix of w, as well. □

Proof of Theorem (continued). Let $w = a_0a_1\ldots$, with $a_i \in \Sigma$, and set

$$p(n) = \min\{d \mid \exists v \in \Sigma^+ : |v| = d \text{ and } a_0a_1\ldots a_n = uv^3\}.$$

Now, let $n(3)$ be the constant of the condition (1), and $m > n(3)$. As a crucial point of the proof we show the following implication:

(2) if $p(n) < p(m)$, for $n = n(3), \ldots, m-1$, then $m - n(3) < p(m)$.

To prove (2) we denote $p(m) = p$, and assume that $p(n) < p$ for $n = n(3), \ldots, m-1$. By the definition of p, we can write $a_0\ldots a_m = uv^3$, with $|v| = p$. Therefore

$$a_0 \dots a_{m-p} = uv^2, \quad \text{with} \quad |v| = p.$$

Now, assume contrary to our claim that $m - n(3) \geq p$. Therefore $m - p \geq n(3)$, and so by our assumption, we can write

$$a_0 \dots a_{m-p} = u'x^3, \quad \text{with} \quad |x| = p(m-p) = q < p.$$

There are two cases to be considered.

First, if $q > \frac{2}{3}p$, then v^2 satisfies the conditions of Lemma 6.1, and so we can write $v^2 = sy^3$, with $|y| = p - q$. This, however, is a contradiction with the choice of q, since $p - q \leq \frac{2}{3}p < q$.

Second, if $q \leq \frac{2}{3}p$, then v^2 has as a suffix a cube of a word of length q. Hence, the same holds for the word $a_0 \dots a_m$. This, however, is a contradiction since $q < p = p(m)$. This ends the proof of (2).

Next we apply (2) to conclude that

$$\sup\{p(n) \mid n \geq n(3)\} < n(3). \tag{3}$$

Indeed, if (3) does not hold, we choose the smallest m such that $p(m) \geq n(3)$. Then, by (2), we know that $m - n(3) < p(m)$, and therefore $m < p(m) + n(3) \leq 2p(m)$. This, however, contradicts with the fact that $a_0 \dots a_m = uv^3$ with $|v| = p(m)$. Hence (3) is indeed proved.

Now, we define

$$P = \sup\{p(n) \mid n \geq n(3)\}.$$

By (3), we know that $P \leq n(3)$, and we complete the proof of the theorem by induction on P.

The starting point $P = 1$ is obvious. To prove the induction step there are two possibilities (where actually only the first one relies on induction).

Case I: If there exist only finitely many numbers n such that $p(n) = P$, we can set

$$n(3) := \max\{n \mid p(n) = P\} + 1,$$

and apply induction hypothesis to conclude that w is ultimately periodic.

Case II: If there exist infinitely many integers n such that $p(n) = P$ we proceed as follows. Let the values $n = m_1, m_2, \dots$ be all such values. We shall prove, again by an induction, that, for $i = 1, 2, \dots$, the word

$$a_{n(3)} \dots a_{m_i} \quad \text{has a period} \quad P. \tag{4}$$

The starting point $i = 1$ is clear, since, by Lemma 6.1, $m_1 - n(3) < P$. So assume that the word $a_{n(3)} \dots a_{m_i}$ has a period P, and consider the word $a_{n(3)} \dots a_{m_{i+1}}$. Applying again Lemma 6.1, where $n(3)$ is replaced by m_i, we conclude that $m_{i+1} - m_i < P$.

We write

$$a_{n(3)} \dots a_{m_{i+1}} = uvw,$$

with

$$|w| = m_{i+1} - m_i,$$

and

$$|v| = 2P - 1.$$

Then, by induction hypothesis, uv has a period P. On the other hand, since $|vw| < 3P$ it follows that also vw has a period P. But, since the overlapping factor v is of length at least $P+1$, it is easy to conclude that also uvw has a period P, which completes the latter induction, as well as the whole proof of Theorem 6.3. □

Actually, as shown in [MRS], Theorem 6.3 can be sharpened as follows:

Proposition 6.2. *A word $w \in \Sigma^\omega$ is ultimately periodic if, and only if, it satisfies (1) for $\rho = \varphi^2$, where φ is the number of the golden ratio.*

Recall that $\varphi = \frac{1}{2}(\sqrt{5}+1)$, i.e., the positive root of the equation $\varphi^2 - \varphi - 1 = 0$. It is also shown in [MRS] that Proposition 6.3 is optimal in the sense that the validity of (1) for any smaller ρ than φ^2 does not imply that the word is ultimately periodic. This can be seen from the infinite Fibonacci word w_F considered in Section 8.

Our above considerations deserve two comments. First results extend to 2-way infinite words. Indeed, from the proof of Theorem 6.3 one can directly derive the following characterization.

Theorem 6.4. *A two-way infinite word $w = \ldots a_{i-1}a_ia_{i+1}\ldots$, with $a_i \in \Sigma$ is periodic if, and only if, there exists a constant N such that, for any i, the word $w \ldots a_{i-1}a_i$ contains a cube of length at most N as its suffix.* □

Note that in Theorem 6.4 the requirement that the cubes must be of a bounded length is necessary, as shown by the next example. In Theorem 6.3 this was not needed, since it dealt with only 1-way infinite words.

Example 6.7. We define a nonperiodic two-way infinite word

$$w = \ldots a_{-1}a_0a_1\ldots,$$

which contains a cube as a suffix of any factor $\ldots a_{i-1}a_i$ as follows. We set $w_0 = aaa$ and define

$$w_{2i+1} = \alpha_i w_{2i} a \quad \text{and} \quad w_{2i+2} = \beta_i w_{2i+1}, \quad \text{for} \quad i \geq 0,$$

where $a \in \Sigma$ and the words α_i and β_i are chosen such that both w_{2i+1} and $w_{2i+2}(\mathrm{suf}_{2i}(w_{2i+2}))^{-1}$ are cubes. Clearly, this is possible. It is also obvious that this procedure yields a word of the required form. □

As the second comment we introduce a modification of the above considerations. Surprisingly the results are quite different.

In above we required that repetitions occurred at any position "immediately to the left from that position". Now, we require that they occur at any position such that this position is the center of the repetition. We obtain the following characterization for periodic 2-way infinite words, in terms of local periods, cf. Section 6.3. Note that the notions of *local periods* extend in a natural way to infinite words, as well.

Theorem 6.5. *A two-way infinite word w is periodic if, and only if, there exists a constant N such that the local period of w at any point is at most N.*

Proof. Clearly, the periodicity of w implies the existence of the required N. The converse follows directly from the Critical Factorization Theorem: periods of all finite factors of w are at most N, and hence by the Theorem of Fine and Wilf w indeed is periodic. □

We note that Theorem 6.5, can be seen as a weak variant of the Critical Factorization Theorem, cf. [Du1]. It is also worth noticing that the boundedness of local periods is crucial, the argumentation being the same as in Example 6.7. Finally, the next example shows the optimality of Theorem 6.5 in a certain sense.

Example 6.8. Theorem 6.5 can be interpreted as follows. If a two-way infinite word w contains at any position a bounded square "centered" at this position, then the word is periodic. The word

$$w = {}^{\omega}aba^{\omega}$$

shows that no repetition of smaller order guarantees this. Indeed, for any $\rho < 2$, the word w contains at any position a bounded repetition of order of at least ρ centered at this position. Here, of course, the bound depends on ρ. □

7. Finiteness conditions

In this section we consider partial orderings of finite words and finite languages, and in particular orderings that are finite in either of two natural senses: either each subset contains only finitely many incomparable elements, i.e., each antichain is finite, or each subset contains only finitely many pairwise comparable elements, i.e., each chain is finite. Hence our interest is in two fundamental properties which are dual to each other.

7.1 Orders and quasi-orderings

For the sake of completeness we recall some basic notions on binary relations R over an arbitrary set S.

A binary relation R is a *strict ordering* if it is *transitive*, i.e., $(x,y) \in R$ and $(y,z) \in R$ implies $(x,z) \in R$, and *irreflexive*, i.e., $(x,x) \in R$ holds for no $x \in S$. It is a *quasi-ordering* if it is transitive and *reflexive*, i.e., $(x,x) \in R$ holds for all $x \in S$. It is a *partial ordering* if it reflexive, transitive and *antisymmetric*, i.e., $(x,y) \in \mathrm{R}$ and $(y,x) \in \mathrm{R}$ implies $x = y$ for all $x, y \in S$.

A *total ordering* is a partial ordering $\preceq$ for which $x \preceq y$ or $y \preceq x$ holds for all $x, y \in S$. An element x of a set S (resp. of a subset $X \subseteq S$) ordered by $\preceq$ is *minimal* if for all $y \in S$ (resp. $y \in X$) the condition $y \preceq x$ implies $x = y$. Of course each subset of a totally ordered set has at most one minimal element.

There is a natural interplay between these three notions. With each quasi-ordering $\preceq$ it is customary to associate the equivalence relation defined as $x \sim y$ if, and only if, $x \preceq y$ and $y \preceq x$ holds. This induces a relation $\leq$ on the quotient $S/\sim$

$$[x] \leq [y] \text{ if, and only if, } x \preceq y,$$

which is a partial ordering on S.

Example 7.1. The relation on Σ^* defined by $x \preceq y$, whenever $|x| \leq |y|$, is a quasi-ordering. The equivalence relation associated with it is: $x \sim y$ if, and only if, $|x| = |y|$. □

If $\prec$ is a strict ordering then the relation $\leq$ defined by $x \leq y$ if, and only if, $x \prec y$ or $x = y$, is a partial ordering. If $\preceq$ is a quasi-ordering, then the relation $<$ defined by $x < y$ if, and only if, $x \preceq y$ and $y \not\preceq x$, is a strict ordering.

Two important notions on partial orderings from the viewpoint of our considerations are those of a chain and an antichain. A subset X of an ordered set S is a *chain* if the restriction of $\preceq$ to X is total. It is an *antichain* if its elements are pairwise incomparable. A partial ordering in which every strictly descending chain is finite is *well-founded* or *Noetherian*. If in addition every set of pairwise incomparable elements is finite it is a *well-ordering*. For example, the set of integers ordered by $n|m$ if, and only if, n divides m is well-founded, but is not a well-ordering.

We concentrate on partial orderings over Σ^* and $\mathrm{Fin}(\Sigma^*)$, the family of finite subsets of Σ^*. We already observed how total orderings like lexicographic or alphabetic orderings are crucial, in considerations envolving words, for example for defining Lyndon words and proving the Critical Factorization Theorem.

Partial quasi-orderings can be defined on Σ^* and $\mathrm{Fin}(\Sigma^*)$ in many ways. Without pretending to be exhaustive, here are a few important examples:

alphabetic quasi-ordering: $x \preceq_a y$ iff $\mathrm{alph}(x) \subseteq \mathrm{alph}(y)$,

length ordering: $x \preceq_l y$ iff $|x| < |y|$ or $x = y$,
commutative image quasi-ordering: $x \preceq_c y$ iff $|x|_a \leq |y|_a$ for all $a \in \Sigma$,
prefix ordering: $x \preceq_p y$ iff there exits z with $xz = y$,
factor ordering: $x \preceq_f y$ iff there exits z, t with $zxt = y$,
subword ordering: $x \preceq_d y$ iff there exist $x_1, \ldots, x_n, u_0, \ldots, u_n \in \Sigma^*$ such that

$$x = x_1 x_2 \ldots x_n \text{ and } y = u_0 x_1 u_1 x_2 u_2 \ldots x_n u_n.$$

Similarly, for the family $\text{Fin}(\Sigma^*)$ we define the following orderings. Here the notation R_X denotes the set of relations satisfied by X.

size quasi-ordering: $X \preceq_s Y$ iff $||X|| \leq ||Y||$,
inclusion ordering: $X \preceq_i Y$ iff $X \subseteq Y$,
semigroup quasi-ordering: $X \preceq_m Y$ iff $X^+ \subseteq Y^+$ where X and Y are minimal generating sets,
relation quasi-ordering: $X \preceq_r Y$ iff there exits a bijection $\varphi : X \to Y$ such that $R_{\varphi(X)} \subseteq R_X$.

We summarize into the following table the facts on how the above partial orderings behave with respect to our two finiteness conditions, i.e., whether or not they allow infinite antichains or chains.

Table 7.1. Finiteness conditions of certain quasi-orderings

	$\preceq_a$	$\preceq_l$	$\preceq_p$	$\preceq_f$	$\preceq_d$	$\preceq_s$	$\preceq_i$	$\preceq_m$	$\preceq_r$
no infinite chains	+	−	−	−	−	−	−	−	$\oplus$
no infinite antichains	+	+	−	−	$\oplus$	+	−	−	−

There are two particularly interesting entries in this table, namely those denoted by $\oplus$. These state two fundamental finiteness conditions on words and finite sets of words we shall be studying in more details later. That the other entries are correct is, as the reader can verify, easy to conclude. We only note that the relation ordering $\preceq_r$ is not a well-ordering even in the family of sets of the same size as shown by the family $\{X_i = \{a, a^i b, b\} | i \geq 1\}$.

7.2 Orderings on words

In this subsection we consider orderings on Σ^*, in particular the subword ordering and another one related to it.

The subword ordering is called division ordering in [Lo], but this notion has another use in the literature, where by *division* ordering is meant a partial ordering satisfying the following two conditions for all $x, y, z, t \in \Sigma^*$

(1) $$1 \preceq x$$

(2) $$x \preceq y \text{ implies } zxt \preceq zyt.$$

Observe that, by (1), we have $1 \preceq z$ and $1 \preceq y$, and hence, by (2), $x \preceq zx, x \preceq xy$ and $zx \preceq zxy$, i.e., $x \preceq zx \preceq zxy$. Thus every word is greater than or equal to each of its factors:

(3) $$\text{for all } x, y, z, \text{ the inequality } x \preceq yxz \text{ holds.}$$

Actually, the subword ordering is the smallest ordering satisfying the conditions (1) and (2), i.e., for all $x, y \in \Sigma^*$ the relation $x \preceq_d y$ implies $x \preceq y$. Indeed, we have $1 \preceq_d 1$ and $1 \preceq 1$ by condition (1). Now, consider $x = ax' \preceq_d y = by'$ with $a, b \in \Sigma$ and $x', y' \in \Sigma^*$, and let us proceed by induction on $|x| + |y|$. If $a = b$, then $x' \preceq_d y'$, i.e., $x' \preceq y'$ by induction, and by (2), $x = ax' \preceq y = ay'$. On the other hand, if $a \neq b$, then $x \preceq_d y'$, and thus by induction $x \preceq y'$. Condition (1) yields $1 \preceq a$ and condition (2) yields $y' \preceq y = ay'$, so by the transitivity $x \preceq y$.

Total division orderings have been studied in [Mart]. It is proved under a certain assumption, namely the ordering being "tame", that each division ordering is finer than the *strong commutative image* ordering, which is obtained from $\preceq_c$ by replacing inequalities by the strict inequalities in each component. It is also conjectured that the statement holds true even without this condition. However, when the alphabet is binary, each division ordering is tame, and thus the result holds.

Theorem 7.1. *Let $\preceq$ be a total division ordering on the free monoid generated by $\{a, b\}$ and let u, v be two words. Then*

(4) $$|u|_a < |v|_a \text{ and } |u|_b < |v|_b \text{ implies } u \prec v.$$

Proof. Since $\preceq$ is total, we may assume without loss of generality that $ba \succ ab$ holds. In particular, by commutating the occurrences of a and b in $u \in \Sigma^*$, we have, by equality (2):

(5) $$b^n a^m \succeq u \succeq a^m b^n \text{ with } m = |u|_a \text{ and } n = |u|_b.$$

Now assume that condition (4) is violated: $|u|_a < |v|_a$, $|u|_b < |v|_b$ and $u \succ v$. By setting $|u|_a = m$, $|v|_a = m'$, $|u|_b = n$, $|v|_b = n'$ we have

$$b^n a^m \succeq u \succ v \succeq a^{m'} b^{n'} \succeq a^{m+1} b^{n+1}.$$

Thus we may assume that we have $u = b^n a^m$, $v = a^{m+1} b^{n+1}$ and $u \succ v$. We first observe that $b^n u \succ u b^{n+1}$. Indeed, we have $b^n u = b^n b^n a^m \succ b^n a^{m+1} b^{n+1}$. Now, by (3), we have $a^{m+1} \succ a^m$, i.e., $b^n a^{m+1} b^{n+1} \succ b^n a^m b^{n+1} = u b^{n+1}$.

Assume $b \succ a$ and for all $k > 0$ compute:

$$\begin{aligned} b^{(k+1)n}b^m &\succ b^{(k+1)n}a^m = b^{kn}u \\ &\succ ub^{k(n+1)} \\ &\succ vb^{k(n+1)} = a^{m+1}b^{n+1}b^{k(n+1)} = a^{m+1}b^{(k+1)n+1}b^k \\ &\succ b^{(k+1)n+1}b^k. \end{aligned}$$

This does not hold when $k+1 > m$. Now, if $a \succ b$, a similar argument leads to the same type of contradiction, proving the theorem. □

The author shows that the inequalities of condition (4) must be strict. Indeed, consider the ordering $\preceq$ on $\{a,b\}^*$, where words are ordered by their number of occurrences first, and then lexicographically with $a \succ b$. Then $u = bababa \succ v = abbabba$, but $|u|_a = |v|_a$ and $|u|_b < |v|_b$.

We turn to consider the subword ordering. We already observed that it is right- and left-invariant, cf. (2). Its second major property, solving one nontrivial entry in Table 7.1, is that it is a well-ordering implying that every subset $X \subseteq \Sigma^*$ has finitely many minimal elements.

Theorem 7.2. *The subword ordering $\preceq_d$ over a finitely generated free monoid is a well-ordering.*

Proof. Clearly subword ordering is well-founded. So we have to prove that any antichain of Σ^* with respect to $\preceq_d$ is finite. Assume to the contrary that $F = \{f_i | i \in \mathbb{N}\}$ is an infinite set of incomparable words. Then, in particular, we have

$$\text{if } i < j, \text{ then } f_i \not\preceq_d f_j. \tag{6}$$

Among the sequences satisfying (6) there exist such sequences, where f_1 is the shortest possible. Continuing inductively we conclude that there exists a sequence, say $(g_i)_{i\geq 0}$, which satisfies (6) and none of the sequences $(h_i)_{i\geq 0}$, with $|h_i| < |g_i|$ for some i, satisfies (6).

