6

Similarity relations on words

Vesa Halava, Tero Harju and Tomi Kärki

6.1 Introduction

In this chapter we consider similarity relations on words. These relations are induced by compatibility relations on symbols, i.e., by reflexive and symmetric relations on alphabets. Originally, similarity relations on words were introduced in Halava *et al.* (2007a) in order to generalise the notion of a partial word as presented in (Berstel and Boasson, 1999). Let Σ be a (finite) alphabet. We study similarity of words mainly from two perspectives that are central in combinatorics on words: periodicity and repetition freeness as discussed in Chapters 4 and 5.

One of the often studied similarity relation is that of partial words. In these one has a special symbol $\diamond \in \Sigma$, and the compatibility relation is

$$R = \{(a,a),(\diamond,a),(a,\diamond) \mid a \in \Sigma\}.$$

In partial words the letter $\diamond$ is sometimes interpreted as a joker symbol or a symbol for 'do not know'. It is compatible with all letters of the alphabet (including itself). Partial words were studied in (Berstel and Boasson, 1999) with respect to periodicity. Combinatorial properties of partial words have since been extensively studied; for this we refer to, e.g., the monograph (Blanchet-Sadri, 2007) and the references therein.

A period of a finite or infinite word is a positive integer p such that all letters in the word occurring in positions congruent modulo p are equal. Also see Definition 1.2.6. A basic theorem on periodicity was proved by Fine and Wilf (1965). According to this result if a word w has two periods p and q, and the length of w is at least $p+q-\gcd(p,q)$, then w also has the period $\gcd(p,q)$. Berstel and Boasson (1999) considered a generalisation of this result for partial words that have a restricted number of holes. Our aim is to prove variations of Fine and Wilf's theorem to witness interaction properties between 'pure' periods and relational periods induced by a similarity relation on words. Indeed, we define three types of period

Combinatorics, Words and Symbolic Dynamics, ed. Valérie Berthé and Michel Rigo. Published by Cambridge University Press.

in Section 6.4: global, external and local relational periods. These variants are then analysed in different interaction cases.

It has already been mentioned in the previous chapters that the study of repetition freeness of words goes back a hundred years to two papers (Thue, 1906a, 1912). These two articles can be considered to be the starting point of combinatorics on words. In particular, Thue showed that there exists an infinite word w over a ternary alphabet that does not contain any squares, i.e., factors of the form xx where x is a non-empty word (see Theorem 4.2.2). Moreover, Thue constructed an infinite binary word t, the so-called Thue–Morse word, that does not contain any overlaps $xyxyx$ for any words x and y where x is not empty (see Theorem 4.2.8). Squarefreeness and related notions can be defined for relational words. This topic is studied in Section 6.5 where we give tight bounds for the repetition thresholds.

The basic concepts and theorems on words needed for subsequent sections are introduced in Section 6.2. We define there the main object of this chapter, i.e., similarity of words. In Section 6.3 we discuss shortly an application of similarity relations in coding theory. Sections 6.4 and 6.5 cover the main themes of the chapter, periodicity and repetition freeness.

6.2 Preliminaries

In this section we introduce the basic notions and notation for the rest of the chapter. We mostly concentrate on two classical topics of combinatorics on word, namely periodicity and repetition freeness. Short introductions to these topics are provided in this sections. For further knowledge on combinatorics of words, we refer to the book (Lothaire, 1983), the tutorial (Berstel and Karhumäki, 2003) and the survey (Choffrut and Karhumäki, 1997). Also, Chapter 8 in Lothaire's second book (Lothaire, 2002) is devoted to aspects of periodicity.

6.2.1 Periodicity

Let $w = w_1 w_2 \cdots w_n$, with $w_i \in \Sigma$ for all i, be a word over the alphabet Σ. An integer $p \geq 1$ is a *period* of w if $w_i = w_{i+p}$ for $1 \leq i \leq |w| - p$. In this case, the word w is called *p-periodic*. The smallest integer which is a period of w is called *the (minimal) period* of w and it is denoted by $\pi(w)$. For each w, w is a rational power of the prefix of length p of w. See (1.1).

