CHAPTER 6

Codes

6.0. Introduction

The theory of codes provides some jewels of combinatorics on words that we want to describe in this chapter.

A basic result is the defect theorem (Theorem 6.2.1), which states that if a set X of n words satisfies a nontrivial relation, then these words can be expressed simultaneously as products of at most $n-1$ words. It is the starting point of the chapter. In Chapters 9 and 13, other defect properties are studied in different contexts.

A nontrivial relation is simply a finite word w which ambiguously factorizes over X. This means that X is not a code. The defect effect still holds if X is not an ω-code, i.e., if the nontrivial relation is an infinite, instead of a finite word (Theorem 6.2.4).

The defect theorem implies several well-known properties on words that are recalled in this chapter. For instance, the fact that two words which commute are powers of the same word is a consequence. Another consequence is that a two-element code or more generally an elementary set is an ω-code. The latter property appears to be a crucial step in one of the proofs of the D0L equivalence problem.

A remarkable phenomenon appears when, for a finite code X, neither the set X nor its reversal $\widetilde{X}$ is an ω-code. In this case the defect property is stronger: the n elements of X can be expressed as products of at most $n-2$ words (Theorem 6.3.4). It follows that for codes X with three elements, either X or $\widetilde{X}$ is an ω-code. The proof of this property is rather long. It uses in a very elegant and subtle way techniques of combinatorics on words.

In this chapter, we also present a deep result by Schützenberger about finite maximal codes. It states that if a finite maximal code X is an ω-code, or equivalently if X has bounded decoding delay, then X is a prefix code (Theorem 6.4.1). The original proof is complex. The proof given here is short and elementary.

6.1. X-factorizations

6.1.1. Codes

Let $X \subseteq A^+$. A sequence $(x_1, x_2, \dots, x_n)$ of n words of X is an *X-factorization* of a word $w \in A^*$ if $w = x_1 x_2 \cdots x_n$.

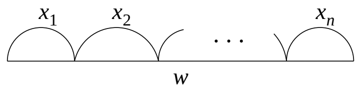

Figure 6.1. An X-factorization of the word w.

A set $X \subseteq A^+$ is a *code* if any word $w \in A^*$ has at most one X-factorization. This definition is equivalent to the one given in Chapter 1.

The simplest codes are *prefix codes* $X \subseteq A^+$. Recall that they are sets such that no word of X is a proper prefix of another word of X. *Suffix codes* are defined symmetrically as sets such that no word of X is a proper suffix of another word of X.

EXAMPLE 6.1.1. Let $A = \{a, b\}$. The set $X = \{a, ab, ba\}$ is not a code because $w = aba$ has two distinct X-factorizations, namely (a, ba), (ab, a).

EXAMPLE 6.1.2. The set $X = \{a, ab, bb\}$ over the alphabet $\{a, b\}$ is a suffix code.

The name of "code" is motivated by the next proposition. Roughly speaking, if the letters of a source alphabet B are put in 1-to-1 correspondence with the words of a code X over a target alphabet A, then a source message $r \in B^*$ is encoded into a coded message $w \in A^*$ by replacing any letter of r by the corresponding word of X. Unique decipherability is insured by the fact that w has exactly one X-factorization.

PROPOSITION 6.1.3. *A set $X \subseteq A^+$ is a code if and only if any morphism $\varphi : B^* \to A^*$ induced by a bijection from B onto X is injective.*

With the notation of the proposition, we say that φ is a *coding morphism for X*.

Proof. Let $\varphi : B^* \to A^*$ be a morphism induced by a bijection from B onto X. Let $r, s \in B^*$ such that $\varphi(r) = \varphi(s)$. Let us prove that $r = s$. Set $r = \alpha_1 \cdots \alpha_n$, $s = \beta_1 \cdots \beta_m$ with $\alpha_i, \beta_j \in B$ and $n, m \geq 0$. Since $\varphi(r) = \varphi(s)$, this word has the two X-factorizations $(\varphi(\alpha_1), \dots, \varphi(\alpha_n))$ and $(\varphi(\beta_1), \dots, \varphi(\beta_m))$. But X is a code, hence $n = m$ and $\varphi(\alpha_i) = \varphi(\beta_i)$ for all i. As φ is injective on B, one has $\alpha_i = \beta_i$ for all i, and $r = s$.

For the converse, let X be a subset of A^+ and $\varphi : B^* \to A^*$ be an injective morphism induced by a bijection from B onto X. Let $w \in A^*$ be a word with X-factorizations $(x_1, \dots, x_n)$, $(y_1, \dots y_m)$. Through the bijection φ, let $x_i = \varphi(\alpha_i)$,

$y_j = \varphi(\beta_j)$ for letters α_i, β_j. Thus $w = \varphi(\alpha_1 \cdots \alpha_n) = \varphi(\beta_1 \cdots \beta_m)$. As φ is injective, one has $\alpha_1 \cdots \alpha_n = \beta_1 \cdots \beta_m$, that is, $n = m$ and $\alpha_i = \beta_i$ for all i. It follows that the two X-factorizations are equal. Thus X is a code. ∎

EXAMPLE 6.1.1 (*continued*). Let $\varphi : B^* = \{\alpha, \beta, \gamma\}^* \to A^*$ be defined by $\varphi(\alpha) = a$, $\varphi(\beta) = ab$ and $\varphi(\gamma) = ba$. It is not injective since $\varphi(\alpha\gamma) = \varphi(\beta\alpha)$.

PROPOSITION 6.1.4. *Let $\varphi : B^* \to A^*$ be an injective morphism. If $Z \subseteq B^+$ is a code, then $\varphi(Z)$ is a code. If $X \subseteq A^+$ is a code, then $\varphi^{-1}(X)$ is a code.*

Proof. Suppose that Z is a code. Consider a word $w \in A^*$ with the $\varphi(Z)$-factorizations $(x_1, \ldots, x_n)$, $(y_1, \ldots, y_m)$ such that $x_i = \varphi(z_i)$, $y_j = \varphi(t_j)$ and $z_i, t_j \in Z$. Then $\varphi(z_1 \cdots z_n) = \varphi(t_1 \cdots t_m)$ and $z_1 \cdots z_n = t_1 \cdots t_m$ since φ is injective. As Z is a code, $n = m$ and $z_i = t_i$ for all i. It follows that $\varphi(z_i) = \varphi(t_i)$ for any i, showing that $\varphi(Z)$ is a code. A similar argument shows that $\varphi^{-1}(X)$ is a code if X is a code. ∎

EXAMPLE 6.1.5. The set $Z = \{\alpha\alpha, \alpha\beta, \alpha\gamma, \beta, \gamma\}$ is a code over the alphabet $B = \{\alpha, \beta, \gamma\}$. Let $\varphi : B^* \to A^*$ be the morphism induced by $\varphi(\alpha) = a$, $\varphi(\beta) = ab$ and $\varphi(\gamma) = bb$. The set $\varphi(Z) = \{aa, aab, abb, ab, bb\}$ is a code.

We end this section with a characterization of codes by a property on the monoid that they generate.

We recall that a submonoid M of A^* has a unique minimal generating set $(M - \varepsilon) - (M - \varepsilon)^2$ (see Chapter 1). For convenience, we call it the *base* of M.

Let $M = X^*$ be the submonoid of A^* generated by a set $X \subseteq A^+$. If X is a code, then X is necessarily the base of M. Otherwise, X contains a word w which belongs to $(M - \varepsilon)^2$. This word has thus two X-factorizations: (w) itself and $(x_1, \ldots, x_n)$, with $n \geq 2$, a contradiction.

However, it may happen that the base of a submonoid M is not a code. We have seen such a base in Example 6.1.1: the monoid $M = \{a, ab, ba\}^*$ is generated by the set $X = \{a, ab, ba\}$ which is also the base of M, but X is not a code.

When the base X of a submonoid M is a code, we say that M is *free*.

A submonoid M of A^* is *stable* if for any $u, w, v \in A^*$,

$$u, wv, uw, v \in M \Rightarrow w \in M.$$

Figure 6.2 gives a pictorial representation of the stability condition.

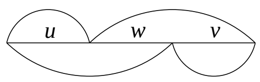

Figure 6.2. The stability condition.

PROPOSITION 6.1.6. *Let M be a submonoid of A^*. Then M is free if and only if M is stable.*

Proof. Suppose that M is free, i.e., the base X of M is a code. Take u, wv, uw, $v \in M = X^*$. Consider the X-factorizations

$$(x_1, \ldots, x_k), (x_{k+1}, \ldots, x_n), (y_1, \ldots, y_\ell), (y_{\ell+1}, \ldots, y_m)$$

of u, wv, uw, v respectively. Since X is a code, the X-factorizations

$$(x_1, \ldots, x_k, x_{k+1}, \ldots x_n), (y_1, \ldots, y_\ell, y_{\ell+1}, \ldots y_m)$$

of $u(wv) = (uw)v$ are equal. Moreover, $\ell \geq k$ since $|uw| \geq |u|$, showing that

$$uw = x_1 \cdots x_k x_{k+1} \cdots x_\ell = u x_{k+1} \cdots x_\ell.$$

Hence, $w = x_{k+1} \cdots x_\ell \in M$, and M is stable.

For the converse, assume that M is stable but its base X is not a code. There exists a word $z \in M$ with X-factorizations $(x_1, \ldots, x_n)$, $(y_1, \ldots, y_m)$ such that $x_1 \neq y_1$. We can suppose that $y_1 = x_1 w$ with w a nonempty word. Hence

$$u = x_1, wv = x_2 \cdots x_n, uw = y_1, v = y_2 \cdots y_m \in M.$$

But M is stable, thus $w \in M$. Consequently, $y_1 = x_1 w \in X \cap (M - \varepsilon)^2$, showing that X is not the base of M. This leads to the contradiction. ■

COROLLARY 6.1.7. *The intersection of an arbitrary family of free submonoids of A^* is a free submonoid of A^*.*

Proof. Take a family of free submonoids M_i of A^*, indexed by a set I. Denote by M the set $\cap_{i \in I} M_i$. It is a submonoid of A^*. Let $u, w, v \in A^*$ be such that $u, wv, uw, v \in M$. As any M_i is free and then stable, the word w belongs to $M_i \subseteq M$. It follows that M is stable and thus free. ■

6.1.2. ω-codes

In this section, infinite instead of finite X-factorizations are studied.

Let $X \subseteq A^+$. An *X-factorization* of a word $w \in A^\omega$ is an infinite sequence $(x_1, x_2, \ldots, x_n, \ldots)$ of elements of X such that $w = x_1 x_2 \cdots x_n \cdots$ (see Figure 6.3).