Now, consider the sequence $(g_i)_{i\geq 0}$. Since Σ is finite there exist a letter a such that, for infinitely many i, we can write $g_i = ag_i'$ with $g_i' \in \Sigma^*$. Say this holds for values $i_1, i_2, \ldots$. Then the sequence

$$g_1, g_2, \ldots, g_{i_1-1}, g'_{i_1}, g'_{i_2}, \ldots$$

satisfies (6) and, moreover, $|g'_{i_1}| < |g_{i_1}|$. This contradicts with the choice of the sequence $(g_i)_{i\geq 0}$. □

This theorem is due to Higman in [Hi], where it is proved in a much more general setting. Subsequently, it has been rediscovered several times, see [Kr] for a complete account. Our proof of Theorem 7.2 is from [Lo]. It is very short and nonconstructive. It is also worth noticing that there is no bound for the size of a maximal antichain in Σ^*, as shown by the antichains $A_n = \{a^i b^{n-i} | i < n\}$ for $n \geq 0$.

We also note that Dickson's Lemma is a consequence of Theorem 7.2. We recall that it asserts that $\mathbb{N}^k$ is well-ordered, where the ordering is the extension of the usual componentwise ordering on $\mathbb{N}$. Indeed, it suffices to interprete the k-tuple $(n_1, \ldots, n_k)$ as the word $a_1^{n_1} \ldots a_k^{n_k}$ over the alphabet $\{a_1, \ldots, a_k\}$.

An interesting formal language theoretic consequence of Theorem 7.2 is the following.

Theorem 7.3. *For each language $L \subseteq \Sigma^*$ the languages $SW(L) = \{w \mid \exists z \in L : w \preceq_d z\}$ and $SW_1(L) = \{w \mid \exists z \in L : z \preceq_a w\}$ are rational.*

Proof. By Theorem 7.2, the set of minimal elements of L with respect to $\preceq_d$ is finite, say F. So, $SW_1(L) = SW_1(F)$, and hence $SW_1(L)$ is rational. A bit more complicated proof for $SW(L)$ is left to the reader. □

Our above considerations on the subword ordering were purely existential. As an example of algorithmic aspects we state a problem motivated by molecular biology. The problem asks to find, for a given finite set $X = \{x_1, \ldots, x_n\}$ of words, a shortest word z such that $x_i \preceq_p z$ for all $i = 1, \ldots, n$. This problem is usually referred to as the *smallest common supersequence problem*, and it is known to be NP-complete, cf. [GJ].

7.3 Subwords of a given word

In this a bit isolated subsection we consider an interesting problem asking to differentiate two words by a shortest possible subword occurring in one but not in the other.

Example 7.2. The word *bbaa* occurs in *abbaab*, but does not occur in *ababab* as a subword. All words of length 3 occur in both words. □

We refer the reader to [Lo] for a full exposition of the problem. In particular it is established that a word of length n is determined by the set of its subwords of length $\lceil \frac{n+1}{2} \rceil$, the pair $a^{m-1}ba^m, a^mba^{m-1}$ showing that the bound is sharp. In [Si] it is proved that a shortest subword distinguishing two given different words can be found in linear time. This is not a priori obvious since there may exist exponentially many subwords of a given length in a word. For instance, $(ab)^n$ contains all words of length n as subwords. The linearity of the algorithm is based on several properties among which the fact that, if two words u and v have the same subwords of length m, then they can be merged in a word having also the same subwords of length m.

An elaboration of this question is to consider the subwords with their multiplicities. In the previous example *baab* occurs twice in *abbaab* but once in *ababab*. Milner (personal communication) defines the *k-spectrum* of a word u as the function that associates with each word of length $0 < k \leq |u|$, the number of its occurrences in u. Given an integer k, consider the maximal integer $n = f(k)$ such that two different words of length n have different k-spectrums. The question is to find a reasonable upper bound on n.

Example 7.3. Define 3 sequences of words by

$$u_0 = aba, v_0 = 1, w_0 = baa,$$
$$u_1 = ab, v_1 = 1, w_1 = ba,$$
$$u_{k+2} = w_k u_{k+1}, v_{k+2} = u_{k+1} u_k, u_{k+2} = u_{k+1} w_k \text{ if } k \text{ is even},$$
$$u_{k+2} = u_{k+1} w_k, v_{k+2} = u_k u_{k+1}, u_{k+2} = w_k u_{k+1} \text{ if } k \text{ is odd}.$$

Then u_k and v_k are different, and have the same k-spectrum. Their common length $\phi(k)$ grows as a "Fibonacci" type function, starting from the values 2 and 5: $\phi(1) = 2$, $\phi(2) = 5$, $\phi(3) = 7$, $\phi(4) = 12$, $\phi(5) = 19$, $\phi(6) = 31, \ldots$.

The exact values for small k are

k	1	2	3	4
$f(k)$	3	5	8	13

,

but for $k = 5$, $\phi(5) = 19$ is far from being optimal due to the following two words of length 16 having the same 5-spectrum:

$$u = abbaaaaabbbaaaab \text{ and } v = baaaabbbaaaaabba. \qquad \square$$

The same questions substituting "factor" for "subword" can be posed. It is shown in [Lo], Exercise 6.2.11, that whenever u is not of the form $(xy)^n x$ with $n \geq 2$, then it is uniquely determined by its factors of length $\lceil \frac{|u|}{2} + 1 \rceil$. If this restriction is relaxed, then the word can not be determined by its proper factors: $(ab)^n a$ and $(ba)^n b$ have the same factors $(ab)^n$ and $(ba)^n$ as occurrences. It is also possible to define the *k-factor* spectrum of a word u which associates with each word of length k the number of its occurrences in u. To our knowledge no nontrivial bounds are known.

7.4 Partial orderings and an unavoidability

In this section we state a generalization of Higman's Theorem. This result is based on the notion of an unavoidable set of words, which *is not* connected to the unavoidability of Section 8. We also consider some other problems connected to this notion of an unavoidability.

We say that a set $X \subseteq \Sigma^*$ is *unavoidable*, if there exists a constant k such that each word $w \in \Sigma^k$ contains a word of X as a factor. For example, the set $X = \{aa, ba, bb\}$ is unavoidable over the free monoid $\{a, b\}^*$, since avoiding a^2 and b^2 obliges a word to be a sequence of a and b alternatively.

This definition was given in [EHR] in connection with an attempt to characterize the rational languages among the context-free ones. In particular, unavoidable subsets are used for extending Theorem 7.2 showing that the subword ordering $\preceq_d$ on words is a *well-ordering*. Actually, saying that a word u is subword of v means that v can be obtained from u by inserting letters. Instead of inserting letters we can insert words from a fixed subset. Given $X \subseteq \Sigma^*$ define $\prec_X$ as the reflexive and transitive closure of the relation

$$\{(u_1u_2, u_1xu_2) | x \in X, u_1, u_2 \in \Sigma^*\}.$$

For instance, if $X = \{ab\}$, then we get $1 \prec_X ab \prec_X aabb \prec_X aabbab$.

Then the following is proved in [EHR].

Proposition 7.1. *The ordering $\prec_X$ is a well-ordering if, and only if, X is unavoidable.*

We continue with some elementary properties of unavoidability. It is clear from the definition that from each unavoidable set we can extract a finite unavoidable subset, so the study can be reduced to finite unavoidable sets. It is also easy to verify that a set X is unavoidable if, and only if, it avoids all one-way infinite words if, and only if, it avoids all two-way infinite words. Indeed, let us verify, e.g., that if X is unavoidable, then every two-way infinite word $\dots a_{-1}a_0a_1\dots$ has a factor in X. By hypothesis, there are infinitely many words avoiding X, so there are infinitely many such words of even length. Now, say a word x is a *central occurrence* of a word y, if $y = y_1xy_2$ with $|y_1| = |y_2|$. An infinite two-way word avoiding X is constructed as follows. For some $(a_0, b_0) \in \Sigma \times \Sigma$ there are infinitely many words having $x_0 = a_0b_0$ as a central factor and avoiding X. Now, for some $(a_1, b_1) \in \Sigma \times \Sigma$ there are infinitely many words having $x_1 = a_1a_0b_0b_1$ as a central factor and avoiding X, and so on. The infinite word $\dots a_2a_1a_0b_0b_1b_2\dots$ thus defined avoids X.

Testing the unavoidability of X can be done in different ways. We may construct a finite automaton recognizing $\Sigma^*X - \Sigma^*X\Sigma^+$, and then check whether or not there is a loop in the automaton. Another approach is more combinatorial and consists in simplifying X as much as possible. For example, assume that $\{babba, bbb\}$ are elements of a set X. We claim that by substituting $babb$ for $babba$ the set of two-way infinite words that are avoided is unchanged. Indeed, if an infinite word contains $babb$, then this occurrence is either followed by a, and then the word contains $babba$, or it is followed by b, but then it contains the occurrence bbb. The point here is that the occurrence $babbb$ has a suffix in $X - \{babba\}$. This leads to the following definitions.

A set X *immediately left-* (resp. *right-*) *simplifies* to the set Y, if either $Y = X - \{x\}$, where x has a proper factor in X, or $Y = X - \{wa\} \cup \{w\}$ (resp.$Y = X - \{aw\} \cup \{w\}$), where $wa \in X$ (resp.$aw \in X$) with $w \in \Sigma^*$ and $a \in \Sigma$, such that the following holds:

> for all $b \in \Sigma, b \neq a$, wb has a suffix (resp. bw has a prefix) in $X - \{wa\}$ (resp. $X - \{aw\}$).

Further a set X *simplifies* (resp. *left-, right-*) *simplifies* to the set Y, if there exists a sequence of $n \geq 0$ sets $X_0 = X, \dots, X_n = Y$ such that X_i immediately simplifies (resp. left-, right-simplifies) to X_{i+1}, with the convention

$X = Y$, if $n = 0$. Finally, a set X is *simple* (resp. *left-, right-simple*), if there is no $Y \neq X$ such that X simplifies (resp. left-, right-simplifies) to Y.

Above simplifications can be used to test unavoidability, as shown in [Ros2] and also known to J.-P. Duval (private communication).

Proposition 7.2. *A subset X is unavoidable if, and only if, it simplifies (resp. left-simplifies, right-simplifies) to the set consisting of the empty word only.*

Example 7.4. As an illustration, when the above is applied to the set $\{aaa, aba, bb\}$ the following sequence of sets is obtained: $X_0 = \{aaa, \underline{aba}, bb\}$, $X_1 = \{\underline{aaa}, ab, bb\}$, $X_2 = \{aa, \underline{ab}, bb\}$, $X_3 = \{\underline{aa}, a, bb\}$, $X_4 = \{a, \underline{bb}\}$, $X_5 = \{a, \underline{b}\}$, $X_6 = \{\underline{a}, 1\}$, $X_7 = \{1\}$. □

Actually, a more general problem was solved in [Ros2] by showing that for all finite subsets X there exists a unique simple set Y equivalent to X, in the sense that it avoids the same set of infinite words. Furthermore, Y can be obtained by first right-simplifying X as long as possible and then left-simplifying it. More precisely, for each X denote by $\overline{X}$ (resp. $\overline{X}^r$, $\overline{X}^l$) any simple (resp. right-, left-simple) subset which is the last element in a chain of simplification (resp. right-, left-simplification) starting from X. The following asserts a property of confluence saying that the result of a maximal sequence of simplification does not depend on the intermediate choices.

Proposition 7.3. *For all X, there exists a unique simple subset equivalent to it, namely $\overline{X} = \overline{\overline{X}^r}^l = \overline{\overline{X}^l}^r$*

Now we come to the problem that motivated the study of unavoidable sets. Haussler conjectured that every unavoidable set of words X can be extended in the sense that there exists an element $u \in X$ and a letter $a \in \Sigma$ such that substituting ua for u in X yields a new unavoidable set. For instance, in the previous example, the word ba can be replaced by bab (but not by baa, a^2 or b^2 as is easily verified). This conjecture held for some time and was supposed to be true. It was a nontrivial statement since, extending a word need not preserve the avoidability, but all computed examples confirmed that there always existed an extendable word. In [CC] some equivalent statements to the conjecture were given and some particular cases were settled. In fact, the conjecture turned out to be wrong, though it needed some clever efforts to exhibit the following counter-example (with the minimal possible number of elements) from [Ros1]:

$$X = \{aaa, bbbb, abbbab, abbab, abab, bbaabb, baabaab\}.$$

The reader may run the above procedure to check that X is unavoidable, as well as to use an exhaustive case study to show that no word can be extended.

Finally, [Ros2] introduces another interesting notion. Two subsets X and Y are *weakly equivalent*, written $X \sim_w Y$, if the sets of infinite periodic words, i.e., of the form $\ldots uu \ldots$ for some $u \in \Sigma^+$, avoiding them are equal. This notion seems to deserve further research. In particular the proof of the fact that two words $u \neq v \in \{a,b\}^*$, satisfy $u \sim_w v$ if, and only if, $\{u,v\} = \{a^n b, ba^n\}$ or $\{u,v\} = \{b^n a, ab^n\}$ is rather long and should be simplifiable.

7.5 Basics on the relation quasi-ordering $\preceq_{\mathrm{r}}$

We turn to consider orderings on finite sets of words, in particular that of the quasi-ordering $\preceq_r$. By definition it was associated with relations satisfied by words of X, and hence with solutions of equations. This leads us to consider systems of equations with a finite number of unknowns and without constants.

Let Ξ be a finite set of unknowns and

$$S : u_i = v_i \text{ with } u_i, v_i \in \Xi^*, \text{ for } i \in I,$$

be a system of equations over Ξ. We are interested in all solutions of such a system in a given free monoid Σ^*, i.e., all morphisms $h : \Xi^* \to \Sigma^*$ satisfying $h(u) = h(v)$ for all $u = v$ in S. We are going to show that any system S is equivalent to one of its finite subsystems, i.e., any solution of this finite subsystem is also a solution of the whole S. Clearly, this states a fundamental compactness property of systems of equations over free monoids, and hence also of words. This property was conjectured by A. Ehrenfeucht at the beginning of 70's in a slightly different form, as we shall see in a moment, cf. also [Ka3].

Let us start with a simple example.

Example 7.5. Consider systems of equations with only two unknowns. Then, by the defect theorem, each solution $h : \{x,y\}^* \to \Sigma^*$ is periodic. Therefore the set of all solutions of a given nontrivial equation consists of morphisms satisfying one of the following conditions:

(i) $h(x) = h(y) = 1$;
(ii) $\exists k \in \mathbb{Q}_+ \cup \{\infty\} : |h(x)|/|h(y)| = k$ and h is periodic;
(iii) $h(x), h(y) \in z^*$ for some $z \in \Sigma^+$, i.e., h is periodic.

Actually, condition (ii) consists of infinitely many different conditions, one for each choice of k. It follows straightforwardly that the set of all solutions of a given system of equations over $\{x,y\}$ is determined by at most two equations. For example, if S contains equations of type (ii) for two different k's, then the only common solution is that of (i), and hence these two equations constitute an equivalent subsystem of two equations. □

It is interesting to note that no similar analysis is known to work in the case of three unknowns. Indeed, no upper bound for the size of an equivalent finite subsystem is known. This is despite of the fact that there exists a finite classification for sets of all equations satisfied by a given morphism $h : \{x, y, z\}^* \to \Sigma^*$, cf. [Sp1].

As we already mentioned the original *Ehrenfeucht's Conjecture* was stated in a slightly different form, more in terms of formal languages. In order to formulate it let us say that two morphisms $h, g : \Sigma^* \to \Delta^*$ *agree* on a word w if $h(w) = g(w)$. Motivated by research on questions when two morphisms agree on all words of a certain language, for more details cf. [Ka3], he conjectured that

$$\forall L \subseteq \Sigma^*, \exists \text{ finite } F \subseteq L : \forall h, g : \Sigma^* \to \Delta^* :$$
$$h(w) = g(w) \text{ for all } w \in L \Leftrightarrow h(w) = g(w) \text{ for all } w \in F.$$

In other words, the conjecture states that, for any language L, there exists a *finite subset* F of L such that to test whether two morphisms agree on words of L it is enough to do that on words of F. Such a finite F is called a *test set* for L. In terms of equations the conjecture states that any system of equations of the form

$$S : u_i = \bar{u}_i, \text{ for } i \in I,$$

where $\bar{u}_i$ is an isomorphic copy of u_i in a disjoint alphabet, is equivalent to one of its finite subsystems. As was first noted in [CuKa1], cf. also [HK2], this restricted formulation of the conjecture is actually equivalent to the general one.

As a result related to Example 7.5 we show next that all languages over a binary alphabet has a very small test set.

Theorem 7.4. *Each binary language possesses a test set of size at most three.*

Proof. The proof is based on Theorem 3.2. Here we present the main ideas of it, but omit a few technical details which can be found in [EKR2].

Let $L \subseteq \{a, b\}^*$ be a binary language. We define the *ratio* of $w \in \{a, b\}^+$ as the quantity $r(w) = |w|_a/|w|_b$. Hence, $r(w) \in \mathbb{Q}_+ \cup \{\infty\}$. A simple length argument shows that no two morphisms h, g, with $h \neq g$, can agree on two words with a different ratio. Consequently, if L contains two words with a different ratio, then they constitute a two-element test set for L.

So we assume that, for some k, $r(w) = k$ for all w in L. Now, each word w in L can be factorized as $w = w_1 \dots w_n$, where, for each i, $r(w_i) = k$ and, for each prefix w_i' of w_i, we have $r(w_i') \neq k$. Let L_k be the set of all factors in the above factorizations of all words of L. It follows that if L_k has a test set of cardinality at most three, so has L: take a subset of L containing all words of the test set of L_k in the above factorizations.