We shall consider *interaction properties of periods*. These are properties of words where two different periods occurring in a finite word implies a third period. The most fundamental theorem of this type is the *theorem of Fine and Wilf*, which was originally proved in connection with real continuous functions (Fine and Wilf, 1965). In its discrete form on words it is as follows.

Theorem 6.2.1 (Fine and Wilf) *If a word w has periods p and q, and $|w| \geq p + q - \gcd(p,q)$, then w also has the period $\gcd(p,q)$.*

Example 6.2.2 The word $w = aabaabaaabaabaa$ has periods $7, 10, 13$ and 14, and, trivially, all integers $n \geq |w| = 15$. Hence, the minimal period is $\pi(w) = 7$. Indeed, in this example $w = (aabaaba)^2 a$. Note that $|w| = 15 = 7 + 10 - \gcd(7,10) - 1$. By Theorem 6.2.1, there are no binary words u of length $|u| \geq 15$ with periods 7 and 10 apart from the unary words, since $\gcd(7,10) = 1$.

An integer p is a period of an infinite word $w = w_0 w_1 \cdots$ if, for all non-negative integers i, it holds $w_i = w_{i+p}$, i.e., $w_i = w_j$ whenever $i \equiv j \pmod{p}$, and in this case

$$w = v^{\omega} = vvv \cdots,$$

where v is the prefix of w of length p.

6.2.2 Similarity relations of words

Relations of symbols in several matching problems are known in the literature. In *non-standard string matching* problem a given pattern p is matched with a text τ up to some relation; see (Muthukrishnan and Ramesh, 1995). In an instance of this problem, we are given a many-to-many matching relation between symbols together with the pattern p. Then we search for positions in a text τ at which the pattern occurs under the relation. Also, the string matching problem with symbols for *'don't care' symbols* was introduced by Fischer and Paterson (1974). There a special symbol is present that matches all other symbols. In *distance matching* problems, a distance function d is given between pairs of symbols. Symbols a and b match if $d(a,b) \leq k$ for some specified constant k. In the latter two cases the relation on the letters is a compatibility relation inducing a similarity relation on words. Thus, a pattern which matches the text is, using our terminology, similar to the text.

We consider binary relations $R \subseteq X \times X$ on a set X. We often write xRy instead of $(x,y) \in R$. The restriction of R on a subset $Y \subseteq X$, i.e., $R \cap (Y \times Y)$ is denoted by R_Y. A relation R on X is an *equivalence relation* if it is *reflexive*, *symmetric* and *transitive*, i.e., it satisfies the following conditions

(E1) $\forall x \in X: \quad xRx,$

(E2) $\forall x, y \in X: \quad xRy \implies yRx,$

(E3) $\forall x, y, z \in X: \quad xRy,\ yRz \implies xRz.$

A relation satisfying (E1) and (E2) is called a *compatibility relation*. If $R \subseteq X \times X$ is a compatibility relation and xRy for $x, y \in X$, we say that x is *R-compatible* with y, or simply, that x and y are *compatible* if R is clear from the context.

Example 6.2.3 As an example of a compatibility relation on integers $\mathbb{Z}$, consider the relation D_n for a positive integer n:

$$x\, D_n\, y \iff |x - y| \leq n.$$

This relation is clearly reflexive and symmetric but it is not transitive. Indeed, for instance, $0\, D_n\, 1$ and $-n\, D_n\, 0$ whereas 1 and $-n$ are not related.

The *identity relation* $\iota_X = \{(x,x) \mid x \in X\}$ on a set X is an equivalence relation. For an alphabet Σ, let $\Omega_\Sigma = \Sigma \times \Sigma$. The *universal (similarity) relation* on Σ^* is then $\Omega_{\Sigma^*} = \{(x,y) \in \Sigma^* \times \Sigma^* \mid |x| = |y|\}$. Also this relation is an equivalence relation. The subscripts X and Σ^* are often omitted when they are clear from the context.

