CHAPTER 8

Periodicity

8.0. Introduction

Periodicity is an important property of words that has applications in various domains. The first significant results on periodicity are the theorem of Fine and Wilf and the critical factorization theorem. These two results refer to two kinds of phenomena concerning periodicity: the theorem of Fine and Wilf considers the simultaneous occurrence of different periods in one finite word, whereas the critical factorization theorem relates local and global periodicity of words. Starting from these basic results the study of periodicity has grown along both directions. This chapter contains a systematic and self-contained exposition of this theory, including very recent results.

In section 8.1 we analyze the structure of the set of periods of one finite word. This section includes a proof of the theorem of Fine and Wilf and also a generalization of this result to words having three periods. We next give the characterization of Guibas and Odlyzko concerning those sets of integers that yield the periods that can simultaneously occur in a single finite word. Another property is further investigated (similar to the one stated by the theorem of Fine and Wilf) in which the occurrence of two periods in a word of a certain length forces the word to have a shorter period only in a prefix (or suffix) of the word. The golden ratio appears in such a result as an extremal value of a parameter involved in the property. This section also contains some results concerning the squares that can appear as factors in a word. This is a prelude to next section, since squares describe a special kind of local periodicity.

In section 8.2 we investigate the relation between local and global periodicity. Local periodicity is described in terms of repetitions. A repetition occurring in a word is not in general a factor of the word, nor it is necessarily a square, but it may be of rational (not only integer) order. Moreover the repetition is referred to a "point" of the word and it is important to consider the relative positions of the repetition and that of the point at which the repetition is detected. Thus we distinguish between central repetitions and left (or right) repetitions. The study of central repetitions leads to the critical factorization theorem, of which we here report a new short proof. In such a result we need repetitions of order greater than or equal to 2 and the value 2 is proved to be tight. The study of left repetitions leads to a similar result, but in this case we need repetitions of order greater than φ^2, where φ denotes the golden ratio, and the value φ^2 is tight for such a result.

The last section is devoted to infinite words. By extending the ideas and the results of section 8.2, we characterize recurrence, periodicity and eventual periodicity of infinite words in terms of local periods. In some of these results the golden ratio again plays a central role.

8.1. Periods in a finite word

8.1.1. Definitions and basic properties

Let $w = a_1 a_2 \cdots a_n$ be a word of length n over the alphabet A.

Recall from Section 1.2.1 that a positive integer $p \le |w|$ is a *period* of w if $a_{i+p} = a_i$ for $i = 1, \ldots, n-p$. The smallest period p of w is called *the* period of w and it is denoted by $p(w)$. From the definition it follows that, if v is a factor of w, then $p(v) \le p(w)$.

The positive rational number $|w|/p(w)$ is called the *order* of w and it is denoted by $\mathrm{ord}(w)$. If u is the prefix of length $p(w)$ of w, we can write $w = u^\rho$ where $\rho = \mathrm{ord}(w)$, and we say that w is a *rational power* of u. Notice that a rational power u^ρ is defined only if $|u|\rho$ is an integer. For instance, $p(abaaba) = 3$, $\mathrm{ord}(abaaba) = 2$ and the word $abaaba$ can be uniquely written $abaaba = (aba)^2$. As another example, $p(ababaaba) = 5$, $\mathrm{ord}(ababaaba) = 1.6$ and the word $ababaaba$ can be written in a unique way as $ababaaba = (ababa)^{1.6}$.

A word v that is both a prefix and a suffix of another word w, with $v \ne w$, is called a *border* of w. It is easy to see that if v is a border of w, then $|w| - |v|$ is a period of w and, conversely, if p is a period of w, then the prefix v of w of length $|w| - p$ is a border of w. The empty string ε is a border of any string w. If there exists a nonempty border v of w then w is called *bordered*, otherwise it is called *unbordered*.

It is easy to verify that a word is unbordered if and only if $\mathrm{ord}(w) = 1$, or, equivalently, if and only if $|w| = p(w)$.

The following three lemmas will be often used in this chapter. We invite the reader to spend some time in reading the proof of these lemmas together with Problem 8.1.1, to get acquainted with the basic tools and ideas used in this chapter.

Lemma 8.1.1. *Let w be a word having two periods p and q, with $q < p \le |w|$. Then the suffix and the prefix of w of length $|w| - q$ have both period $p - q$.*

Proof. We prove only that the prefix of w of length $|w| - q$ has period $p - q$, the proof for the suffix being analogous. Since $|w| - q \ge p - q$, we have to prove that

$$a_{i+p-q} = a_i \qquad i = 1, \ldots, n-p\,.$$

Let i be such that, $1 \leq i \leq n-p$. Thus $1 \leq i+p-q \leq n-q$. Since w has period p, one has that $a_i = a_{i+p}$. Since w has period q and $1 \leq i+p-q \leq n-q$, one has that $a_{i+p-q} = a_{i+p}$. ∎

LEMMA 8.1.2. *Let u, v, w be words such that uv and vw have period p and $|v| \geq p$. Then the word uvw has period p.*

Proof. Let $uvw = a_1 \cdots a_n$, $u = a_1 \cdots a_l$, $v = a_{l+1} \cdots a_j$, $w = a_{j+1} \cdots a_n$. By the hypothesis $j - l \geq p$. Let i be an integer with $1 \leq i \leq n-p$. We have to prove that $a_i = a_{i+p}$.

If $i \leq j-p$, since uv has period p, then $a_i = a_{i+p}$. If $i > j-p$, since $j-l \geq p$, then $i \geq l+1$. Since vw has period p, then $a_i = a_{i+p}$. ∎

LEMMA 8.1.3. *Suppose that w has period q and that there exists a factor v of w with $|v| \geq q$ that has period r, where r divides q. Then w has period r.*

Proof. Let $w = a_1 \cdots a_n$ and let $v = a_h \cdots a_k$, with $1 \leq h < k \leq n$ and $k-h+1 \geq q$. Let us suppose that $i \equiv j \pmod r$, $1 \leq i, j \leq n$. We have to prove that $a_i = a_j$. Since, by hypothesis $k-h+1 \geq q$, for any integers i, j, there exist i', j' with $h \leq i', j' \leq k$ such that $i \equiv i' \pmod q$ and $j \equiv j' \pmod q$. Since $i \equiv j \pmod r$ and since r divides q, then $i' \equiv j' \pmod r$. But w has period q and thus $a_i = a_{i'}$ and $a_j = a_{j'}$. Finally, since v has period r, one gets $a_{i'} = a_{j'}$ and the lemma is proved. ∎

8.1.2. The theorem of Fine and Wilf

This section is devoted to the theorem of Fine and Wilf and some generalizations. It is a classical and basic result on periodicity and one of its proofs is reported in Lothaire, 1983. The proof we give here can be considered as a first step of the proof of a more general result that is stated and proved in the next section.

The proof reported here is closely related to Euclid's algorithm computing the greatest common divisor of two integers. In particular, we use the fact that, given two positive integers p, q, with $q < p$, $\gcd(p, q) = \gcd(p-q, q)$.

THEOREM 8.1.4 (Fine and Wilf). *Let w be a word having periods p and q, with $q \leq p$. If $|w| \geq p + q - \gcd(p, q)$, then w has also period $\gcd(p, q)$.*

Proof. Set $r = \gcd(p, q)$. The proof is by induction on the integer $n = (p+q)/r$. For $n = 2$, $q = p = r$ and the statement is trivially verified.

Let us consider the case $n > 2$. This, in particular implies that $q < p$. Suppose that the statement holds for all integers smaller than n. Consider a word w having periods p and q and such that $|w| \geq p + q - r$. Denote by u the prefix of w of length q and set $w = uv$.

By Lemma 8.1.1, v has period $p - q$ and, since v is a factor of w, v has also period q. Moreover one has

$$|v| = |w| - q \geq (p-q) + q - r = (p-q) + q - \gcd(p-q, q).$$

By the inductive hypothesis, v has also period $\gcd(p-q,q) = \gcd(p,q) = r$. Since $p-q \geq r$, then $|v| = |w| - q \geq (p-q) + q - r \geq q$. By Lemma 8.1.3, the word w also has period r and this concludes the proof. ■

Remark 8.1.5. The bound given in Theorem 8.1.4 is tight, as shown by the word $w = abaababaaba$. It has period 5 and period 8, length $5+8-2 = 11$ and it has not period $\gcd(5,8) = 1$. An infinite family of words proving that the bound is tight is the family of central words considered in Chapter 2.

Theorem 8.1.4 can be generalized to words having three periods. As in the previous case, the statement and the proof are closely related to Euclid's algorithm.

Let $p = (p_1, p_2, p_3)$ be a triple of non negative integers. If $p_1 \leq p_2 \leq p_3$, we call p an *ordered triple*. We denote by O the operator which, given an arbitrary triple, returns the corresponding ordered triple. We define two additional operators R and S on ordered triples by

$$R(p) = \begin{cases} (p_1, p_2 - p_1, p_3 - p_1), & \text{if } p_1 \neq 0, \\ (0, p_2, p_3 - p_2), & \text{if } p_1 = 0 \end{cases}$$

and $S(p) = O(R(p))$.

Given an ordered triple p, let us consider the sequence $(p^{(k)})_{k \geq 0}$ of ordered triples defined recursively as follows:

$$p^{(0)} = p, \quad p^{(k+1)} = S(p^{(k)}), \quad k \geq 0\,.$$

The elements of the triple $p^{(k)}$ are denoted by:

$$p^{(k)} = (p_1^{(k)}, p_2^{(k)}, p_3^{(k)})\,.$$

Let us denote by $|p| = p_1 + p_2 + p_3$ the sum of the elements of the triple p. Set

$$m(p) = \min\{k \mid p_1^{(k)} = 0\}, \quad M(p) = \min\{k \mid p_1^{(k)} = p_2^{(k)} = 0\}.$$

With these notations, given the triple $p = (p_1, p_2, p_3)$, one has

$$\gcd(p_1, p_2, p_3) = |p^{(M(p))}|.$$

We also define a function h which plays an important role in the next result by

$$h(p_1, p_2, p_3) = |p^{(m(p))}|.$$

By definition, the function h satisfies the condition

$$h(p_1, p_2, p_3) = h(p_1, p_2 - p_1, p_3 - p_1).$$

Example 8.1.6. Consider the triple $p = (7, 11, 13)$. Euclid's algorithm gives the following sequence of triples: $p^{(0)} = (7, 11, 13)$, $p^{(1)} = (4, 6, 7)$, $p^{(2)} = (2, 3, 4)$, $p^{(3)} = (1, 2, 2)$, $p^{(4)} = (1, 1, 1)$, $p^{(5)} = (0, 0, 1)$. We get $\gcd(7, 11, 13) = h(7, 11, 13) = 1$.

Given an ordered triple $p = (p_1, p_2, p_3)$ of non negative integers, we introduce the function:

$$f(p_1, p_2, p_3) = \frac{1}{2}[p_1 + p_2 + p_3 - 2\gcd(p_1, p_2, p_3) + h(p_1, p_2, p_3)].$$

Notice that $f(p_1, p_2, p_3)$ is greater than or equal to $\gcd(p_1, p_2, p_3)$.

THEOREM 8.1.7. *Let w be a word over the alphabet A having three periods p_1, p_2 and p_3, with $p_1 \leq p_2 \leq p_3$. If $|w| \geq f(p_1, p_2, p_3)$, then w also has period* $\gcd(p_1, p_2, p_3)$.

REMARK 8.1.8. The statement of this theorem includes, as a particular case, the statement of the theorem of Fine and Wilf. Indeed, the condition that a word w has periods p, q, with $p \leq q$ corresponds to the triple $(0, p, q)$. Since, by definition, $h(0, p, q) = p + q$, it follows that

$$f(0, p, q) = \frac{1}{2}[p + q - 2\gcd(p, q) + p + q] = p + q - \gcd(p, q).$$

Proof. We shall prove the theorem by induction on the integer

$$n = p_1(p_1 + p_2 + p_3).$$

The case $n = 0$ corresponds to the classical Fine and Wilf theorem (see also Remark 8.1.8).

Let us now suppose that the statement is true for all ordered triples $q = (q_1, q_2, q_3)$ such that $m = q_1(q_1 + q_2 + q_3) < n$ and consider an ordered triple $p = (p_1, p_2, p_3)$ such that $p_1(p_1 + p_2 + p_3) = n$.

Let w be a word having periods p_1, p_2 and p_3 and length $|w| \geq f(p_1, p_2, p_3)$. Let u be the prefix of w of length p_1 :

$$w = uv, \qquad \text{with } |u| = p_1, \qquad |v| = |w| - p_1.$$

By Lemma 8.1.1, the word v has periods p_1, $p_2 - p_1$, $p_3 - p_1$ and length $|v| = |w| - p_1$. Since $|w| \geq f(p_1, p_2, p_3)$, one has

$$\begin{aligned}|v| &= |w| - p_1 \geq f(p_1, p_2, p_3) - p_1 \\ &= \tfrac{1}{2}[p_1 + p_2 + p_3 - 2\gcd(p_1, p_2, p_3) + h(p_1, p_2, p_3)] - p_1 \\ &= \tfrac{1}{2}[p_1 + (p_2 - p_1) + (p_3 - p_1) \\ &\quad -2\gcd(p_1, p_2 - p_1, p_3 - p_1) + h(p_1, p_2 - p_1, p_3 - p_1)] \\ &= f(p_1, p_2 - p_1, p_3 - p_1)\,.\end{aligned}$$

By the inductive hypothesis v also has period

$$\gcd(p_1, p_2 - p_1, p_3 - p_1) = \gcd(p_1, p_2, p_3).$$

By Lemma 8.1.3, the word $w = uv$ has period $\gcd(p_1, p_2, p_3)$. This concludes the proof. ∎

Remark 8.1.9. The bound given in Theorem 8.1.7 is tight, as shown by the following example. An infinite family of words proving the tightness is considered in Problem 8.1.8 (see also Remark 8.1.5).

Example 8.1.10. Consider the word

$$w = abacabaabacaba.$$

The word w has length $|w| = 14$ and periods $7, 11, 13$. Since $\gcd(7,11,13) = h(7,11,13) = 1$, (see Example 8.1.6) one has:

$$f(7,11,13) = \frac{1}{2}(7+11+13-2+1) = 15.$$

Then w is a word having periods $7, 11, 13$; its length is $|w| = f(7,11,13)-1$ and w has no period $\gcd(7,11,13)$.

8.1.3. Structure of the periods of a word

In this section, we give the structure of the set of periods of a single finite word. As a consequence, we obtain, for any word w, a word w' over a binary alphabet that has exactly the same set of periods as w.