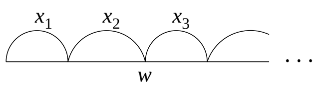

Figure 6.3. An X-factorization of the infinite word w.

A set $X \subseteq A^+$ is an *ω-code* if any word of A^ω has at most one X-factorization. Any ω-code is a code as two distinct X-factorizations $(x_1, \ldots, x_n)$, $(y_1, \ldots, y_m)$ of a word $w \in A^+$ lead to two distinct X-factorizations of the word $w^\omega \in A^\omega$: $(x_1, \ldots, x_n, x_1, \ldots, x_n, \ldots)$, $(y_1, \ldots, y_m, y_1, \ldots, y_m, \ldots)$. The converse is false, as shown in Example 6.1.2.

Example 6.1.2 (*continued*). The code $X = \{a, ab, bb\}$ is not an ω-code because $(a, bb, bb, \ldots)$ and $(ab, bb, bb, \ldots)$ are two distinct X-factorizations of the word ab^ω.

Let $u, v \in A^*$ be two words. We write $u \leq v$ when u is prefix of v.

A code $X \subseteq A^+$ has a *bounded decoding delay* if there is an integer $d \geq 0$ such that for any $x, x', y_1, \ldots y_d \in X$ and $z \in X^*$,

$$xy_1 \cdots y_d \leq x'z \Rightarrow x = x'.$$

In other words, the knowledge of a prefix $xy_1 \cdots y_d \in X^{d+1}$ of a word $x'z \in x'X^*$ does not allow the situation depicted in Figure 6.4. The smallest integer d in

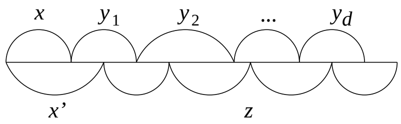

Figure 6.4. An impossible situation for a decoding delay d.

the previous definition is called the *decoding delay* of the code.

This notion extends in a natural way the concept of prefix code, since prefix codes have a decoding delay $d = 0$.

Example 6.1.8. The code $X = \{a, ab\}$ is not prefix, but it has a decoding delay $d = 1$. More generally, the code $X = \{a, a^d b\}$ has a decoding delay d.

Example 6.1.2 (*continued*). The suffix code $X = \{a, ab, bb\}$ does not have a bounded decoding delay, because for any $d \geq 0$, $a(bb)^d$ is prefix of $ab(bb)^d$.

The next proposition shows the relationship between ω-codes and codes with bounded decoding delay (see also Problem 6.1.3).

Proposition 6.1.9. *Any code with bounded decoding delay is an ω-code. Conversely, any finite ω-code has a bounded decoding delay.*

Proof. Let $X \subseteq A^+$ be a code with decoding delay $d \geq 0$. Assume that X is not an ω-code. Then, there exists $w \in A^\omega$ with two X-factorizations $(x, y_1, y_2, \ldots, y_m, \ldots)$, $(x', x_1, x_2, \ldots, x_n, \ldots)$ such that $x \neq x'$. This is impossible since $xy_1 \cdots y_d \leq x'x_1 \cdots x_n$ for some $n \geq 0$.

For the converse, assume that X does not have a bounded decoding delay. Then for any $d \geq 0$, there exist $n \geq 0$ and $x, x', y_1, \ldots, y_d, x_1, \ldots, x_n \in X$ such that

$$xy_1 \cdots y_d \leq x'x_1 \cdots x_n \quad \text{and} \quad x \neq x'.$$

As X is finite, for a large enough d, there exists a suffix w of a word of X which is repeated as follows (see Figure 6.5):

$$xu = x'vw, \;\; xuu' = x'vv'w$$

with $u = y_1 \cdots y_c$, $u' = y_{c+1} \cdots y_d$, $v = x_1 \cdots x_\ell$, $v' = x_{\ell+1} \cdots x_m$, $\ell < m \leq n$. It

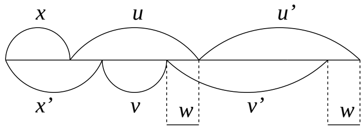

Figure 6.5. The suffix w is repeated.

follows that the word $xuu'^\omega = x'vv'^\omega$ has two distinct X-factorizations, showing that X is not an ω-code. ■

The concepts of coding morphism, free and stable submonoid carry over to ω-codes as follows. The related propositions remain true, with similar proofs.

PROPOSITION 6.1.10. *A set $X \subseteq A^+$ is an ω-code if and only if any morphism $\varphi : B^\infty \to A^\infty$ induced by a bijection from B onto X is injective on B^ω.* ■

We say that φ is a *coding morphism for X*.

PROPOSITION 6.1.11. *Let $\varphi : B^\infty \to A^\infty$ be a morphism injective on B^ω. If $Z \subseteq B^+$ is an ω-code, then $\varphi(Z)$ is an ω-code. If $X \subseteq A^+$ is an ω-code, then $\varphi^{-1}(X)$ is an ω-code.* ■

We recall that a binoid $M \subseteq A^\infty$ is *finitary* if its minimal generating set X is a subset of A^+. In particular $M = X^\infty$ (see Chapter 1). For convenience we call X the *base* of M.

If the base X of a finitary binoid $M \subseteq A^\infty$ is an ω-code, we say that M is *free*.

Freeness for a submonoid M of A^* means that its base X does not satisfy any *nontrivial relation* over finite words. Freeness for a finitary binoid of A^∞ means that its base does not satisfy any *nontrivial relation* over infinite words. A nontrivial relation on finite words implies a nontrivial relation on infinite words because any ω-code is a code.

A binoid M of A^∞ is *stable* if for any $u, w \in A^*$ and $v \in A^\omega$,

$$u, wv, uw, v \in M \Rightarrow w \in M$$

(see Figure 6.6).

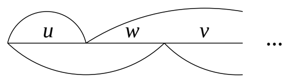

Figure 6.6. The stability condition for infinite words.

PROPOSITION 6.1.12. *Let M be a finitary binoid of A^∞. Then M is free if and only if M is stable.* ■

EXAMPLE 6.1.2 (*continued*). Let $M = X^\infty$ be the finitary binoid generated by $X = \{a, ab, bb\}$. It is not free since its base X is not an ω-code. It is not stable because the words $u = a$, $wv = b(bb)^\omega$, $uw = ab$, $v = (bb)^\omega$ belong to M, but $w = b$ does not belong to M. Nevertheless the submonoid X^* generated by X is free because X is a code. It is stable by Proposition 6.1.6.

6.2. Defect

This section is devoted to the defect theorem which is a basic result on sets of words. It states that if a set X of n words satisfies a nontrivial relation, then these words can be expressed simultaneously as products of at most $n-1$ words. Originally the defect theorem relies on a nontrivial relation over A^*. It still works with a nontrivial relation over A^∞. We also show that the $n-1$ words can be taken in a code associated to X, which is "minimal" in a certain sense.

6.2.1. Defect theorem

Given a set $X \subseteq A^+$, we consider the family $\mathcal{F}$ of all the free submonoids of A^* containing X. It is not empty since A^* belongs to $\mathcal{F}$. It is closed for arbitrary intersection by Corollary 6.1.7. It follows that the *smallest* free submonoid of A^* containing X exists: it is equal to $\cap_{M\in\mathcal{F}} M$. This set is called the *free hull* of X. In particular, the base Y of the free hull of X is a code and $X \subseteq Y^*$.

Let $X \subseteq Y^*$, with Y a code over the alphabet A. As any word of X has a unique Y-factorization, Y can be viewed as an alphabet for X. We denote by $\mathrm{Alph}_Y(X)$ the set of words of Y appearing in the Y-factorization of the elements of X. We define $\mathrm{First}_Y(X)$ as the set of all y_1 such that there exists an Y-factorization $(y_1, \ldots, y_n)$ of some $x \in X$. We define $\mathrm{Last}_Y(X)$ similarly. This notation is much used below in Section 6.3.2. We simply write $\mathrm{Alph}(X)$, $\mathrm{First}(X)$ and $\mathrm{Last}(X)$ when $Y = A$.

THEOREM 6.2.1 (Defect theorem). *Let $Y \subseteq A^+$ be the base of the free hull of a finite set $X \subseteq A^+$. If X is not a code, then*

$$\mathrm{Card}(Y) \le \mathrm{Card}(X) - 1.$$

Proof. Suppose that Y is the base of the free hull of X. Then Y is a code such that $X \subseteq Y^*$. Any element x of X has thus a unique Y-factorization $(y_1, \ldots, y_n)$. Hence the function $\alpha : X \to Y$ such that $\alpha(x) = y_1 = \mathrm{First}_Y(x)$ is well-defined. It is not injective. Indeed, there exists a word $w \in A^*$ with two X-factorizations $(x_1, \ldots, x_k)$, $(x'_1, \ldots, x'_\ell)$ such that $x_1 \neq x'_1$, because X is not a code. But w has only one Y-factorization over the code Y. This implies that $\alpha(x_1) = \alpha(x'_1)$. If α is not surjective, then there is a word y in $Y - \alpha(X)$. Consider the set $Z = (Y - y)y^*$. Clearly $X \subseteq Z^* \subsetneq Y^*$. Moreover it is not difficult to check that Z is a code (see also Example 6.2.9 below and Section 6.2.4). This is impossible because Y^* is the smallest free submonoid which contains X. Consequently, $\alpha : X \to Y$ is surjective, not injective, and $\mathrm{Card}(Y) \leq \mathrm{Card}(X) - 1$. ■

The proof of Theorem 6.2.1 is based on the following property of the free hull.

PROPOSITION 6.2.2. *Let $X \subseteq A^+$ and Y be the base of the free hull of X. Then*

$$\mathrm{First}_Y(X) = Y.$$ ■

We end this section with a nice link between the free hull of X and its dependency graph. The *dependency graph* G_X of a finite set $X \subseteq A^+$ is an undirected graph with X as set of vertices and edges $(x, x') \in X \times X$ whenever there exist $z, z' \in X^*$ such that $xz = x'z'$.

PROPOSITION 6.2.3. *Let $Y \subseteq A^+$ be the base of the free hull of a finite set $X \subseteq A^+$. If X is not a code, then*

$$\mathrm{Card}(Y) \leq c(X) \leq \mathrm{Card}(X) - 1$$

where $c(X)$ is the number of connected components of the dependency graph G_X of X.