So it remains to be shown that L_k has a test set of size at most three. If $||L_k|| \leq 2$, there is nothing to be proved. So, assume that $||L_k|| \geq 3$. Now, we use the partial characterization of binary equality sets proved in Theorem 3.2. Such a set is always of one of the following forms:

(i) $X_r = \{w | r(w) = r\}$ with $r \in \mathbb{Q}_+ \cup \{\infty\}$,
(ii) $\{\alpha, \beta\}^*$ for some words $\alpha, \beta \in \Sigma^*$,
(iii) $(\alpha\beta^*\gamma)^*$ for some words $\alpha, \beta, \gamma \in \Sigma^*$.

For morphisms having an equality set of form (i) any one-element subset of L_k works as a test set. For morphisms having an equality set of form (ii) any two-element subset of L_k works, since no word in L_k is a product of words having the same ratio. Finally, morphisms having an equality set of the form (iii) (if there are any!) are most complicated to handle. In this case one can show, cf. [EKR2], that if an equality set of form (iii) contains two elements of L_k, then these two elements determine this equality set uniquely. Consequently, for morphisms having equality sets of form (iii) any two-element subset of L_k works for all other pairs of morphisms except for those having as an equality set the one determined by these two words. And for those this two-element set can be extended to a three-element test set by adding a third word from L_k.

Consequently, in all the cases three words are enough. □

Of course, even in Theorem 7.4 a test set cannot be found effectively, in general. However, our above proof indicates that under a rather mild assumptions on L this can be done, cf. [EKR2].

7.6 A compactness property

In this section we prove the compactness property conjectured by Ehrenfeucht, and will later interprete it as a finiteness condition on finite sets of words, as well as consider its consequences.

Theorem 7.5. *Each system of equations with a finite number of unknowns over a free monoid is equivalent to one of its finite subsystems.*

Proof. Let Ξ be a finite set of unknowns in the equations

$$S : u_i = v_i \qquad \text{for } i \in I,$$

and Σ^* a free monoid, where these equations are solved. We exclude the case $\|\Sigma\| = 1$, since this is a trivial exercise in linear algebra. We also note that due the embeddings of Section 3.2 it does not matter what $\|\Sigma\|$ is – it can be even nondenumerable. Let us fix $\Sigma = \{a_0, \ldots, a_{n-1}\}$ with $n \geq 2$.

The basic idea is that we convert equations on words into polynomial equations on numbers. This is possible simply because a word w can be interpreted as the number it presents in n-ary notations.

More precisely, consider an equation

(1) $$u = v \text{ with } u, v \in \Xi^+.$$

Define two copies of Ξ, say Ξ_1 and Ξ_2, and associate with (1) the following pair of polynomial equations

$$(2) \qquad \begin{cases} l(u) - l(v) &= 0, \\ n(u) - n(v) &= 0, \end{cases}$$

where $l, n : \Xi^* \to (\Xi_1 \cup \Xi_2)^*$ are mappings defined recursively as

$$(3) \qquad \begin{cases} l(a) &= a_1, & \text{for } a \in \Xi, \\ l(wa) &= l(w)a_1, & \text{for } a \in \Xi \text{ and } w \in \Xi^+, \\ n(a) &= a_2, & \text{for } a \in \Xi, \\ n(wa) &= n(w)l(a) + n(a), & \text{for } a \in \Xi \text{ and } w \in \Xi^+. \end{cases}$$

Equations (2) are well-defined, and they are polynomial equations over the set $\Xi_1 \cup \Xi_2$ of *commuting* unknowns. In fact, coefficients of the monomials in (2) are from the set $\{-1, 0, 1\}$. Note also that the function n, as is obvious by induction, satisfies the relation

$$(4) \qquad n(w_1 w_2) = n(w_1)l(w_2) + n(w_2), \text{ for all } w_1, w_2 \in \Xi^+.$$

Now, let $w = a_{i_{k-1}} \dots a_{i_0}$, with $a_{i_j} \in \Sigma$, be a word in Σ^+. We associate with it two numbers

$$\sigma(w) = a_{i_0} + a_{i_1} n + \dots + a_{i_{k-1}} n^{k-1}$$

and

$$\sigma_0(w) = n^k.$$

Hence $\sigma(w)$ is the value of w as the n-ary number and $\sigma_0(w)$ is the value $n^{|w|}$. This guides us to set $\sigma(1) = 0$ and $\sigma_0(1) = n^0 = 1$.

Obviously, the correspondence $w \leftrightarrow (\sigma_0(w), \sigma(w))$ is one-to-one, and we use it to show:

> $h : \Xi^* \to \Sigma^*$, i.e., $(h(a_0), \dots, h(a_{n-1}))$, is a solution of (1),

if, and only if,

> the $2n$-tuple $(\sigma_0(h(a_0), \dots, \sigma_0(h(a_{n-1})), \sigma(h(a_0)), \dots, \sigma(h(a_{n-1})))$ is a solution of (2).

To prove this equivalence , let us denote $s = (h(a_0), \dots, h(a_{n-1}))$, $s_1 = \sigma_0(s)$ and $s_2 = \sigma(s)$, where σ_0 and σ are applied to s componentwise. Then, if $h(u) = h(v)$, we conclude that

$$l(u)\Big|_{s_1} = n^{|h(u)|} = n^{|h(v)|} = l(v)\Big|_{s_1},$$

i.e., s_1 is a solution of the equation $l(u) - l(v) = 0$. Similarly, factorizing $u = u_1 u_2$, with $h(u_1), h(u_2) \neq 1$, we conclude from (4) that

$$\begin{aligned} n(u)\Big|_{s_1,s_2} &= n(u_1)\Big|_{s_1,s_2} \cdot l(u_2)\Big|_{s_1,s_2} + n(u_2)\Big|_{s_1,s_2} \\ &= \sigma(h(u_1))n^{|h(u_2)|} + \sigma(h(u_2)) = \sigma(h(u_1 u_2)) = \sigma(h(u)), \end{aligned}$$

where the second equality is due to induction. The above holds also, as the basis of induction, when u does not have the above factorization. Symmetrically, $n(v)\Big|_{s_1,s_2} = \sigma(h(v))$, so we have shown that (s_1, s_2) is a solution of (2). On the other hand, if in above notations $s = (s_1, s_2)$ is a solution of (2) the above calculations show that h is a solution of (1), proving the equivalence.

Now, assume that S is our given system of equations, with Ξ as the set of unknowns, consisting of equations $u_i = v_i$ for $i \in I$. Let

$$p_j(\Xi_1, \Xi_2) = 0 \text{ for } j \in J$$

be a set of polynomial equations, with $\Xi_1 \cup \Xi_2$ as the set of unknowns, consisting of those equations which are obtained in (2) when i ranges over I. For simplicity let $p_j = p_j(\Xi_1, \Xi_2)$ and $\mathcal{P} = \{p_j | j \in J\}$. By Hilbert's Basis Theorem, cf. [Co], $\mathcal{P}$ is finitely based, i.e., there exists a finite subset $\mathcal{P}_0 = \{p_j | j \in J_0\} \subseteq \mathcal{P}$ such that each $p \in \mathcal{P}$ can be expressed as a linear combination of polynomials in $\mathcal{P}_0$:

$$p = \sum_{j \in J_0} g_j p_j \text{ with } g_j \in \mathbb{Z}(\Xi_1 \cup \Xi_2).$$

Consequently, the systems "$\mathcal{P}_j = 0$ for $j \in J$" and "$\mathcal{P}_j = 0$ for $j \in J_0$" have exactly the same solutions. Therefore, by the equivalence we proved, our original system S is equivalent to its finite subsystem containing only those equations of S needed to construct $\mathcal{P}_0$. □

The proof of Theorem 7.5 deserves a few comments. There are several proofs of this important compactness result, however, all of those rely on Hilbert's Basis Theorem. The two original ones are those by Albert and Lawrence in [AL1] and Guba in [MS]. Our proof is modelled from ideas of Guba presented in [McN2] and [Sal3], cf. also [RoSa2], using n-ary numbers. The other simple possibility of proving this result is to use embeddings of Σ^* into the ring of 2×2-matrices over integers, cf. [Per] or [HK2]. The advantage of the above proof is that it uses only twice as many unknowns as there are in the original system.

It is also worth noticing that we did not need above the full power of Hilbert's Basis Theorem. Indeed, we only needed the fact that the common roots of the polynomials $\mathcal{P}_j$, for $j \in J_0$, are exactly the same as those of the polynomials $\mathcal{P}_j$, for $j \in J$, which is not the Hilbert's Basis Theorem, but only its consequence. Note also that the reduction from word equations to polynomial equations goes only in one direction. Indeed, the existence of a solution of an equation is decidable for word equations, as shown by Makanin, while it is undecidable for polynomial equations, as shown by Matiyasevich, cf. [Mat] and also [Da].

Finally, let us still emphasize one peculiar feature of the above proof. The original problem is, without any doubts, a problem in a very noncommutative algebra, while its solution relies – unavoidably according to the current knowledge – on a result in a commutative algebra.

Of course, a finite equivalent subsystem for a given system of equations cannot be found effectively, in general. However, in several restricted cases this goal can be achieved. The proofs are normally direct combinatorial proofs not relying, for example, on Hilbert's Basis Theorem. We present one such example needed in our later considerations, for other such results we refer to [ACK], [Ka3], [HK2], [KRJ] or [KPR].

We recall that a system of equations in unknowns Ξ is called *rational* if it is a rational relation of $\Xi^* \times \Xi^*$, cf. [Be1].

Theorem 7.6. *For each rational system of equations in a finite number of unknowns one can effectively find an equivalent finite subsystem.*

Proof. Of course, the formulation of Theorem 7.6 silently assumes that the system S is effectively given, for example, defined by a finite transducer τ, cf. [Be1]. Let n be the number of states of τ. Set

$S_0 = \{u = v \in S \mid (u,v)$ has an accepting computation in τ of length at most $2n\}$.

We claim that S_0 is equivalent to S. Assume the contrary that $h : \Xi^* \to \Sigma^*$ is a solution of S, but not of S_0. Choose an equation $u = v$ from S such that $h(u) \neq h(v)$, and moreover, (u,v) is minimal in the sense that there is no such equation in S which would have a shorter computation in τ than what is the shortest one for (u,v).

By the choice of S_0, words u and v factorize as $u = u_1u_2u_3u_4$ and $v = v_1v_2v_3v_4$ such that in τ we have

$$i \xrightarrow{u_1,v_1} q \xrightarrow{u_2,v_2} q \xrightarrow{u_3,v_3} q \xrightarrow{u_4,v_4} t$$

for some states i, q and t, with i initial and t final. It follows from the minimality of (u,v) that

$$(5) \qquad \begin{cases} h(u_1u_2) &= h(v_1v_2), \\ h(u_1u_2u_4) &= h(v_1v_2v_4) \text{ and} \\ h(u_1u_3u_4) &= h(v_1v_3v_4). \end{cases}$$

We apply to these identities the following implication on words, the proof of which is straightforward and left to the reader: for any words u, v, w, z, u', v', w', $z' \in \Sigma^*$ we have

$$(6) \qquad \begin{cases} uv &= u'v' \\ uwv &= u'w'v' \\ uzv &= u'z'v' \end{cases} \Rightarrow uwzv = u'w'z'v'.$$

Now, conditions (5) and (6) imply that $h(u) = h(v)$, a contradiction. □

We note that although our above proof does not imply that S_0 can be chosen "small", a more elaborated proof in [KRJ] shows that it can be chosen to be of the size $\mathcal{O}(n)$, where n denotes the number of transitions in τ.

Possibilities of generalizing the fundamental compactness property of Theorem 7.5 are considered in [HKP], cf. also [HK2].

7.7 The size of an equivalent subsystem

Theorem 7.5 leaves it open how large a smallest equivalent subsystem for a given system can be. This is the problem we consider here. Consequently, this section is closely connected to Section 4.4.

Recall that a system S in unknowns Ξ is *independent* if it is not equivalent to any of its proper subsystems. Our problem is to estimate the maximal size of an independent system of equations. Very little seems to be known on this problem. Indeed, we do no know whether the maximal size can be bounded by any function on $||\Xi||$.

What we can report here are some nontrivial lower bounds achieved in [KaPl1]. First we note that Example 4.8 introduces an independent system of equations over a *free semigroup* Σ^+ consisting of n^3 equations in $3n$ unknowns. Therefore a lower bound for the maximal size of independent system of equations over a free semigroup is $\Omega(||\Xi||^3)$.

Our next example shows that we can do better in a *free monoid.*

Example 7.6. Let $\Xi = \{y_i, x_i, u_i, v_i, \bar{y}_i, \bar{x}_i, \bar{u}_i, \bar{v}_i, \tilde{y}_i, \tilde{x}_i, \tilde{u}_i, \tilde{v}_i \mid i = 1, \ldots, n\}$ and S a system consisting of the following equations

$$S : y_i x_j u_k v_l \bar{x}_j \bar{u}_k \bar{v}_l \tilde{x}_j \tilde{u}_k \tilde{v}_l = x_j u_k v_l \bar{x}_j \bar{u}_k \bar{v}_l \tilde{x}_j \tilde{u}_k \tilde{v}_l y_i \text{ for } i, j, k, l = 1, \ldots, n.$$

Therefore $||\Xi|| = 12n$ and $||S|| = n^4$. Let us fix the values i, j, k and l and denote the corresponding equation by $e(i, j, k, l)$. In order to prove that S is independent we have to construct a solution of the system $S - \{e(i, j, k, l)\}$ which is not a solution of $e(i, j, k, l)$. Such a solution is given as follows:

$$\begin{cases} y_i & = ababa, \\ x_j & = u_k = v_l = ab, \\ \bar{x}_j & = \bar{u}_k = \bar{v}_l = a, \\ \tilde{x}_j & = \tilde{u}_k = \tilde{v}_l = ba, \\ z & = 1, \text{ for all other unknowns.} \end{cases}$$

This is not a solution of the equation $e(i, j, k, l)$, since

$$ababa.ab \ldots \neq ab.ab.ab \ldots \quad .$$

However, it is a solution of any other equation since the alternatives are

$y_i \neq ababa$, when the equations become an identity, or
$y_i = ababa$ and $0, 1$ or 2 of the words x_j, u_k and $v_l \neq ab$, when the corresponding relations are:

$$ababa = ababa,$$
$$ababa.ab.a.ba = ab.a.ba.ababa,$$
$$ababa.ab.ab.a.ba.ba = ab.ab.a.a.ba.ba.ababa.$$

Finally, we emphasize that this example uses heavily the empty word 1. □

We summarize the above considerations to

Theorem 7.7. *(i) A system of equations with n unknowns may contain $\Omega(n^4)$ independent equations over a free monoid.*

(ii) A system of equations with n unknowns may contain $\Omega(n^3)$ independent equations over a free semigroup without the unit element.

A natural problem arises.

Problem 7.1. Does there exist an independent system of equations over a free semigroup or a free monoid consisting of exponentially many equations with respect to the number of unknowns?

We note that if the above question is posed in free groups then the answer is affirmative, although our compactness property is still valid, cf. [HK2]. Even more strongly, in [AL2] it is shown that systems of independent equations in three unknowns over a free group may be unboundedly large.

7.8 A finiteness condition for finite sets of words

In this section we interpret the above compactness result in terms of orderings. We consider relation quasi-ordering $\preceq_r$ defined on finite set of words by the condition

$$X \preceq_r Y \Leftrightarrow \exists \text{ bijection } \varphi : X \to Y \text{ such that } R_{\varphi(X)} \subseteq R_X.$$

Consequently, a finite set X is here considered as a solution of a system of equations, and Y is larger than X if X satisfies all equations Y does.

Now, we obtain as a direct interpretation of Theorem 7.5 our second nontrivial finiteness condition of Table 7.1 in Section 7.1.

Theorem 7.8. *Each chain with respect to relation ordering $\preceq_r$ is finite.* □

Note that Theorem 7.8 states that $\preceq_r$ is well-founded, and moreover, that also the reverse of $\preceq_r$ is well-founded. We also want to emphasize that our two nontrivial finiteness conditions, namely those stated in Theorems 7.2 and 7.8, are different in the sense that in Theorem 7.2 arbitrarily large, although always finite, antichains are known to exist, while it is not known whether there exist arbitrary large chains with respect to $\preceq_r$.

Two natural questions connected to the ordering $\preceq_r$ are to decide, for two given finite sets X and Y of the same cardinality, whether $X \preceq_r Y$ or whether $X = Y$ with respect to $\preceq_r$. These problems have very natural interpretations in terms of questions considered in Section 3.1. The latter asks whether F-semigroups X^+ and Y^+ are *isomorphic*, and the former asks (essentially) whether an F-semigroup can be *strongly embedded* into another one, i.e., whether there exists an injective morphism mapping generators to generators. Recall, as we showed in Section 3.1, that an F-semigroup X^+ can always be embedded into any Y^+ containing two words which do not commute.

As the answer to the above questions we prove, cf. [HK1].

Theorem 7.9. *Given two finite sets $X, Y \subseteq \Sigma^+$ it is decidable whether the F-semigroups X^+ and Y^+ are isomorphic.*

Proof. We may assume that $||X|| = ||Y||$, and restrict our considerations to a fixed bijection $\varphi : X \to Y$. We have to decide whether the extension of $\varphi : X^+ \to Y^+$ is an isomorphism, i.e., whether X and $\varphi(X)$ satisfies exactly the same relations. Let the sets of these relations be R_X and $R_{\varphi(X)}$ having a common set Ξ of unknowns, respectively.

It is an easy exercise to conclude that R_X and $R_{\varphi(X)}$ are rational relations, cf. constructions in Example 2.1. Now, deciding whether $R_X = R_{\varphi(X)}$ would solve our problem, but unfortunately the equivalence problem for rational relations is undecidable, cf. [Be1]. So we have to use some other method. Such a method can be found, when noticing that we are asking considerably less than whether R_X and $R_{\varphi(X)}$ are equal, namely we are asking only whether $Y = \varphi(X)$ satisfies R_X, and vice versa. To test this is not trivial, but by Theorem 7.5 it reduces to testing whether Y satisfies a finite subsystem of R_X, and moreover, by Theorem 7.6 such a finite subsystem can be found effectively. Hence, indeed, we have a method to test whether X^+ and Y^+ are isomorphic. □

Note that the proof of Theorem 7.9 does not need the full power of Theorem 7.5. Only its effective validity for rational systems is needed, and this was easy to prove by direct combinatorial arguments.