Let $\Pi(w)$ denote the set of all periods of w, with 0 included. For instance, if $\hat{w} = abcabcadefgabcabca$, then $\Pi(\hat{w}) = \{0, 11, 14, 17, 18\}$ and if $\underline{w} = aaaaaaaaaaaaaaaaaa$, then $\Pi(\underline{w}) = \{0, 1, \ldots, 18\}$.

Clearly, if p is a period of w then any multiple of p that is smaller than or equal to the length of w is also a period of w. Indeed, since 1 is a period of $\underline{w}$ in the example above, then all positive integers smaller than or equal to $|\underline{w}|$ are also periods of $\underline{w}$.

If $\Pi(w) = \{0 = p_0 < p_1 < \cdots < p_s = |w|\}$, we define the sequence of differences

$$\delta_h = p_h - p_{h-1}, \;\; 1 \le h \le s$$

If p is a period of w and q is a period of the suffix of length $|w| - p$ of w then $p + q$ is also a period of w. Therefore (Problem 8.1.1), for any positive integer k such that $p + kq \le |w|$, the integer $p + kq$ is a period of w. This fact and Lemma 8.1.1 imply (Problem 8.1.9) that the sequence of differences δ_h is non-increasing.

In the example above, the integers 11 and 14 are periods of $\hat{w}$. By Lemma 8.1.1, the integer $14 - 11 = 3$ is a period of the suffix of length $|\hat{w}| - 11 = 7$ of $\hat{w}$. Therefore $17 = 11 + 6$ must also be a period of $\hat{w}$.

We know that the sequence of the differences δ_h is non-increasing, but some more conditions must be added in order to characterize the set of periods of a single word. They appear as conditions 3 and 4 in next theorem.

Theorem 8.1.11. *Let $\Pi = \{0 = p_0 < p_1 < \cdots < p_s = n\}$ be a set of integers and let $\delta_h = p_h - p_{h-1}$, $1 \le h \le s$. Then the following conditions are equivalent:*

1. *There exists a word w over a two-letter alphabet with $\Pi(w) = \Pi$.*
2. *There exists a word w with $\Pi(w) = \Pi$.*
3. *For each h, such that $\delta_h \leq n - p_h$, one has*
   (a) *$p_h + k\delta_h \in \Pi$, for $k = 1, \cdots, \lfloor (n - p_h)/\delta_h \rfloor$ and*
   (b) *if $\delta_{h+1} < \delta_h$, then $\delta_h + \delta_{h+1} > n - p_h + \gcd(\delta_h, \delta_{h+1})$.*
4. *For each h, such that $\delta_h \leq n - p_h$, one has*
   (a) *$p_h + \delta_h \in \Pi$ and*
   (b) *if $\delta_h = k\delta_{h+1}$, for some integer k then $k = 1$.*

Proof. Trivially condition 1 implies condition 2. Let us prove that condition 2 implies condition 3.

Let us now suppose that there exists a word $w = a_1 \cdots a_n$ with $\Pi(w) = \Pi$ and that $\delta_h \leq n - p_h$. By Lemma 8.1.1, δ_h is a period of the suffix of w of length $n - p_{h-1}$. Hence for any $k > 0$ and $i > p_h$ such that $i + k\delta_h \leq n$, one has that $a_i = a_{i+k\delta_h}$. Since p_h is a period of w, for any $i > p_h$, $a_{i-p_h} = a_i$. Setting $j = i - p_h$ we have that for any $j > 0$ and for any k such that $j + p_h + k\delta_h \leq n$, one has that $a_j = a_{j+p_h+k\delta_h}$, i.e., $p_h + k\delta_h \in \Pi$, for $k = 1, \cdots, \lfloor (n - p_h)/\delta_h \rfloor$ and 3.a is proved.

Let us suppose by contradiction that $\delta_h \leq n - p_h$, that $\delta_{h+1} < \delta_h$ and that $\delta_h + \delta_{h+1} \leq n - p_h + \gcd(\delta_h, \delta_{h+1})$. By Lemma 8.1.1, δ_h is a period of the suffix u of w of length $n - p_{h-1}$ and δ_{h+1} is a period of the suffix v of w of length $n - p_h$. Clearly v is a suffix of u. Hence, both δ_h and δ_{h+1} are periods of v. Since $|v| = n - p_h \geq \delta_{h+1} + \delta_h - \gcd(\delta_h, \delta_{h+1})$, by Theorem 8.1.4 v has also period $r = \gcd(\delta_h, \delta_{h+1})$. Since $|v| = n - p_h \geq \delta_h$ and since u has period δ_h, by Lemma 8.1.3, u has also period r This fact implies that $p_{h-1} + r$ is a period of w. But $p_{h-1} + r < p_{h-1} + \delta_h = p_h$ contradicting the fact that p_h is the smallest period greater than p_{h-1}.

Let us prove that condition 3 implies condition 4. Trivially condition 3.a implies condition 4.a.

Let us suppose by contradiction that $\delta_h \leq n - p_h$ and that $\delta_h = k\delta_{h+1}$, for some integer $k > 1$. Then $\delta_{h+1} < \delta_h$ and also $\delta_{h+1} = \gcd(\delta_h, \delta_{h+1})$. Therefore, by condition 3.b, $\delta_h + \delta_{h+1} > n - p_h + \gcd(\delta_h, \delta_{h+1}) = n - p_h + \delta_{h+1}$ i.e., $\delta_h > n - p_h$ that is a contradiction.

Let us finally prove that condition 4 implies condition 1. We prove that if condition 4 holds, then there exists a binary string w, such that $\Pi(w) = \Pi$.

Let $\Pi_h = \{p - p_h \mid p \in \Pi \text{ and } p \geq p_h\}$, $0 \leq h \leq s$. For instance, if $\Pi = \{0, 11, 14, 17, 18\}$, then $\Pi_4 = \{0\}$, $\Pi_3 = \{0, 1\}$, $\Pi_2 = \{0, 3, 4\}$, $\Pi_1 = \{0, 3, 6, 7\}$. Clearly, $\Pi_0 = \Pi$.

We prove by induction on $h = s, \ldots, 0$ that there exist binary strings w_h, such that $\Pi(w_h) = \Pi_h$. Then, w_0 is the required string with $\Pi(w_0) = \Pi$. Notice that $|w_h| = n - p_h$.

For the basis of the induction we have that $w^s = \epsilon$, since $\Pi_s = \{0\}$. Assume now that there exists a string w_h, such that $\Pi(w_h) = \Pi_h$. There are two cases.

CASE 1. $\delta_h > n - p_h$. We claim that there exists a sequence $a_1, \ldots, a_{\delta_h - |w_h|}$ of letters in the same binary alphabet of w_h such that the word $w_{h-1} = w_h a_1 \cdots a_{\delta_h - |w_h|} w_h$ has no periods of length smaller than δ_h.

The proof of this claim is by induction on $\delta_h - |w_h|$. Suppose $\delta_h - |w_h| = 1$. Consider the two words $w_h x w_h$ and $w_h y w_h$ with $x \neq y$ the two different letters in the binary alphabet of w_h. If, by contradiction, both words $w_h x w_h$ and $w_h y w_h$ have a period smaller than $\delta_h = |w_h x| = |w_h y|$ then, by problem 8.1.4, they must be equal, which is impossible because they differ in the central position.

Inductive step of the claim: Suppose that the binary alphabet is $\{x, y\}$. Suppose that the claim is true for $\delta_h - |w_h| = n - 1$ and suppose also that, by contradiction, putting a letter x or a letter y between positions $a_{\lceil n/2 \rceil}$ and $a_{\lceil n/2 \rceil + 1}$ we get two different words that both have a period smaller than or equal to $|w_h| + \lceil n/2 \rceil + 1$. These two periods cannot be equal, because of the different letter in the same position, and, so, one must be smaller than $|w_h| + \lceil n/2 \rceil + 1$.

These words have length $2|w_h| + n + 1$ and the sum of the two periods is smaller than or equal to $2|w_h| + n + 1$. By Problem 8.1.4, they must be equal, which is impossible because they differ in the central position. This concludes the proof of the claim.

By the claim one has $\Pi(w_{h-1}) = \{\delta_h + p \mid p \in \Pi_h\} \cup \{0\} = \Pi_{h-1}$, and the inductive step is proved.

CASE 2. If $\delta_h \leq n - p_h$, then let $w_{h-1} = a_1 \cdots a_{\delta_h} w_h$ where $a_1 \cdots a_{\delta_h}$ is the prefix of length δ_h of w_h . Since $p_h + \delta_h \in \Pi$, we get that $\delta_h \in \Pi_h$ and, by inductive hypothesis, δ_h is a period for w_h. Hence δ_h is also a period of w_{h-1}. Consequently, by Problem 8.1.1 and Lemma 8.1.1, $\delta_h + p$ is a period of w_{h-1}, for some integer $p \geq 0$ if and only if p is a period of w_h. This is equivalent to saying, by the inductive hypothesis, that $\Pi(w_{h-1}) \cap \{\delta_h, \ldots, |w_{h-1}|\} = \{\delta_h + p \mid p \in \Pi_h\}$.

Assume, by contradiction, that $\Pi(w_{h-1}) \neq \Pi_{h-1}$ and, consequently, there exists a period $t < \delta_h$ in $\Pi(w_{h-1}) \setminus \Pi_{h-1}$. We have that both t and $\delta_h - t$ must be periods of w_h. The first is a period because w_h is a suffix of w_{h-1}. The second is a period because both δ_h and t are periods of w_{h-1} and because, by Lemma 8.1.1, the suffix of w_{h-1} of length $|w_{h-1}| - t > |w_{h-1}| - \delta_h = |w_h|$ has period $\delta_h - t$ and contains w_h as factor.

Since by the induction hypothesis, δ_{h+1} is the shortest non-zero period of w_h, $\delta_{h+1} \leq t$ and $\delta_{h+1} \leq \delta_h - t$, and, consequently, $\delta_{h+1} + t \leq \delta_h$ and $\delta_{h+1} + \delta_h - t \leq \delta_h$.

By the theorem of Fine and Wilf and by the minimality of δ_{h+1}, $\delta_{h+1} = \gcd(\delta_{h+1}, t)$ and $\delta_{h+1} = \gcd(\delta_{h+1}, \delta_h - t)$, i.e., δ_{h+1} divides t and also divides $\delta_h - t$. Hence δ_{h+1} also divides $t + \delta_h - t = \delta_h$ and this contradicts condition 4.b because $\delta_{h+1} \leq t < \delta_h$. ■

Let us now give an example of how to construct a word w' over a binary alphabet $\{a, b\}$ such that $\Pi(w') = \{0, 11, 14, 17, 18\}$.

Notice that the set $\Pi = \{0, 11, 14, 17, 18\}$ satisfies condition 4 of the previous theorem. It is indeed the set of periods of the word $\hat{w} = abcabcadefgabcabca$.

We have that $\Pi_4 = \{0\}$, $\Pi_3 = \{0, 1\}$, $\Pi_2 = \{0, 3, 4\}$, $\Pi_1 = \{0, 3, 6, 7\}$. Clearly, always $\Pi_0 = \Pi$.

Moreover we know that $\delta_4 = 18 - 17 = 1$, $\delta_3 = 17 - 14 = 3$, $\delta_2 = 14 - 11 = 3$, $\delta_1 = 11 - 0 = 11$.

We inductively construct words w_h for $h = 4, 3, 2, 1, 0$, where $w_0 = w'$ is the required word. Recall that $|w_h| = 18 - p_h$.

Word $w_4 = \epsilon$ and $\Pi(w_4) = \Pi_4 = \{0\}$. Since $\delta_4 = 1 > 18 - p_4 = 18 - 18 = 0$ we are in the first case of previous proof, and so there exists a sequence of $\delta_4 - |w_4| = 1 - 0 = 1$ letter(s) $a_1, \ldots, a_{\delta_4 - |w_4|}$ (in this case just one letter a_1) such that $w_3 = w_4 a_1 w_4$. In this case both letters a and b can be chosen to be a_1 in order to have $\Pi(w_3) = \Pi_3 = \{0, 1\}$. Let us choose $a_1 = a$, and, so, $w_3 = a$.

Since $\delta_3 = 3 > 18 - p_3 = 18 - 17 = 1$, we are in the first case of previous proof, and so there exists a sequence of $\delta_3 - |w_3| = 3 - 1 = 2$ letters $a_1, \ldots, a_{\delta_3 - |w_3|}$ (in this case two letters, a_1 and a_2) such that $w_2 = w_3 a_1 a_2 w_3 = a a_1 a_2 a$ has no period smaller than δ_3. Letter a_1 must be chosen such that $w_3 a_1 w_3$ has no period smaller than $|w_3| + 1 = 2$. There is only one possibility, which is that $a_1 = b$. Letter a_2 must be chosen such that $w_3 a_1 a_2 w_3 = aba_2a$ has no period smaller than $|w_3| + 2 = 3$. In this case both letters a and b can be chosen in order to have $\Pi(w_2) = \Pi_2 = \{0, 3, 4\}$. Let us choose $a_1 = b$, and so $w_2 = abba$.

Notice that the order that we use to choose the letters in the general sequence $a_1, \ldots, a_{\delta_h - |w_h|}$ is not the usual order and follows the inductive proof of always choosing the "central" letter, i.e., $a_1, a_{\delta_h - |w_h|}, a_2, a_{\delta_h - |w_h| - 2}, \ldots, a_{\lceil (n - p_h)/2 \rceil}$ if $(n - p_h)$ is odd, while if it is even the last letter to be chosen is $a_{\lceil (n - p_h)/2 \rceil + 1}$. In previous situation, when there are only two letters, this order coincides with the usual one.

Since $\delta_2 = 3 \leq 18 - p_2 = 18 - 14 = 4$, we are in the first case of the previous proof and, so, $w_1 = abbw_2 = abbabba$, because abb is the prefix of length δ_2 of w_2 (whose length is $18 - p_2$). Indeed $\Pi(w_1) = \Pi_1 = \{0, 3, 6, 7\}$.

Since $\delta_1 = 11 > 18 - p_1 = 18 - 11 = 7$, we are in the first case of the previous proof and, so, there exists a sequence of $\delta_1 - |w_1| = 11 - 7 = 4$ letters a_1, a_2, a_3, a_4 such that $w_0 = w_1 a_1 a_2 a_3 a_4 w_1 = a a_1 a_2 a$ has no period smaller than δ_1. For brevity we do not perform the inductive steps for choosing this sequence of letters, that can be chosen to be a, a, a, a, i.e., $w' = w_0 = abbabbaaaaaabbabba$.

It is easy to verify that $\Pi(w') = \{0, 11, 14, 17, 18\}$.

8.1.4. Golden ratio and periodicity

We present here a result that has some analogies with the theorem of Fine and Wilf and relates periodicity and the golden ratio. This result plays an important role in Section 8.2.2.