Proof. The inequality $c(X) \leq \mathrm{Card}(X) - 1$ holds because X is not a code. To show the first inequality, we define a function α from the set of connected components of G_X into Y as follows. Let C be a connected component and $x \in X$ which belongs to C. Then $\alpha(C) = y$ such that $\mathrm{First}_Y(x) = y$. This function is well-defined because if (x, x') is an edge of G_X, then $xz = x'z'$ for some $z, z' \in X^*$. Hence $xz = x'z'$ has a unique Y-factorization beginning with y. By Proposition 6.2.2, α is surjective, showing that $\mathrm{Card}(Y) \leq c(X)$. ■

EXAMPLE 6.1.1 (*continued*). The set $X = \{a, ab, ba\}$ is not a code. The smallest free submonoid M containing X is equal to $\{a, b\}^*$. Indeed M is stable by Proposition 6.1.6. Since $u = a, wv = ba, uw = ab, v = a \in M$, then $b \in M$. This shows that $a, b \in M$. Hence the base of the free hull of X is $Y = \{a, b\}$ and $\mathrm{First}_Y(X) = Y$. The dependency graph G_X of X has two connected components, namely $\{a, ab\}$ and $\{ba\}$.

6.2.2. Infinite words

All the arguments of Section 6.2.1 can be repeated in the context of infinite words.

Given a set $X \subseteq A^+$, the smallest free finitary binoid of A^∞ containing X is called the *ω-free hull* of X. This binoid always exists. Indeed let M be the intersection of all free finitary binoids of A^∞ containing X. Then M is a stable binoid because the intersection of stable binoids is again a stable binoid. If M is not finitary, then $M' = (M \cap A^*)^\infty$ is a stable (and thus free) finitary binoid which is strictly included in M and contains X. This is impossible. Note that the base Y of the ω-free hull of X is an ω-code and $X \subseteq Y^*$.

The defect theorem remains true. The proof is similar.

THEOREM 6.2.4 (Defect theorem). *Let $Y \subseteq A^+$ be the base of the ω-free hull of a set $X \subseteq A^+$. Then $\mathrm{First}_Y(X) = Y$. If X is a finite set which is not an ω-code, then*

$$\mathrm{Card}(Y) \leq \mathrm{Card}(X) - 1.$$

Proof. Clearly $\mathrm{First}_Y(X) \subseteq Y$. As in the proof of Theorem 6.2.1, we define $\alpha : X \to Y$ such that $\alpha(x) = \mathrm{First}_Y(x)$, that is, $\alpha(x)$ is the first element y_1 in the Y-factorization $(y_1, \dots, y_n)$ of x. This function is surjective. Otherwise considering $y \in Y - \alpha(X)$, we get a set $Z = (Y - y)y^*$ which is an ω-code (see Example 6.2.9 and Proposition 6.2.10) and such that $X \subseteq Z^\infty \subsetneq Y^\infty$. This is impossible. It follows that $\mathrm{First}_Y(X) = Y$. If X is a finite set which is not an ω-code, then α is not injective and $\mathrm{Card}(Y) \leq \mathrm{Card}(X) - 1$. ∎

EXAMPLE 6.1.1 (*continued*). Recall that the set $X = \{a, ab, bb\}$ is a code but not an ω-code. The base of its free hull is X itself because X^* is a free submonoid. The ω-free hull M of X is the free finitary binoid $M = \{a, b\}^\infty$. Indeed M is stable by Proposition 6.1.12. Since $u = a, wv = b(bb)^\omega, uw = ab, v = (bb)^\omega \in X^\infty \subseteq M$, then $b \in M$.

Propositions 6.2.3 also holds in the context of infinite words. The definition of the graph G_X is slightly different. There is an edge $(x, x') \in X \times X$ if and only if there exist $z, z' \in X^\omega$ (instead of X^*) with $xz = x'z'$.

6.2.3. Consequences

We state three corollaries of the defect theorem.

COROLLARY 6.2.5. *If two words commute, or more generally satisfy a nontrivial relation on finite or infinite words, then they are powers of the same word.*

Proof. Let $x, y \in A^+$ satisfying a nontrivial relation. Then $X = \{x, y\}$ is either not a code or not an ω-code, depending on whether the relation is on finite or infinite words. By the defect theorem (Theorems 6.2.1 or 6.2.4), the base of the free or ω-free hull of X is a one-element set $\{z\} \subseteq A^+$ such that $\{x, y\} \subseteq z^*$. ■

Recall that any ω-code is a code. The converse is true for codes with two elements as shown in the next corollary. This property is false for larger codes (see Example 6.1.2).

COROLLARY 6.2.6. *Any two-element code is an ω-code.*

Proof. Let $X = \{x, y\}$ be a code. If X is not an ω-code, then $X \subseteq z^*$ where $\{z\}$ is the base of the ω-free hull of X (Theorem 6.2.4). This is impossible because X is a code. ■

The last corollary deals with elementary sets. A finite set X is *simplifiable* if $X \subseteq Y^*$ with $\mathrm{Card}(Y) \leq \mathrm{Card}(X) - 1$. Otherwise it is *elementary.*

COROLLARY 6.2.7. *Any elementary set is an ω-code.*

Proof. By the defect theorem, if X is not an ω-code, then $X \subseteq Y^*$ with Y an ω-code such that $\mathrm{Card}(Y) \leq \mathrm{Card}(X) - 1$. Hence X is not an elementary set. ■

The converse of this corollary is false, as shown by the next example.

EXAMPLE 6.2.8. The set $X = \{a, abc, abcbc\}$ is an ω-code. However X is simplifiable because $X \subseteq \{a, bc\}^*$.

6.2.4. Composition of codes

This section is devoted to the operation of composition of codes. It clarifies the construction of the code $(Y - y)y^*$ in the proof of Theorems 6.2.1 and 6.2.4.

Take two sets $Y \subseteq A^+$, $Z \subseteq B^+$ with

$$B = \mathrm{Alph}(Z).$$

We say that Y, Z are *composable* if there exists a bijection φ from B onto Y. The set

$$X = \varphi(Z) \subseteq Y^*$$

is obtained by replacing the letters of Z by the corresponding (through φ) words of Y. The set X resulting of the *composition* of Y and Z is denoted by

$$X = Y \circ_\varphi Z,$$

or more simply by

$$X = Y \circ Z.$$

EXAMPLE 6.1.5 (*continued*). The sets $Y = \{a, ab, bb\}$, $Z = \{\alpha\alpha, \alpha\beta, \alpha\gamma, \beta, \gamma\}$ are composable, thanks to the bijection φ such that $\varphi(\alpha) = a$, $\varphi(\beta) = ab$ and $\varphi(\gamma) = bb$. The set $X = Y \circ_\varphi Z$ is equal to $\{aa, aab, abb, ab, bb\}$.

EXAMPLE 6.2.9. Let Y be a set over A and $Z = (B - b)b^*$ over B, with b is a particular letter of B. Let φ be a bijection from B onto Y. Denote by y the word of Y equal to $\varphi(b)$. Then $X = Y \circ_\varphi Z$ is the set $(Y - y)y^*$.

The operation of composition conserves the code or the ω-code property.

PROPOSITION 6.2.10. *Let $Y \subseteq A^+$, $Z \subseteq B^+$ be two composable sets. If Y, Z are codes (resp. ω-codes), then $X = Y \circ Z$ is a code (resp. ω-code).*

Proof. Let $\varphi : B = \mathrm{Alph}(Z) \to Y$ be the function used for the composition of Y and Z. Suppose that Y, Z are codes. The function φ extends into a morphism injective on B^* by Proposition 6.1.3. It follows by Proposition 6.1.4 that $X = \varphi(Z)$ is a code. The case of ω-codes is solved similarly thanks to Propositions 6.1.10 and 6.1.11. ∎

Example 6.1.5 illustrates this proposition for codes.

EXAMPLE 6.2.9 (*continued*). The $Z = (B - b)b^*$ is a code and an ω-code. If Y is a code (resp. ω-code), then X is also a code (resp. ω-code). This property is used in the proof of Theorems 6.2.1 and 6.2.4.

Given a finite set X, the length $\sum_{x \in X} |x|$ of X is denoted by $\mathrm{Lg}(X)$.

The composition of codes is an associative operation, that is

$$X \circ (Y \circ Z) = (X \circ Y) \circ Z. \tag{6.2.1}$$

Note also that if $X = Y \circ Z$, then

$$\mathrm{Card}(Z) = \mathrm{Card}(X) \text{ and } \mathrm{Lg}(Z) \leq \mathrm{Lg}(X) \tag{6.2.2}$$

with $\mathrm{Lg}(Z) = \mathrm{Lg}(X)$ if and only if $Y = \mathrm{Alph}(X)$.

The next proposition gives conditions such that a set $X \subseteq A^+$ is written as $Y \circ Z$ for a given code $Y \subseteq A^+$.

PROPOSITION 6.2.11. *Let $X, Y \subseteq A^+$ such that Y is a code. If*

$$X \subseteq Y^* \text{ and } \mathrm{Alph}_Y(X) = Y,$$

then $X = Y \circ Z$ for some $Z \subseteq B^+$. Moreover, if X is a code, then Z is a code.

Informally, as Y is a code such that $X \subseteq Y^*$ and $\mathrm{Alph}_Y(X) = Y$, the set X can be viewed as written over the alphabet Y, instead of the alphabet A. This is the definition of the set Z.

Proof. Set $Z = \varphi^{-1}(X)$ where $\varphi : B^\infty \to A^\infty$ is a coding morphism for Y. By definition, φ induces a bijection from B onto Y. As $\mathrm{Alph}_Y(X) = Y$, we get $\mathrm{Alph}(Z) = B$. Thus, $X = Y \circ_\varphi Z$. If X is a code, then $Z = \varphi^{-1}(X)$ is also a code by Proposition 6.1.4. ∎

6.3. More defect

6.3.1. ξ-codes

In this section, we study sets $X \subseteq A^+$ such that both X and its reversal $\widetilde{X}$ are ω-codes. These sets are called *ξ-codes*. Roughly speaking, no right-infinite word and no left-infinite word is "ambiguously" factorized by words of X (see Figure 6.7).

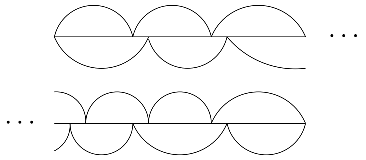

Figure 6.7. The two forbidden situations for ξ-codes.

Given a set $X \subseteq A^+$, in the same way we have defined the free hull of X and its ω-free hull, we can define the *ξ-free hull* of X. It is the smallest finitary binoid M of A^∞ containing X such that M and $\widetilde{M}$ are free. This binoid always exists. Its base Y is a ξ-code such that $X \subseteq Y^*$.

We get the next two properties as a consequence of Theorem 6.2.4.