Theorem 7.9 and its proof have the following two interesting consequences:

Theorem 7.10. *Given finite sets $X, Y \subseteq \Sigma^+$ it is decidable whether the F-semigroup X^+ is strongly embeddable into the F-semigroup Y^+.* □

Theorem 7.11. *For finite sets $X, Y \subseteq \Sigma^+$ it is decidable whether (i) $X \preceq_r Y$ or (ii) $X = Y$ with respect to $\preceq_r$.* □

The proof of Theorem 7.9 is not difficult, however, it contains quite a surprising feature: it does not seem to be extendable to rational subset of Σ^+. This is interesting to note since for many problems finite and rational sets behave in a similar way – due to the fact that rational sets are finite via their syntactic monoids. For instance, in a special case of the above isomorphism problem asking only whether a given F-semigroup X^+ is free, there is no essential difference whether X is finite or rational, cf. [BePe]. In the general isomorphism problem it is not only so that the method of Theorem 7.9 does not seem to work, but we have an open problem:

Problem 7.2. Is it decidable whether two rational subsets of Σ^+ generate isomorphic semigroups?

We conclude this section by considering how equations can be used to describe subsemigroups of Σ^+. These considerations are connected to the validity of Ehrenfeucht's Conjecture.

Let Σ be a fixed finite alphabet and Ξ a denumerable set of unknowns. We say that a system S of equations, with a finite number of unknowns from Ξ, *F-presents* an F-semigroup X^+ if, and only if, the following holds

(i) X is a solution of S; and
(ii) S is equivalent to R_X.

Intuitively this means that X satisfies the equations of S, but nothing else in the sense that any other equation e satisfied by X is dependent on S, i.e., S and $S \cup \{e\}$ are equivalent.

Example 7.7. Consider the following three singleton sets of equations

$$S_1 : xy = zx \ ; \ S_2 : xy = yx \ ; \ S_3 : xyy = yxxx \ .$$

The first one is an F-presentation of X_1^+ with $X_1 = \{a, ba, ab\}$, for example. Indeed, denoting these words by x, y and z in this order, we see that the minimal nontrivial relations of X_1 are $xy^n = z^n x$ for $n \geq 1$. But this set of nontrivial relations is equivalent to the equation $xy = zx$:

$$xy^n = xyy^{n-1} = zxy^{n-1} = zz^{n-1}y = z^n y.$$

On the other hand, S_2 is not an F-presentation. Indeed, assume that $X = \{x, y\}$ satisfies S_2. Then there is a word $z \in \Sigma^+$ and integers p and q such that $x = y^p$ and $y = z^q$. The cases $p = 0$ or $q = 0$ are easy to rule out. In the remaining case R_X is equivalent to the equation $x^q = y^p$, which is not equivalent to S_2. Finally, the above argumentation shows that S_3 is an F-presentation of the semigroup $\{a, aa\}^+$. □

We did not require in the definition of an F-presentation that the set S is neither finite nor independent. However, such an F-presentation can always been found for any finitely generated F-semigroup.

Theorem 7.12. *For each finite $X \subseteq \Sigma^+$ the F-semigroup X^+ has a finite F-presentation consisting of an independent set of equations. Moreover, such an F-presentation can be found effectively.*

Proof. It is the proof of Theorem 7.6 which allows us to find a finite F-presentation S for X^+. It follows trivially that some of the equivalent subsets of S is independent, and hence a required F-presentation. To find it effectively we proceed as follows. By employing Makanin's algorithm we can test whether two finite systems of equations are equivalent, cf. Section 5. Hence a required F-presentation can be found by an exhaustive search. □

The problem of characterizing those systems of equations which are F-presentations seems to be so far a neglected research area. Our Example 7.7 shows that not all finite systems are F-presentations. As a related question we state

Problem 7.3. Is it decidable whether a given finite system of equations is an F-presentation?

We note that Problem 7.3 is semidecidable, i.e., if we know that a given finite S is an F-presentation, then an F-semigroup X^+ having S as an F-presentation can be effectively found. This follows by an exhaustive search and arguments presented in the proof of Theorem 7.12.

8. Avoidability

The goal of this section is to give a brief survey on most important results of the theory of avoidable words, or as its special case of repetition-free words. A typical question of this theory asks: does there exist an infinite word over a given finite alphabet which avoids a certain pattern (repetition, resp.), that is does not contain as a factor any word of the form of the pattern (any repetition of that order, resp.). If the pattern is xx all squares must be avoided. It should be clear that, contrary to many other fragments of formal language theory, results of this theory depend on the size of the alphabet.

8.1 Prelude

The theory of avoidable words is among the oldest in formal language theory. A systematic study was carried out by A. Thue at the beginning of this century, see [T1], [T2], [Be6] and [Be8] for a survey of Thue's work. Later these problems have been encountered several times in different connections, and many important results, including most of Thue's original ones, have been discovered or rediscovered, cf. Chapter 3 in [Lo]. The topic has been under a very active research since early 80's, and it is no doubt that this revival is due to a few important papers, such as [BEM], and papers emphasizing a close connection of this theory to the theory of fixed points of iterated morphisms, cf. [Be2] and [CS].

Some basic results of the theory have already been published in details in books like [Lo] and [Sal2]. For survey papers we refer to [Be4] and [Be5]. Finally, applications of the theory especially to algebra, are discussed in [Sap].

To start with our presentation we recall that the basic notions were already defined in Section 2.3. The theory, at its present form, is closely related to an iteration of a morphism $h : \Sigma^* \to \Sigma^*$. For convenience we consider only 1-*free prolongable* morphisms, i.e., 1-free morphisms h satisfying $h(a) = a\alpha$ for some $a \in \Sigma$ and $\alpha \in \Sigma^+$. Then obviously, for each i, $h^{i+1}(a)$ is a proper prefix of $h^i(a)$, so that the unique word

$$w_h = \lim_{i\to\infty} h^i(a)$$

is obtained. Consequently, w_h is a *fixed point* of h, i.e., $h(w_h) = w_h$. Since it is defined by iterating morphism h (at point a) we say that w_h is obtained as a fixed point of *iterated morphism* h. This mechanism, often generalized by a possibility of mapping w_h by another morphism, is by far the most commonly used method to construct avoidable infinite words.

As an illustration let us consider morphisms

$$T : \begin{cases} a \to ab \\ b \to ba \end{cases} \quad \text{and} \quad F : \begin{cases} a \to ab \\ b \to a \end{cases} .$$

The words they define as iterated morphisms at a are

$$w_T = abbabaabbaababbabaababbaabbabaab \ldots$$

and

$$w_F = abaababaabaababaababaabaababa \ldots$$

The first one played an important role in the considerations of Thue, and later it was made well-known by Morse, cf. [Mor1] and [Mor2]. Therefore it is usually referred to as *Thue–Morse word*, although it was actually considered by Prouhet already in 1851, cf. [Pr]. The latter one is normally referred to as *Fibonacci word*, due to the fact that the lengths of the words $F^i(a)$ form the famous Fibonacci sequence. Accordingly, the morphisms T and F are called *Thue–Morse* and *Fibonacci morphisms.*

It is striking to note that these two words are among the most simple ones obtained by iterated morphisms, and still they have endless number of interesting combinatorial properties. In fact they seem to be the most commonly used counterexamples. For instance, prefixes of w_T of length 2^n show that factors of a word w of length n with multiplicities do not determine w uniquely, cf. Section 7.3. Similarly, w_F can be used to show that Proposition 6.2 is optimal, as well as that prefixes of lengths $p+2$, $q+2$, with p and q consecutive Fibonacci numbers, can be used to show the optimality of the Theorem of Fife and Wilf.

As an illustration of another way of defining repetition-free words we note that w_T can be defined recursively by formulas

$$\begin{cases} u_0 = a, \\ v_0 = b, \end{cases} \qquad \begin{cases} u_{n+1} = u_n v_n & \text{for } n \geq 0, \\ v_{v+1} = v_n u_n & \text{for } n \geq 0, \end{cases}$$

since then $T^n(a) = u_n$, as is easy to verify.

8.2 The basic techniques

The following two examples illustrate the basic techniques of proving that a fixed point of an iterated morphism avoids a certain pattern or a certain type of a repetition. In principal, the techniques is very simple, namely that of the infinite descending already used by Fermat, but its implementation might lead to a terrifying case analysis.

Example 8.1. We claim that the fixed point w_h of the iterated morphism

$$h: \begin{array}{l} a \rightarrow aba \\ b \rightarrow abb \end{array}$$

is 3^--free, in other words, does not contain any cube, but does contain repetitions of any order smaller than 3. The latter statement is trivial since any word of the form

$$uuu(\mathrm{suf}_1(u))^{-1}$$

is mapped under h to a word of the same form, and as the starting point w_h contains a factor aab.

To prove the second sentence, assume that w_h contains a cube $v = uuu$, with $|u| = n \geq 2$. Now we consider the four cases depending on the prefix u_2 of u of length 2, and analyse the cutpoints in $\{h(a), h(b)\}$-interpretations of u. It is due to a favourable form of h that, with the exception of the prefix ba, such a cutpoint in u_2 is unique, as depicted in Figure 8.1.

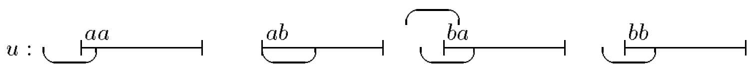


Figure 8.1. Cutpoints inside u_2

In the cases aa, ab and bb the three prefixes in different occurrences of u have exactly the same cutpoints. Consequently, in the case of ab there exists a word u' such that $h(u') = u$, and in the other two cases there exists a word u' such that $h(u') = \mathrm{suf}_k(u)u\,\mathrm{suf}_k(u)^{-1}$, for $k = 1$ or 2, i.e., $h(u')$ is obtained from u by a shift as illustrated in Figure 8.2 for the prefix aa.

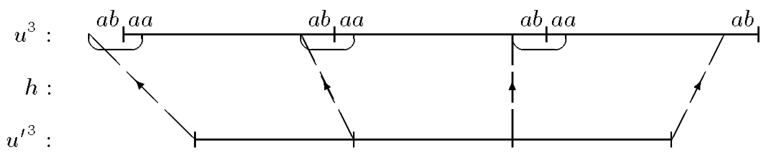


Figure 8.2. The case $u_2 = aa$

In the case ba is the prefix of u, if the ba prefixes of the first and the second u have the same cutpoint, so have the third one as well, by the length argument. Hence, the above considerations apply. On the other hand, if the first and the second prefix have a different cutpoint, then the third one has, again by the length argument, still a different one. This, however, is not possible.

From above we conclude that, if w_h contains a cube longer than 6, then it contains also a shorter cube, and hence inductively a cube of length at most 6. That this is not the case is trivial to check. □

Our second example deals with abelian repetitions, and is due to [Dek]. The basic idea of the proof is as above, only the details are more tedious.

Example 8.2. Let w_h be the word defined by the iterated morphism

$$h : \begin{array}{l} a \to abb \\ b \to aaab. \end{array}$$

We intend to show that w_h is abelian 4-free, i.e., does not contain 4 consecutive commutatively equivalent factors. The idea of the proof is that illustrated in Figure 8.2. Starting from an abelian 4-repetition, we conclude that its small modification by a shift is an image under h of a shorter abelian 4-repetition. Now, the 4 consecutive blocks are only commutatively equivalent, so that it is not clear how to do the shifting. This means that h must possess some strong additional properties. To formalize these we associate with a word $u \in \{a,b\}^*$ a *value* in the group $\mathbb{Z}_5$ (of integers modulo 5) by a morphism $\mu : \{a,b\}^* \to \mathbb{Z}_5$ defined as

$$\mu(a) = 1 \quad \text{and} \quad \mu(b) = 2.$$

It follows that

(i) $$\mu(h(w)) = 0 \quad \text{for all} \quad w \in \{a,b\}^*.$$

Now assume that $B_1B_2B_3B_4$ is an abelian 4-repetition in w_h. We illustrate this, as well as an $\{h(a), h(b)\}$-interpretation of it in Figure 8.3.

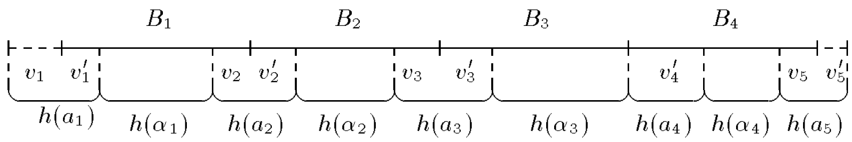


Figure 8.3. $\{h(a), h(b)\}$-interpretation of $B_1B_2B_3B_4$

Formally, the above means that

$$h(a_1\alpha_1 \ldots \alpha_4 a_5) = v_1 B_1 B_2 B_3 B_4 v_5' \quad \text{with} \quad a_i \in \Sigma,\ \alpha_i \in \Sigma^*,$$

where, for $i = 1, \ldots, 5$ and $j = 1, \ldots, 4$,

$$h(a_i) = v_i v_i' \quad \text{and} \quad B_j = v_j' h(\alpha_j) v_{j+1} \quad \text{with} \quad v_i \in \Sigma^*,\ v_i' \in \Sigma^+.$$

Since μ is a morphism we obtain from (i) that, for $j = 1, \ldots, 4$,

$$\begin{aligned}\mu(v_{j+1}) &= \mu(B_j) - \mu(h(\alpha_j)) - \mu(v'_j)\\ &= \mu(B_j) + \mu(v_j) = g + \mu(v_j),\end{aligned}$$

where g, due to the commutative equivalence of B_j's, denotes a constant element of $\mathbb{Z}_5$. Therefore the sequence

(1) $$\mu(v_1), \mu(v_2), \mu(v_3), \mu(v_4), \mu(v_5)$$

is an arithmetic progression in $\mathbb{Z}_5$. We want to allow only trivial arithmetic progressions, which guides us to require that

(ii) $$S = \{a \in \mathbb{Z}_5 \mid \exists z \in \mathrm{pref}\{h(a), h(b)\} : a = \mu(z)\}$$

is 5-*progression free*, i.e., does not contain any subset $\{a + ng | n = 0, \ldots, 4\}$ with $g \neq 0$. That our morphism h satisfies this condition is easy to see: indeed, we have

(2) $$(\mu(a), \mu(ab)) = (1, 3) \ \text{ and } \ (\mu(a), \mu(aa), \mu(aaa)) = (1, 2, 3),$$

so that $S = \{0, 1, 2, 3\}$, while in $\mathbb{Z}_5$ any arithmetic progression of length 5, with $g \neq 0$, equals the whole $\mathbb{Z}_5$.

Since v_i's in (1) are prefixes of $h(a)$ or $h(b)$ we can write the arithmetic progression (1) in the form

(3) $$\mu(v_1) = \mu(v_2) = \mu(v_3) = \mu(v_4) = \mu(v_5).$$

What we would need, in order to have a shift, is that from (3) we could conclude that either the words v_i or the words v'_i are equal. This is our next condition imposed for h and μ. We say that μ is *h-injective*, if for all factorizations $v_i v'_i \in \{h(a), h(b)\}$, with $i = 1, \ldots, 5$, we have

(iii) $$\mu(v_1) = \cdots = \mu(v_5) \Rightarrow v_1 = \cdots = v_5 \ \text{ or } \ v'_1 = \cdots = v'_5.$$

From our computations in (2) we see that the only case to be checked is the case when $v_1 = ab$ and $v_2 = aaa$. And then indeed $v'_1 = b = v'_2$, so that our μ is h-injective.

We are almost finished. We know that the words v_i (or symmetrically the words v'_i) coincide. Consequently, the four abelian repetitions can be shifted to match with the morphism h: instead of B_i's we now consider the commutatively equivalent blocks $D_i = v_i B_i v_i^{-1}$ (or $D_i = v_i'^{-1} B_i v'_i$), for $i = 1, \ldots, 4$. Then there are words C_i such that

(4) $$h(C_i) = D_i \quad \text{with} \quad \pi(D_i) = \pi(D_j) \quad \text{for} \quad i, j = 1, \ldots, 4\ ,$$

where π gives the commutative image of a word. If we would know that C_i's were commutatively equivalent, we would be done. Indeed, then by an inductive argumentation w_h would contain either $aaaa$ or $bbbb$ as a factor, and this is clearly not the case.

So to complete the proof we still impose one requirement for h, namely that

$$(iv) \qquad M(h) = \begin{pmatrix} |h(a)|_a & |h(a)|_b \\ |h(b)|_a & |h(b)|_b \end{pmatrix} \quad \text{is } \textit{invertible}.$$

Then, by (4), we would have $\pi(C_i) \cdot M(h) = \pi(D_i)$, or equivalently $\pi(C_i) = \pi(D_i) \cdot M(h)^{-1}$, for $i = 1, \ldots, 4$, so that C_i's would be commutatively equivalent. That $M(h)$ is indeed invertible is clear, since it equals to $\begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix}$. □

It is worth noticing that conditions (i)–(iv) in the above proof are general ones, which can be used to prove similar results for different values of the size of the alphabet and/or the order of the repetition.

The argumentation of Examples 8.1 and 8.2 was already used by Thue in order to conclude

Theorem 8.1. *The Thue-Morse word w_T is 2^+-free, i.e., does not contain any overlapping factors.* □

When applied to the Fibonacci word w_F, the above argumentation, with rather difficult considerations, yields the result that it is $(2+\varphi)^-$-free, where φ is the number of the golden ratio, i.e., $\frac{1}{2}(1+\sqrt{5})$, cf. [MP].