The theorem of Fine and Wilf states, roughly speaking, that if a word w has two periods p and q and it is long enough ($|w| \geq p + q - \gcd(p, q)$), then it has a shorter period ($\gcd(p, q)$).

In next result, we start from a weaker hypothesis on the length of the word, in which the golden ratio appears, and we derive a weaker conclusion.

Indeed, denoting by φ the golden ratio, we suppose that $|w| > \varphi \max\{p, q\}$ and we derive that there exists only a suffix (and a prefix) having "shorter" period.

Let us recall that the golden ratio, denoted by φ, is the real positive root of the equation $x^2 - x - 1 = 0$, i.e., $\varphi = (\sqrt{5} + 1)2 = 1.618\ldots$. Since, by definition,

$\varphi^2 = \varphi + 1$, in the sequel of the chapter we shall sometimes interchange φ^2 with $\varphi + 1$ without mention.

THEOREM 8.1.12. *Let x and y be nonempty words and let ρ and σ be positive rational numbers such that $\varphi < \rho < \sigma$. If $x^\rho = y^\sigma$, then there exist a nonempty word z and a rational number $\tau \geq \rho + 1$ such that z^τ is a suffix (a prefix) of $x^\rho = y^\sigma$.*

Proof. Set $w = x^\rho = y^\sigma$. If $\sigma \geq \rho + 1$, then the statement is trivially satisfied with $z = y$ and $\tau = \sigma$. Let us now suppose that $\sigma < \rho + 1$. Let $|x| = p$ and $|y| = q$. Then

$$|w| = p\rho = q\sigma < q(\rho + 1).$$

If $\rho \geq \varphi$, then, by the definition of the golden ratio φ, $\rho + 1 < \rho^2$. From the inequality $p\rho < q(\rho + 1)$, we derive $p\rho < q\rho^2$ and then $p - \rho q < 0$. By adding $\rho p - q$ to both sides of the last inequality one has

$$p - \rho q + \rho p - q < \rho p - q$$

which can be rewritten

$$(p - q)(\rho + 1) < \rho p - q.$$

By definition, w has periods p and q, with $q < p$. By Lemma 8.1.1, the suffix (prefix) v of w of length $|w| - q$ has period $p - q$. Denoting by τ the order of v, we have

$$\tau = \mathrm{ord}(v) = \frac{|v|}{p - q} = \frac{|w| - q}{p - q} = \frac{\rho p - q}{p - q}.$$

By the previous inequality, $\tau > \rho + 1$, i.e., $v = z^\tau$, with $\tau > \rho + 1$. ∎

REMARK 8.1.13. The number φ is tight in Theorem 8.1.12. Recall from Section 1.2 that $(F_n)_{n\geq 1}$ denotes the sequence of Fibonacci numbers and that the sequence of Fibonacci words$(f_n)_{n\geq 1}$ is defined by the inductive rules: $f_1 = b$, $f_2 = a$, and $f_{n+1} = f_n f_{n-1}$.

For any $n \geq 1$ consider the word v_{n+2} defined as the prefix of length $|f_{n+2}| - 2$ of f_{n+2}, i.e.,

$$v_{n+2}xy = f_{n+2},$$

where x, y are letters, $x \neq y$. It is known that v_{n+2} has periods F_n and F_{n+1}. Then v_{n+2} can be written:

$$v_{n+2} = f_n^{\rho_n} = f_{n+1}^{\rho'_n},$$

where

$$\rho_n = \frac{F_{n+2} - 2}{F_n}, \quad \rho'_n = \frac{F_{n+2} - 2}{F_{n+1}},$$

and $\rho'_n < \varphi$.

One can prove that there does not exist any prefix (or suffix) of v_{n+2} having order $\geq \varphi + 1$ (Problem 8.2.6). The tightness of Theorem 8.1.12 is a consequence of the well-known property of Fibonacci numbers stating that, for any $\epsilon > 0$, there exists an integer m such that for any $n \geq m$

$$\varphi - \epsilon < \rho'_n.$$

Also notice that the statement of Theorem 8.1.12 is sharp in the sense that the prefix (or suffix) of w of the form z^τ, with $\tau \geq \rho + 1$, cannot in general coincide with the whole word w. Indeed the word

$$w = a^6 b a^6$$

has periods 7 and 8 and it can be written

$$w = (a^6 b)^\rho = (a^6 b a)^\sigma,$$

with $\sigma = 13/8 = 1.625$ and $\rho = 13/7$, i.e., $\varphi < \sigma < \rho$. The word w satisfies the hypothesis of the lemma and it has as a prefix (and a suffix) the word a^6 according to the lemma. However w cannot be written as $w = z^\gamma$ with $\gamma \geq \sigma + 1$.

8.1.5. Squares in a word

In this section we study squares that appear as prefixes or as factors in words. Recall that a *square* is a word w of the form $w = v^2$.

The occurrence of a square as a factor in a word can be considered as a particular kind of *local period* occurring in the word. A more general notion of local period will be discussed in Section 8.2. Here we study the squares that can appear as prefixes or as factors of a word.

The first result we prove states a fundamental inequality on the lengths of different squares that can occur as prefixes of the same word. It is used in the second result and also in Chapter 12, Section 12.1.7.

The second result gives a bound on the number of squares occurring as factors of a word.

In what follows we write $w < w'$ to denote that the word w is a prefix of w'. Recall also that a word w is called primitive if it is not a power of another word, i.e., there exists no word z such that $w = z^k$ for some integer k greater than 1.

LEMMA 8.1.14. *Let u, v, w be words such that u is primitive, $v \notin u^*$, and $uu < vv < ww$. Then $|u| + |v| \leq |w|$.*

Proof. By contradiction we assume that $|v| < |w| < |vu|$. First assume that $2|u| \leq |v|$. Then we have $uu \leq v < w$. Since $vv < ww$, the (second) word w has a prefix ru, where $1 \leq |r| < |u|$. Hence u is a internal factor of uu, which is impossible since u is a primitive word (cf. Problem 8.1.6).

Therefore we may assume that $u < v < uu$. Let $v = uz$. Then we have $zu = uy$ for some non-empty word y. Let x be the primitive root of z. Then we can write $z = x^\beta$ and $u = x^\alpha r$ for some integers $\alpha \geq \beta \geq 1$ and some non-empty word $r < x$. Hence we have $v = x^\alpha r x^\beta < w$ and $s x^\beta < w$, where

s is a proper suffix of $x^\alpha r$. If $|s| \geq |x|$, then $x \leq s$. Hence $|s| \geq |x|$ implies $s = x^\gamma r$ for some integer $\gamma < \alpha$. Therefore $x^\gamma rx < x^{\alpha+1}$, which is impossible since $1 \leq |r| < |x|$. Thus, s is a proper prefix of x. If $|s| \leq |x^{\alpha-1}r|$, then $sx \leq x^\alpha r < x^{\alpha+1}$, which again is a contradiction. Therefore we may assume that $\alpha = \beta = 1$ and $|r| < |s|$. We have $u = xr$, $v = xrx$, and $w = xrxt$ for some t such that $ts = xr$. In particular, $1 < t < x$. Hence, $txr < xrx < xrxr$ and this means $tu < uu$, which is impossible since u is primitive. ∎

Remark 8.1.15. Notice that the statement of previous lemma is sharp. Let

$$u = abbab,\ v = abbababb,\ \text{and}\ w = abbababbabbab.$$

The lengths are 5, 8, and 13. The words u, v, w are primitive and $uu < vv < ww$. The general scheme for such an example is

$$u = pr,\ v = prp,\ \text{and}\ w = prppr$$

where p and r are words such that $r < p$.

We now study how many squares a word can contain.

Let $w = a_1 \cdots a_n$. For $i = 1, \ldots, n$, let $s_i(w)$ be the number of squares that are prefixes of $a_i \cdots a_n$ and never appear as prefixes of $a_j \cdots a_n$, with $i < j \leq n$. For example in the word $w = abaababaababa$ we have $s_3(w) = 1$ since $aababaabab$ is a square beginning in position 3 which do not appear later on. The square aa also begin in position 3 but it also appears in position 8 and, so, it does not affect the value of $s_3(w)$.

Theorem 8.1.16. *For every nonempty word x of length n, $s_i(x) \leq 2$ for all $i \in \{1, \ldots, n\}$.*

Proof. Suppose, by contradiction, that there exists a word x of length n such that $s_i(x) \geq 3$, for some $i \in \{1, \ldots, n\}$.

Then $x_i = a_i a_{i+1} \cdots a_n$ has three prefixes u^2, v^2, w^2, $uu < vv < ww$ which do not occur elsewhere in the word x. Let $p = |w|, q = |v|$ and $r = |u|$. If $p \geq 2r$, then u^2 occurs again at position $i + p$, because w occurs there, that is impossible. Thus $p < 2r$, and we have that $p > q > r > p/2$ and also $q < p < 2r < q + r < 2q$. It follows from Lemma 8.1.14 that u is non-primitive.

Therefore there exists a primitive word y such that $u = y^k$ for some integer $k \geq 2$. If we set $r_1 = |y|$, then $r = kr_1$. Now we have that $yy < uu < vv$ with y primitive, so by Lemma 8.1.14 that $r_1 + q \leq p$. Note that r_1 is also a period of u^2 and, since $p < 2r$, w is a prefix of u^2 and, so, r_1 is a period of w. Since $p < 2q$, w is a prefix of v^2 and, so, q is also a period of w. Since $r_1 + q \leq p$ we can apply Theorem 8.1.4 and obtain that $\gcd(p, q) = d$ is also a period of w. Since the word y is primitive and d divides r_1, we have $d = r_1$. Hence r_1 divides q. Now $r = kr_1$, and, consequently, $q = (k+h)r_1$ for some integer $h \geq 1$. Therefore v^2 has length $2(t + s)r_1$ and u^2 has length $2tr_1$. It follows that u^2 also appears at position $r_1 + 1$, which is a contradiction. ∎

Notice that, since there are no squares beginning at the last position, $s_n(x) = 0$ for any word x of length n.

For any word $x = a_1 \cdots a_n$ let us denote by $SQ(x)$ the cardinality of the set of squares that are factors of x. It is easy to verify by definition and by previous Notice that $SQ(x) = \Sigma_{i=1}^{n-1} s_i(x)$.

Hence, we have the following corollary of Theorem 8.1.16.

COROLLARY 8.1.17. *For any word x of length n, $SQ(x) \leq 2n - 2$.*

8.2. Local versus global periodicity

In this section we investigate the relationships between local and global periodicity of words. In section 8.1.5 we have taken into account a particular kind of local period, i.e., squares occurring as factors in a word. Here we introduce a very general notion of local period in terms of *repetitions.* A repetition occurring in a word w is not in general a factor of w, nor it is necessarily a square, but its order can be an arbitrary rational number ρ. Moreover the repetition is here referred to a "point" of the word w and it is important to consider the relative positions of the repetitions and that of the point of the word w at which the repetition is detected.

In order to give the formal definitions, we first introduce the notion of pointed word. This is the appropriate notion to define local properties of a word.

Let $w = a_1 a_2 \cdots a_n$ be a word over the alphabet A. A *pointed word* is a pair $(a_1 \cdots a_i, a_{i+1} \cdots a_n)$, $1 \leq i < n$. The pointed word is also denoted by (w, i) and we refer to (w, i) as the word w at the point (or the position) i.

Let (x, y) be a pair of words. The pair (x, y) *matches* the pointed word (w, i), or simply matches the word w at the point i, if

$$A^*x \cap A^*a_1 \cdots a_i \neq \emptyset \quad \text{and} \quad yA^* \cap a_{i+1} \cdots a_n A^* \neq \emptyset.$$

Notice that the word $z = xy$ is not in general a factor of the word w and that the pair (x, y) specifies the relative position of the word z and the point i. So we can distinguish between *central repetitions* and *left* (or right) *repetitions.*

8.2.1. Central repetitions

A word w contains a *repetition* of order ρ having as *center* the point (or position) i, or shortly a *central repetition* of order ρ at the point i, if there exists a non-empty word z of order $\mathrm{ord}(z) = \rho$ and a factorization $z = xy$, with $|x| = |y|$, such that the pair (x, y) matches w at the point i. This means that the point i is *central* with respect to the repetition z. The word z is called a central repetition of (w, i) and must have even length. This central repetition is *proper* (or *internal*) if x is a suffix of $a_1 \cdots a_i$ and y is a prefix of $a_{i+1} \cdots a_n$. It is *left external* if $a_1 \cdots a_i$ is a proper suffix of x. It is *right external* if $a_{i+1} \cdots a_n$ is a proper prefix of y.

Central repetitions of order 2 play an important role in this theory. By definition, a central repetition of order 2 at the point i of w is a word z of the form $z = x^2$ such that the pair (x, x) matches w at the point i. We say that w has a *square* having its center in the position i.

EXAMPLE 8.2.1. Given the word

$$w = abaababaabaaba$$

the pointed word $(w, 8)$ is the pair

$$(abaababa, abaaba).$$

The pair (aba, aba) matches the pointed word $(w, 8)$, and so the word $abaaba$ is a central repetition of w at the point 8. It has order 2 and period 3. In the point 8 of w there is another central repetition of order 2, or a square having its center in it. It is the word aa and it has period 1. Both these repetitions are proper. The pointed word $(w, 7)$ is the pair

$$(abaabab, aabaaba).$$

The word $aabaababaabaabab$ is a central repetition of $(w, 7)$ of order 2 and period 8. It is both left and right external. Since the pair $(abab, aaba)$ matches w at the point 7, the word $ababaaba$ is a proper central repetition of $(w, 7)$ of order 1.6 and period 5.

As shown in the previous example, a word can have different central repetitions of same order at a given point. We are interested, for a given order, in detecting the central repetition of minimal period. This leads to the notion of minimal central repetition and of central local period.

For any real $\alpha > 1$, $c_\alpha(w, i)$ denotes the *central local period* (of order α) of the pointed word (w, i),

$$c_\alpha(w, i) = \min\{p(z) \mid z \text{ is a central repetition of } (w, i) \text{ of order } \geq \alpha\}.$$

The central repetition z of (w, i) such that $p(z) = c_\alpha(w, i)$ is called the *minimal* central repetition (of order α) of w at the point i.

It is immediate to verify that, if $\alpha < \beta$, then $c_\alpha(w, i) \leq c_\beta(w, i)$ and that for any α and any $i \geq 1$, $c_\alpha(w, i) \leq p(w)$.