Proposition 6.3.1. *Let $X \subseteq A^+$ and Y be the base of the ξ-free hull of X. Then*

$$\mathrm{First}_Y(X) = Y \text{ and } \mathrm{Last}_Y(X) = Y. \qquad \blacksquare$$

Theorem 6.3.2. *Let $Y \subseteq A^+$ be the base of the ξ-free hull of a finite subset X of A^+. If X is not a ξ-code, then*

$$\mathrm{Card}(Y) \le \mathrm{Card}(X) - 1. \qquad \blacksquare$$

Example 6.3.3. The set $X = \{ab, ababbb, bab\}$ is a ξ-code. Thus it is equal to the base of its ξ-free hull. The set $Y = \{ab, ababbb, bbbb\}$ is not a ξ-code since $(ab, ab, bbbb, bbbb, \ldots)$ and $(ababbb, bbbb, bbbb, \ldots)$ are two distinct X-factorizations of the word $abab(bb)^\omega$. The base of its ξ-free hull is the set $\{ab, bb\}$ which is also the base of its ω-free hull.

Theorem 6.3.2 leads to a refinement of Corollaries 6.2.6 and 6.2.7. Namely, any two-element code is a ξ-code; any elementary set is a ξ-code.

If in Theorem 6.3.2, the hypothesis that "X is not a ξ-code" is replaced by "X is a code, but neither X nor $\widetilde{X}$ is an ω-code", we get a stronger defect. This defect effect is remarkable.

Theorem 6.3.4. *Let $Y \subseteq A^+$ be the base of the ξ-free hull of a finite set $X \subseteq A^+$. If X is a code, but neither X nor $\widetilde{X}$ is an ω-code, then*

$$\mathrm{Card}(Y) \le \mathrm{Card}(X) - 2.$$

EXAMPLE 6.3.5. The set $X = \{a, ab, bbab, bbbb\}$ is a code, but neither X nor $\widetilde{X}$ is an ω-code. Indeed, $(a, bbbb, bbbb, \ldots)$, $(ab, bbbb, bbbb, \ldots)$ are X-factorization of the word ab^ω, and $(ba, bbbb, bbbb, \ldots)$, $(babb, bbbb, bbbb, \ldots)$ are $\widetilde{X}$-factorizations of the word bab^ω. The base $Y = \{a, b\}$ of the ξ-free hull of X has cardinality $\mathrm{Card}(Y) = \mathrm{Card}(X) - 2$.

The proof of Theorem 6.3.4 is based on Proposition 6.3.6. The proof of this proposition needs several steps. Section 6.3.2 below is completely dedicated to it.

PROPOSITION 6.3.6. *If X is a finite code satisfying*

$$\mathrm{Card}(\mathrm{Alph}(X)) = \mathrm{Card}(X) - 1$$

and

$$\mathrm{Alph}(X) = \mathrm{First}(X) = \mathrm{Last}(X),$$

then either X or $\widetilde{X}$ is an ω-code.

EXAMPLE 6.1.2 (*continued*). The code $X = \{a, ab, bb\}$ satisfies the condition of Proposition 6.3.6 since $\mathrm{Alph}(X) = \{a, b\} = \mathrm{First}(X) = \mathrm{Last}(X)$. It is not an ω-code, but its reversal $\widetilde{X} = \{a, ba, bb\}$ is an ω-code.

Proof of Theorem 6.3.4. Assume the contrary: let $X \subseteq A^+$ be a code with a minimal $\mathrm{Lg}(X)$ such that X, $\widetilde{X}$ are not ω-codes and the base Y of the ξ-free hull of X has cardinality

$$\mathrm{Card}(Y) \geq \mathrm{Card}(X) - 1.$$

We will show that X is a code satisfying the hypotheses of Proposition 6.3.6 but not the thesis. The main tool is the operation of composition of codes, in particular Propositions 6.2.10, 6.2.11 and Properties (6.2.1), (6.2.2).

By Theorem 6.3.2, $\mathrm{Card}(Y) \leq \mathrm{Card}(X) - 1$. Hence

$$\mathrm{Card}(Y) = \mathrm{Card}(X) - 1. \tag{6.3.1}$$

Let us prove that

$$Y = \mathrm{Alph}(X). \tag{6.3.2}$$

As Y^∞ is the ξ-free hull of X, one has $X \subseteq Y^*$. By Proposition 6.3.1, $\mathrm{Alph}_Y(X) = Y$. It follows by Proposition 6.2.11 that

$$X = Y \circ X' \tag{6.3.3}$$

for some finite code X'. Neither X' nor $\widetilde{X'}$ is an ω-code (as X and $\widetilde{X}$ are not; see Proposition 6.2.10). By (6.3.3), $\mathrm{Lg}(X') \leq \mathrm{Lg}(X)$. Suppose that $\mathrm{Lg}(X') < \mathrm{Lg}(X)$. Thus by definition of X, the base Y' of the ξ-free hull of X' satisfies

$$\mathrm{Card}(Y') \leq \mathrm{Card}(X') - 2. \tag{6.3.4}$$

As for X, write X' as the composition

$$X' = Y' \circ Z'.$$

Considering $X = Y \circ X' = Y \circ (Y' \circ Z') = (Y \circ Y') \circ Z'$, we get $X \subseteq (Y \circ Y')^* \subseteq Y^*$ and then

$$(Y \circ Y')^\infty \subseteq Y^\infty.$$

Being the composition of two ξ-codes, $Y \circ Y'$ is also a ξ-code. Hence, since Y^∞ is the ξ-free hull of X, we have $Y^\infty = (Y \circ Y')^\infty$, and then

$$Y = Y \circ Y'. \tag{6.3.5}$$

So

$$\mathrm{Card}(X) - 1 \overset{(6.3.1)}{=} \mathrm{Card}(Y) \overset{(6.3.5)}{=} \mathrm{Card}(Y') \overset{(6.3.4)}{\leq} \mathrm{Card}(X') - 2 \overset{(6.3.3)}{=} \mathrm{Card}(X) - 2$$

which is impossible. It follows that $\mathrm{Lg}(X) = \mathrm{Lg}(X')$ and by (6.3.3) $Y = \mathrm{Alph}(X)$.

We end the proof. By (6.3.1) and (6.3.2), $\mathrm{Card}(X) - 1 = \mathrm{Card}(\mathrm{Alph}(X))$. By Proposition 6.3.1, $\mathrm{First}(X) = \mathrm{Alph}(X) = \mathrm{Last}(X)$. Consequently, either X or $\widetilde{X}$ is an ω-code by Proposition 6.3.6. This brings the contradiction. ∎

6.3.2. A particular class of codes

The proof of Proposition 6.3.6 is rather long. It uses in an elegant way techniques of combinatorics on words.

Let $u, v \in A^\infty$. We use the notation $u \leq v$ when u is prefix of v, and $u < v$ when u is a proper prefix of v. The longest common prefix of u and v is denoted by $u \wedge v$. The words u, v are called *incomparable* if neither $u \leq v$ nor $v \leq u$. The set of prefixes of words in X is denoted by $\mathrm{Pref}(X)$. For suffixes, we use the notation $\mathrm{Suff}(X)$.

In a first step, we study the following particular class of codes $X \subseteq A^+$:

HYPOTHESIS 6.3.7. *$X \subseteq A^+$ is a finite code, it is not an ω-code and*

$$\mathrm{Card}(\mathrm{First}(X)) = \mathrm{Card}(X) - 1.$$

In this hypothesis, since X is not an ω-code, there is an infinite word w with two distinct X-factorizations. Hence all words of X begin with distinct letters, except for two words $x, y \in X$ such that

$$x < y.$$

We denote by Amb_X the set of words *ambiguously covered*, i.e.,

$$\mathrm{Amb}_X = \mathrm{Pref}(xX^\omega) \cap \mathrm{Pref}(yX^\omega).$$

Note that Amb_X contains w and all its prefixes.

We now prove several lemmas (6.3.8–6.3.11) under Hypothesis 6.3.7. We begin with a technical one.

LEMMA 6.3.8. *Let $u, v \in A^+$ be a pair of incomparable words.*

1. *If $u, v \in \mathrm{Pref}(xX^\omega)$ (resp. $u, v \in \mathrm{Pref}(yX^\omega)$), then $u \wedge v \in xX^*$ (resp. $u \wedge v \in yX^*$).*
2. *If $u \in \mathrm{Pref}(xX^\omega)$ and $v \in \mathrm{Pref}(yX^\omega)$, or the contrary, then $u \wedge v \in X^*$.*

Proof. Define $\mathcal{A}$ as the set of pairs (u, v) of incomparable words such that either $u, v \in \mathrm{Pref}(xX^\omega)$ and $u \wedge v \notin xX^*$, or $u, v \in \mathrm{Pref}(yX^\omega)$ and $u \wedge v \notin yX^*$. Define also $\mathcal{B}$ as the set of pairs (u, v) of incomparable words such that either $u \in \mathrm{Pref}(xX^\omega)$, $v \in \mathrm{Pref}(yX^\omega)$ and $u \wedge v \notin X^*$, or $u \in \mathrm{Pref}(yX^\omega)$, $v \in \mathrm{Pref}(xX^\omega)$ and $u \wedge v \notin X^*$. Let us prove that both $\mathcal{A}$ and $\mathcal{B}$ are empty.

(a) Let us show that if $(u, v) \in \mathcal{A}$, then there exists $(u', v') \in \mathcal{B}$ with $|u' \wedge v'| < |u \wedge v|$.

Suppose that $u, v \in \mathrm{Pref}(xX^\omega)$ and $u \wedge v \notin xX^*$. There exist $x_1, \ldots, x_n$, $y_1, \ldots, y_m \in X$, $n, m \geq 1$, such that

$$xx_1 \cdots x_{n-1} < u \leq xx_1 \cdots x_n,$$
$$xy_1 \cdots y_{m-1} < v \leq xy_1 \cdots y_m.$$

Let i maximum such that $x_j = y_j$, for all $j \in \{1, \ldots, i\}$. Let $z = xx_1 \cdots x_i$. Then $z \leq u \wedge v$. As $u \wedge v \notin xX^*$, z is a proper prefix of $u \wedge v$. Consider $x_{i+1} \neq y_{i+1}$. By Hypothesis 6.3.7, $\{x_{i+1}, y_{i+1}\} = \{x, y\}$. Hence the thesis holds for the pair (u', v') such that $u' = z^{-1}u$ and $v' = z^{-1}v$. See Figure 6.8.

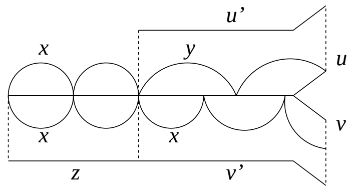

Figure 6.8. $(u, v) \in \mathcal{A}$ and $(u', v') \in \mathcal{B}$.