From Theorem 8.1 we easily obtain

Theorem 8.2. *There exists a 2-free infinite word in the ternary alphabet.*

Proof. Define the morphism $\varphi : \{a, b, c\}^* \to \{a, b\}^*$ by setting $\varphi(a) = abb$, $\varphi(b) = ab$ and $\varphi(c) = a$. Since φ has a bounded delay, the word $\varphi^{-1}(w)$ for $w \in \{a, b\}^\omega$, if defined, is unique, and since it is defined for each w containing no three consecutive b's, it follows that the word

$$(5) \qquad w_2 = \varphi^{-1}(w_T) = abcacbabcbacabca \ldots$$

is well-defined. Moreover, it is 2-free since w_T is 2^+-free, and each of the words $\varphi(d)$, with $d \in \{a, b, c\}$, starts with a. □

The word w_2 can be obtained also as the fixed point of the iterated morphism h defined as $h(a) = abc$, $h(b) = ac$ and $h(c) = b$.

For the sake of completeness we state the result of Example 8.2, and its modification for abelian 3-free words in the ternary alphabet, also due to [Dek], as the following theorem.

Theorem 8.3. *(i) There exists an infinite abelian 4-free word in the binary alphabet.*

(ii) There exists an infinite abelian 3-free word in the ternary alphabet.

□

Table 8.1. Lengths of maximal words avoiding integer repetitions and abelian repetitions

$\|\|\Sigma\|\|$ \ k	2	3	...
2	3	∞	
3	∞		
⋮			

Word case

$\|\|\Sigma\|\|$ \ k	2	3	4	...
2	3	7	∞	
3	9	∞		
4	∞			
⋮				

Abelian case

Now, with the calculations in Section 2.3 we can summarize all avoidable integer repetitions and abelian repetitions to the following table. Here k tells the order of the repetition, and the value of each entry the length of the longest word avoiding this repetition in the considered alphabet.

We note that special cases of (ii) in Theorem 8.3 was solved earlier. The first step was taken in [Ev], where it was shown that the 25th abelian powers were avoidable in the binary case. This was improved to 5 in [Pl] using an iterated *uniform morphism h of size* 15, i.e., $|h(a)| = 15$ for each letter a. Later the same result was shown in [Ju] using uniform morphisms of size 5.

Finally, the problem whether abelian squares can be avoided in the 4-letter alphabet, sometimes referred to as Erdös' Problem, was open for a long time, until it was solved affirmatively in [Ke2]. The proof is an interesting combination of a computer checking and of a mathematical reasoning showing that an abelian 4-free word can be obtained as the fixed point of an iterated uniform morphism of size 85. Moreover, it is shown that no smaller uniform morphism works here!

By Table 8.1, all 2-free words in the binary alphabet are finite, while by Theorem 8.1, there exists an infinite 2^+-free binary word. This motivates us to state the following notion explicitly defined in [Bra], cf. also [Dej]. For each $n \geq 2$, the *repetitiveness treshold* in the alphabet of n letters is the number $T(n)$ satisfying:

(i) there exists a $T(n)^+$-free infinite word in the n-letter alphabet; and
(ii) each $T(n)$-free word in the n-letter alphabet is finite.

It follows from the fact that for any irrational number r, the notions of r-free and r^+-free coincide, that the repetitiveness treshold is always rational, if it exists. And it is known to exist for $n \leq 11$: As we noted $T(2) = 2$. The value of $T(3)$ was solved in [Dej], by showing that each ternary $\frac{7}{4}$-free word is finite, and by constructing an infinite $\frac{7}{4}^+$-free ternary word as the fixed point of a uniform morphism of size 19. She also conjectured the values of $T(n)$ correctly up to the current knowledge, which is shown in Table 8.2. For 4 the problem was solved in [Pan1] and for the values from 5 up to 11 in [Mou].

Table 8.2. The repetitiveness tresholds and the lengths $\max(n)$ of longest $T(n)$-free words in the n-letter alphabet

$\lVert\Sigma\rVert$	2	3	4	5	6	7	8	9	10	11
$T(n)$	2	7/4	7/5	5/4	6/5	7/6	8/7	9/8	10/9	11/10
$\max(n)$	3	38	122	6	7	8	9	10	11	12

It is interesting to note that, for all $k \geq 2$, only very short words can avoid repetitions of order $\frac{k}{k-1}$. Indeed, any word of length $k+2$ in the k-letter alphabet Σ_k either contains a factor of length k in a $(k-1)$-letter alphabet or an image of the word $1 \ldots k12$ under a permutation of Σ_k. Consequently, such a word contains either a repetition of order at least $\frac{k}{k-1}$ or $\frac{k+2}{k}$, and both of these are at least $\frac{k}{k-1}$, for $k \geq 2$. Consequently, assuming the first line of Table 8.2 the second one follows for k at least 5 by noting that words $1 \ldots k1$ are $\frac{k}{k-1}$-free.

8.3 Repetition-free morphisms

As we have seen, constructions of repetition-free words rely typically on iterated morphisms, which preserve this property when started from a letter a, or in general, from a word having this property. This guides us to state the following definition. A morphism $h : \Sigma^* \to \Delta^*$ is said to be *k-free* if it satisfies:

$$\text{whenever } \; w \in \Sigma^+ \; \text{ is } k\text{–free, so is } \; h(w).$$

Note that the definition of the k-freeness does not require k to be a number – it can also be α^+ or α^- for some number α. For example, the Thue-Morse morphism is 2^+-free. Similarly, a morphism can be abelian k-free, for an integer k, as in Example 8.2.

The problem of deciding, for a given k and a morphism h, whether h is k-free is very difficult. Indeed, even for integer values of k it seems to be still open, cf. [Ke1] for partial solutions. On the other hand, computationally feasible sufficient conditions for the k-freeness, with $k \in \mathbb{N}$, are known, an example being the following result from [BEM].

Proposition 8.1. *Let k be an integer ≥ 2. A morphism $h : \Sigma^+ \to \Delta^+$ is k-free if it satisfies the following conditions*

(i) h *is* k*-free on* k*-free words of length at most* $k+1$*;*
(ii) *whenever* $h(a) \in F(h(b))$*, with* $a, b \in \Sigma$*, then* $a = b$*; and*
(iii) *whenever* $h(b)h(c) = uh(a)v$*, with* $a, b, c \in \Sigma$*, then* $u = 1$ *and* $a = b$*, or* $v = 1$ *and* $a = c$*.*

The first complete characterization of 2-free morphisms was achieved in [Be3]. Later in [Cr] it was extended to the following sharp form, where $M(h)$ and $m(h)$ denote the maximal and minimal lengths of $h(a)$, when a ranges over the domain alphabet of h.

Proposition 8.2. *(i) A morphism* $h : \Sigma^+ \to \Delta^+$ *is 2-free if, and only if, it is 2-free on 2-free words of length at most* $\max\{3, (M(h)-3)/m(h)\}$*.*

(ii) A morphism $h : \{a,b,c\}^+ \to \{a,b,c\}^+$ *is 2-free if, and only if, it is 2-free on 2-free words of length at most 5.*

A characterization similar to (ii) – requiring to check words up to length 10 – was shown for 3-free morphism over the binary alphabet in [Ka1]. Note here that not only the decidability of the k-freeness of a morphism, in general, but also the decidability of the 3-freeness in the arbitrary alphabet seems to be open.

We conclude these considerations with two more sharp characterization results. The first one was already known to Thue, cf. also [Harj]. The second one, due to [LeC], considers the problem whether a given morphism $h : \Sigma^+ \to \Delta^+$ is k-free, for all integer values of $k \geq 2$, in other words is *power-free.*

Proposition 8.3. *A binary morphism* $h : \{a,b\}^+ \to \{a,b\}^+$ *is* 2^+*-free if, and only if, it is of the form* T^k *or* $T^k \circ \mu$*, where* T *is the Thue-Morse morphism,* μ *is the permutation of* $\{a,b\}$ *and* k *is an integer* ≥ 1*.*

Proposition 8.4. *A morphism* $h : \Sigma^+ \to \Delta^+$ *is power-free if, and only if,*

(i) h *is 2-free; and*
(ii) $h(a^2)$ *is 3-free for each* $a \in \Sigma$*.*

8.4 The number of repetition-free words

In this subsection we study the number of repetition-free words in some special cases. More precisely we consider 2^+- and 3-free words in the binary case and 2-free words in the ternary case. Let us denote by $SF_n(3)$ the set of all 2-free words of length n over the ternary alphabet, where n is allowed to be ∞, as well. Similarly, let $S^+F_n(2)$ and $CF_n(2)$ denote the corresponding sets of 2^+- and 3-free words over the binary alphabet.

We shall show the following result of [Bra], cf. also [Bri].

Theorem 8.4. $\|SF_n(2)\|$ *is exponential, i.e., there exist constants* A*,* B*,* ρ *and* σ*, with* $A, B > 0$ *and* $\rho, \sigma > 1$*, such that*

$$A\rho^n \leq \|SF_n(3)\| \leq B\sigma^n \quad \textit{for all} \quad n.$$

Proof. The existence of B and σ are clear. The crucial point in proving the lower bound is to find a 2-free morphism $h : \Sigma^+ \to \{a,b,c\}^+$, with $||\Sigma|| > 3$. As shown in [Bra] such a morphism exists for each value of $||\Sigma||$, and moreover, can be chosen uniform. For small values of $||\Sigma||$ it is not difficult to find such a morphism using Proposition 8.2.

Now, let $h : \{a,b,c,\bar{a},\bar{b},\bar{c}\}^+ \to \{a,b,c\}^+$ be a uniform 2-free morphism. As shown in [Bra] the smallest size of such a morphism is 22, which means that after having it, the checking of its 2-freeness is computationally easy. Next we define a finite substitution $\tau : \{a,b,c\}^+ \to \{a,b,c,\bar{a},\bar{b},\bar{c}\}^+$ by setting

$$\tau(x) = \{x, \bar{x}\} \quad \text{for} \quad x \in \{a,b,c\}.$$

We fix a 2-free word w_k of length k, which by Theorem 8.2 exists, and consider the set $h(\tau(w_k))$ of words. Clearly, words in this set are 2-free, and of length $22k$. Moreover, $||h(\tau(w_k))||$ contains 2^k words, since h must be injective, or even a prefix code, by its 2-freeness. So we have concluded that, for each $n \geq 2$, the cardinality of $SF_{22n}(3) \geq 2^n$. This implies that ρ can be chosen to be $2^{\frac{1}{22}} \sim 1,032$. □

Theorem 8.4 stimulates for a few comments. First of all, a closer analysis of the problem shows that the constants can be chosen such that

$$6 \cdot 1,032^n \leq ||SF_n(3)|| \leq 6 \cdot 1,38^n.$$

Moreover, the 20 smallest values of the number of 2-free words of length n over $\{a,b,c\}$ are: 3, 6, 12, 18, 30, 42, 60, 78, 108, 144, 204, 264, 342, 456, 618, 798, 1044, 1392, 1830, 2388.

Second, the above proof immediately extends to infinite words. Starting from a fixed infinite 2-free word over the ternary alphabet Σ_3, say w_2 of Theorem 8.2, τ creates nondenumerably many of those over a six-letter alphabet Σ_6, and h being injective also on Σ_6^ω brings equally many back to Σ_3^ω. So we have

Theorem 8.5. *$SF_\infty(3)$ is nondenumerable.*

Finally, the above ideas can be applied to estimate the number of 3-free words over the binary alphabet Σ_2, if a uniform 3-free morphism $h : \Sigma^+ \to \Sigma_2^+$, with $||\Sigma|| > 2$, is found. Again, as shown in [Bra], such morphisms exist for each value of $||\Sigma|| > 2$. Therefore, since the uniformity and the 3-freeness imply a bounded delay, and hence the injectivity on Σ^ω, we obtain

Theorem 8.6. *$CF_n(2)$ is exponential, and $CF_\infty(2)$ is nondenumerable.*

The bounds given for the number of 3-free words of length n in the binary case are

$$2 \cdot 1,08^n \leq ||CF_n(2)|| \leq 2 \cdot 1,53^n.$$

For 2^+-free words the results are not quite the same as the above for 2- and 3-free words. The result stated as Proposition 8.5 follows from the characterizations of finite and infinite 2^+-free binary words presented in the next subsection.

Proposition 8.5. *$S^+F_n(2)$ is polynomial, while $S^+F_\infty(2)$ is nondenumerable.*

Recently, it was shown in [Car2], using the morphism of [Ke2], that the number of abelian 2-free words over the 4-letter alphabet grows exponentially, as well as that of abelian 2-free infinite words is nondenumerable. This seems to be the only estimate for the number of abelian repetition-free words. For repetition-free words over partially commutative alphabets we refer to [CF].

At this point the following remarks are in order. As we saw in all the basic cases the sets of repetition-free infinite words form a nondenumerable set. Consequently, "most" of such words cannot be algorithmically defined. In particular, the by far most commonly used method using iterated morphisms can reach only very rare examples of such words. In the case of 2^+-free words the situation is even more striking: as shown in [See] the Thue-Morse word is the only binary 2^+-free word which is the fixed point of an iterated morphism.

8.5 Characterizations of binary 2^+-free words

In this subsection we present structural characterizations of both finite and infinite binary 2^+-free words. These are obtained by analysing how a given 2^+-free word can be extended preserving the 2^+-freeness. In order to be more precise, let us recall that the recursive definition of the Thue-Morse word was based on two sequences $(u_n)_{n\geq 0}$ and $(v_n)_{n\geq 0}$ of words satisfying

$$\begin{array}{lll} u_0 = a & v_0 = b & \\ u_{n+1} = u_n v_n & v_{n+1} = v_n u_n & \text{for } \ n \geq 0. \end{array}$$

Let us call words u_n and v_n *Morse blocks* of order n, and set $U_n = \{u_n, v_n\}$ and $U = \bigcup_{n=1}^{\infty} U_n$. Clearly, the lengths of Morse blocks are powers of 2, and for instance $v_3 = baababba$.

Now, a crucial lemma in the characterizations is the following implication:

$$(1) \qquad uvwx \in S^+F_{3\cdot 2^n+1}(2),\ u,v \in U_n,\ |w| = 2^n,\ x \in \Sigma \Rightarrow w \in U_n.$$

This means that, if a product of two Morse blocks of the same order, can be extended to the right, preserving the 2^+-freeness, by a word which is longer than these blocks, then the extension starts with a Morse block of the same order than the original ones.

The proof of (1) is by induction on n. For $n = 0$ there is nothing to be proved. Further the induction step can be concluded from the illustration of Figure 8.4. Indeed, the possible extensions of length 2^n for $u_{n+1}v_{n+1}$ are, by induction hypothesis, words u_n and v_n, and of the two potential extensions of these of length $|v_n|$ one is ruled out in both the cases, since the word must

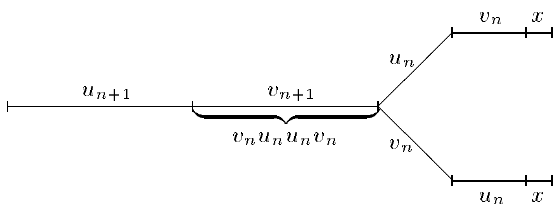


Figure 8.4. The proof of (1) for $u_{n+1}v_{n+1}$

remain 2^+-free. Consequently, for $u_{n+1}v_{n+1}$, the word w is either u_{n+1} or v_{n+1} as claimed. Similarly, one can prove the other cases of the products.

Based on (1), and a bit more detailed analysis, the following characterization is obtained for 2^+-free finite words in [ReSa]: for each 2^+-free word w there exists a constant k such that w can be written *uniquely* in the form

$$(2) \qquad w = l_0 \dots l_{k-1} u r_{k-1} \dots r_0 \ \text{ with } l_i \in L_i,\ r_i \in R_i \ \text{ and } \ u \in \bigcup_{i=1}^{12} U_k^i,$$

where $k = \mathcal{O}(|w|)$ and the sets L_i and R_i, for $i = 0, \dots, k-1$, are of the cardinality 15.

Denoting $n = |w|$ we obtain from (2) that

$$||S^+F_n(2)|| \leq || \bigcup_{i=1}^{12} U_k^i || \cdot 15^{2k} = \mathcal{O}(n^\alpha),$$

for some $\alpha > 0$. Actually, as computed in [ReSa], α can be chosen to be $\log_2 15 < 4$. Hence, the first sentence of Proposition 8.5 holds.

Note that (2) gives only a necessary condition, and hence only a partial characterization, for finite 2^+-free words. Later a more detailed analysis has improved estimates for the number of binary 2^+-free words of length n, cf. [Kob], [Car1], [Cas1] and [Lep2]. The strictest bounds are given in [Lep2], where, as well as in [Car1], a complete characterization of all finite 2^+-free words is achieved:

$$A \cdot n^{1,22} \leq ||S^+F_n(2)|| \leq B \cdot n^{1,37}.$$

On the other hand, in [Cas1] it is shown that the limit

$$\lim_{n \to \infty} \frac{||S^+F_n(2)||}{n^\alpha}$$

does not exist for any α, meaning that the number of 2^+-free binary words of length n behaves irregularly, when n grows.

Now, let us move to a characterization of 1-way infinite binary 2^+-free words. This remarkable result was proved in [F], while our automata-theoretic presentation is from [Be7]. Let us recall that U_n denoted the set of Morse blocks of order n and U the set of all Morse blocks. Further for each binary w let $\bar{w}$ denote its complement, i.e., word obtained from w by interchanging each of its letters to the other. The crucial notion here is the so-called *canonical decomposition* of a word $w \in \Sigma^* U_1$, which is the factorization

$$w = zy\bar{y},$$

where $\bar{y}$ is chosen to be the longest possible $\bar{y}$ in U such that w ends with $y\bar{y}$. Next, three mappings, interpreted as left actions, $\alpha, \beta, \gamma : \Sigma^* U_1 \to \Sigma^* U_1$ are defined based on the canonical decompositions of words:

$$(3) \qquad \begin{cases} w \circ \alpha = zy\bar{y} \circ \alpha = zy\bar{y}yy\bar{y} = wyy\bar{y} \\ w \circ \beta = zy\bar{y} \circ \beta = zy\bar{y}y\bar{y}\bar{y}y = wy\bar{y}\bar{y}y \\ w \circ \gamma = zy\bar{y} \circ \gamma = zy\bar{y}\bar{y}y = w\bar{y}y. \end{cases}$$

We consider

$$A = \{\alpha, \beta, \gamma\}$$

as a ternary alphabet. The mappings α, β and γ extend a word $w = zy\bar{y}$ from the right by words $yy\bar{y}$, $y\bar{y}\bar{y}y$ and $\bar{y}y$, respectively. The use of the canonical decompositions makes these mappings well-defined. It also follows from the fact that w is a proper prefix of $w \circ \delta$, for any $\delta \in A$, that any infinite word $\omega \in A^\omega$ defines a unique word $w \circ \omega \in \Sigma^\omega$. Such an ω is called the *description* of $w \circ \omega$. Of course, the description can be finite, as well.