In the special case $\alpha = 2$ one has

$$c_2(w, i) = \min\{|x| \mid x \neq \varepsilon \text{ and } (x, x) \text{ matches } w \text{ in the position } i\}.$$

EXAMPLE 8.2.2. $w = abaababaabaaba$

w	a	b	a	a	b	a	b	a	a	b	a	a	b	a
i		1	2	3	4	5	6	7	8	9	10	11	12	13
$c_2(w,i)$		2	3	1	5	2	2	8	1	3	3	1	3	2
$c_{1.6}(w,i)$		2	3	1	5	2	2	5	1	3	3	1	3	2

We denote by $P_\alpha(w)$ the maximum of the central local periods (of order α) of w

$$P_\alpha(w) = \max\{c_\alpha(w,i) \mid 1 \leq i < |w|\}.$$

A point (or position) i is *critical* if $c_\alpha(w,i) = P_\alpha(w)$. We denote by $C_\alpha(w)$ the set of critical points of w

$$C_\alpha(w) = \{i \mid 1 \leq i < |w| \text{ and } c_\alpha(w,i) = P_\alpha(w)\}.$$

We denote further by $Z_\alpha(w)$ and $S_\alpha(w)$ the minimum and the maximum respectively of the critical points:

$$Z_\alpha(w) = \min C_\alpha(w), \quad S_\alpha(w) = \max C_\alpha(w).$$

EXAMPLE 8.2.2 (*continued*). For $w = abaababaabaaba$, $P_2(w) = 8$, $C_2(w) = \{7\}$, $Z_2(w) = S_2(w) = 7$, $P_{1.6}(w) = 5$, $C_{1.6}(w) = \{4, 7\}$, $Z_{1.6}(w) = 4$, $S_{1.6}(w) = 7$.

Notice that the notion of critical point introduced in this chapter slightly differs from the one used in the literature, where a critical point i usually denotes a position where the local period of order 2, $c_2(w,i)$, is equal to the global period $p(w)$. The difference is motivated by the fact that we here take in account also repetitions of an arbitrary order $\alpha > 1$.

It is easy to verify that $c_\alpha(w,i) \leq p(w)$ for $\alpha > 1$ and $i = 1, \cdots, |w| - 1$, i.e., the central local periods are smaller than or equal to the period. On the other hand, if α is sufficiently large, i.e., $\alpha \geq 2|w|$, it is possible to prove that $c_\alpha(w,i) = p(w)$ for all i, as stated in particular in next proposition.

PROPOSITION 8.2.3. *Let $k = \lceil \alpha/2 \rceil$. If the period of w is smaller than or equal to k then in every position of i, one has $c_\alpha(w,i) = p(w)$. Hence, if $\alpha \geq 2|w|$ then every position is critical of order α.*

Proof. Let i, $1 \leq i < |w|$, be a position in $|w|$. If the central repetition of order α at the point i is both left and right external then $c_\alpha(w,i)$ is also a period of w and, consequently, $p(w) \leq c_\alpha(w,i)$ and, by previous remark, the thesis follows.

Suppose now that the central repetition of order α at the point i is either left or right internal or both. Suppose that it is left internal. We claim that $c_\alpha(w,i)$ divides $p(w)$. Indeed if $c_\alpha(w,i) = 1$ there is nothing to prove. Suppose that $c_\alpha(w,i) > 1$. In this case the part v of the central repetition of order α at the point i that is at the left of point i has, by hypothesis, length greater than or equal to $2p(w)$. Then v has periods $p(w)$ and $c_\alpha(w,i)$. We can apply the theorem of Fine and Wilf and obtain that it has period $d = \gcd(p(w), c_\alpha(w,i))$. But d cannot properly divide $c_\alpha(w,i)$ by the minimality of $c_\alpha(w,i)$. Hence $d = c_\alpha(w,i)$ and the claim is proved. We can now apply Lemma 8.1.3 and obtain that $c_\alpha(w,i)$ is also a period of w and, consequently, $p(w) \leq c_\alpha(w,i)$ and, by previous remark, the thesis follows. ∎

The critical factorization theorem in particular states that for $\alpha = 2$ there exists at least a point such that the central local period detected at this point coincides with the (global) period of the word, i.e., there exists an integer j, $1 \leq j < |w|$, such that $c_{2(w,j)} = p(w)$.

An important step in the proof of the critical factorization theorem is the following proposition.

Proposition 8.2.4. *If $z = x^2$ is the square of minimal length having its center in position j of w, $1 \leq j < |w|$, then x is unbordered.*

Proof. If there exists a nonempty border t of x, i.e., t is both prefix and suffix of x, then t^2 is a square having its center in the position j of w that is shorter than x^2, contradicting the definition of x. ■

Theorem 8.2.5 (Critical Factorization Theorem). *Let w be a word having length $|w| \geq 2$. In every sequence of $l \geq \max(1, p(w) - 1)$ consecutive positions there is a critical one and, moreover, $P_2(w) = p(w)$.*

Proof. The proof is by induction on $P_2(w)$. Suppose that $P_2(w) = 1$. Since for all natural numbers i, $1 \leq i < |w|$, $c_{2(w,i)} = 1$, then $a_i = a_{i+1}$. If $a = a_1$ and $n = |w|$, then w is of the form $w = a^n$, $p(w) = 1 = P_2(w)$ and all positions are critical.

Let us suppose that the statement of the proposition holds true for all words w' such that $P_2(w') \leq k - 1$, $k > 1$. Let w be a word having $P_2(w) = k$. We prove the following properties:

i) If j is a critical position and $j + 1, \ldots, j + l$ are not critical then $P_2(w) > l + 1$.
ii) If j is a critical position and $j - l, \ldots, j - 1$ are not critical then $P_2(w) > l + 1$.

As an immediate consequence of previous two properties one has that

iii) Every sequence of at least $P_2(w) - 1$ consecutive positions contains a critical one.

Let us recall that for any position j of w one has $c_2(w, j) \leq p(w)$, and, so, $P_2(w) \leq p(w)$. Hence property *iii* implies the first part of the theorem.

In order to prove *i*) let us consider the word $u = a_{j+1} \cdots a_{j+l} a_{j+l+1}$. Since any central repetition at point $j + i$ of w is a repetition having its center at point i of u one has

$$c_2(u, i) \leq c_2(w, j + i) \qquad i = 1, \ldots, l.$$

Since no position $j + i$ of w, with $i = 1, \ldots, l$, is a critical position, one has that

$$c_2(u, i) < k \qquad i = 1, \ldots, l.$$

As a consequence, $c_2(u, i) < k$ for $i = 1, \ldots, l$, and then $P_2(u) < k$. By inductive hypothesis $p(u) = P_2(u) < k$.

Let $z = x^2$ be the square of minimal length having its center at position j of w. Since by hypothesis position j is critical, one has that $c_2(w, j) = P_2(w) = k$, and $|x| = k$.

By Proposition 8.2.4 the word x is unbordered. If x is a prefix of the word $u = a_{j+1} \cdots a_{j+l}a_{j+l+1}$ then $p(x) \leq p(u) < k$, that is a contradiction. Hence u is a proper prefix of x and, consequently, $P_2(w) = k = |x| > |u| = l + 1$.

The proof of *ii)* is analogous by taking $u = a_{j-l} \cdots a_{j-1}$. In order to complete the proof of the theorem we must prove that

iv) $p(w) = P_2(w)$.

As noticed above, we have always that $P_2(w) \leq p(w)$. It remains to prove that $P_2(w) \geq p(w)$. Let i be a position such that $1 \leq i < i + P_2(w) \leq |w|$. By property *iii)* there exists a critical position j in the set $\{i, \ldots, i + P_2(w) - 1\}$. There exists then a square x^2 having its center at position j with $|x| = P_2(w)$. Note that $a_i \cdots a_{i+P_2(w)}$ is a factor of x^2, and, consequently, $a_i = a_{i+P_2(w)}$. Therefore $P_2(w)$ is a period of w and then $P_2(w) \geq p(w)$, and this concludes the proof. ∎

COROLLARY 8.2.6. *Let w be a word of length $|w| \geq 2$ and $p(w) > 1$. We have that $Z_2(w) < P_2(w)$, i.e., the central repetition at point $Z_2(w)$ is left external. We have also that $|w| - P_2(w) < S_2(w)$, i.e., the central repetition at point $S_2(w)$ is right external.*

COROLLARY 8.2.7. *Let $w = a_1 \cdots a_n$, be a word of length n. Given i, j, $1 \leq i < j \leq n$, if $c_2(w, h) < c_2(w, j)$ for any h such that $i \leq h < j$, then $c_2(w, j) > j - i + 1$.*

Proof. Let $v = a_i \cdots a_j$. On has that $c_2(v, h) \leq c_2(w, h) < c_2(w, j)$, for $i \leq h < j$. According to Theorem 8.2.5 we have that $p(v) < c_2(w, j)$. Let u^2 be the square of length $2c_2(w, j)$ having its center at position j of w. According to Proposition 8.2.4, u is an unbordered word. Hence u cannot be a suffix of v longer than $p(v)$. Therefore v is a proper suffix of u and $|u| = c_2(w, j) > j - i + 1$. ∎

In Example 8.2.2 $P_2(w)$ is, according to Theorem 8.2.5, exactly the period of w. Moreover the unique critical point of w is 7 and it satisfies the conditions (iii) and (iv) in the proof of the theorem. The same example shows that the theorem does not hold true for $\alpha = 1.6$. Indeed $P_{1.6}(w) = 5 \neq p(w) = 8$. The following example shows that the value $\alpha = 2$ is tight.

EXAMPLE 8.2.8. For any $\epsilon > 0$, consider the word $ba^{m-1}ba^mb$, with m such that $2m/(m+1) \geq 2 - \epsilon$. The unique critical point of order 2 is the point $m+1$, corresponding to the pair $(ba^{m-1}b, a^mb)$. The minimal central repetition of order 2 at such a point is the word $a^mba^{m-1}ba^mba^{m-1}b$, which has period $2m+1$, according to the critical factorization theorem. However the minimal central repetition of order $2 - \epsilon$ at the same point is the word $u = a^{m-1}ba^m$. Indeed u has period $m+1$ and order $2m/(m+1) > 2 - \epsilon$. It is easy to verify that such a point is also a critical point of order $2 - \epsilon$, and then

$$P_{2-\epsilon}(ba^{m-1}ba^mb) = m + 1 \neq p(ba^{m-1}ba^mb) = 2m + 1.$$

Statements (ii), (iii) and (iv) in the proof of Theorem 8.2.5 as given are sharp. Indeed the word $a^m ba^m ba^m$, $m \geq 1$, has period $m+1$ and exactly four critical points, m, $m+1$, $2m+1$ and $2m+2$, corresponding to the pairs $(a^m, ba^m ba^m)$, $(a^m b, a^m ba^m)$, $(a^m ba^m, ba^m)$ and $(a^m ba^m b, a^m)$ respectively.

8.2.2. Left repetitions

In the previous section we considered central repetitions, i.e., we required the repetition occurring at the point i of a word w to be such that this point is the center of the repetition

In this section we take into account a new notion of repetition, in which we require that the repetition occurs at a given point "immediately to the *left* from that point". The symmetric case of repetitions occurring "immediately to the right from a given point" is similar and it is not explicitly considered here.

For technical reasons it is convenient, in the case of left repetitions, to change a bit the definition of a pointed word (w,i) and to allow that the positive integer i ranges from 1 to $|w|$ (in the case of central repetitions i ranges from 1 to $|w|-1$).

A word $w = a_1 \cdots a_n$ contains a *left repetition* of order ρ at the point i, $(1 \leq i \leq n)$, if there exists a word z of order $\mathrm{ord}(z) = \rho$ such that

$$A^* z \cap A^* a_1 \cdots a_i \neq \emptyset.$$

The word z is called left repetition of (w,i). It is *external* if $a_1 \cdots a_i$ is a proper suffix of z. It is *proper* (or *internal*) if z is a suffix of $a_1 \cdots a_i$.

Example 8.2.9. Given the word

$$w = abaababaabaaba,$$

$aabaab$ is an external left repetition of $(w,5)$. It has order 2 and period 3. The words aa and $baabaa$ are both proper left repetitions of $(w,12)$ of order 2 and periods 1 and 3 respectively. The word $abab$ is a proper left repetition of $(w,7)$ of order 2 and period 2, whereas the word $ababaababaabab$ is an external left repetition of $(w,7)$ of order 2.8 and period 5.

The previous example shows that a word can have different left repetitions of the same order at a given point. As in the case of central repetitions, we are interested, for a given order, to detect the left repetition of minimal period.

For any real $\alpha > 1$, $l_\alpha(w,i)$ denotes the *left local period* (of order α) of the pointed word (w,i):

$$l_\alpha(w,i) = \min\{p(z) \mid z \text{ is a left repetition of } (w,i) \text{ of order } \geq \alpha\}.$$

The left repetition z of (w,i) such that $p(z) = l_\alpha(w,i)$ is called the *minimal* left repetition (of order α) of w at the point i.

It is immediate to verify that, if $\alpha < \beta$, then $l_\alpha(w,i) \leq l_\beta(w,i)$.

EXAMPLE 8.2.10. $w = abaababaabaaba$

w	a	b	a	a	b	a	b	a	a	b	a	a	b	a
i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$\ell_2(w,i)$	1	2	2	1	3	3	2	2	1	5	3	1	3	3
$\ell_{2.65}(w,i)$	1	2	2	3	3	3	5	5	5	5	5	8	3	3

We denote by $Q_\alpha(w)$ the maximum of the left local periods (of order α) of w:

$$Q_\alpha(w) = \max\{l_\alpha(w,i) \mid 1 \le i \le |w|\}.$$

A point (or position) i is (left) *critical* if $l_\alpha(w,i) = Q_\alpha(w)$. We denote by $K_\alpha(w)$ the set of (left) critical points of w:

$$K_\alpha(w) = \{i \mid 1 \le i \le |w| \text{ and } l_\alpha(w,i) = Q_\alpha(w)\}.$$

We denote further by $T_\alpha(w)$ and $R_\alpha(w)$ the minimum and the maximum of the (left) critical points:

$$T_\alpha(w) = \min K_\alpha(w), \quad R_\alpha(w) = \max K_\alpha(w).$$

EXAMPLE 8.2.10 (*continued*). For $w = abaababaabaaba$, $Q_2(w) = 5$, $K_2(w) = \{10\}$, $T_2(w) = R_2(w) = 10$, $Q_{2.65}(w) = 8$, $K_{2.65}(w) = \{12\}$, $T_{2.65}(w) = R_{2.65}(w) = 12$.

It is easy to verify that $l_\alpha(w,i) \le p(a_1 \cdots a_i) \le p(w)$ for $\alpha > 1$ and $i = 1, \ldots, |w|$, i.e., the left local periods are smaller than or equal to the period.

Contrary to the case of central repetitions, it is not possible, for left repetitions, to determine a fixed value of the parameter α (not depending on the length of the word w) such that the period of w coincides with the left local period of order α detected at some point i of w. The following example illustrates this fact and the differences between central local periods and left local periods.