(b) Let us show that if $(u, v) \in \mathcal{B}$, then there exists $(u', v') \in \mathcal{A}$ with $|u' \wedge v'| \leq |u \wedge v|$.

By Hypothesis 6.3.7, there exists an infinite word $w \in \mathrm{Amb}_X$. Either $w \wedge (u \wedge v) < u \wedge v$ or $u \wedge v \leq w$ (see Figure 6.9).

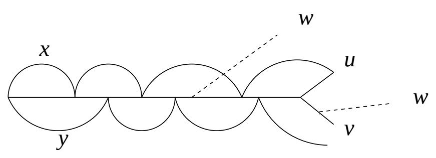

Figure 6.9. The two possible cases for w.

In the first case, let $u' \leq u \wedge v$ and $v' < w$ such that u', v' are incomparable. Then $u', v' \in \mathrm{Amb}_X$. The word $u' \wedge v'$ cannot belong to $xX^* \cap yX^*$ because X is a code. One then checks that the pair (u', v') belongs to $\mathcal{A}$.

In the second case, as u, v are incomparable, $u \wedge v = u \wedge w$ or $u \wedge v = w \wedge v$ (see Figure 6.9). We only consider the first case, the second one is handled similarly. Define $u' = u$ and $v' < w$ such that u', v' are incomparable. Since $v' \in \mathrm{Amb}_X$, it follows that $(u', v') \in \mathcal{A}$.

(c) We conclude the proof. If $\mathcal{A}$ is not empty, there exists a pair $(u, v) \in \mathcal{A}$ with minimal $|u \wedge v|$. Apply (a) and then (b) to this pair. We get a new pair $(u', v') \in \mathcal{A}$ such that $|u' \wedge v'| < |u \wedge v|$. This is impossible. A similar argument shows that $\mathcal{B}$ is empty. ■

As X is not an ω-code, we know that there exists an infinite word in Amb_X. The proposition below states that this word is unique. It constitutes a nice combinatorial property of finite codes X which are not ω-codes and such that $\mathrm{Card}(\mathrm{First}(X)) = \mathrm{Card}(X) - 1$.

PROPOSITION 6.3.9. *There exists a unique word $\sigma_X \in A^\omega$ such that*

$$\mathrm{Amb}_X = \mathrm{Pref}(\sigma_X).$$

Proof. Assume the contrary: there exist two incomparable words $u, v \in \mathrm{Amb}_X$. These words both belong to $\mathrm{Pref}(xX^\omega)$ and $\mathrm{Pref}(yX^\omega)$. Since X is a code, $u \wedge v \notin xX^*$ or $u \wedge v \notin yX^*$. In both cases, we get a contradiction with Lemma 6.3.8, part 1. ■

The next result is also interesting.

LEMMA 6.3.10. *If $u, v \in X^*$, then $u \wedge v \in X^*$.*

Proof. Suppose the contrary and take two words $u, v \in X^*$ such that $u \wedge v \notin X^*$ and $|u \wedge v|$ is minimal. It follows that u, v are incomparable words, and by minimality of $|u \wedge v|$, $u \in \mathrm{Pref}(xX^\omega)$, $v \in \mathrm{Pref}(yX^\omega)$, or the contrary. This is impossible in view of Lemma 6.3.8, part 2. ■

The following lemma shows that σ_X has two distinct X-factorizations which are eventually periodic.

LEMMA 6.3.11. *There exist $r, s, r', s' \in X^*$ such that*

$$\sigma_X = rs^\omega = r's'^\omega$$

and $\mathrm{First}_X(rs) \neq \mathrm{First}_X(r's')$. Moreover, for any $t \in X^*$,

$$\sigma_X \neq t^\omega.$$

Proof. By hypothesis, σ_X has two X-factorizations, one beginning with x and the other with y

$$(x, x_1, x_2, \ldots),\ (y, y_1, y_2, \ldots).$$

As X is a finite set, there exists a word $w \in \mathrm{Suff}(X)$ which is repeated in the following way, for some $u = x_1 \cdots x_k \in X^*$, $u' = x_{k+1} \cdots x_n \in X^+$, $v = y_1 \cdots y_\ell \in X^*$, $v' = y_{\ell+1} \cdots y_m \in X^+$:

$$xu = yvw, \ \ xuu' = yvv'w$$

(see Figure 6.10). It follows that $xuu'^\omega = yvv'^\omega = \sigma_X$ by unicity of σ_X (Propo-

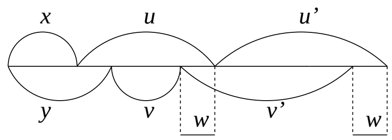

Figure 6.10. The suffix w appears twice.

sition 6.3.9). The first part of the lemma is proved.

Assume now that $\sigma_X = t^\omega$ for some $t \in X^+$. Either $\mathrm{First}_X(t) \neq \mathrm{First}_X(rs)$ or $\mathrm{First}_X(t) \neq \mathrm{First}_X(r's')$. Suppose that we are in the first case. By replacing t and s by adequate powers, say $t^i, s^j \in X^+$, we can make the assumption that $|t| = |s|$ and $|t| > |r|$. Looking at Figure 6.11, one sees that $rs = tr$ with

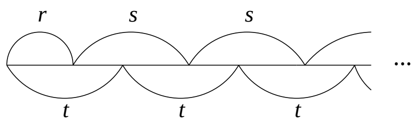

Figure 6.11. $\sigma_X = rs^\omega = t^\omega$.

$\mathrm{First}_X(rs) \neq \mathrm{First}_X(tr)$. But X is a code: contradiction. ∎

The next lemma is the last step before proving Proposition 6.3.6. There is no contradiction between this lemma and the previous one: Lemma 6.3.12 states that $\sigma_X = p^\omega$ for some $p \in A^+$; Lemma 6.3.11 states that $p \notin X^+$.

Lemma 6.3.12. *If X and $\widetilde{X}$ both satisfy Hypothesis 6.3.7, then σ_X and $\sigma_{\widetilde{X}}$ are periodic.*

Proof. By Proposition 6.3.9,

$$\mathrm{Amb}_X = \mathrm{Pref}(\sigma_X).$$

By Lemma 6.3.11,

$$\sigma_X = rs^\omega = r's'^\omega \tag{6.3.6}$$

where $r, s, r', s' \in X^+$ and $\mathrm{First}_X(r) \neq \mathrm{First}_X(r')$. With adequate powers of s, s', one can suppose that $|s| = |s'|$ and $|r|, |r'| \leq |s|$.

Since X is a code and $\mathrm{First}_X(r) \neq \mathrm{First}_X(r')$, one gets $r \neq r'$; let us say that $r'w = r$ for some $w \neq \varepsilon$. Again because X is a code, $w \notin X^*$. Consider

$$\widetilde{u} = \widetilde{r} \wedge \widetilde{s}.$$

Hence w is suffix of u (see Figure 6.12). By Lemma 6.3.10 applied to $\widetilde{X}$, $u \in X^+$ showing that $w \neq u$. Let

$$\widetilde{u}' = \widetilde{r}' \wedge \widetilde{s}'.$$

We have $u'w = u$ and $u' \in X^+$ by Lemma 6.3.10 again (see Figure 6.12).

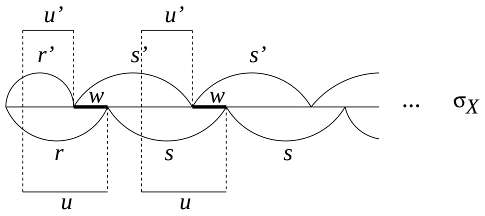

Figure 6.12. $u'w = u$ with $u, u' \in X^+$, $w \notin X^*$.

Let us factorize u, u' as $u = tr_1$, $u' = tr_1'$ with $t, r_1, r_1' \in X^*$ and t of maximal length. As $w \notin X^*$, $r_1' \neq \varepsilon$. It follows that $\mathrm{First}_X(r_1) \neq \mathrm{First}_X(r_1')$. Therefore $r_1 s^\omega = r_1' s'^\omega \in \mathrm{Amb}_X$ (see Figure 6.13). This word is equal to σ_X

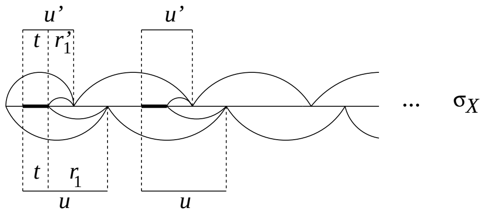

Figure 6.13. σ_X is a proper suffix of itself.

by Proposition 6.3.9. Thus observe on Figure 6.13 that σ_X is a proper suffix of σ_X. It follows that σ_X is periodic.

The same argument applied to $\widetilde{X}$ instead of X shows that $\sigma_{\widetilde{X}}$ is periodic. ∎

Remark 6.3.13. In the previous proof, if r, r' are chosen with minimal length in (6.3.6), then $r = r_1$, $r' = r_1'$ and r is suffix of s, r' is suffix of s' (see Figure 6.14).

We are now ready to prove Proposition 6.3.6. The proof is *ad absurdum*. In particular, X, $\widetilde{X}$ both satisfy Hypothesis 6.3.7. The key results to get the contradiction are Proposition 6.3.9 and Lemmas 6.3.10-6.3.12. The proof goes

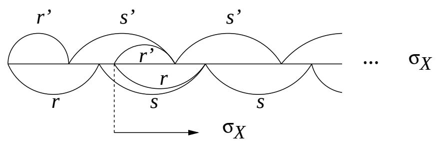

Figure 6.14. The words r, r' are chosen with minimal length.

back and forth between X and $\widetilde{X}$. To avoid any confusion, we systematically use the reversal notation, e.g. $\widetilde{v} \in \widetilde{X}^*$, whenever we work with $\widetilde{X}$.

Proof of Proposition 6.3.6. Assume that X and $\widetilde{X}$ are not ω-codes. As in the beginning of the previous proof, take

$$\sigma_X = rs^\omega = r's'^\omega$$

with $r, s, r', s' \in X^+$ and $|s| = |s'| \geq |r|, |r'|$. Recall that $r'w = r$ with

$$w \notin X^*.$$

We also choose r, r' with minimal length as in Remark 6.3.13. Consequently, $\mathrm{Last}_X(r) \neq \mathrm{Last}_X(s)$ and $\mathrm{Last}_X(r') \neq \mathrm{Last}_X(s')$. By Remark 6.3.13, $\widetilde{r} \leq \widetilde{s}$ and $\widetilde{r'} \leq \widetilde{s'}$.