By definition, $R \subseteq \Sigma^* \times \Sigma^*$ is a *similarity relation* if, and only if, it is a compatibility relation that satisfies

$$u_1u_2\cdots u_m R v_1v_2\cdots v_n \iff m = n \text{ and } u_i R v_i \text{ for all } i = 1,2,\ldots,m \tag{6.1}$$

for all $u_i, v_j \in \Sigma$. In particular, if uRv then $|u| = |v|$. It also follows that a similarity relation satisfies the *simplifiability property*

$$uu'Rvv' \text{ with } |u| = |v| \implies uRv \text{ and } u'Rv' \tag{6.2}$$

as well as the *multiplicativity property*

$$uRv \text{ and } u'Rv' \implies uu'Rvv'. \tag{6.3}$$

Remark 6.2.4 A similarity relation $R \subseteq \Sigma^* \times \Sigma^*$ does not need to be an equivalence relation, and thus it may fail to be a congruence relation on the monoid Σ^*. In fact, the most interesting similarity relations of this chapter are not transitive.

A similarity relation R on Σ is generated by its restriction R_Σ on letters, and therefore R can be represented by the list of the pairs $\{a,b\}$, $a \neq b$, such that $(a,b) \in R_\Sigma$. Henceforth we use the notation

$$R = \langle r_1, r_2, \ldots, r_n \rangle,$$

where $r_i = (a_i, b_i) \in \Sigma \times \Sigma$ for $i = 1,2,\ldots,n$, to denote that R is the similarity relation generated by the reflexive and symmetric closure of $\{r_1, r_2, \ldots, r_n\}$.

Put differently, R can be represented as a *graph* $G_R = (\Sigma, E)$, where Σ is the vertex set and $E = \{\{a,b\} \mid aRb \text{ with } a,b \in \Sigma,\ a \neq b\}$ is the set of the edges.

Example 6.2.5 Consider the similarity relation $R = \langle (a,c),(b,c),(a,d) \rangle$. Then the corresponding graph is given in Figure 6.1.

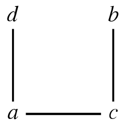

Figure 6.1 The graph G_R representing $R = \langle (a,c),(b,c),(a,d) \rangle$.

A similarity relation R is clearly a compatibility relation, since the generating relation is reflexive and symmetric. Hence, it is justified to use expressions *R-compatible* and *R-similar* side by side.

6.2.3 Partial words

Partial words were introduced in the present context in Berstel and Boasson (1999).

Definition 6.2.6 Let Σ be an alphabet. A partial function $w : \{1,2,\ldots,n\} \to \Sigma$ is called a *partial word* of length n. The domain $D(w)$ of w is the set of positions $p \in \{1,2,\ldots,n\}$ for which $w(p)$ is defined. The set $H(w) = \{1,2,\ldots,n\} \setminus D(w)$ is the set of *holes* of w.

With each partial word we may associate a *companion word* $w_\diamond$ over the extended alphabet $\Sigma_\diamond = \Sigma \cup \{\diamond\}$ as follows:

$$w_\diamond(p) = \begin{cases} w(p), & \text{if } p \in D(w); \\ \diamond, & \text{if } p \in H(w). \end{cases}$$

Clearly, partial words are in one-to-one correspondence with the words over $\Sigma_\diamond$, and we will also call the words over $\Sigma_\diamond$ partial words.

Let x and y be partial words of equal length. Then y is said to *contain* x if $D(x) \subseteq D(y)$ and $x(i) = y(i)$ for all $i \in D(x)$. Two partial words x and y are *compatible*, denoted by $x \uparrow y$, if there exists a partial word z containing both x and y.