EXAMPLE 8.2.11. Let $\alpha \ge 2$ be a real number and let $w = ba^t$, with $t > \alpha$.

The period of w is $p(w) = t+1$. Let $k = \lceil \alpha \rceil$ be the smallest integer greater than or equal to α. The following table gives the values of $c_\alpha(w,i)$ and $l_\alpha(w,i)$ respectively

w	b	a	a	a	$\cdot\ \cdot$	a	a	a	a $\cdot\ \cdot\ \cdot$	a	a	a
i		1	2	3	$\cdot\ \cdot\ \cdot$	$k-1$	k	$k+1$	$\cdot\ \cdot\ \cdot\ \cdot$	$t-1$	t	$t+1$
$c_\alpha(w,i)$		$t+1$	$\cdot$	$\cdot$	$\cdot\ \cdot\ \cdot$	$t+1$	1	1	1 $\cdot\ \cdot\ \cdot$	1	1	
$\ell_\alpha(w,i)$		1	2	3	$\cdot\ \cdot\ \cdot$	$k-1$	1	1	1 $\cdot\ \cdot\ \cdot$	1	1	1

In the previous example $P_\alpha(w) = t+1 = p(w)$, according to the critical factorization theorem, but $Q_\alpha(w) = k-1 \neq p(w)$.

In spite of this, the main theorem of this section states that, for a suitable value of the parameter α, $Q_\alpha(w)$ is equal to the period of the prefix of w of length $R_\alpha(w)$. For instance, in the previous example, $R_\alpha(w) = k-1$ and the prefix of w of length $R_\alpha(w)$ is ba^{k-2}. Its period is $k-1$ and it coincides with $Q_\alpha(w)$. Notice that $\alpha = 2$ does not suffice to establish this relationship. Indeed consider the word $w' = aw$, where w is the word in Example 8.2.2. The values of the function $l_2(w', i)$ are given in the following table.

w'	a	a	b	a	a	b	a	b	a	a	b	a	a	b	a	
i		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$\ell_2(w', i)$		1	2	3	3	1	3	3	2	2	1	5	3	1	3	3

$Q_2(w') = 5$ and the prefix of w' of length $R_2(w') = 11$ has period 8, as one can easily verify.

The parameter required for the main theorem of this section is related to the golden ratio (see Section 8.1.4).

The following theorem states in particular that for $\alpha = \varphi^2 = \varphi + 1 = 2.618\ldots$ the maximal values of local periods $Q_\alpha(w)$ coincide with the global period of the prefix of w of length $R_\alpha(w)$. For convenience of notation the subscript φ^2 will be often omitted. So by $l(w,i), K(w), T(w), R(w)$ we will refer to $l_{\varphi^2}(w,i), K_{\varphi^2}(w), T_{\varphi^2}(w), R_{\varphi^2}(w)$ respectively. In the same way, in the sequel, by left repetitions we will refer to left repetitions of order φ^2.

A fundamental step in the proof of next theorem is given by the following lemma, which uses Theorem 8.1.12 and explains the role of the golden ratio φ in the result.

LEMMA 8.2.12. *If the minimal left repetition at the point i is proper, then* $l(w, i - l(w,i)) \geq l(w,i)$.

Proof. Let z be the minimal repetition at the point i of order $\geq \varphi^2$. Then $z = x^\gamma$, with $|x| = p(z)$ and γ, the order of z, a rational number greater than φ^2. By definition, $l(w,i) = |x|$.

Since this repetition is proper, z is a suffix of $a_1 \cdots a_i$. Consider the word $a_1 \cdots a_j$, with $j = i - |x| = i - l(w,i)$. Since x^γ is a suffix of $a_1 \cdots a_i$, $x^{\gamma-1}$ is a suffix of $a_1 \cdots a_j$. From the inequality $\gamma > \varphi^2 = \varphi + 1$, it follows that $\gamma - 1 = \rho > \varphi$.

Let t be the minimal left repetition at the point j of order $\geq \varphi^2$. By definition

$$l(w, i - l(w,i)) = l(w,j) = p(t).$$

Since x^ρ is a suffix of $a_1 \cdots a_j$, then, either t is a suffix of x^ρ or x^ρ is a suffix of t.

If t is a suffix of x^ρ, then t is a *proper* suffix of $x^{\rho+1} = x^\gamma = z$. Hence t is also a suffix of $a_1 \cdots a_i$, i.e., t is a left repetition at the point i of order $\geq \varphi^2$, with $p(t) < p(z)$. This contradicts the minimality of z. Then x^ρ is a suffix of t.

Assume now that $l(w, i - l(w,i)) < l(w,i)$. Then $p(t) < p(z) = |x|$ and the word x^ρ can also be written $x^\rho = y^\sigma$, for some word y and some rational σ, with $|y| = p(t) < |x|$ and then $\sigma > \rho > \varphi$. The word $x^\rho = y^\sigma$ satisfies the conditions of Theorem 8.1.12. It follows that x^ρ has a suffix of the form u^τ, with $0 < |u| < |x|$ and $\tau > \varphi^2$, contradicting the hypothesis that $z = x^\gamma = x^{\rho+1}$ is the minimal left repetition at the point i of order $\geq \varphi^2$. ■

THEOREM 8.2.13. *Let $w = a_1 a_2 \cdots a_n$ be a nonempty word. On has:*
(i) $p(a_1 \cdots a_{R(w)}) = Q(w)$.
(ii) *If r, s $(r < s)$ are consecutive elements of $K(w)$, then $s - r \leq Q(w)$.*
(iii) $T(w) < \varphi^2 Q(w)$.

Proof. Let us first prove (iii), which states that $T(w) < \varphi^2 Q(w)$, i.e., that the minimal repetition at the point $T(w)$ is external. Indeed, if we assume that this repetition is proper, then, by Lemma 8.2.12,

$$l(w, T(w) - Q(w)) \geq l(w, T(w)) = Q(w),$$

contradicting the fact that $T(w)$ is the least critical point.

Let us now prove (ii). Let us consider two consecutive elements r, s of $K(w)$, with $r < s$. If the minimal left repetition at the point s is proper, then, by Lemma 8.2.12, $s - Q(w)$ is also an element of $K(w)$. Since r and s are consecutive elements of $K(w)$, then $r \geq s - Q(w)$, i.e., $s - r \leq Q(w)$.

In the opposite case, i.e., if the minimal left repetition at the point s is left external, also the minimal left repetition at the point r is left external. One has that $p(a_1 \cdots a_s) = l(w,s) = Q(w)$. Indeed, by definition, $l(w,s) \leq p(a_1 \cdots a_s)$. Moreover, if the minimal left repetition at the position s is external, then $l(w,s)$ is a period of $a_1 \cdots a_s$, i.e., $p(a_1 \cdots a_s) \leq l(w,s)$. Analogously, one has the same result for position r, i.e., $p(a_1 \cdots a_r) = l(w,r) = Q(w)$.

Let us suppose, by contradiction, that $s - r > Q(w)$. Then the word $a_1 \cdots a_{r+Q(w)}$ is a prefix of $a_1 \cdots a_s$ and it has as prefix the word $a_1 \cdots a_r$. Hence $Q(w) = p(a_1 \cdots a_r) \leq p(a_1 \cdots a_{r+Q(w)}) \leq p(a_1 \cdots a_s) = Q(w)$, i.e., the word $a_1 \cdots a_{r+Q(w)}$ has period $Q(w)$. It follows that $a_1 \cdots a_r$ is a suffix of $a_1 \cdots a_{r+Q(w)}$ and then $l(w,r) \leq l(w, r + Q(w))$. Since $l(w,r) = Q(w)$, then $r + Q(w) \in K(w)$. This contradicts the hypothesis that r, s are consecutive elements of K(w) and $s - r > Q(w)$.

Let us finally prove statement (i). The proof is by induction on the integer $m = \mathrm{Card}(K(w))$. We first prove the statement for $m = 1$. In this case $T(w) = R(w)$ and, as a consequence of (iii), the minimal left repetition at the point $T(w)$ is left external, i.e., $p(a_1 \cdots a_{R(w)}) = Q(w)$.

We now have to prove the inductive step. Let us suppose that the statement (i) is true for all words v such that $\mathrm{Card}(K(v)) = m$, and consider a word

$w = a_1 \cdots a_n$ such that $\mathrm{Card}(K(w)) = m + 1$. Let

$$\hat{R}(w) = \max\{i \mid i \in K(w) \text{ and } i < R(w)\}$$

denote the greatest among the critical points of w smaller than $R(w)$. By the inductive hypothesis

$$p(a_1 \cdots a_{\hat{R}(w)}) = Q(w).$$

By statement (ii), $R(w) - \hat{R}(w) \leq Q(w)$. On the other hand, since the minimal left repetition of order $\geq \varphi^2$ at the point $R(w)$ has period $Q(w)$, we have that $Q(w)$ is a period of $a_{\hat{R}(w)-Q(w)+1} \cdots a_{R(w)}$. Set

$$\begin{aligned} u_1 &= a_1 \cdots a_{\hat{R}(w)-Q(w)}, \\ u_2 &= a_{\hat{R}(w)-Q(w)+1} \cdots a_{\hat{R}(w)}, \\ u_3 &= a_{\hat{R}(w)+1} \cdots a_{R(w)}. \end{aligned}$$

We know that u_1u_2 and u_2u_3 have period $Q(w)$ and that $|u_2| = Q(w)$. By Lemma 8.1.2, $u_1u_2u_3 = a_1 \cdots a_{R(w)}$ has period $Q(w)$. This concludes the proof of the theorem. ■

Remark 8.2.14. Notice that, in the case of central repetitions, the least critical point $Z_2(w)$ is bounded above by $P_2(w)$, whereas, for left repetitions, the least critical point $T(w)$ is bounded above by $Q(w)$ times φ^2. The following example shows that $Q(w)$ is not an upper bound for $T(w)$. Consider the word $w = a^{m-1}ba^mba^mb$ $(m > 3)$. The least element of $K(w)$ is $T(w) = 2m + 1$ whereas the maximal local left period is $Q(w) = m + 1$. The last two critical positions of w are $2m + 3$ and $3m + 2$. Their distance is $m - 1 = Q(w) - 2$ and this is the best possible. Indeed it is possible to improve statement (ii) of Theorem 8.2.13 by proving the following tight one. In every sequence of $d \geq \max(1, Q(w) - 2))$ consecutive positions between $T(w)$ and $R(w)$ there is a left critical one (of order φ^2) (Problem 8.2.10).

Remark 8.2.15. The constant φ^2 is tight in Theorem 8.2.13. Indeed, for any $\epsilon > 0$ it is possible to prove that there exist $\bar{n}$ and a constant $D(\epsilon)$ such that for any $n > \bar{n}$ the maximum $Q_{\varphi^2-\epsilon}(f_n)$ of the left local periods of order $\varphi^2 - \epsilon$ of the n-th Fibonacci word (defined in Remark 8.1.13) is smaller than $D(\epsilon)$. Moreover, for any ϵ, the sequence of the maximum of left critical points in f_n $(R_{\varphi^2-\epsilon}(f_n))_{n\in\mathbb{N}}$ is not bounded. Let w_n be the prefix of f_n of length $R_{\varphi^2-\epsilon}(f_n)$. The sequence of periods of w_n, $(p(w_n))_{n\in\mathbb{N}}$ is not bounded (Problem 8.2.11).

8.3. Infinite words

In this section we will consider applications of the results of previous section to the case of one-sided and two-sided infinite words. In particular characterizations of recurrent, periodic and eventually periodic infinite words are given.

8.3.1. Recurrence and periodicity

Recall from Section 1.2 that a one-sided infinite word $x = x_0x_1 \cdots$ is *periodic* if there exists a positive integer p such that $x_i = x_{i+p}$, for all $i \in \mathbb{N}$. The smallest p satisfying previous condition is called the period of x.

A one-sided infinite word $x = x_0x_1 \cdots$, is *eventually periodic* if there exist two positive integers k, p such that $x_i = x_{i+p}$, for all $i \geq k$. An infinite word is *aperiodic* if it is not eventually periodic.

A one-sided infinite word x is *recurrent* if any factor occurring in x has an infinite number of occurrences.

Notice that a one-sided infinite word is periodic if and only if it is recurrent and eventually periodic.

A two-sided infinite word $x = \cdots x_{-1}x_0x_1 \cdots$ is *periodic* if there exists a positive integer p such that $x_i = x_{i+p}$, for all $i \in \mathbb{Z}$. The smallest p satisfying previous condition is called the period of x.

As to concerns the notion of local period, the definitions of previous sections extend to one-sided and to two-sided infinite words but there are some natural differences.

In the case of a one-sided infinite word x, for any order α, there could exist integers j such that there are no central repetitions of order α at position j. In this case the value of $c_\alpha(x, j)$ is $+\infty$.

Notice further that any central repetition cannot be right external. As an example consider the one-sided word $x = x_0x_1x_2x_3 \cdots$ with $x_i \in \{a, b\}$ defined by $x_0 = a$ and for any $i \geq 1$, $x_i = b$ (i.e., $x = abbbbbbbb \cdots$). At position 0 of x, for any $\alpha > 1$ there exists no central repetition of order α, and, consequently, $c_\alpha(x, 0) = +\infty$.

A more sophisticated example is the following one.

EXAMPLE 8.3.1. Let f be the infinite word of Fibonacci (see Section 1.2). For any position j, $c_2(f, j)$ is finite and the square of minimal length having its center in position j is external if and only if $j = F_n - 2$ for some Fibonacci number F_n as proved in next proposition.

PROPOSITION 8.3.2. *In the infinite word of Fibonacci f, there exists a square having its center in any position and the square of minimal length having its center in position j is external if and only if $j = F_n - 2$ for some Fibonacci number F_n. Moreover, when $j = F_n - 2$, the minimal length $c_2(f, j)$ of a square having its center in position j is F_n.*

Proof. Recall that if F_n, $n \in \mathbb{N}$, is the sequence of Fibonacci numbers defined by $F_0 = 1$, $F_1 = 1$, $F_{n+1} = F_n + F_{n-1}$ for $n \geq 1$, one has that $|f_n| = F_n$ for any $n > 0$.

We will prove that for any $n \geq 2$ and for any position $j \leq F_n - 2$ one has that there exists a square having its center in j and the square of minimal length having its center in position j is external if and only if $j = F_k - 2$ for some Fibonacci number F_k, $k \leq n$. Moreover, when $j = F_k - 2$, the value of $c_2(f, j)$ is F_k.

The proof of this is by induction on n. The base of the induction is easily verified for $n = 2, 3$. Let us suppose previous statement is true for $n > 3$ and let us prove it for $n+1$. By inductive hypothesis the statement is true for any j up to $F_n - 2$.