The mappings α, β and γ are chosen so that, given the canonical description $zy\bar{y}$ of w, they add to the end of w two Morse blocks of the same order as $\bar{y}$ in all possible ways the condition (1) allows this to be done preserving the 2^+-freeness. Actually, in the case of α such a block would be yy, but now also one extra $\bar{y}$ is added, since the next block of this length would be $\bar{y}$ in any case, again by (1). Similarly β adds istead of $y\bar{y}$ the word $y\bar{y}\bar{y}y$.

It follows from these considerations that *each* 1*-way infinite binary* 2^+*-free word* has a description, which moreover, by (3), *is unique.* Which of the descriptions actually define a 2^+-free infinite word is the contents of the characterization we are looking for. In order to state the characterization we set

$$I = \{\alpha, \beta\}(\gamma^2)^*\{\beta\alpha, \gamma\beta, \alpha\gamma\},$$

and consider the following sets of infinite words over A:

$$F = A^\omega - A^* I A^\omega \quad \text{and} \quad G = \beta^{-1} F.$$

Now, we are ready for the characterization known as Fife's Theorem.

Proposition 8.6. *Let* $w \in \Sigma^\omega$.

(i) A word $w \in ab\Sigma^\omega$ *is* 2^+*-free if, and only if, its description is in* F*;*
(ii) A word $w \in aab\Sigma^\omega$ *is* 2^+*-free if, and only if, its description is in* G.

The detailed proof of this result is not very short, cf. e.g. [Be7]. On the other hand, the result provides a very nice example of the usefulness of finite automata in combinatorics. Namely, the set of all descriptions of binary 2^+-free infinite words can be read from the finite automaton of Figure 8.5 accepting any infinite computation it allows.

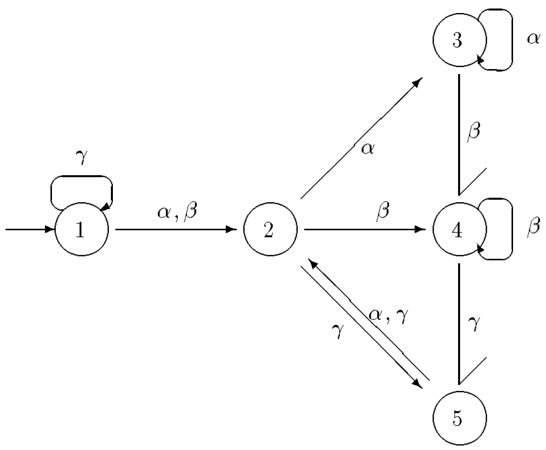


Figure 8.5. Fife's automaton $\mathcal{A}_F$

Now the second sentence of Proposition 8.5 stating that there exist denumerably many infinite 2^+-free words over the binary Σ is obvious. Indeed, the automaton contains two loops in state 3, for example.

We conclude our discussion on 2^+-free words by recalling a characterization of 2-way infinite binary 2^+-free words. This characterization has interesting interpretations in the theory of symbolic dynamics, cf. [MH].

Proposition 8.7. *A two-way infinite binary word w is 2^+-free if, and only if, there exists a two-way infinite word w' such that $w = T(w')$, where T is the Thue-Morse morphism.*

This characterization was already known to Thue, and it is much easier to obtain than that of Proposition 8.6, by using standard tools presented at the beginning of this section.

We note that no characterization of 2-free words – either finite or infinite – over the three letter alphabet is known. Some results in that direction are obtained in [She], [ShSo1] and [ShSo2]. For example it is shown that the set of such infinite words is perfect in the sense of topology implying immediately Theorem 8.5.

8.6 Avoidable patterns

In this last subsection we consider an interesting problem area introduced in [BEM], and also in [Z], namely that of the avoidability of general patterns. We defined this notion already in Section 2.3, and moreover have used it implicitly several times. Indeed, Theorem 8.2 says that the pattern xx is avoidable in the ternary alphabet, i.e., there exists an infinite ternary word having no square as a factor. It is trivially unavoidable in the binary alphabet, while the pattern $xyxyx$, as shown in Theorem 8.1, is avoidable in this alphabet.

It follows, as expected, that the avoidability of a pattern *depends* on the size of the alphabet considered – contrary to many other problems in formal language theory. More precisely, the pattern $P_2 = xx$ separates the binary and ternary alphabets.

It turned out much more difficult to separate other alphabets of different sizes. A pattern separating 3- and 4-letter alphabets was given in [BMT]. The pattern, containing 7 different letters and being of length 14, is as follows:

$$P_3 = ABwBCxCAyBAzAC.$$

It was shown that any word over $\{a, b, c\}$ of length 131293 (which, however, is not the optimal bound) contains a morphic image of P_3 under a 1-free morphism into $\{a, b, c\}^+$ as a factor. On the other hand, the infinite word obtained – again – as the fixed point of a morphism avoids the pattern P_3. Such a morphism is given by $h(a) = ab$, $h(b) = cb$, $h(c) = ad$ and $h(d) = cd$, i.e., can be chosen uniform of size 2.

We summarize the above as follows.

Proposition 8.8. *For each $i = 1, 2, 3$ there exists a pattern P_i which is unavoidable in the i-letter alphabet, but avoidable in the $(i+1)$-letter alphabet.*

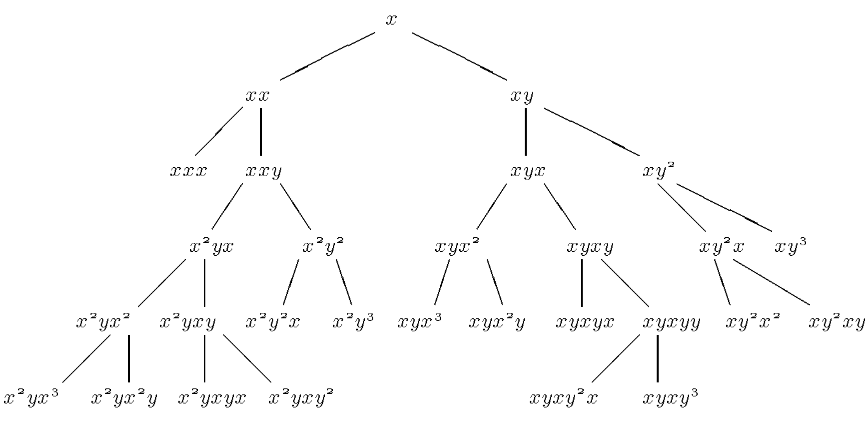


Figure 8.6. Avoidable and unavoidable binary patterns

It is an open question to settle whether Proposition 8.8 extends to larger alphabets.

As we saw, the problem of settling whether a pattern is avoidable in a given alphabet is not easy at all. However, the case where both the pattern and the alphabet are binary, is completely solved. By a *binary* pattern we, of course, mean a pattern consisting of two letters only, say x and y.

The research leading to this interesting result was initiated in [Sc], continued and almost completed in [Rot], and finally completed in [Cas2].

The result is summarized in Figure 8.6. There the labels of the leaves, and hence also any word obtained as their extensions, are avoidable, while those of inside nodes are unavoidable. Note that the tree covers all the words starting with x, and hence up to the renaming all binary patterns, and yields

Proposition 8.9. *Each binary pattern of length at least* 6 *is avoidable in the binary alphabet.*

Each of these avoidable patterns was shown to be so by constructing an infinite word avoiding the pattern as the fixed point of an iterated morphism, or as a morphic image of the fixed point of an iterated morphism. For each unavoidable pattern α let $\max(\alpha)$ be the length of the longest finite binary words avoiding α. The values of $\max(\alpha)$, for all unavoidable patterns omitting symmetrical cases, are listed in Table 8.3.

Table 8.3. Unavoidable patterns and maximal lengths of binary words avoiding those

α :	x	xy	x^2	x^2y	xyx	x^2yx	xy^2x	x^2y^2	$xyxy$	x^2yx^2	x^2yxy
$\max(\alpha)$:	0	1	3	4	4	9	10	11	18	18	38

In accordance with Theorem 8.6 and Proposition 8.5 we note the result of [GV] showing that any avoidable binary pattern is avoided by nondenumerably many infinite words.

9. Subword complexity

In this final section we consider a problem area which has attracted quite a lot of attention in recent years, and which provides a plenty of extremely challenging combinatorial problems. A survey of this topic can be found in [Al].

9.1 Examples and basic properties

Let $w \in \Sigma^\omega$ be an infinite word. We define its *subword complexity*, or briefly *complexity*, as the function $g_w : \mathbb{N} \to \mathbb{N}$ by

$$g_w(n) = \|\{u \in \Sigma^n \mid u \in F(w)\}\|.$$

Consequently, $g_w(n)$ tells the number of different factors of length n in w. A very related notion can be defined for languages (consisting of finite words) instead of infinite words in a natural way.

Two problems are now obvious to be asked:

(i) Given a $w \in \Sigma^\omega$, compute its complexity g_w.
(ii) Given a complexity g, find a word having g as its complexity.

In both of these cases one can work either with the *exact* complexity or with the *asymptotic* complexity, i.e., identifying complexities g and g' if they satisfy $g(n) = \theta(g'(n))$. The above problems are natural to call the *analysis problem* and the *synthesis problem* for complexities of infinite words. Mostly only asymptotic versions of these problems are considered here.

We start with two examples.

Example 9.1. Let $w_K \in \{1,2\}^\omega$ be the famous *Kolakoski word*, cf. [Kol], [Lep1] or [CKL],

$$w_K = 2211212212211211221211 21\ldots$$

defined by the rule: w_K consists of consecutive blocks of 1's and 2's such that the length of each block is either 1 or 2, and the length of the ith block is equal to the ith letter of w_K. Hence, odd blocks consists of 2's and even ones of 1's. The word is an example of a *selfreading* infinite word, cf. [Sl]. The answer to the analysis problem of w_K is not known, in fact it is not even known whether $g_{w_K}(n) = \mathcal{O}(n^2)$. □

Example 9.2. As an example of the case when the complexity of a word is precisely known we consider the Fibonacci word w_F defined as the fixed point of the Fibonacci morphism: $F(a) = ab$, $F(b) = a$. We show that its complexity satisfies

$$g_{w_F}(n) = n + 1 \quad \text{for} \quad n \geq 1. \tag{1}$$

This is seen inductively by showing that, for each n, there exists just one word w of length n such that both wa and wb are in $F(w_F)$. Let us call such factors *special*. For $n = 1$ and $n = 2$ the sets of factors of these lengths are $\{a, b\}$ and $\{aa, ab, ba\}$, where a and ba are the special ones. Now consider a factor w of length $n+1$, with $n \geq 2$. If w ends with b, then, by the form of the morphism F, w admits only the continuation by a, i.e., the a-extension. If $w = xw'a$, with $x \in \{a, b\}$, then by the induction hypothesis of the words $w'a$, with $|w'| = n - 1$, only one is special. Therefore, we are done, when we show that of the words $aw'a$ and $bw'a$, with $|w'| \geq 1$, only one is special. Indeed, one is special since w_F is obtained by iterating a morphism so that any factor appears arbitrary late.

Assume to the contrary that both of these words are special. Then all words $aw'aa$, $aw'ab$, $bw'aa$ and $bw'ab$ are in $F(w_F)$. From the form of F it follows that the $\{F(a), F(b)\}$-interpretations of all of these words match with the word w', i.e. w' is an image of a unique word w'' under F. But then both of the words $aF^{-1}(w'')$ and $bF^{-1}(w'')$ are special, a contradiction with the induction hypothesis. □

Binary words satisfying (1) are so-called infinite *Sturmian words*. Such words have several equivalent definitions, cf. [MH] and [Ra] emphasizing different aspects of these words, and [Bro] containing a brief survey. Their properties has been studied extensively, cf. [CH], [DG], [Mi] and [Ra], in particular recent works in [BdL], [dL] and [dLM] have revealed their fundamental importance in the theory of combinatorics of words.

Our next simple result, noted already in [CH], shows that the complexity of Sturmian words is the smallest unbounded complexity. In particular, the Fibonacci word is an example of a word achieving this.

Theorem 9.1. *Let $w \in \Sigma^\omega$ with $\|\Sigma\| \geq 2$. If g_w is not bounded, then $g_w(n) \geq n+1$ for all $n \geq 1$.*

Proof. We prove that, if for some $n \geq 1$, $g_w(n+1) = g_w(n)$, then w is ultimately periodic, and therefore g_w is bounded. Consequently, Theorem 9.1 follows from the fact that the complexity of a word is a nondecreasing function.

Assume now that $g_w(n_0+1) = g_w(n_0)$. This implies that each factor u of w of length n_0 admits *one and only one* way to extend it by one symbol on the right such that the result is in $F(w)$. Let the function $E : \Sigma^{n_0} \to \Sigma$ define such extensions. Let now $u_0 \in \Sigma^{n_0}$ be a factor of w, say αu_0 is a prefix of w. We define recursively

$$u_{i+1} = u_i \cdot E(\mathrm{suf}_{n_0}(u_i)) \ , \quad \text{for} \quad i \geq 0.$$

Then, by the definition of E, αu_i is a prefix of w for all i, implying that $w = \lim_{i\to\infty} \alpha u_i$. But by the pigeon hole principle and the fact that E is a function $\lim_{i\to\infty} \alpha u_i$ is ultimately periodic. □

Theorem 9.1 states that there exists a gap $(\theta(1), \theta(n))$ in the family of complexities of finite words. According to the current knowledge this is the only known gap. We also note that Theorem 9.1 can be reformulated as

Corollary 9.1. *Let $w \in \Sigma^\omega$ with $\|\Sigma\| \geq 2$. Then w is ultimately periodic if, and only if, g_w is bounded.* □

Above corollary yields a simple criterium to test whether the complexity of a given word is bounded. Unfortunately, however, it is not trivial to verify this criterium. Indeed, even for fixed points of iterated morphisms the verification is not obvious, although can be done effectively, cf. [HL] and [Pan3].

We continue with another example where the asymptotic complexity can be determined. This is a special case of so-called *Toeplitz words* considered in [CaKa].

Example 9.3. We define an infinite word $w_t \in \{1,2\}^\omega$ as follows. Let $p = 1?2?2$ be a word over the alphabet $\{1,2,?\}$, and define recursively

$$\begin{aligned} w_0 &= p^\omega \\ w_{i+1} &= t(w_i) \quad \text{for} \quad i \geq 0, \end{aligned}$$

where $t(w_i)$ is obtained from w_i by substituting w_0 to the positions of w_i filled by the letter ? . Consequently,

$$w_1 = (112?2122?2122121?2221?222)^\omega,$$

and the word $w_t = \lim_{i\to\infty} w_i$ is well-defined over $\{1,2\}$. The word w can be defined as a selfreading word like the Kolakoski word as follows. For each $i \geq 1$, replace the ith occurrence of ? in w_0 by the ith letter of the word so far defined. Clearly, this yields a unique word w', and moreover $w' = w_t$. These two alternate mechanisms to generate w are referred to as *iterative* and *selfreading*, respectively.

In order to compute g_{w_t} we consider factors of w_t of length $5n$. Each such factor is obtained, by the selfreading definition of w_t, from a conjugate of $u_n = (1?2?1)^n$ by substituting a factor v_n of length $3n$ to the positions filled by ?'s. Therefore,

$$g_{w_t}(5n) \leq 5g_{w_t}(3n). \tag{2}$$

It is a straightforward to see that u_n is different from any of its conjugates for $n \geq 2$. Moreover, when different v_n's are substituted to a given conjugate of u_n, different factors of w_t are obtained. Therefore (2) would become the equality, if we could show that each factor v which occurs in w_t occurs in any position modulo 3, i.e., the length of the prefix immediately preceding v can be any number modulo 3. This, indeed, can be concluded from the iterative definition of w_t. First, for each i, the word w_i is periodic over $\{1,2,?\}$ with a period 5^i. Second, each factor v of w_t is a factor of w_{i_0} for some i_0. Consequently, since 3 and 5 are coprimes, the above v occurs in all positions modulo 3 in the prefix of length $3 \cdot 5^{i_0}$ of w_t.

So far we concluded the formula

$$g_{w_t}(5n) = 5g_{w_t}(3n)\ , \quad \text{for} \quad n \geq 2.$$

It is a simple combinatorial exercise to derive from this that $g_{w_t}(n) = \theta(n^r)$ with $r = \log 5/(\log 5 - \log 3)$. □

Next we say a few words about the synthesis problem.

Example 9.4. We already saw how the smallest unbounded complexity of binary words could be achieved. The largest one is even easier to obtain. Indeed, the complexity of the word $w_{\rm bin} = {\rm bin}(1){\rm bin}(2)\ldots$, where ${\rm bin}(i)$ is the binary representation of the number i, equals the exponent function $g(n) = 2^n$. □

Example 9.5. In [Cas3] the synthesis problem is elegantly solved to all linear function $f(n) = an + b$, with $(a, b) \in \mathbb{N} \times \mathbb{Z}$. Namely, it is shown that such a function is the complexity of an infinite word if, and only if, $a + b \geq 1$ and $2a+b \leq (a+b)^2$, and in the affirmative case a word w having this complexity is constructed as a morphic image of the fixed point of an iterated morphism. □

9.2 A classification of complexities of fixed points of iterated morphisms

The rest of this section is devoted to a classification of the asymptotic complexities of words obtained as fixed points of iterated morphisms. This research was initiated in [ELR], later continued in [ER2], [ER3], [ER4], and finally completed in [Pan2]. The classification is based on the structure of the morphism, and it allows to decide the asymptotic complexity of such a word, i.e., to solve the analysis problem for iterated morphisms. It also allows an easy way to solve the asymptotic synthesis problem for those complexities which are possible as fixed points of iterated morphisms. As we shall see there exist only five different such possibilities.