Define $\widetilde{v} \in \mathrm{Pref}(\widetilde{X}^\omega)$ of maximal length such that (see Figure 6.15)

$$\begin{aligned} \widetilde{u} &= \widetilde{r}\,\widetilde{v} \leq \widetilde{s}^\omega, \\ \widetilde{u}' &= \widetilde{r}'\widetilde{v} \leq \widetilde{s}'^\omega. \end{aligned} \tag{6.3.7}$$

As $\mathrm{Last}_X(r) \neq \mathrm{Last}_X(s)$, $\mathrm{Last}_X(r') \neq \mathrm{Last}_X(s')$, it follows that $\widetilde{u}, \widetilde{u}' \in \mathrm{Amb}_{\widetilde{X}}$, hence

$$\widetilde{u}, \widetilde{u}' \in \mathrm{Pref}(\sigma_{\widetilde{X}}).$$

If $\widetilde{u}$ is an infinite word, then $\widetilde{u} = \sigma_{\widetilde{X}} = \widetilde{s}^\omega$, with $\widetilde{s} \in \widetilde{X}^+$, a contradiction with Lemma 6.3.11. Thus $\widetilde{u}$ is a finite word. In the same way, $\widetilde{u}'$ is a finite word. This situation is summarized in Figure 6.15.

Let us go further. By Lemma 6.3.10, as $\widetilde{u} = \widetilde{s}^\omega \wedge \sigma_{\widetilde{X}}$ and $\widetilde{u}' = \widetilde{s}'^\omega \wedge \sigma_{\widetilde{X}}$, then

$$\widetilde{u}, \widetilde{u}' \in \widetilde{X}^+.$$

However

$$\widetilde{v} \notin \widetilde{X}^*.$$

Otherwise, let $a \in \mathrm{Alph}(\widetilde{X})$ such that $\widetilde{r}\widetilde{v}a < \widetilde{s}^\omega$ (see Figure 6.15). As $\mathrm{First}(\widetilde{X}) = \mathrm{Alph}(\widetilde{X})$, there exists $\widetilde{z} \in \widetilde{X}$ with $\mathrm{First}(\widetilde{z}) = a$. Therefore $\widetilde{v}a$ is a word longer than $\widetilde{v}$ such that $\widetilde{v}a \in \mathrm{Pref}(\widetilde{X}^\omega)$ and $\widetilde{r}\widetilde{v}a < \widetilde{s}^\omega$, a contradiction with the definition (6.3.7) of $\widetilde{v}$.

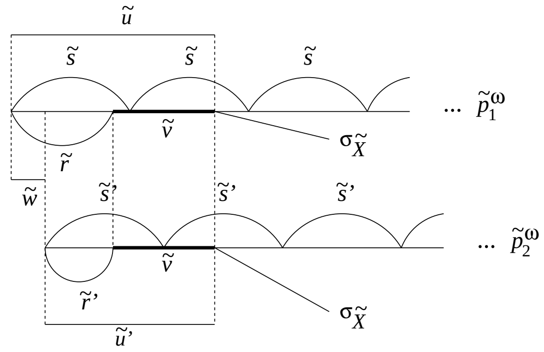

Figure 6.15. $\widetilde{u}, \widetilde{u}' \in \widetilde{X}^+$ but $\widetilde{v} \notin \widetilde{X}^*$.

Now, by Lemma 6.3.12, there exist primitive words $p, \widetilde{q} \in A^+$ such that $\sigma_X = p^\omega$, $\sigma_{\widetilde{X}} = \widetilde{q}^\omega$. We can suppose that $|p| \geq |\widetilde{q}|$. By Fine and Wilf's Theorem (Proposition 1.2.1), as $\sigma_X = p^\omega = rs^\omega = r's'^\omega$, we have

$$s \in p_1^+, s' \in p_2^+$$

where p_1, p_2 and p are conjugate words.

Observe that the word $\widetilde{u}$ cannot be too long

$$|\widetilde{u}| < |p| + |q|. \tag{6.3.8}$$

Indeed, if $|\widetilde{u}| \geq |p| + |q|$, as $\widetilde{u} \in \mathrm{Pref}(\widetilde{s}^\omega) \cap \mathrm{Pref}(\sigma_{\widetilde{X}}) = \mathrm{Pref}(\widetilde{p_1}^\omega) \cap \mathrm{Pref}(\widetilde{q}^\omega)$, we get $\widetilde{p_1} = \widetilde{q}$ by Fine and Wilf's Theorem. In particular, $\widetilde{s} \in \widetilde{q}^+$. This implies that $\sigma_{\widetilde{X}} = \widetilde{s}^\omega$ with $\widetilde{s} \in \widetilde{X}^+$. This is impossible by Lemma 6.3.11.

To get the final contradiction, we show that Inequality (6.3.8) never holds. For this, we come back to σ_X in the following way. Decompose $u, u' \in X^+$ as

$$u = tr_1, u' = tr_1'$$

where $t, r_1, r_1' \in X^*$ and t is of maximal length (see Figure 6.16). We have $r_1' \neq \varepsilon$ because $w \notin X^*$, and $r_1' \neq r'$ because $v \notin X^*$. By definition of t, $\mathrm{First}_X(r_1) \neq \mathrm{First}_X(r_1')$ and then $r_1 s^\omega = r_1' s'^\omega = \sigma_X$ (Proposition 6.3.9). We have $|r'| < |r_1'|$ by minimality of $|r'|$ and because $r' \neq r_1'$. So r' is a proper suffix of r_1' as depicted in Figure 6.16. Hence σ_X is a proper suffix of σ_X, that is $\sigma_X = t_1 \sigma_X$ with $v = tt_1$. Thus $\sigma_X = p^\omega = t_1^\omega$. As p is primitive, it follows that

$$|v| \geq |p|$$

(see Figure 6.16). Now let us look again at Figure 6.15. The words $\widetilde{u}'$ and $\widetilde{u} = \widetilde{w}\widetilde{u}'$ are prefix of $\widetilde{q}^\omega$ because $\sigma_{\widetilde{X}} = \widetilde{q}^\omega$. Since $|\widetilde{u}'|, |\widetilde{u}| \geq |v| \geq |p| \geq |q|$ and q is primitive, we deduce that $|\widetilde{w}| \geq |\widetilde{q}|$. Hence $|\widetilde{u}| \geq |\widetilde{w}| + |\widetilde{v}| \geq |\widetilde{q}| + |\widetilde{p}|$. This ends the proof. ∎

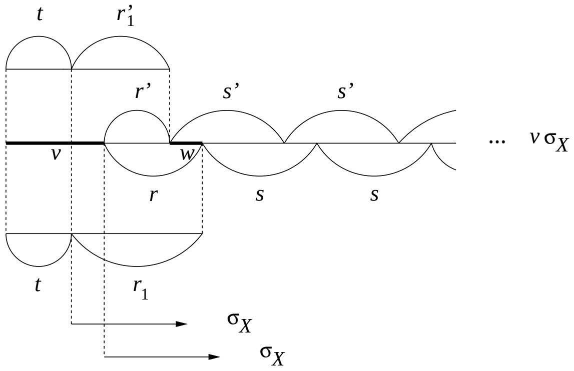

Figure 6.16. Back to σ_X.

6.3.3. Three-element codes

Theorem 6.3.4 leads to the following nice property of codes with three elements. This property is no longer true for larger codes (see Example 6.3.5).

COROLLARY 6.3.14. *Let $X = \{x, y, z\}$ be a code. Then X or $\widetilde{X}$ is an ω-code.*

Proof. If not, by Theorem 6.3.4, $X \subseteq t^*$ where t^∞ is the ξ-free hull of X. This is impossible since X is a code. ∎

Another property which is characteristic of three-element codes, but does not hold for arbitrary ones, is given in the next proposition.

PROPOSITION 6.3.15. *For a three-element code $X = \{x, y, z\}$, there exists at most one infinite word σ with two X-factorizations $(x_1, x_2, \ldots)$, $(y_1, y_2, \ldots)$ such that $x_1 \neq y_1$.*

Proof. If X is an ω-code, such an infinite word σ cannot exist. Suppose that X is not an ω-code. If $\mathrm{Card}(\mathrm{First}(X)) = \mathrm{Card}(X) - 1$, we have done by Proposition 6.3.9. Otherwise let

$$u = x^\omega \wedge y^\omega \wedge z^\omega \neq \varepsilon.$$

By Fine and Wilf's Theorem (Proposition 1.2.1)

$$|u| < \min\{|x| + |y|, |y| + |z|, |z| + |x|\}.$$

Indeed, if $|u| \geq |x| + |y|$ for instance, then $x^i = y^j$ with $i, j \geq 1$. This is impossible because X is a code.

We show that for any $v \in \mathrm{Pref}(X^*)$

$$|u| \leq |v| \Rightarrow u \leq v. \tag{6.3.9}$$

Let v' be the prefix of v such that $|u| = |v'|$. If $v' \in \mathrm{Pref}(x^\omega)$, then $u = v'$ by definition of u. If $v' = x^n y'$ with $y' \in \mathrm{Pref}(y) - \varepsilon$ and $n \geq 1$, then $v' = u$ because $y' < u < x^\omega$. The other cases are proved similarly.

By (6.3.9), u is prefix of each word xu, yu, zu. We define a new three-element code $X' = \{x', y', z'\}$ such that $x' = u^{-1}xu$, $y' = u^{-1}yu$, $z' = u^{-1}zu$. By (6.3.9), $w \in A^\omega$ has an X-factorization $(x_1, x_2, \ldots)$ if and only if $u^{-1}w$ has an X'-factorization $(u^{-1}x_1u, u^{-1}x_2u, \ldots)$. So X' is not an ω-code. Moreover the condition $\mathrm{Card}(\mathrm{First}(X')) = \mathrm{Card}(X') - 1$ is now satisfied. Consequently, there exists exactly one infinite word $\sigma_{X'}$ with two distinct X'-factorizations (Proposition 6.3.9). The conclusion follows for X with $\sigma = u\sigma_{X'}$. ∎

Example 6.3.16. The code $X = \{ba, bab, bb\}$ is such that $\mathrm{Card}(\mathrm{First}(X)) < \mathrm{Card}(X) - 1$. As in the previous proof, we construct the new code $X' = b^{-1}Xb = \{ab, abb, bb\}$, with $\sigma_{X'} = ab^\omega$. Thus there is a unique infinite word $\sigma = bab^\omega$ with two X-factorizations $(ba, bb, bb, \ldots)$ and $(bab, bb, bb, \ldots)$ that begin with distinct words of X.

The situation is much more complex for larger codes.