We have that $f_{n+1}f_{n+1}$ is a prefix of f and, moreover, we have that $f_{n+1} = f_n f_{n-1} = f_{n-1}f_{n-2}f_{n-2}f_{n-3}$. Hence (f_{n-2}, f_{n-2}) matches position $F_n - 1$. Moreover, since f_{n-2} is a prefix of $f_{n-3}f_{n+1}$, then in any position j, $F_n - 1 \leq j \leq F_n + F_{n-2}$ one has that there is a repetition of order 2 of length $2F_{n-2}$ having its center in position j, i.e., $c_2(f, j) \leq F_{n-2}$.

Let us consider now j such that $F_n + F_{n-2} \leq j \leq F_n + F_{n-1} - 2 = F_{n+1} - 2$. These positions belong to the occurrence of f_{n-1} of the prefix $f_{n+1}f_{n+1} = f_n f_{n-1} f_{n+1}$ of f. Since $f_{n-1}f_{n-1}$ is also a prefix of f and $f_n f_{n-1} f_{n-1}$ a prefix of $f_n f_{n-1} f_{n+1}$, the inductive hypothesis gives us the information that in any position j, with $F_n + F_{n-2} \leq j \leq F_{n+1} - 2$, there exists an internal square having its center in j, with the exception of the position $\underline{j} = F_n + F_{n-1} - 2 = F_{n+1} - 2$. Moreover, again by inductive hypothesis, we know that in both positions there is "almost" a square centred in it. More precisely there would be a square if the $(j+1)$-th letter of f would be equal to the $(\underline{j} - F_{n-1} + 1)$-th letter of f. And it is not difficult to prove that it is false. By inductive hypothesis there is no central square in $\underline{j}$ of length $\leq 2F_{n-1}$. If there was an internal central square, by a classical result, it would have length a Fibonacci number, i.e F_n, because $F_{n+1} > j + 1$. But again it is not difficult to prove that the $(\underline{j} + 1)$-th letter of f is different from the $(\underline{j} - F_n + 1)$-th letter of f, and so in $\underline{j} = F_{n+1} - 2$ there are no internal central squares. But, since $f_{n+1}f_{n+1}$ is a prefix of f, it is easy to see that $c_2(f, \underline{j}) = F_{n+1}$, and this concludes the proof. ∎

Also in the case of a two-sided infinite word x, "a fortiori" there could exist integers j such that there are no central repetitions of order α at position j, i.e., $c_\alpha(x, j) = +\infty$. However, in the case of a two-sided infinite word x all the central repetitions at every position j such that $c_\alpha(x, j)$ is finite, are internal. As an example consider the two-sided infinite word $y = \cdots y_{-2}y_{-1}y_0y_1y_2y_3\cdots$ with $y_i \in \{a, b\}$ defined by $y_0 = a$ and for any $i \neq 0$, $y_i = b$ (i.e., $y = \cdots bbbbabbbb \cdots$). At position 0 and position -1 of y, for any $\alpha > 1$ there exist no central repetitions of order α and for all other positions there exist a square having it for a center.

The following proposition is an easy consequence of Lemma 8.1.3 and its proof is left to the reader.

PROPOSITION 8.3.3. *Suppose that x is an infinite word that has period q and that there exists a factor v of x with $|v| \geq q$ that has period d, where d divides q. Then x has period d.*

The periodicity of an infinite word x strongly depends on whether the sequence $c_\alpha(x, j)$ of local periods is bounded or not. Let

$$M_\alpha(x) = \sup\{c_\alpha(x, j) \mid j \in \mathbb{N}\}.$$

The following theorem is a consequence of the critical factorization theorem.

Theorem 8.3.4. *An infinite word x is periodic if and only if $M_2(x)$ is finite. Moreover the period of x is equal to $M_2(x)$.*

Proof. If x is periodic then trivially in any position there exists a square having it for a center and the sequence of local periods $(c_2(x,i))_{i\in\mathbb{N}}$ is bounded by the period P of x, i.e., $M_2(x) \le P$. If $M_2(x) < P$ then take a factor v of length $2P$ of x. Clearly P is a period of v and $P_2(v) \le M_2(x) < P$. By the critical factorization theorem $P_2(v)$ is a period of v and $P_2(v) < P$. By the theorem of Fine and Wilf v also has period $d = \gcd(P, P_2(v)) < P$. Since d divides P, by Proposition 8.3.3, d is also a period of x, contradicting the minimality of P.

Let us prove the "if" part of the proposition. Let Z be a position where the sequence $(c_2(x,i))_{i\in\mathbb{N}}$ reaches its maximum $M_2(x)$. We have to prove that for any $i \in \mathbb{N}$, $x_i = x_{i+M_2(x)}$. Take the factor $v = x_r \cdots x_s$ of x where

$$r = \min\{i, Z - M_2(x)\} \text{ and } s = \max\{i + M_2(x), Z + M_2(x)\}.$$

In previous definition if $r < 0$ we consider v defined as $v = x_0 \cdots x_s$.

It is easy to see that position Z is also a critical position for v and its central local period is again $M_2(x)$. This implies that $M_2(x) = P_2(v)$. Hence, by the critical factorization theorem, $p(v) = P_2(v) = M_2(x)$ and, so, $x_i = x_{i+M_2(x)}$. ∎

The proof of the following theorem is analogous to the proof of Theorem 8.3.4 and it is left to the reader.

Theorem 8.3.5. *A two-sided infinite word x is periodic if and only if $M_2(x)$ is finite. Moreover the period of x is equal to $M_2(x)$.*

In both theorems 8.3.4 and 8.3.5 the constant 2 is tight. Indeed, for any $\epsilon > 0$, we can construct one-sided and two-sided infinite words that are non-periodic and in any position have a central repetition of order $2 - \epsilon$, as shown in next example.

Example 8.3.6. For any $\epsilon > 0$ let m be a positive integer such that for any $n \ge m$, $\epsilon > 2/n$. Let v_n be the finite word defined by $v_n = a^{n^2}b^n$, and let y_m the infinite word obtained by concatenating $v_m, v_{m+1}, v_{m+2}, \cdots$.

In any position of y_m there is a central repetition of order $2 - \epsilon$. Indeed, if the square aa or the square bb are not central in position j then either j is a position between the concatenation of v_n and v_{n+1} for some n or j is the position between the sequence of a's and b's inside a word v_n for some n.

In the first case the pair $(a^{n^2-n}b^n, a^{n^2})$ matches position j and in the second case the pair $(b^{n-1}a^{n^2}, b^n a^{n^2-1})$ matches position j. Both $a^{n^2-n}b^n a^{n^2}$ and $b^{n-1}a^{n^2}b^n a^{n^2-1}$ have period n^2+n. The first has length $2n^2$ and the second $2n^2+2n-2 \ge 2n^2$. In both cases the order α of this repetition is greater than or equal to

$$\frac{2n^2}{n^2+n} = \frac{2n}{n+1} = 2 - \frac{2}{n+1} > 2 - \epsilon.$$

One can define a two-sided infinite word $x_m = \cdots x_{-1}x_0x_1\cdots$ starting from the previously defined one-sided infinite word $y_m = y_0y_1\cdots$ by the rule $x_i = y_{|i|}$. It is easy to check that also in any position of x there is a central repetition of order $2-\epsilon$.

THEOREM 8.3.7. *Let x be a one-sided infinite word. If x is recurrent then in any position there is a central repetition of order α, for any α such that $1 < \alpha \leq 2$. Conversely, for any $\alpha \geq 2$, if in any position there is a central repetition of order α, then x is recurrent. In particular x is recurrent if and only if in any position there is a central repetition of order 2.*

Proof. Let us suppose that x is recurrent. If we prove that in every position j there exists a square having it for a center then, "a fortiori", there is a central repetition of order α for any $\alpha < 2$. Let $k > 0$ be the position where the prefix $x_0\cdots x_j$ occurs for the second time, i.e., $x_0\cdots x_j = x_k\cdots x_{k+j}$. If we set $v = x_{j+1}\cdots x_{k+j}$ then it is not difficult to see that (v,v) matches position j and, so, $z = v^2$ is a square having its center in position j.

Suppose now that in every position of x there exists a central repetition of order $\alpha \geq 2$. In particular, in every position of x there exists a square having it for a center. If the sequence of central local periods is bounded then, by Theorem 8.3.4, x is periodic and, so, recurrent. If the sequence of central local periods is not bounded then there exists a sequence $(j_i)_{i\in\mathbb{N}}$ of positions such that for any i, $c_2(x,j_i) > c_2(x,h)$ for any $h < j_i$. For any i consider the finite word $v = x_0\cdots x_{j_i+c_2(x,j_i)}$. It is not difficult to prove that position j_i is the least critical position for v and its central local period is again $c_2(x,j_i)$, i.e., $j_i = Z_2(v)$ and $c_2(x,j_i) = c_2(v,j_i)$. By Corollary 8.2.6 the minimal square $z = u^2$ having its center in j_i is left external in v. This means that the prefix $x_0\cdots x_{j_i}$ is a suffix of u and then it is also suffix of $x_0\cdots x_{j_i+c_2(x,j_i)}$, i.e., the prefix $x_0\cdots x_{j_i}$ occurs a second time in x. Since the sequence $(j_i)_{i\in\mathbb{N}}$ of positions is an increasing sequence, we find a sequence of prefixes of x of increasing length that have a second occurrence in x. This fact easily implies that any prefix of x has a second occurrence, i.e., x is recurrent. ∎

The value 2 is tight in both directions of previous theorem. For any $\epsilon > 0$ it is known that there exists a one-sided recurrent infinite word x that is $(1+\epsilon)$-power free. For $\alpha \geq 2+2\epsilon$, the word x has no central repetition of order α in any position.

Conversely, for fixed m, the word $y = a^m baaaaa\cdots$, i.e., the word $y = y_0y_1\cdots$ with $y_i = a$ if $i \neq m$ and $y_m = b$ has in every position a central repetition of order $2-(1/m)$ and it is not recurrent.

For two-way infinite words there are no similar characterizations. Any square-free recurrent two-sided infinite word obviously fails to have a square having its center in any position. Notice that in a square-free recurrent one-sided infinite word, in any position the minimal central repetition of order 2 exists (according to Theorem 8.3.7) and it must be left external.

In the following characterizations of periodic infinite words, the local property here considered is related to unbordered words.

Let x be an infinite word (one or two sided). Denote by $U(x)$ the maximal length of unbordered factors of x if such length exists, infinite otherwise:

$$U(x) = \sup\{|v| \mid v \text{ is an unbordered factor of } x\}.$$

PROPOSITION 8.3.8. *Let x be a one-sided infinite word. If $U(x)$ is finite then x is recurrent.*

Proof. Let f be a finite prefix of x. We prove, by induction on $|f|$, that if f has no second occurrence in x then $U(x) = \infty$.

If $|f| = 1$, i.e., if $f = a$ with $a \in A$, then all the prefixes of x are unbordered and, consequently, $U(x) = \infty$.

Let us suppose that the statement holds for $|f| = n \geq 1$ and consider the case $|f| = n+1$. The prefix f can be written as $f = av$, with $a \in A$ and $|v| = n$. Let W be the set of prefixes of x having v as suffix. We distinguish two cases.

CASE 1. W is finite. Let w be an element of W of maximal length and let w' be such that $w = w'v$. Let y be the infinite word defined by the relation $x = w'y$. The prefix of y of length n is v and v has no second occurrence in y. By the inductive hypothesis $U(y) = \infty$. Since $U(y) \leq U(x)$, the statement follows.

CASE 2. W is infinite. Let w be an arbitrary element of W having length greater than $2n + 2$. By definition we can write

$$w = avw'v.$$

Let au be the longest unbordered prefix of av. The word au is not a suffix of the word $g = avw'u$, since the last letter of w' is different from a (otherwise $f = av$ would have a second occurrence in w and hence in x). Let $s(g)$ denote the shortest border of g. It is easy to verify that $s(g)$ is unbordered.

$s(g)$ cannot be a proper prefix of au. Indeed, in this case, $s(g)$ is also a suffix of u, contradicting the hypothesis that au is unbordered.

$s(g)$ is not equal to au, since au is not a suffix of g. Moreover it is not possible that $s(g)$ is a prefix of av and that has as prefix au, since au is the longest unbordered prefix of av.

It follows that av is a prefix of $s(g)$.

Since av has not a second occurrence in g, we have that $s(g) = g$, i.e., g is unbordered. We conclude that there exist unbordered factors of x of increasing length, i.e., $U(x) = \infty$. ∎

THEOREM 8.3.9. *Let x be a one-sided infinite word. Then x is periodic if and only if $U(x)$ is finite. Moreover the period of x is $U(x)$.*

Proof. Suppose that $U(x)$ is finite. By Proposition 8.3.8 x is recurrent. By Proposition 8.3.7 in every position of x there exists a square having its center in it. Let $z_i = u_i^2$ be the square of minimal length having its center in position

i. By Proposition 8.2.4 applied to the finite prefix w of x having length $i + |u_i|$, we know that u_i is unbordered. By hypothesis, it follows that $|u_i| \leq U(x) < \infty$, i.e., the sequence of central local periods is bounded by $U(x)$. Therefore, by Theorem 8.3.4, x is periodic and the period is smaller than or equal to $U(x)$. The period P of x cannot be smaller than $U(x)$ otherwise, if v is an unbordered factor of x of length $|v| = U(x)$, v has period $P < |v|$ and hence it is bordered, a contradiction.

If x is periodic with period P then, by Theorem 8.3.4, $P = M_2(x)$, i.e., there exists a position j such that the minimal square $z = uu$ having its center in position j is such that $|u| = M_2(x) = P$. By Proposition 8.2.4, the word u is unbordered. Since z is right internal, u is a factor of x, i.e., $U(x) \geq P$. But, as proved above, the period P of x cannot be smaller than $U(x)$ and, consequently, $P = U(x)$. ∎

COROLLARY 8.3.10. *Let $x = \cdots a_{-1}a_0a_1\cdots$ be a two-sided infinite word. One has that x is periodic if and only if $U(x)$ is finite. Moreover the period of x is $U(x)$.*

Proof. Suppose that $U(x)$ is finite. Let i be an integer. We have to prove that $a_i = a_{i+U(x)}$. Pick an unbordered factor $v = a_j \cdots a_{j+U(x)-1}$ of x having length $U(x)$ and define $m = \min\{i, j\}$. The one-sided infinite word $x' = a_m a_{m+1} a_{m+2} \cdots$ is such that $U(x') = U(x)$ because it has v as a factor, and, by Theorem 8.3.9 it has period $U(x)$. Since $i \geq m$, one has $a_i = a_{i+U(x)}$.