Let $h : \Sigma^* \to \Sigma^*$ be a morphism which need not be 1-free, but is assumed to satisfy the condition $a \in {\rm pref}(h(a))$, for some a, in order to yield the unique word

$$w_h = \lim_{i\to\infty} h^i(a).$$

Consequently, w_h may be finite or infinite. In the former case the complexity of w_h is $\mathcal{O}(1)$. Of course, we assume here that Σ is minimal, i.e., all of its letters occur in w_h.

The classification of morphisms is based on their growth properties as presented in [SaSo] and [RoSa1]. For a letter $a \in \Sigma$ we consider the function $h_a : \mathbb{N} \to \mathbb{N}$ defined by

$$h_a(n) = |h^n(a)| \quad \text{for} \quad n \geq 0.$$

It follows that there exists a nonnegative integer e_a and an algebraic real number ρ_a such that

$$h_a(n) = \theta(n^{e_a}\rho_a^n),$$

the pair (e_a, ρ_a) being referred to as the *growth index* of a in h.

The set Σ_B of so-called bounded letters plays an important role in the classification. A letter a is called *bounded* if, and only if, the function h_a is so, i.e., its growth index equals either to $(0, 0)$ or $(0, 1)$. We say that h is

nongrowing, if there exists a bounded letter in Σ;
quasi-uniform, if $\rho_a = \rho_b > 1$ and $e_a = e_b = 1$ for each $a, b \in \Sigma$;
polynomially diverging, if $\rho_a = \rho_b > 1$ for each $a, b \in \Sigma$, and $e_a > 1$ for some $a \in \Sigma$;
exponential diverging, if $\rho_a > 1$ for each $a \in \Sigma$, and $\rho_a > \rho_b$ for some $a, b \in \Sigma$.

It is not difficult to conclude, cf. [SaSo] or [RoSa1], that this classification is both exhaustive and unambiguous, i.e., each morphism is in exactly one of these classes. In particular, if we call a morphism *growing*, whenever h_a is unbounded for each $a \in \Sigma$, then the three last notions define a partition on the set of growing morphisms.

The above classification is constructive in the sense that for a given morphism we can decide which of the above types it is. Indeed, the growth index of a letter a can be effectively computed, as well as the questions "$q_a > 1$?" and "$\rho_a > \rho_b$?" can be effectively answered. Details needed to conclude these observations can be found in [SaSo].

Now we are ready for the classification proved in [Pan2]. Unfortunately, it does not depend only on the type of the morphism h, but also on the distribution of the bounded letters in w_h. Even worsely, the complexity can be the smallest possible, namely $\theta(1)$, in each of the four cases, since ultimately periodic words can be fixed points of morphisms of any of the above types. However, as we already mentioned, it is decidable whether an iterated morphism defines an ultimately periodic word.

Proposition 9.1. *Let h be a growing iterated morphism. Then if w_h is not ultimately periodic, its complexity is either $\theta(n)$, $\theta(n \log\log n)$ or $\theta(n \log n)$ depending on whether h is quasi-uniform, polynomially diverging or exponentially diverging, respectively.*

The case of nongrowing morphisms is more complicated, essentially due to the fact that this notion is defined existentially, i.e., a morphism is nongrowing whenever there exists a bounded letter.

Proposition 9.2. *Let h be a nongrowing (not necessarily 1-free) iterated morphism generating a non-ultimately periodic word w_h. Then*

(i) if w_h contains arbitrarily long factors over Σ_B, the complexity of w_h is $\theta(n^2)$;

(ii) if all factors of w_h over Σ_B are shorter than a constant K, the complexity of w_h is that of one of the cases in Proposition 9.1, namely $\theta(n)$, $\theta(n \log\log n)$ or $\theta(n \log n)$, and moreover it is decidable which of these it is.

Propositions 9.1 and 9.2 together with our earlier remarks yield immediately the following important results.

Corollary 9.2. *The asymptotic analysis problem for (not necessarily 1-free) iterated morphisms is decidable.* □

Corollary 9.3. *The asymptotic synthesis problem for the complexities $\theta(1)$, $\theta(n)$, $\theta(n \log\log n)$, $\theta(n \log n)$ and $\theta(n^2)$ can be solved.* □

Detailed proofs of Propositions 9.1 and 9.2 can be found in [Pan2]. Here we outline two basic observations of the proofs, as well as give an example of a morphism of each of the above types, and compute the complexities of the corresponding words.

A proof of the fact that the complexity of w_h, for any h, is at most quadratic is not difficult, cf. [ELR]. To see this let us fix n and consider a factor v of w_h of length n. First assume that v is derived in one step from a word v' containing at least one unbounded letter, i.e., the considered (occurrence of) v is a factor in $h(v')$. Let v' be as short as possible and denote $v_0 = v$ and $v_1 = v'$. Obviously, v_1 satisfies automatically our requirement for v, so that we can define inductively $v_0, v_1, v_2, \ldots$ up to v_k with $v_k \in \Sigma$. It follows that $k \leq \|\Sigma\| \cdot n$. Therefore all the factors of length n satisfying our above restriction can be found among the factors of $h(v)$, for $j = k, \ldots, 1$. There are at most $\mathcal{O}(n^2)$ such factors. To cover all the factors of length n, it is enough to note that, for any $v \in \Sigma_B^*$, the language $\{h^i(v) | i \geq 0\}$ contains at most K words for some finite K independent of v. Therefore $\mathcal{O}(n^2)$ is also a valid upper bound for all factors of length n.

Our second remark concerns case (ii) in Proposition 9.2. In [Pan2] this is concluded as follows. Now the factors of w_h in Σ_B^* are shorter than a fixed constant, say K. In particular, each factor v of w_h longer than K contains a growing letter, and therefore for some i independent of v, the word $h^i(v)$ is longer than K. Hence, replacing h by its suitable power, and considering that as a morphism which maps factors of lengths from $K+1$ to $2K$ into words of factors of these lengths, we can eliminate the bounded letters. Let h' be a new morphism constructed in this way. It follows that h' is growing, and moreover, generates as an iterated morphism a word which consists of certain consecutive factors of w_h. Hence, the original w_h can be recovered from the word $w_{h'}$ by using a 1-free morphism mapping the above factors to the corresponding words of Σ^*. Consequently, the word $w_{h'}$ is nothing but a representation of w_h in a larger alphabet, and therefore the asymptotic complexities of w_h and $w_{h'}$ coincide. This explains how case (ii) in Proposition 9.2 is reduced to Proposition 9.1.

As we already said instead of proving Proposition 9.1 and case (i) in Proposition 9.2, we only analyse one example in each of the complexity classes. First, any ultimately periodic word is a fixed point of an iterated morphism yielding the complexity $\theta(1)$. Second, the Fibonacci word w_F of Example 9.2 has the complexity $\theta(n)$, and indeed the morphism is quasi-uniform with $\rho_a = \rho_b = \frac{1}{2}(1+\sqrt{5})$. The remaining cases are covered in Examples 9.6–9.8.

Example 9.6. Consider the morphism h defined by $h(a) = aba$ and $h(b) = bb$. Now h is polynomially diverging since

$$|h^i(a)| = (\frac{1}{2}i+1)2^i \quad \text{and} \quad |h^i(b)| = 2^i \quad \text{for} \quad i \geq 0.$$

To prove that $g_{w_h}(n) = \theta(n \log\log n)$ we first note that under the interpretation $a \leftrightarrow 0$ and $b^{2^i} \leftrightarrow i+1$ the word $h^i(a)$ equals to the so-called ith *sesquipower* s_i defined recursively by

$$\begin{aligned} s_0 &= 0, \\ s_{i+1} &= s_i(i+1)s_i \quad \text{for} \quad i \geq 0. \end{aligned}$$

This means that $h^i(a)$ can be described as

$$h^i(a) = \underbrace{\underbrace{\overbrace{s_1 2 s_1 3 s_1 2 s_1}^{s_3} 4 s_3}_{s_4} 5 s_4 \ldots s_{i-2}}_{s_{i-1}} \; i s_{i-1}.$$

We fix integer $n \geq 2$ and choose $i_0 = \lceil \log n - \log\log n \rceil + 2$, where logarithms are at base 2. Then we have

$$|s_{i_0}| \leq i_0 2^{i_0} \leq (\log n + 3)2^{\log n - \log\log n + 3} \leq (\log n + 3)\frac{8n}{\log n} \leq 32n.$$

Consider now factors of length n occurring in w_h such that they overlap with, or contain as a factor, the first occurrence of i, i.e., b^{2^i}, in w_h. Clearly, any factor of w_h of length n is among these factors for some $i \leq \lfloor \log n \rfloor$. Since, for each i, there are at most $n + 2 \cdot 2^i$ such factors we have

$$g_{w_h}(n) \leq |s_{i_0}| + \sum_{i=i_0+1}^{\lfloor \log n \rfloor} (n + 2\cdot 2^i) \leq 32n + \sum_{i=i_0+1}^{\lfloor \log n \rfloor} 3n = \mathcal{O}(n \log\log n).$$

On the other hand, of the above factors at least $n - 2^i$, for $i = i_0, \ldots, \lceil \log n \rceil$, are such that they do not occur earlier in w_h. Therefore we also have

$$g_{w_h}(n) \geq \sum_{i=i_0}^{\lceil \log n \rceil} (n - 2^i) \geq \sum_{i=i_0}^{\lfloor \log n \rfloor - 1} \frac{n}{2} = \Omega(n \log\log n).$$

So we have proved that $g_{w_h}(n) = \theta(n \log\log n)$. □

Example 9.7. Consider the morphism defined by $h(\$) = \$ab$, $h(a) = aa$ and $h(b) = bbb$. Then $|h^i(a)| = 2^i$, $|h^i(b)| = 3^i$ and $\rho_{\$} > 1$, so that h is exponentially diverging. Denote

$$\alpha(i) = \$aba^2b^3 \ldots a^{2^i}b^{3^i} \in \mathrm{pref}(w_h).$$

Clearly each factor of w_h of length n occurs in $\alpha(\lfloor \log_3(n) \rfloor)$, so we obtain

$$g_{w_h}(n) \leq |\alpha(\lfloor \log_3 n \rfloor)| = 1 + \sum_{i=0}^{\lfloor \log_3 n \rfloor} (2^i + 3^i) = \mathcal{O}(n \log n).$$

On the other hand, for $i = \lceil \log_3 n \rceil, \ldots, \lfloor \log_2 n \rfloor$, w_h contains at least

$$\sum_{i=\lceil \log_3 n \rceil}^{\lfloor \log_2 n \rfloor} (n - 2^i) \geq \Omega(n \log n)$$

different factors in $b^+ a^+ b^* \cup b^* a^+ b^+$. Therefore we have concluded that $g_{w_h}(n) = \Omega(n \log n)$. □

Example 9.8. Finally consider the word

$$w = abcbccbccc \ldots bc^n \ldots,$$

which is the fixed point of the morphism defined as $h(a) = abc$, $h(b) = bc$ and $h(c) = c$. So h is nongrowing and w contains unboundedly long factors in b^*. Let $\alpha(i) = h^i(a)$. Now, all the factors of w of length n occur in the prefix $\alpha(n+1)$. On the other hand, all factors of $\alpha(\lceil \frac{n}{2} \rceil 1)$ of length n are different. Therefore the estimate

$$|\alpha(i)| = 1 + \sum_{j=0}^{i} (1 + j) = \theta(i^2)$$

shows that $g_{w_h}(n) = \theta(n^2)$. □

Our above classification can be straightforwardly modified to D0L languages, i.e., to the language of the form $\{h^i(w) \mid i \geq 0\}$, where h is a morphism and w is a finite word, cf. [RoSa1]. Indeed each iterated morphism h, with $a \in \text{pref}(h(a))$, defines a D0L language via the pair (h, a), and each pair (h, w) determines an iterated morphism h' as an extention of h defined by $h'(\$) = \$w$, where \$ is a new letter. The classification of complexities of D0L languages leads exactly to the above five classes – although the transformation $(h, w) \to h'$ might change the class.

Acknowledgement. The authors are grateful to J. Berstel, S. Grigorieff, T. Harju, F. Mignosi, J. Sakarovitch and A. Restivo for useful discussions during the preparation of this work.

References

[Ab1] H. Abdulrab, Résolution d'équations en mots: étude et implémentation LISP de l'algorithme de Makanin, Ph.D. Thesis, Université de Rouen, 1987.

[Ab2] H. Abdulrab, Implementation of Makanin's algorithm, Springer LNCS **572**, 1991.

[ACK] J. Albert, K. Culik II and J. Karhumäki, Test sets for context-free languages and algebraic systems of equations in a free monoid, Inform. Control **52**, 172–186, 1982.

[AL1] M.H. Albert and J. Lawrence, A proof of Ehrenfeucht's Conjecture, Theoret. Comput. Sci. **41**, 121–123, 1985.

[AL2] M.H. Albert and J. Lawrence, The descending chain condition on solution sets for systems of equations in groups, Proc. Edinburg Math. Soc. **29**, 69–73, 1985.

[Al] J.-P. Allouche, Sur la complexité des suites infinies, Bull. Belg. Math. Soc. **1**, 133–143, 1994.

[BMT] K.A. Baker, G.F. McNulty and W. Taylor, Growth problems for avoidable words, Theoret. Comput. Sci. **69**, 319–345, 1989.

[BEM] D.R. Bean, A. Ehrenfeucht and G.F. McNulty, Avoidable patterns in strings of symbols, Pacific J. Math. **85**, 261–294, 1979.

[Be1] J. Berstel, Transductions and Context-Free Languages, Teubner, 1979.

[Be2] J. Berstel, Mots sans carré et morphismes itérés, Discr. Math. **29**, 235–244, 1979.

[Be3] J. Berstel, Sur les Mots sans carrés définis par morphisme, Springer LNCS **71**, 16–25, 1979.

[Be4] J. Berstel, Some recent results on squarefree words, Springer LNCS **166**, 17–25, 1984.

[Be5] J. Berstel, Properties of infinite words, Springer LNCS **349**, 36–46, 1989.

[Be6] J. Berstel, Axel Thue's work on repetitions in words, Proc. 4th FPSAC, Montreal, 1992; also LITP Report **70**, 1992.

[Be7] J. Berstel, A rewriting of Fife's Theorem about overlap-free words, Springer LNCS **812**, 19–29, 1994.

[Be8] J. Berstel, Axel Thue's papers on repetitions in words: a translation, Publications du Laboratoire de Combinatoire et d'Informatique Mathematique, Université du Québec à Montréal **20**, 1995.

[BdL] J. Berstel and A. de Luca, Sturmian words, Lyndon words and trees, LITP Report **24**, 1995.

[BePe] J. Berstel and D. Perrin, Theory of Codes, Academic Press, 1985.

[BPPR] J. Berstel, D. Perrin, J.F. Perrot and A. Restivo, Sur le Théorème du défaut, J. Algebra **60**, 169–180, 1979.

[BePo] J. Berstel and M. Pocchiola, Average cost of Duval's algorithm for generating Lyndon words, LITP Report **23**, 1992.

[Bra] F.-J. Brandenburg, Uniformly growing k-th powerfree homomorphisms, Theoret. Comput. Sci. **23**, 69–82, 1989.

[Bri] J. Brinkhuis, Non-repetitive sequences on three symbols, Quart. J. Math. Oxford **34**, 145–149, 1983.

[Bro] T.C. Brown, Descriptions of the characteristic sequence of an irrational, Canad. Math. Bull. **36**, 15–21, 1993.

[BS] R. Büchi and S. Senger, Coding in the existential theory of concatenation, Arch. Math. Logik **26**, 101–106, 1986/87.

[Car1] A. Carpi, Overlap-free words and finite automata, Theoret. Comput. Sci. **115**, 243–260, 1993.

[Car2] A. Carpi, On the number of abelian square-free words on four letter alphabet, manuscript, 1994.

[Cas1] J. Cassaigne, Counting overlap-free binary words, Springer LNCS **665**, 216–225, 1993.

[Cas2] J. Cassaigne, Unavoidable binary patterns, Acta Informatica **30**, 385–395, 1993.

[Cas3] J. Cassaigne, Motifs évitables et régularité dans les mots, Thèse de Doctorat, Université Paris VI, 1994.

[CaKa] J. Cassaigne and J. Karhumäki, Toeplitz words, generalized periodicity and periodically iterated morphisms, Springer LNCS **959**, 244–253, 1995; also J. Eur. Comb. (to appear).

[CV] Y. Césari and M. Vincent, Une caractérisation des mots périodiques, C.R. Acad. Sci. Paris **286** A, 1175–1177, 1978.

[Ch] C. Choffrut, Bijective sequential mappings of a free monoid onto another, RAIRO Theor. Inform. Appl. **28**, 265–276, 1994.

[CC] C. Choffrut and K. Culik II, On Extendibility of Unavoidable Sets. Discr. Appl. Math. **9**, 125–137, 1984.

[Co] P.M. Cohn, Algebra, Vol 2, John Wiley and Sons, 1989.

[CF] R. Cori and M.R. Formisano, On the number of partially abelian square-free words on a three-letter alphabet, Theoret. Comput. Sci. **81**, 147–153, 1991.

[CH] E.M. Coven and G.A. Hedlund, Sequences with minimal block growth, Math. Syst. Theory **7**, 138–153, 1973.

[Cr] M. Crochemore, Sharp characterizations of squarefree morphisms, Theoret. Comput. Sci. **18**, 221–226, 1982.

[CP] M. Crochemore and D. Perrin, Two-way string matching, J. ACM **38**, 651–675, 1991.

[CR] M. Crochemore and W. Rytter, Text Algorithms, Oxford University Press, 1994.

[CuKa1] K. Culik II and J. Karhumäki, Systems of equations and Ehrenfeucht's conjecture, Discr. Math. **43**, 139–153, 1983.