Example 6.3.17. We associate with the code $X = \{ba, bab, abaa, aabaab\}$ the graph of Figure 6.17. An edge $u \to v$ means that there exists $x \in X$ such that

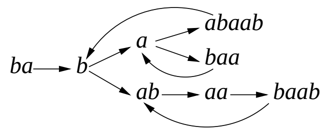

Figure 6.17. Graph of infinite words with two X-factorizations.

$u = xv$ or $uv = x$. The infinite paths beginning with ba thus describe words of A^ω with two distinct X-factorizations. For instance the infinite path

$$ba \to b \to a \to baa \to a \to abaab \to b \to a \to baa \to a \to \ldots$$

describes the overlappings of Figure 6.18.

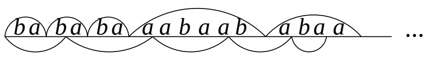

Figure 6.18. From a path to the associated pair of Y-factorizations.

6.4. A theorem of Schützenberger

In this last section, we want to describe a remarkable property of finite maximal codes. A code $X \subseteq A^+$ is *maximal* if it cannot be strictly included in a code

Y over the same alphabet A. The next theorem states two extremal behaviors of finite codes $X \subseteq A^+$ which are maximal: either X is not an ω-code or X is a prefix code. If one recalls that finite ω-codes are exactly finite codes with bounded decoding delay (Proposition 6.1.9), another interpretation of this result is: the decoding delay of a finite maximal code is either null or not bounded.

THEOREM 6.4.1 (Schützenberger's Theorem). *Let $X \subseteq A^+$ be a finite maximal code. If X is an ω-code, then X is a prefix code.*

EXAMPLE 6.4.2. The set $X = \{a, ab, bb\}$ is a maximal code for the following reasons. One can easily prove (by induction on $|w|$) that any $w \in \{a,b\}^*$ belongs to $X^* \cup bX^*$. Assume that $Y = X \cup \{w\}$ is a code for some word $w \notin X$. Then $aw \in X^*$. This shows that aw has two distinct Y-factorizations: one X-factorization and the Y-factorization (a, w). Contradiction. Hence X is a maximal code. The set X is not a prefix code, and thus not an ω-code by Theorem 6.4.1. The reversal code $\widetilde{X}$ is prefix.

The next two examples show that the hypotheses of Theorem 6.4.1 are necessary.

EXAMPLE 6.4.3. The infinite code $X = ab^*$ is a maximal code. It is an ω-code which is not prefix.

EXAMPLE 6.4.4. The set $X = \{a, aba\}$ is a code which is not maximal. For instance, $X \cup \{bb\}$ is still a code. Both X and $\widetilde{X}$ are ω-codes without being prefix codes.

Proof of Theorem 6.4.1. (1) We first show that if $X \subseteq A^+$ is a maximal code, then X is *complete*, i.e.,

$$\forall w \in A^*, X^*wA^* \cap X^* \neq \emptyset \tag{6.4.1}$$

(see also Problem 6.4.1). This is trivially true for any alphabet $A = \{a\}$. Suppose that $\mathrm{Card}(A) \geq 2$ and some $w \in A^*$ satisfies

$$X^*wA^* \cap X^* = \emptyset. \tag{6.4.2}$$

In this equality, we can suppose that w is *unbordered*, i.e., if $w \in uA^+ \cap A^+u$, then $u = \varepsilon$. Indeed, if $\mathrm{First}(w) = a$, replace w by the unbordered word $wab^{|w|}$ such that $b \in A$, $b \neq a$.

The set $Y = X \cup \{w\}$ is not a code by hypothesis. So there exists a word $z \in Y^*$ with two Y-factorizations $(x_1, \ldots, x_n)$, $(y_1, \ldots, y_m)$ such that $x_1 \neq y_1$. As X is a code and by (6.4.2), w must appear among the x_i's *and* the y_j's. Consider the first occurrences of w

$$x_1, \ldots, x_{i-1} \in X,\ \ x_i = w,$$
$$y_1, \ldots, y_{j-1} \in X,\ \ y_j = w.$$

Again by (6.4.2), x_i and y_j must overlap and thus coincide since w is unbordered. Therefore, as $x_1 \neq y_1$, we have $i, j \geq 2$ and the word $x_1 \cdots x_{i-1}$ has two distinct X-factorizations $(x_1, \ldots, x_{i-1})$, $(y_1, \ldots, y_{j-1})$. This is impossible because X is a code.

(2) We now make the assumption that X is an ω-code which is not prefix. By Proposition 6.1.9, X has a decoding delay $d > 0$. Hence, define $t \in \mathrm{Pref}(X^*)$ of maximal length such that

$$xtA^* \cap yX^* \neq \emptyset, \text{ with } x, y \in X, x \neq y \tag{6.4.3}$$

(see Figure 6.19). Note that t is well defined since $0 < d < \infty$ and X is finite.

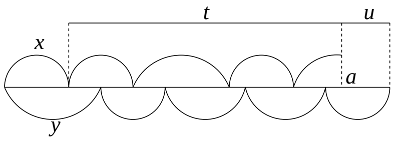

Figure 6.19. X is an ω-code which is not prefix.

Let $u \in A^*$ such that $xtu \in yX^*$. We can suppose that $u \neq \varepsilon$. Indeed, if $u = \varepsilon$, replace u by any element of X. We denote by a the first letter of u.

Let $w = zta$ such that z is a word of X of maximal length (recall that X is finite). By (6.4.1), there exist $x_1, \ldots, x_n, y_1, \ldots, y_m \in X$ and $v \in A^*$ such that

$$x_1 \cdots x_n wv = y_1 \cdots y_m$$

(see Figure 6.20). As $|t|$ is maximal for property (6.4.3), $x_k = y_k$, for all $k \in$

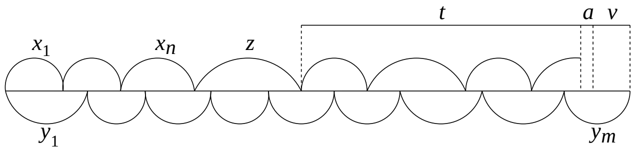

Figure 6.20. X is a maximal code.

$\{1, \ldots, n\}$. By definition of z, we have $y_{n+1} \leq z$. Assume that $y_{n+1} < z$. Define $t' \in \mathrm{Pref}(y_{n+2} \cdots y_m)$ such that $y_{n+1}t' = zt$ (see Figure 6.21). It follows that t' satisfies (6.4.3) with $|t'| > |t|$. This is impossible. Hence $y_{n+1} = z$ showing that ta belongs to $\mathrm{Pref}(y_{n+2} \cdots y_m) \subseteq \mathrm{Pref}(X^*)$. We have again a contradiction (see Figure 6.19): ta satisfies (6.4.3) since a is the first letter of u and $|ta|$ is longer than $|t|$. The conclusion is that X is necessarily a prefix code. ■

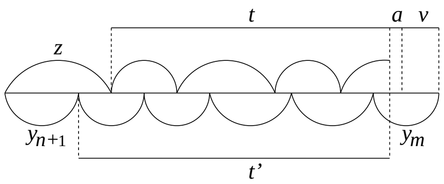

Figure 6.21. y_{n+1} is a proper prefix of z.

Problems

Section 6.1

6.1.1 Show that a subset X of A^+ is a code if and only if for any $x \in X^+$, x^ω has only one X-factorization.

6.1.2 Show that a submonoid M of A^* is stable if and only if for any $u, v \in M$

$$uv, u, vu \in M \Rightarrow v \in M.$$

6.1.3 a. Show that for a rational code $X \subseteq A^+$, X has a bounded decoding delay if and only if X is an ω-code and $X^\omega \cap X^* \operatorname{Adh}(X) = \emptyset$, where

$$\operatorname{Adh}(X) = \{w \in A^\omega \mid \operatorname{Pref}(w) \subseteq \operatorname{Pref}(X)\}$$

(see Proposition 6.1.9).

b. Give an example of a rational ω-code which has not a bounded decoding delay.

6.1.4 Find an example of a family $(X_n)_{n\in\mathbb{N}}$ of codes with bounded decoding delay such that the base of $\cap_{n\in\mathbb{N}} X_n^*$ is a code which has not a bounded decoding delay. Compare with Corollary 6.1.7.

6.1.5 Let $A^{\mathbb{Z}}$ be the set of two-sided infinite words over A and A^ς the set of the equivalence classes on $A^{\mathbb{Z}}$ under the shift (observe the difference between A^ς and A^ζ as defined in Chapter 1). We denote by X^ς the subset of A^ς equal to

$$X^\varsigma = \{\cdots x_{-1} x_0 x_1 x_2 \cdots \mid x_n \in X, n \in \mathbb{Z}\}$$

Given $X \subseteq A^+$, an *X-decomposition* of $w \in A^{\mathbb{Z}}$ is a strictly increasing sequence $(i_n)_{n\in\mathbb{Z}}$ of integers such that $i_0 \geq 0$, $i_{-1} < 0$ and $w_{[i_n, i_{n+1}[} \in X$. An *X-factorization* of $w \in A^\varsigma$ is a sequence $(x_n)_{n\in\mathbb{Z}}$ of words of X such that

$$w = \cdots x_{-1} x_0 x_1 x_2 \cdots.$$

A set $X \subseteq A^+$ is called a *$\mathbb{Z}$-code* (resp. *ς-code*) if any element of $A^{\mathbb{Z}}$ (resp. A^ς) has at most one X-decomposition (resp. X-factorization).

a. Show that any $\mathbb{Z}$-code is a ς-code. Give an example of a ς-code which is not a $\mathbb{Z}$-code.

b. Let φ be a bijection between B and $X \subseteq A^+$. Prove that φ is injective on B^ς if and only if X is a ς-code. Prove that φ is injective on B^ς and for any $r \in B^+$

$$r \text{ primitive } \Rightarrow \varphi(r) \text{ primitive}$$

if and only if X is a $\mathbb{Z}$-code.

Section 6.2

*6.2.1 Let X be a subset of A^+. Define a sequence $(M_n)_{n\in\mathbb{N}}$ of submonoids of A^* by

$$M_0 = X^*, M_{n+1} = \left(M_n^{-1}M_n \cap M_nM_n^{-1}\right)^*.$$

Set $M = \cup_{n\in\mathbb{N}} M_n$.

a. Show that M is the free hull of X.

b. Prove that M is rational if X is rational.

6.2.2 Find a method for constructing the ω-free hull of a subset X of A^+.

6.2.3 a. Find a definition for the $\mathbb{Z}$-free hull of a finite set X.

b. If Y is the base of the $\mathbb{Z}$-free hull of X, prove that

$$\mathrm{Card}(Y) \le \mathrm{Card}(X)$$

and that no better upper bound exists.

c. Prove that

$$\mathrm{Card}(Y) \le \mathrm{Card}(X) - 1$$

if there exists a not periodic word $w \in A^{\mathbb{Z}}$ with two X-decompositions.