Suppose that x is periodic with period P and let $x' = a_0a_1a_2\cdots$ be the one-sided infinite word that is the right suffix of x starting from position 0. The word x' has also period P (see also Problem 8.3.1) and, by Theorem 8.3.9, $P = U(x')$. Since x is periodic any unbordered word v that is a factor of x is also a factor of x'. It follows that $U(x) = U(x') = P$. ∎

The following theorem summarizes some of the results presented in this section.

THEOREM 8.3.11. *Let x be a (one-sided or two-sided) infinite word. The following conditions are equivalent.*

1. *x is periodic and P is its minimal period.*
2. *$M_2(x) = P$.*
3. *$U(x) = P$.*

8.3.2. Characterizations of eventually periodic words

We now give a characterization of one-sided infinite eventually periodic words. This characterization property is similar to the property characterizing one-sided recurrent infinite words as described in Theorem 8.3.7. Now we require something less (for all "large enough" positions j there is a square having j for a center) and something more (the minimal central repetition at position j is internal if j is "large enough").

Notice further that here we do not explicitly require, as in Theorem 8.3.4, that the sequence $(c_2(x,j))_{j\in\mathbb{N}}$, is bounded. This condition is actually obtained (see Lemma 8.3.13 below) as a consequence of the existence of an internal repetition for any large enough position j.

THEOREM 8.3.12. *A one-sided infinite word $x = x_0x_1x_2\cdots$ is eventually periodic if and only if there exists a number k such that for any $j \geq k$ there exists a suffix of $x_0 \cdots x_j$ that is also a prefix $x_{j+1}x_{j+2}\cdots$, i.e., at any position $j \geq k$ there exists a proper central repetition of order* 2.

The proof of this theorem is based on the following lemma.

LEMMA 8.3.13. *If there exists a number k such that in any position $j \geq k$ there exists a proper central repetition of order 2, then the sequence of local periods $(c_2(x,j))_{j\geq k}$ is bounded.*

Proof. The proof is by contradiction. Let us suppose that the sequence of central local periods at positions $j \geq k$ is not bounded. By hypothesis $(j - c_2(x,j)) > 0$ for any $j \geq k$.

Let j_1 be such that $(j_1 - c_2(x,j_1))$ assumes the minimal value between all $j \geq k$ and let j_2 be the least position greater than j_1 such that $c_2(x,j_2) > c_2(x,j_1)$.

Consider the word $v = x_{j_1-c_2(x,j_1)+1}x_{j_1-c_2(x-j_1)+2}\cdots x_{j_2+c_2(x-j_2)}$.

In the sequel of this proof, for simplicity, with abuse of notation, we will denote by t the position $t - (j_1 - c_2(x,j_1))$ of v. It is easy to verify that $c_2(x,j_1) = c_2(v,j_1)$.

Also notice that the minimal square having its center in position j_2 of v is not right external. This square cannot be left external by the minimality of j_1 and the fact that, for any position t of v, $c_2(x,t) \geq c_2(v,t)$. Since this square is not left external then $c_2(v,j_2) = c_2(x,j_2) > c_2(x,j_1) \geq c_2(v,j_1)$.

Since for any position t of v, $c_2(x,t) \geq c_2(v,t)$, and since j_2 is the least position greater than j_1 such that $c_2(x,j_2) > c_2(x,j_1) = c_2(v,j_1)$, one has that for any position t of v with $j_1 \leq t < j_2$, $c_2(v,t) \leq c_2(v,j_1)$. If $j_1 - c_2(x,j_1) + 1 \leq t < j_1$, we also have, by construction of v, that $c_2(v,t) \leq c_2(v,j_1)$.

By Corollary 8.2.7, $c_2(v,j_2) > j_2 - (j_1 - c_2(x,j_1) + 1) + 1$, i.e., $j_2 - c_2(v,j_2) < j_1 - c_2(x,j_1)$, contradicting the minimality of j_1. ∎

Proof of Theorem 8.3.12. If x is eventually periodic then we can write $x = wy$, where y is a one-sided infinite word that is periodic with period P. Hence, if we set $k = |w| + P$, it is easy to check that at any position $j \geq k$ of x there exists a central repetition of order 2 that is internal.

Let us prove the "if" part. Let us write $x = uy$, where $|u| = k$. Let us consider a position i of y and the corresponding position $i + k$ of x. Since y is a suffix of x one has that in any position $i \geq 0$ there exists a central repetition of order 2 and $c_2(y,i) \leq c_2(x,i+k)$. Hence, by Lemma 8.3.13 the sequence of local periods $(c_2(y,i))_{i\in\mathbb{N}}$ is bounded and, by Theorem 8.3.4, y is periodic. ∎

By the fact that an infinite word is periodic if and only if it is recurrent and eventually periodic, and by Theorem 8.3.7, one has

COROLLARY 8.3.14. *A one-sided infinite word x is periodic if and only if at any position there is a central repetition of order 2 and this repetition is external only for finitely many positions.*

REMARK 8.3.15. Notice that, in previous theorem, one cannot bound the period of y as function of k, as shown by the one-sided infinite word $y_n = ba^nba^nba^n \cdots$ where n is any positive natural number. In this word the number k is 2 and this word has period n.

EXAMPLE 8.3.6 (*continued*). The word y_1 previously defined with $m = 1$ shows that the number 2 is tight in Lemma 8.3.13. Indeed for any ϵ it is easy to see that there exists a constant $k(\epsilon)$ such that at any position $j \geq k(\epsilon)$ there exists a central repetition of order $2 - \epsilon$, but the sequence of local periods at positions $j \geq k(\epsilon)$ is not bounded.

The same word y_1 also shows that the constant 2 is tight in Theorem 8.3.12, because y_1 is not eventually periodic.

A more sophisticated example, showing that the constant 2 is tight in Theorem 8.3.12, is given by the infinite word of Fibonacci f defined in Example 8.3.1. The word f is not eventually periodic but it is possible to prove that for any ϵ there exists a constant $k(\epsilon)$ such that at any position $j \geq k(\epsilon)$ there exists a central repetition of order $2 - \epsilon$, and that the sequence of local periods at positions $j \geq k(\epsilon)$ is bounded (Problem 8.3.3).

An analogy of Theorem 8.3.12 does not hold for two-sided infinite words, as shown by next example.

EXAMPLE 8.3.16. For any $\alpha > 1$ one can construct a non-periodic two-sided infinite word $x_\alpha = \cdots x_{-1}x_0x_1 \cdots$ such that at any position there exists a central repetition of order α.

Consider the sequence of all integers $0, -1, 1, -2, 2, -3, 3, \cdots -i, i, \cdots$. Our construction inductively fixes letters in the word x_α in order to have a central repetition at position n_j, where $n_j = (-1)^j \lceil (j/2) \rceil$ is the j-th element of previous sequence.

First, let $k = \lceil (\alpha/2) \rceil$ be the smallest integer greater than or equal to $\alpha/2$ and set $x_{-k} = x_{-k+1} = \cdots = x_0 = x_1 = \cdots = x_{k-1} = a$ and $x_k = b$.

By construction at position 0 there is a central repetition of order α. Suppose that we have fixed letters from position s_j up to position t_j such that at all positions $n_0, \cdots n_j$ there exists a central repetition of order α that is internal to the word $x_{s_j} \cdots x_{t_j}$. Since position n_{j+1} is adjacent to position n_{j-1} then $s_j \leq n_{j+1} \leq t_j$.

Let us denote $u = x_{s_j} x_{s_j+1} \cdots x_{n_{j+1}}$ and $v = x_{n_{j+1}+1} \cdots x_{t_j}$. Suppose that $w = uv$ has period P and that a is the letter such that wa has period greater than P. Set $x_{t_j+1} = a$. Now assign letters from the position $s_{j+1} = n_{j+1} - k|vau|$ to the position $t_{j+1} = n_{j+1} + 1 + k|vau|$ so that

$$(vau)^k = x_{s_{j+1}} x_{s_{j+1}+1} \cdots x_{n_{j+1}} \text{ and } (vau)^k = x_{n_{j+1}+1} \cdots x_{t_{j+1}} .$$

Notice that this assignment is compatible with previous assignment and that at position n_{j+1} there exists a central repetition of order α. Notice further that, since $x_{s_{j+1}} x_{s_{j+1}+1} \cdots x_{t_{j+1}}$ has wa as factor, its period is strictly greater than the period P of w. Using this property it is not difficult to prove that the infinite word x_α is non-periodic.

As regards left repetitions, in the case of one-sided infinite words x, for any order α, the minimal left repetition of order α is defined at any position j, i.e., $l_\alpha(x, j)$ is defined for any $\alpha > 1$ and for any $j \in \mathbb{N}$. In the case of two-sided infinite words x, there could exist integers j such that there are no left repetitions of order α at position j and, consequently, l_α is not defined in position j. We can define, analogously to the central case,

$$L_\alpha(x) = \sup\{l_\alpha(x, j) \mid j \in \mathbb{N}\}.$$

THEOREM 8.3.17. *Let $x = a_1 a_2 \cdots$ be a one-sided infinite word. $L_{\varphi^2}(x)$ is finite if and only if the word x is eventually periodic, i.e., $x = wy$, with y periodic. Moreover, if P is the period of y then $P \leq L_{\varphi^2}(x)$.*

Proof. If x is eventually periodic then we can write $x = wy$ where y is a one-sided infinite word that is periodic with period P. Hence, if we set $k = |w| + 3P$, it is easy to check that at any position $j \geq k$ of x there exists a left repetition of order $3 > \varphi^2$ that is internal and that $l_{\varphi^2}(x, j) \leq P$. Hence $L_{\varphi^2}(x) \leq \max\{P, Q_{\varphi^2}(w)\}$, where $Q_{\varphi^2}(w)$ is the maximum of the left local periods of order φ^2 of the prefix w of x of length k.

Let us prove the "only if" part. The proof is by induction on $Q = L_{\varphi^2}(x)$.

If $Q = 1$ then for any $j \geq 3$, $a_1 \cdots a_j$ ends with a cube of period 1, i.e., it ends with a^3 where a is a letter. Trivially the infinite word x is periodic with period 1.

Let us suppose the statement is true for any Q', $1 \leq Q' \leq Q$. Two cases are possible.

CASE 1. There are infinitely many positions j such that $l_{\varphi^2}(x, j) = Q$. We want to prove that x has period Q. Let us consider position i, $1 \leq i$. We have to prove that $a_i = a_{i+Q}$. Take j such that $j \geq i+Q$ and $l_{\varphi^2}(x, j) = Q$. By Theorem 8.2.13 (i), the word $a_1 \cdots a_j$ has period Q and, consequently $a_i = a_{i+Q}$.

CASE 2. There are finitely many number j such that $l_{\varphi^2}(x, j) = Q$. Let s be the greatest of these numbers and let $x' = a_{s+1} a_{s+2} \cdots$ be the suffix of x that starts at position $s + 1$. Since the maximum of the sequence $(l_{\varphi^2}(x, i))_{i \in \mathbb{N}}$ is a number $Q' < Q$ then x' satisfies the inductive hypothesis and consequently it is eventually periodic with period $P \leq Q' < Q$. Since x' is a suffix of x, also x is eventually periodic with period $P < Q$. ∎

We now give another characterization of one-sided infinite eventually periodic words that is very similar to the characterization given in Theorem 8.3.12.

Theorem 8.3.18. *A one-sided infinite word $x = a_0a_2\cdots$ is eventually periodic if and only if there exists a number k such that for any $j \geq k$ there exists a suffix of $a_0\cdots a_j$ of order greater than φ^2.*

Proof. If the one-sided infinite word $x = a_0a_2\cdots$ is eventually periodic then there exist natural numbers $M > 0$, $Q > 0$, such that the infinite word $x' = a_Ma_{M+1}\cdots$ has period Q. Since $\varphi^2 < 3$, in any position j, $j \geq M + 3Q$, there exists a left repetition of order φ^2 that is internal.

Let us suppose now that there exists a number k such that in any position $j \geq k$ there exists a left repetition of order φ^2 that is internal. If the sequence $(l_{\varphi^2}(x,i))_{i\in\mathbb{N}}$ is not bounded then there exists an increasing subsequence of positions $(j_i)_{i\in\mathbb{N}}$ such that for any $i \in \mathbb{N}$ and any position $s < j_i$ one has that $l_{\varphi^2}(x,j_i) > l_{\varphi^2}(x,s)$. By Theorem 8.2.13 (*iii*) applied to the words $a_1\cdots a_{j_i}$, the minimal left repetition of order φ^2 at position j_i is external, a contradiction whenever $j_i > k$. Therefore the sequence $(l_{\varphi^2}(x,i))_{i\in\mathbb{N}}$ is bounded and by applying Theorem 8.3.17 x is eventually periodic. ∎

Remark 8.3.19. By the statement of Theorem 8.3.17 one has that any suffix y of x that is periodic has period P where P is bounded by $L_{\varphi^2}(x)$. But one cannot bound the length of the shortest prefix w of x such that $x = wy$ with y periodic, as shown by the one-sided infinite word $x_n = (aaab)^n aaaaaaaaaaaa\cdots$ where n is any positive natural number. In this word $L_{\varphi^2}(x_n) = 4$ and the word $w = (aaab)^n$ has length $4n$.

By the proof of Theorem 8.3.18 one can see that any suffix y of x that is periodic has period P where P is bounded by k where k is the number in the statement of the theorem but one cannot bound the length of the shortest prefix w such that $x = wy$ with y periodic. Indeed in the same previous example $x_n = (aaab)^n aaaaaaaaaaaaa\cdots$ where n is any positive natural number, one can take $k = 14$ and the word $w = (aaab)^n$ has length $4n$.

The following example shows that in Theorem 8.3.18 the number φ^2 is tight.

Example 8.3.1 (*continued*). Let f be the infinite word of Fibonacci. For any $\epsilon > 0$ there exists a constant $\bar{n}$ such that for any $n \geq \bar{n}$ there exists a left repetition of order $\varphi^2 - \epsilon$ at position n and f is not eventually periodic (Problem 8.3.4).

An analogous of Theorem 8.3.18 does not hold for two-sided infinite words, as shown by next example.

Example 8.3.16 (*continued*). For any $\beta > 1$, take $\alpha = 2\beta$. The word x_α has, by construction, at any position a left repetition of order β, that is obviously internal.

The following theorem summarizes some of the results presented in this section.