[CuKa2] K. Culik II and J. Karhumäki, On the equality sets for homomorphisms on free monoids with two generators, RAIRO Theor. Informatics **14**, 349–369, 1980.

[CKL] K. Culik II, J. Karhumäki and A. Lepistö, Alternating iteration of morphisms and the Kolakoski sequence, in: G. Rozenberg and A. Salomaa (eds.): Lindenmayer Systems, Springer, 93–106, 1992.

[CS] K. Culik II and A. Salomaa, On infinite words obtained by iterating morphisms, Theoret. Comput. Sci. **19**, 29–38, 1982.

[Da] M. Davis, Hilbert's tenth problem is undecidable, Amer. Math. Monthly **80**, 233–269, 1973.

[Dej] F. Dejean, Sur un Théorème de Thue, J. Comb. Theor, Ser. A **13**, 90–99, 1972.

[Dek] F.M. Dekking, Strongly non-repetitive sequences and progression-free sets, J. Comb. Theor., Ser. A **27**, 181–185, 1979.

[dL] A. de Luca, Sturmian words: Structure, combinatorics and their arithmetics, Theoret. Comput. Sci. (to appear).

[dLM] A. de Luca and Filippo Mignosi, Some combinatorial properties of Sturmian words, Theoret. Comput. Sci. **136**, 361–385, 1994.

[Do] E. Domenjoud, Solving Systems of Linear Diophantine Equations: An Algebraic Approach, Springer LNCS **520**, 1991.

[DG] S. Dulucq and D. Gougou-Beauchamps, Sur les facteurs des suites de Sturm, Theoret. Comput. Sci. **71**, 381–400, 1990.

[Du1] J.P. Duval, Périodes et répétitions des mots de monoïde libre, Theoret. Comput. Sci. **9**, 17–26, 1979.

[Du2] J.P. Duval, Factorizing words over an ordered alphabet, J. Algorithms **4**, 363–381, 1983.

[EHR] A. Ehrenfeucht, D. Haussler and G. Rozenberg, On Regurality of Context-free Languages. Theoret. Comput. Sci. **27**, 311–322, 1983.

[EKR1] A. Ehrenfeucht, J. Karhumäki and G. Rozenberg, The (generalized) Post Correspondence Problem with lists of consisting of two words is decidable, Theoret. Comput. Sci. **21**, 119–144, 1982.

[EKR2] A. Ehrenfeucht, J. Karhumäki and G. Rozenberg, On binary equality languages and a solution to the test set conjecture in the binary case, J. Algebra **85**, 76–85, 1983.
[ELR] A. Ehrenfeucht, K.P. Lee and G. Rozenberg, Subword complexities of various classes of deterministic developmental languages without interactions, Theoret. Comput. Sci. **1**, 59–75, 1975.
[ER1] A. Ehrenfeucht and G. Rozenberg, Elementary homomorphisms and a solution to the D0L sequence equivalence problem, Theoret. Comput. Sci. **7**, 169–183, 1978.
[ER2] A. Ehrenfeucht and G. Rozenberg, On the subword complexity of square-free D0L-languages, Theoret. Comput. Sci. **16**, 25–32, 1981.
[ER3] A. Ehrenfeucht and G. Rozenberg, On the subword complexity of D0L-languages with a constant distribution, Inform. Proc. Letters **13**, 108–113, 1981.
[ER4] A. Ehrenfeucht and G. Rozenberg, On the subword complexity of locally catenative D0L-languages, Inform. Proc. Letters **16**, 121–124, 1983.
[Ei] S. Eilenberg, Automata, Languages and Machines, vol. A, Academic Press, 1974.
[Ev] A.A. Evdokimov, Strongly asymmetric sequences generated by a finite number of symbols, Dokl. Akad. Nauk. SSSR **179**, 1268–1271, 1968 (English transl. Soviet Math. Dokl. **9**, 536–539, 1968).
[F] E.D. Fife, Binary sequences which contain no BBb, Trans. Amer. Math. Soc. **261**, 115–136, 1980.
[FW] N.J. Fine and H.S. Wilf, Uniqueness theorem for periodic functions, Proc. Am. Math. Soc. **16**, 109–114, 1965.
[GJ] M.R. Garey and D.S. Johnson, Computers and Intractrability: A Guide to the Theory of NP-Completeness, Freeman, 1979.
[GK] P. Goralčik and V. Koubek, On discerning words by automata, Springer LNCS **226**, 116–122, 1986.
[GV] P. Goralčik and T. Vaniček, Binary patterns in binary words, Intern. J. Algebra Comput. **1**, 387–391, 1991.
[GR] S. Ginsburg and G.F. Rosen, A characterization of machine mappings, Can. J. Math. **18**, 381–388, 1966.
[GO] L.J. Guibas and A.M. Odlyzko, Periods in strings, J. Comb. Theory, Ser. A **30**, 19–42, 1981.
[HW] G.H. Hardy and E.M. Wright, An Introduction to the Theory of Numbers, 1959.
[Harj] T. Harju, On cyclically overlap-free words in binary alphabets, in: G. Rozenberg and A. Salomaa (eds.), The Book of L, Springer, 123–130, 1986.
[HK1] T. Harju and J. Karhumäki, On the defect theorem and simplifiability, Semigroup Forum **33**, 199–217, 1986.
[HK2] T. Harju and J. Karhumäki, Morphisms, in this Handbook.
[HKK] T. Harju, J. Karhumäki and D. Krob, Remarks on generalized Post Correspondence Problem, Springer LNCS **1046**, 39–48, Springer-Verlag, 1996.
[HKP] T. Harju, J. Karhumäki and W. Plandowski, Compactness of systems of equations in semigroups, Springer LNCS **944**, 444–454, 1995; also Int. J. Algebra Comp (to appear).
[HL] T. Harju and M. Linna, On the periodicity of morphisms on free monoids, RAIRO Theor. Inform. Appl. **20**, 47–54, 1986.
[Harr] M. Harrison, Introduction to Formal Language Theory, Addison-Wesley, 1978.
[Hi] G. Higman, Ordering with divisibility in abstract algebras, Proc. London Math. Soc. **3**, 326–336, 1952.

[Hm] Y.I. Hmelevskii, Equations in free semigroups, Proc. Steklov Inst. Math. **107**, 1971; Amer. Math. Soc. Translations, 1976.
[HU] J.E. Hopcroft and J.D. Ullman, Introduction to Automata Theory, Languages and Computation, Addison-Wesley, 1979.
[Jo] J.H. Johnson, Rational equivalence relations, Springer LNCS **226**, 167–177, 1986.
[Ju] J. Justin, Characterization of the repetitive commutative semigroups, J. Algebra **21**, 87–90, 1972.
[Ka1] J. Karhumäki, On cube-free ω-words generated by binary morphisms, Discr. Appl. Math. **5**, 279–297, 1983.
[Ka2] J. Karhumäki, A note on intersections of free submonoids of a free monoid, Semigroup Forum **29**, 183–205, 1984.
[Ka3] J. Karhumäki, The Ehrenfeucht Conjecture: A compactness claim for finitely generated free monoids, Theoret. Comput. Sci. **29**, 285–308, 1984.
[KRJ] J. Karhumäki, W. Rytter and S. Jarominek, Efficient construction of test sets for regular and context-free languages, Theoret. Comput. Sci. **116**, 305–316, 1993.
[KaPl1] J. Karhumäki and W. Plandowski, On the size of independent systems of equations in semigroups, Theoret. Comput. Sci. (to appear).
[KaPl2] J. Karhumäki and W. Plandowski, On defect effect of many identities in free semigroups, in: G. Paun(ed.), Mathematical Aspects of Natural and Formal Languages, World Scientific, 225–232, 1994.
[KPR] J. Karhumäki, W. Plandowski and W. Rytter, Polynomial size test sets for context-free languages, J. Comput. System Sci. **50**, 11–19, 1995.
[Ke1] V. Keränen, On the k-freeness of morphisms on free monoids, Ann. Acad. Sci. Fenn. Ser. A I Math. Dissertationes **61**, 1986.
[Ke2] V. Keränen, Abelian squares are avoidable on 4 letters, Springer LNCS **623**, 41–52, 1992.
[Kob] Y. Kobayashi, Enumeration of irreducible binary words, Discr. Appl. Math. **20**, 221–232, 1988.
[Kol] W. Kolakoski, Self generating runs, Problem 5304, Amer. Math. Monthly **72**, 674, 1965; Solution by N. Ücoluk, Amer. Math. Monthly **73**, 681–682, 1966.
[KoPa] A. Koscielski and L. Pacholski, Complexity of Makanin's algorithm, JACM **43**, 1996.
[Kos] M. Koskas, Complexités de suites de Toeplitz, Discr. Math. (to appear).
[Kr] J.B. Kruskal, The Theory of Well-Quasi-Ordering: A Frequently Discovered Concept, J. Combin. Theory Ser. A **13**, 297–305, 1972.
[La] G. Lallement, Semigroups and Combinatorial Applications, Wiley ,1979.
[LeC] M. Le Conte, A charcterization of power-free morphisms, Theoret. Comput. Sci. **38**, 117–122, 1985.
[Len] A. Lentin, Equations dans les Monoides Lïbres, Gauthier-Villars, 1972.
[LeSc] A. Lentin and M.P. Schützenberger, A combinatorial problem in the theory of free monoids, in: R.C. Bose and T.E. Dowling (eds.), Combinatorial Mathematics, North Carolina Press, Chapel Hill, 112–144, 1967.
[Lep1] A. Lepistö, Repetitions in Kolakoski sequence, in: G. Rozenberg and A. Salomaa (eds.), Developments in Language Theory, World Scientific, 130–143, 1994.
[Lep2] A. Lepistö, Master's thesis, University of Turku, 1995
[Lev] F.W. Levi, On semigroups, Bull. Calcuta Math. Soc. **36**, 141–146, 1944.
[Lo] M. Lothaire, Combinatorics on Words, Addison-Wesley, 1983.
[LySc] R.C. Lyndon and M.P. Schützenberger, The equation $a^m = b^n c^p$ in a free group, Michigan Math J. **9**, 289–298, 1962.

[McN1] R. MacNaughton, A decision procedure for generalized mappability-onto of regular sets, manuscript.
[McN2] R. MacNaughton, A proof of the Ehrenfeucht conjecture, Informal memorandum, 1985.
[Mak] G.S. Makanin, The problem of solvability of equation in a free semigroup, Mat. Sb. **103**, 147–236, 1977 (English transl. in Math USSR Sb. **32**, 129–198).
[Marc] S.S. Marchenkov, Undecidability of the ∀∃-positive Theory of a free Semigroup, Sibir. Mat. Journal **23**, 196–198, 1982.
[Mart] U. Martin, A note on division orderings on strings. Inform. Proc. Letters **36**, 237–240, 1990.
[Mat] Y. Matiyasevich, Enumerable sets are diophantine, Soviet Math. Doklady **11**, 354–357, 1970 (English transl. Dokl. Akad. Nauk. SSSR **191**, 279–282, 1971).
[MN] H.A. Maurer and M. Nivat, Rational bijections of rational sets, Acta Informatica **13**, 365–378, 1980.
[McC] E.M. McCreight, A space-economical suffix tree construction algorithm, J. ACM **23**, 262–272, 1976.
[Mi] F. Mignosi, On the number of factors of Sturmian words, Theoret. Comput. Sci. **82**, 71–84, 1991.
[MP] F. Mignosi and G. Pirillo, Repetitions in the Fibonacci infinite word, RAIRO Theor. Inform. Appl. **26**, 199–204, 1992.
[MRS] F. Mignosi, A. Restivo and S. Salemi, A periodicity theorem on words and applications, Springer LNCS **969**, 337–348, 1995.
[Mor1] M. Morse, Recurrent geodesics on a surface of negative curvature, Trans. Am. Math. Soc. **22**, 84–100, 1921.
[Mor2] M. Morse, A solution of the problem of infinite play in chess, Bull. Am. Math. Soc. **44**, 632, 1938.
[MH] M. Morse and G. Hedlund, Symbolic dynamics, Am. J. Math. **60**, 815–866, 1938.
[Mou] J. Moulin-Ollagnier, Proof of Dejean's conjecture for alphabets with 5, 6, 7, 8, 9, 10 and 11 letters, Theoret. Comput. Sci. **95**, 187–205, 1992.
[MS] A.A. Muchnik and A.L. Semenov, Jewels of Formal Languages (Russian translation of [Sal2]), Mir, Moskow, 1986.
[Ne1] J. Néraud, Elementariness of a finite set of words is co-NP-complete, RAIRO Theor. Inform. Appl. **24**, 459–470, 1990.
[Ne2] J. Néraud, On the rank of the subset a free monoid, Theoret. Comput. Sci. **99**, 231–241, 1992.
[Ne3] J. Néraud, Deciding whether a finite set of words has rank at most two, Theoret. Comput. Sci. **112**, 311–337, 1993.
[Ni] J. Nielsen, Die Isomorphismengruppe der freien Gruppen, Math. Ann. **91**, 169–209.
[Pan1] J.-J. Pansiot, A propos d'une conjecture de F. Dejean sur les répétitions dans les mots, Discr. Appl. Math. **7**, 297–311, 1984.
[Pan2] J.-J. Pansiot, Complexité des facteurs des mots infinis engendrés par morphismes itérés, Springer LNCS **172**, 380–389, 1984.
[Pan3] J.-J. Pansiot, Decidability of periodicity for infinite words, RAIRO Theor. Inform. Appl. **20**, 43–46, 1986.
[Pav] V.A. Pavlenko, Post Combinatorial Problem with two pairs of words, Dokladi AN Ukr. SSR **33**, 9–11, 1981.
[Pec] J.P. Pécuchet, Solutions principales et rang d'un système d'équations avec constantes dans le monoïde libre, Discr. Math. **48**, 253–274, 1984.
[Per] D. Perrin, On the solution of Ehrenfeucht's Conjecture, Bull. EATCS **27**, 68–70, 1985.

[Pl] P.A. Pleasants, Non repetitive sequences, Mat. Proc. Cambridge Phil. Soc. **68**, 267–274, 1970.
[Pr] M.E. Prouhet, Mémoire sur quelques relations entre les puissances des numbres, C.R. Acad. Sci. Paris. **33**, Cahier **31**, 225, 1851.
[Ra] G. Rauzy, Mots infinis en arithmétique, Springer LNCS **95**, 165–171, 1984.
[ReSa] A. Restivo and S. Salemi, Overlap-free words on two symbols, Springer LNCS **192**, 198–206, 1984.
[Rob1] J.M. Robson, Separating Strings with Small Automata, Inform. Proc. Letters **30**, 209–214, 1989.
[Rob2] J.M. Robson, Separating Words with Machines and Groups, (to appear), 1995.
[Ros1] L. Rosaz, Making the inventory of unavoidable sets of words of fixed cardinality, Ph.D. Thesis, Université de Paris 7, 1992.
[Ros2] L. Rosaz, Unavoidable Languages, Cuts and Innocent Sets of Words, RAIRO Theor. Inform. Appl. **29**, 339–382, 1995.
[Rot] P. Roth, Every binary pattern of length six is avoidable on the two-letter alphabet, Acta Informatica **29**, 95–107, 1992.
[RoSa1] G. Rozenberg and A. Salomaa, The Mathematical Theory of L Systems, Academic Press, 1980.
[RoSa2] G. Rozenberg and A. Salomaa, Cornerstones of Undecidability, Prentice Hall, 1994.
[Sal1] A. Salomaa, Formal Languages, Academic Press, 1973.
[Sal2] A. Salomaa, Jewels of Formal Languages, Computer Science Press, 1981.
[Sal3] A. Salomaa, The Ehrenfeucht Conjecture: A proof for language theorists, Bull. EATCS **27**, 71–82, 1985.
[SaSo] A. Salomaa and M. Soittola, Automata-Theoretic Aspects of Formal Power Series, Springer, 1978.
[Sap] M. Sapir, Combinatorics on words with applications, LITP Report **32**, 1995.
[Sc] U. Schmidt, Avoidable patterns on two letters, Theoret. Comput. Sci. **63**, 1–17, 1985.
[See] P. Séébold, Sequences generated by infinitely iterated morphisms, Discr. Appl. Math. **11**, 93–99, 1985.
[Sei] S. Seibert, Quantifier Hierarchies and Word Relations, Springer LNCS **626**, 329–338, Springer-Verlag, 1992.
[She] R. Shelton, Aperiodic words on three symbols I, J. Reine Angew. Math **321**, 195–209, 1981.
[ShSo1] R. Shelton and R. Soni, Aperiodic words on three symbols II, J. Reine Angew. Math **327**, 1–11, 1981.
[ShSo2] R. Shelton and R. Soni, Aperiodic words on three symbols III, J. Reine Angew. Math **330**, 44–52, 1982.
[Shy] H.J. Shyr, Free Monoids and Languages, Hon Min Book Company, Taiwan, 1991.
[Si] I. Simon, An Algorithm to Distinguish Words Efficiently by Their Subwords, manuscript, 1983.
[SkSe] D. Skordev and Bl. Sendov, On equations in words, Z. Math. Logic Grundlagen Math. **7**, 289–297, 1961.
[Sl] N.J.A. Sloane, A Handbook of Integer Sequences, Academic Press, 1973.
[Sp1] J.-C. Spehner, Quelques problèmes d'extension, de conjugaison et de presentation des sous-monoïdes d'un monoïde libre, Ph.D. Thesis, Université Paris VII, 1976.
[Sp2] J.-C. Spehner, Quelques constructions et algorithmes relatifs aux sous-monoïdes d'un monoïde libre, Semigroup Forum **9**, 334–353, 1975.

[T1] A. Thue, Über unendliche Zeichenreihen, Kra. Vidensk. Selsk. Skrifter. I. Mat.-Nat. Kl., Christiana, Nr. **7**, 1906.
[T2] A. Thue, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, Kra. Vidensk. Selsk. Skrifter. I. Mat.-Nat. Kl., Christiana, Nr. **12**, 1912.
[Z] A.I. Zimin, Blocking sets of terms, Math USSR Sb. **47**, 353–364, 1984.