(See Problem 6.1.5).

6.2.4 A submonoid M of A^* is *right-unitary* if

$$u, uw \in M \Rightarrow w \in M.$$

a. Show that $X \subseteq A^+$ is a prefix code if and only if it is the base of a right-unitary submonoid M of A^*.

b. Given a finite subset X of A^+, let Y be the base of the smallest right-unitary submonoid which contains X. Prove that Y exists and that

$$\mathrm{Card}(Y) \le \mathrm{Card}(X),$$

showing that there is no defect.

c. Give an example where the previous inequality becomes an equality.

6.2.5 Given a finite set $X \subseteq A^+$, we define four different ranks: the *free rank* $r_f(X)$ equal to $\mathrm{Card}(Y)$ with Y the base of the free hull of X, similarly the *ω-free rank* $r_\omega(X)$, the *right-unitary rank* $r_u(X)$, and finally the *combinatorial rank* $r(X)$ equal to $\min\{\mathrm{Card}(Y) \mid X \subseteq Y^*\}$.

a. Show that $r(X) \leq r_u(X) \leq r_\omega(X) \leq r_f(X) \leq \mathrm{Card}(X)$.
b. Find an example with strict inequalities.

6.2.6 a. Give a direct proof that any two-element code is an ω-code, i.e., without using the defect theorem.
b. Show that the constant d of the decoding delay can be arbitrarily large.

6.2.7 A morphism $\varphi : B^\infty \to A^\infty$ is *simplifiable* if $\varphi = \psi \circ \phi$ with $\phi : B^\infty \to C^\infty$, $\psi : C^\infty \to A^\infty$ two morphisms such that $\mathrm{Card}(C) \leq \mathrm{Card}(B) - 1$. If no such alphabet C exists, φ is called *elementary*. Show that a finite set $X \subseteq A^+$ is elementary if and only if there exists an elementary morphism $\varphi : B^\infty \to A^\infty$ such that $X = \varphi(B)$.

*6.2.8 Let $X \subseteq A^+$ be an elementary set. Show that the decoding delay of the ω-code X is bounded by

$$\sum_{x \in X} |x| \; - \; \mathrm{Card}(X)$$

(use Problem 6.2.9).

6.2.9 Show that if $X = Y \circ Z$ with Y and Z two codes with decoding delay d_Y, d_Z respectively, then X is a code with decoding delay $d_X \leq d_Y + d_Z$.

*6.2.10 A code X is called *prefix-suffix composed* if

$$X = X_1 \circ X_2 \circ \ldots \circ X_k$$

with each X_n being a prefix or a suffix code.
a. Show that any two-element code is prefix-suffix composed.
b. Verify that the three-element code $X = \{a, aba, babaab\}$ is not prefix-suffix composed.
c. Find a finite maximal code which contains X (see Section 6.4 for the definition of a maximal code).

**6.2.11 a. Prove that every prefix-suffix composed code with n elements, $n \geq 3$, uses at most $2n - 3$ prefix and suffix codes.
b. Show that this upper bound is tight.

Section 6.3

6.3.1 Show that any ς-code is a ξ-code (see Problem 6.1.5 for the definition of a ς-code).

6.3.2 Let X be a subset of A^+ and Y be the base of its ξ-free hull.
a. Find an example of a set X which is not a code and such that $\mathrm{Card}(Y) \leq \mathrm{Card}(X) - 3$.
b. Find another example such that $\mathrm{Card}(Y) = \mathrm{Card}(X) - 2$.
c. Compare with Theorem 6.3.4 and Example 6.3.3.

6.3.3 Lemma 6.3.11 states that there are two pairs (r, s), (r', s') of words $r, s, r', s' \in X^*$ such that $\mathrm{First}_X(rs) \neq \mathrm{First}_X(r's')$ and $\sigma_X = rs^\omega = r's'^\omega$. Take the words r, s, r', s' with minimal length. Show that there exists no other such pair.

*6.3.4 A variant of Theorem 6.3.4 is: If X is a code, but X and $\widetilde{X}$ are not ω-codes, then there exists Y such that $X \subseteq Y^*$ and $\mathrm{Card}(Y) \leq \mathrm{Card}(X) - 2$ (Y is not assumed to be the base of the ξ-free hull of X). Give a proof of this result, using Proposition 6.3.6 and elementary morphisms (see Problem 6.2.7 for the definition of an elementary morphism).

6.3.5 Prove that if $X = \{x, y\}$ is a code with two elements, then $\mathrm{Amb}_X = \mathrm{Pref}(xy \wedge yx)$.

*6.3.6 Let X be a three-element code which is not an ω-code. Prove that there exists no ω-code $Y \subseteq X^*$ such that $X^\omega = Y^\omega$.

Section 6.4

**6.4.1 Prove that a rational code is maximal if and only if it is complete.

*6.4.2 An elementary set $X \subseteq A^+$ is *maximal elementary* if X is not strictly included in any elementary set over A. Prove that an elementary set $X \subseteq A^+$ is maximal elementary if and only if $\mathrm{Card}(X) = \mathrm{Card}(A)$.

Notes

The theory of codes is a well-developed branch of Theoretical Computer Science. We refer the reader to the books of Berstel and Perrin 1985, Shyr 1991, Jürgensen and Konstantinidis 1997.

Codes are investigated in depth for the first time in Schützenberger 1956, which contains Proposition 6.1.6. The notion of ω-code appears in Staiger 1986 under the name of ifl-codes. Codes for two-sided infinite words are introduced in Devolder and Timmerman 1992 (see Problem 6.1.5). It is also possible to consider codes composed with finite and infinite words (see Do Long Van 1982). This approach is not considered here.

The defect theorem stating that if a set X with n elements is not a code, then there exists Y with at most $n-1$ words such that $X \subseteq Y^*$, is folklore. It has been proved under various forms (Skordev and Sendov 1961, Lentin 1972, Makanin 1976, Ehrenfeucht and Rozenberg 1978). A defect theorem for sets X which are not ω-codes appears in Linna 1977. The proof given here for Theorems 6.2.1 and 6.2.4 is from Berstel, Perrin, Perrot, and Restivo 1979. Proposition 6.2.3 appears in Harju and Karhumäki 1986. See the chapter *Combinatorics on words* in the Handbook of Formal Languages for a presentation of the defect theorem and related material, Choffrut and Karhumäki 1997. Defect properties for other structures like two-sided infinite words or trees are studied in Karhumäki, Maňuch, and Plandowski 1998b, Mantaci and Restivo 1999 and Mantaci and Karhumäki 1999.

Elementary sets are related to elementary morphisms, a notion used in an elegant way for one of the proofs of the D0L equivalence problem given in Ehrenfeucht and Rozenberg 1978 (see also Rozenberg and Salomaa 1980).

A measure of the defect can be evaluated thanks to the combinatorial rank of a finite set X defined by $r(X) = \min\{\mathrm{Card}(Y) \mid X \subseteq Y^*\}$. Other notions of rank can be defined such as the cardinality of the base of the free hull of X. See Harju and Karhumäki 1986 for a comparison of different kinds of ranks (see also Problem 6.2.5).

Algorithmic questions related to the defect effect are treated in several papers. The computation of the free hull of X is given in Spehner 1975 when X is finite and in Berstel et al. 1979 when X is rational (see Problem 6.2.1). The complexity results for the rank $r(X)$ are due to Néraud 1990a, Néraud 1993. It is there proved that deciding for a given finite set X and a given number k, whether $r(X) \leq k$ is a NP-complete problem. The choice $k = 2$ makes the problem computationally easy as it can be solved in time $O(n \ln^2 m)$ with $n = \mathrm{Lg}(X)$ and $m = \max\{|x| \mid x \in X\}$.

Theorem 6.3.4 is a variant of a result by Honkala 1988 (where Y is not necessarily equal to the base of the ξ-free hull of X; see Problem 6.3.4). The proofs given in Section 6.3 to get Theorem 6.3.4 are strongly based on the material developed in Karhumäki 1985a, Karhumäki 1985b. It would be nice to have a shorter proof of this remarkable result. Corollary 6.3.14 and Proposition 6.3.15 about three-element codes are due to Karhumäki 1985a, Karhumäki 1985b.

Theorem 6.4.1 is due to Schützenberger 1966, solving a conjecture of Gilbert and Moore 1959. The original proof is tricky and in a certain way magic. The proof given in this chapter comes from Bruyère 1992, where another simple proof is also given which is based on properties of automata. The fact that for rational sets, complete codes are equivalent to maximal codes is a basic result in the theory of codes (see Schützenberger 1956, Berstel and Perrin 1985).

Let us mention some open problems.

Well-known algorithms exist to decide whether a given rational submonoid of A^* is generated by a code (see Berstel and Perrin 1985). Deciding whether a rational subset X of A^ω is equal to Y^ω with Y an (finite) ω-code is still an open question. This problem is solved in Litovsky 1991 with the condition that Y is an ω-code replaced by the condition that Y is a prefix code. Other partial interesting results can be found in Devolder 1999. See also Problem 6.3.6.

Some questions on three-element codes still remain open. One is the existence of a code $X = \{x, y, z\}$ which cannot be included in any finite maximal code. It is conjectured in Restivo, Salemi, and Sportelli 1989 that any three-element code is prefix-suffix composed, which implies that such an example X does not exist. In Derencourt 1996 a family of counter-examples to the conjecture of Restivo et al. 1989 is given, among which the not prefix-suffix composed code $\{a, aba, babaab\}$ of Problem 6.2.10. However all these examples can be included in a finite maximal code.

Another conjecture on three-element codes is proposed in Devolder 1993: a code $X = \{x, y, z\}$ such that

$$w^n \in X^* \Rightarrow w \in X^*$$

is a π-code, that is, any periodic word has at most one X-factorization.

Problem 6.1.1 is from Devolder, Latteux, Litovsky, and Staiger 1994. Problem 6.1.3 is from Devolder et al. 1994. Exercises 6.1.4 and 6.2.1 appear in Berstel et al. 1979. Problem 6.2.3 c. is from Karhumäki et al. 1998b. Problem 6.2.5 is solved in Harju and Karhumäki 1986. Exercise 6.2.8 comes from Rozenberg and Salomaa 1980. Problem 6.2.11 is solved in Derencourt 1996. Problem 6.3.3 comes from Karhumäki 1985a. Problem 6.3.4 is solved in Honkala 1988. Problem 6.3.6 appears in Julia 1996. Problem 6.4.2 is from Néraud 1990b.