Theorem 8.3.20. *Let $x = a_0a_1\cdots$ be a one-sided infinite word. The following conditions are equivalent.*

1. *x is eventually periodic.*
2. *There exists an integer k such that, for any $j \geq k$ there exists a suffix of $a_0 \cdots a_j$ that is equal to a prefix of $a_{j+1}a_{j+2}\cdots$.*
3. *There exists an integer h such that for any $j \geq h$ there exists a suffix of $a_0 \cdots a_j$ of order greater than φ^2.*

Problems

Section 8.1

8.1.1 A word $w = a_1 \cdots a_n$ has period $p \leq n$ if and only if for any integers i, j, $1 \leq i, j \leq n$

$$i \equiv j \pmod p \Rightarrow a_i = a_j.$$

If q is a period of w then for any positive integer k such that $kq \leq n$, kq is a period of w. If p is a period of w and q is a period of the suffix of length $n-p$ of w then $p+q$ is also a period of w. Therefore for any positive integer k such that $p+kq \leq n$, $p+kq$ is a period of w.

8.1.2 Consider the non directed graph $G = (I_{p+q}, E)$, where the set of vertices $I_{p+q} = \{1, 2, 3, \cdots, p+q-1, p+q\}$ is the set of positive integers smaller than or equal to $p+q$ where p, q are integers such that $\gcd(p,q) = 1$. The arc $\{i, j\} \in E$ if and only if $|i-j| \in \{p, q\}$.
Prove that graph G is a cycle. Deduce that, if G_t is the graph obtained by G by eliminating vertex t and all arcs containing t, then G_t is a connected graph.

8.1.3 Let $w = a_1 a_2 \cdots a_{p+q-1}$ be a word that has period p and period q with $\gcd(p,q) = 1$.
Use graph G_i with $i = p+q$ defined in previous problem to give a new proof of Theorem 8.1.4 that works for the case where $\gcd(p,q) = 1$. (Hint. If there is an arc (i,j) in G_{p+q} then $a_i = a_j$).

8.1.4 Let $w = a_1 \cdots a_n$ and $v = b_1 \cdots b_n$ be two words having same length n, such that v has period p and w has period q with $p \neq q$ and $p+q \leq n$. Suppose that there exists a position t, $1 \leq t \leq n$ such that for any position $i \neq t$, $1 \leq i \leq n$ one has that $a_i = b_i$ (i.e., the two words w and v coincide except, maybe, in position t). Then w and v have both period $r = \gcd(p,q)$. Since $\gcd(p,q) \leq \min\{p,q\} \leq \lfloor n/2 \rfloor$, then $w = v$. (Hint. Use the technique developed in previous problems, and in particular, if $\gcd(p,q) = 1$ use graph G_t).

8.1.5 Prove the following statement.
A necessary and sufficient condition for two non-empty words u, v to be power of the same word is that uv and vu contain a common left factor of length $|u| + |v| - \gcd(|u|, |v|)$.
(Hint. If $v < u$ then $uv < u^2$ and $vu < v^k$ for some $k \geq 2$. Use the theorem of Fine e Wilf).

8.1.6 A factor of a word v is internal if it is not a suffix or a prefix of v. Prove that a word u is primitive if and only if u is not an internal factor of uu.

*8.1.7 Prove the equivalence of 1, 2 and 4 in Theorem 1.3.13 by using the theorem of Fine and Wilf.

**8.1.8 A triple (p_1, p_2, p_3) is called a *good triple* if

$$h(p_1, p_2, p_3) = gcd(p_1, p_2, p_3) = 1.$$

Given a good triple (p_1, p_2, p_3) prove that there exists a word w of length $|w| = \frac{1}{2}(p_1 + p_2 + p_3 - 3) = f(p_1, p_2, p_3) - 1$ over an alphabet of 3 letters that has period p_1, p_2 and p_3.

8.1.9 If $\Pi(w) = \{0 = p_0 < p_1 < \cdots < p_s = |w|\}$ and if $\delta_h = p_h - p_{h-1}$, $1 < h \leq s$, then the sequence of the differences δ_h is a non-increasing sequence. (Hint. Use Lemma 8.1.1 and Problem 8.1.1).

Section 8.2

8.2.1 Fix $\alpha > 1$. If j is a critical point for w and i, $j < i < |w|$ (respectively $0 < i < j$) is not a critical position, then the minimal central repetition of order α at position i is not left external (respectively right external).

8.2.2 Let $w = a_1 \cdots a_n$ and let $u = a_i \cdots a_j$ be a factor of w with $1 \leq i < j < n$. Either $c_2(w, j) < p(u)$ or $c_2(w, j) \geq |u| + 1$.

8.2.3 Let $w = a_1 \cdots a_n$ and let $v = a_h \cdots a_k$, $1 < h$, $k \leq n$ be a factor of w such that $p(v) \geq 2$ and the word $v' = a_{h-1}v$ has period $p(v') > p(v)$ (one cannot extend the word v to the left maintaining the period).
Show that if $Z_2(v)$ is the least critical position of v then $c_2(w, Z_2(v)) > p(v)$.

8.2.4 Find examples different from $a^m ba^m ba^m$ that show that statement (iii) in the proof of critical factorization theorem is tight.
(Hint. Look for words in the set of central words defined in Chapter 2).

8.2.5 Let $k = \lceil \alpha \rceil$. If the period of w is smaller than or equal to k then every position of w is left critical (of order α).

*8.2.6 For any natural number n there does not exist any prefix (or suffix) of the n-th Fibonacci word f_n having order $\geq \varphi + 1$.

8.2.7 Prove that if a square uu is a factor of a Fibonacci word then its length is a Fibonacci number and it is a conjugate of some other Fibonacci word.

8.2.8 Fix $\alpha > 1$. If j is a left critical point for w and i, $j < i < |w|$, is not a left critical position, then the minimal left repetition of order α at position i is not left external.

8.2.9 Let $w = a_1 \cdots a_n$ and let $v = a_h \cdots a_k$, $1 < h$, $k \leq n$, be a factor of w such that $p(v) \geq 2$ and the word $v' = a_{h-1}v$ has period $p(v') > p(v)$ (one cannot extend the word v to the left maintaining the period).

Show that if $T_{\varphi^2}(v)$ is the least critical position of the word v then $l_{\varphi^2}(w, T_{\varphi^2}(v)) > p(v)$.

**8.2.10 In every sequence of $d \geq \max(1, Q(w) - 2))$ consecutive positions between $R(w)$ and $T(w)$, there is a left critical one (of order φ^2).
(Hint. Use induction on the period $p(w)$ of w. The base of induction is $p(w) = 1, 2, 3$. In the inductive step suppose that there exist two consecutive critical points $r < s$ such that $s - r > Q(w) - 2$. Find the positions j, $r < j < s$, where $l_{\varphi^2}(w, j)$ reaches the maximum value. Use inductive hypothesis, Theorem 8.2.13 and previous two problems in order to find a contradiction).

**8.2.11 Prove that the constant φ^2 is tight in Theorem 8.2.13.
(Hint. Follows the suggestions in Remark 8.2.15).

**8.2.12 Prove that any word w admits a factorization $w = vu$ such that there exists at most one internal left repetition of $w, |v|$ of order φ^2 and $|u| \leq Cp(v)$, with the real constant C smaller than or equal to 2.

**8.2.13 Prove or disprove the following open conjecture. Let v be a word of length $2n$ such that its prefix u of length n is an unbordered word and such that any of its unbordered factors have length $\leq n$. Then v is a square, i.e., $v = u^2$.

Section 8.3

8.3.1 Let x be a two-sided infinite word and suppose that P is the period of x, and v is a factor of x of length $|v| \geq 2P - 1$ or v is a one-sided infinite word that is a suffix of x. Then P is also the period of v. Find a periodic word x with minimal period P and a factor v of x with $|v| = 2P - 1$ such that the period of v is smaller than P.

8.3.2 Let x be a one-sided eventually periodic infinite word and let p be its (shortest) period. Let $m(1)$, $m(2)$, ... be an increasing sequence of points such that $m(i+1) < 2m(i)$, and denote $p(i) = m(i+1) - m(i)$. Prove that, if $c_2(x, m(i)) = p(i)$, then from a given rank, the sequence $(p(i))_{i \geq 1}$ is constant and equal to the period p of the word x.

*8.3.3 Let f be the infinite word of Fibonacci. For any ϵ there exists a constant $k(\epsilon)$ such that at any position $j \geq k(\epsilon)$ there exists a central repetition of order $2 - \epsilon$, and the sequence of local periods at positions $j \geq k(\epsilon)$ is bounded.

*8.3.4 Let f be the infinite Fibonacci word. For any $\epsilon > 0$ there exists a constant $\bar{n}$ such that for any $n \geq \bar{n}$ there exists a left repetition of order $\varphi^2 - \epsilon$ at position n and f is not eventually periodic.

8.3.5 A two-sided infinite word $y = \cdots a_{-1}a_0a_1 \cdots$ is *eventually periodic to the right* with period P if there exists an integer j such that the one-sided infinite word $a_j a_{j+1} a_{j+2} \cdots$ is periodic with period P. It is *eventually periodic to the left* with period P if there exists an integer j such that the one-sided infinite word $a_j a_{j-1} a_{j-2} \cdots$ is periodic with period P

Prove that if $L_{\varphi^2}(x)$ is finite then x is eventually periodic to the left with period P_1 and it is eventually periodic to the right with period P_2. Moreover $P_1 = L_{\varphi^2}(x) \leq P_2$ and if $P_1 = P_2$ then x is periodic with period P_1.
(Hint. Cf. the proof of Theorem 8.3.17.)

8.3.6 For each non-eventually periodic one-sided infinite word w, there exists a one-sided infinite word w' so that

a) each factor of w' is a factor of w,

b) w' does not begin in any $1+\varphi$ powers.

**8.3.7 Let w be a one-sided infinite word such that all except finitely many prefixes have a square uu as suffix with $|u| \leq 4$. Prove that w is eventually periodic.
Give examples that show that if the condition on the length of u is relaxed to $|u| \leq 5$ above conclusion does not hold anymore.

**8.3.8 Let w be a one-sided infinite word such that all except finitely many prefixes v have a suffix u such that

a) u appears a second time as factor in v.

b) $|u|/|v| \geq c$ for some fixed constant c.

Find the smallest value (or the *inf* of those values) for c so that previous conditions imply that w is eventually periodic.

**8.3.9 Let w be a recurrent word and let $R(n)$ be the smallest integer such that any factor of w of length $R(n)$ contains all factors of w of length n. $R(n)$ is called the recurrence function of w. Prove that, if for all except finitely many n $R(n)/n < 2+\varphi$, then w is periodic.

Notes

The original reference for Theorem 8.1.4 is Fine and Wilf 1965. Another proof can be found in Lothaire 1983. A good reference to Euclid's algorithm is Knuth 1988. The ideas used in Problem 8.1.3 can be found in Choffrut and Karhumäki 1997 and also independently in Giancarlo and Mignosi 1994. A solution to Problem 8.1.5 can be found in Lentin and Schützenberger 1967. A solution to Problem 8.1.4 can be found in Berstel and Boasson 1999. Theorem 8.1.7 is from Castelli et al. 1999, where it is also possible to find a solution to Problem 8.1.8. The words described in this problem are words of Arnoux and Rauzy and can be considered as a generalization of Sturmian words to a three letter alphabet (Chapter 2). A generalizations of these results to the case of more than three periods is in Justin 2000. A notion of quasi-periodicity for words has been introduced in Apostolico and Ehrenfeucht 1993. For recent contributions on this subject cf. Régnier and Mouchard 2000 and Brodal and Pedersen 2000 and references therein. For other extensions of the notion of periodicity cf. Carpi and Luca 2000 and references therein. A solution of Problem 8.1.7 can be found in Epifanio, Koskas, and Mignosi 1999. Many results and applications concerning generalizations to the multidimensional case of Theorem 8.1.4 have been developed, starting from the seminal works Amir and Benson 1992, Amir and Benson 1998.

The equivalence of points *1)*, *2)* and *3)* of Theorem 8.1.11 is proved in Guibas and Odlyzko 1981. Point *4)* as well as some of the proofs are from Breslauer 1995. A simple proof of the equivalence of only *1)* and *2)* can be found in Halava, Harju, and Ilie 2000. They also describe a linear time algorithm which, given a word, computes a binary one with the same set of periods. Some related results can be found in E. Rivals 2001 and in Régnier and Mouchard 2000.

Theorem 8.1.16 and its corollary are from Fraenkel and Simpson 1998. See also Crochemore, Hancart, and Lecroq 2000. Lemma 8.1.14 is from Crochemore and Rytter 1995a. The simple and elegant proof given here is due to V. Diekert (Chapter 12). Short surveys on related problems can be found in the introductions of the beautiful papers Kolpakov and Kucherov 1999a, Kolpakov and Kucherov 1999b. See also Crochemore and Rytter 1994, Czumaj and Gasieniec 2000 and references therein. A solution to Problem 8.1.6 can be found in Crochemore and Rytter 1995a.

A weak form of the critical factorization theorem was first conjectured in Schützenberger 1976 and settled in Césari and Vincent 1978. Subsequent improvements in Duval 1979 lead to the actual formulation (see also Duval 1982, Duval 1998). The proof reported here is from Duval, Mignosi, and Restivo 2001. Among the applications of the critical factorization theorem we cite Crochemore and Perrin 1991 and Breslauer, Jiang, and Jiang 1997. A solution to Problem 8.2.6 can be found in Mignosi and Pirillo 1992. A solution to Problem 8.2.7 can be found in Séébold 1985 and Pirillo 1997 for further improvements. A solution to Problem 8.2.12 can be found in Mignosi, Restivo, and Salemi 1995.

Theorem 8.2.13 is proved in Mignosi et al. 1995 and in Mignosi, Restivo, and Salemi 1998

Problem 8.2.13 states an old standing open conjecture in Duval 1982. This is the latest and strongest of a sequence of three conjectures stated by different authors. To our knowledge, no weaker versions of it have even been proved.

Theorem 8.3.9 and Corollary 8.3.10 are in Ehrenfeucht and Silberger 1979 (see also Assous and Pouzet 1979). The proof here reported, as well as Proposition 8.3.8, is inspired by Duval 1982. Almost all the remaining results in the last section are from Mignosi et al. 1995 and Mignosi et al. 1998, if they concern left repetitions and from Duval et al. 2001 otherwise.

Solutions to Problems 8.3.6 and 8.3.7 can be found respectively in Holton and Zamboni 2000 and in Karhumäki, Lepistö, and Plandowski 1998a. See also Lepistö 1999 for related results. The article Holton and Zamboni 2000 is the last, for the moment, of a long sequence of articles and results concerning periodicities and Sturmian words (Chapter 2). It is worth to notice that in Sturmian words, periodicities are strongly related to the recurrence function defined in Problem 8.3.9, as shown by the two beautiful and independent papers Cassaigne 1999 and Vandeth 2000.

Problem 8.3.8 was posed, as a conjecture for a fixed constant c, in Shallit and Breibart 1996 and settled in Cassaigne 1997. Problem 8.3.9 was stated as conjecture in Rauzy 1983. It is linked in some way to the previous problem, as proved in Allouche and Bousquet-Mélou 1995.