Example 6.2.7 Let $\Sigma = \{a,b,c,d\}$, and choose the partial words $x = ab\diamond bc\diamond$ and $y = a\diamond\diamond bca$. We align these words letter-to-letter to obtain

$$\begin{array}{cccccc} a & b & \diamond & b & c & \diamond \\ a & \diamond & \diamond & b & c & a. \end{array}$$

This shows that $x \uparrow y$. Indeed, here $z = ab\diamond bca$ is a partial word containing both x and y.

The compatibility relation on partial words is a similarity relation. Indeed, partial words with compatibility relation $\uparrow$ can be seen as (total) words over the alphabet $\Sigma_\diamond$ with the similarity relation

$$R_\uparrow = \langle \{(\diamond, a) \mid a \in \Sigma\} \rangle. \tag{6.4}$$

6.2.4 Example: similarity relations in biology

Motivation for the research on partial words comes partly from the study of biological sequences such as DNA (deoxyribonucleic acid), RNA (ribonucleic acid) and proteins; see, e.g., (Blanchet-Sadri, 2004a). Partial words are a special case of words with a similarity relation, it is evident that bioinformatics gives a suitable background for applications of similarity relations as well. We briefly describe the molecular biological setting and some basic operations where relations can be employed in theoretical models. See also Section 5.6.

Proteins are macromolecules made of amino acids playing an essential role in living organisms by participating in all central processes within cells. The 20 amino

Table 6.1 *The 20 amino acids and their three-letter abbreviations.*

Name	**Abbreviation**	**Name**	**Abbreviation**
Alanine	Ala	Leucine	Leu
Arginine	Arg	Lysine	Lys
Asparagine	Asn	Methionine	Met
Aspartic acid	Asp	Phenylalanine	Phe
Cysteine	Cys	Proline	Pro
Glutamine	Gln	Serine	Ser
Glutamic acid	Glu	Threonine	Thr
Glycine	Gly	Tryptophan	Trp
Histidine	His	Tyrosine	Tyr
Isoleucine	Ile	Valine	Val

acids and their three-letter abbreviations are given in Table 6.1. The length of the genes depends upon the length of the amino acid chain it encodes.

In the cytoplasm, the messenger RNA (mRNA) binds to ribosomes for *translation*. Ribosomes read the coding of mRNA by moving from one codon to the next one, interpreting the present codon as one of the 20 amino acids until the corresponding protein is formed. Every mRNA sequence begins with trinucleotide AUG implying that all amino acid chains begin with methionine. The subsequent amino acids of the protein are coded according to Table 6.2. Finally, at the end of the mRNA, the ribosome translates UGA, UAA or UAG as a 'stop' symbol.

As an example of the use of similarity relations within the framework of protein synthesis, we consider codons as letters forming a 64-letter alphabet. Codons that encode the same amino acid are then similar. For instance, both GAA and GAG correspond to glutamic acid (Glu) whereas UUA, CUA and CUU are code for leucine (Leu). Thus, the following two codon strings stand for the same amino acid sequence (Glu-Leu-Leu) despite the mutations (underlined changes) in the genetic code and therefore the sequences are similar.

$$\underbrace{\mathrm{GAA}}_{\mathrm{Glu}}\underbrace{\mathrm{UUA}}_{\mathrm{Leu}}\underbrace{\mathrm{CUU}}_{\mathrm{Leu}} \quad \overset{mutation}{\rightsquigarrow} \quad \underbrace{\mathrm{GA}\underline{\mathrm{G}}}_{\mathrm{Glu}}\underbrace{\underline{\mathrm{C}}\mathrm{UA}}_{\mathrm{Leu}}\underbrace{\mathrm{CU}\underline{\mathrm{A}}}_{\mathrm{Leu}}$$

Table 6.2 *The genetic code table (5′ to 3′ direction).*

First position	Second position				Third position
	A	C	G	U	
A	Lys	Thr	Arg	Ile	A
	Asn	Thr	Ser	Ile	C
	Lys	Thr	Arg	**Met**	G
	Asn	Thr	Ser	Ile	U
C	Gln	Pro	Arg	Leu	A
	His	Pro	Arg	Leu	C
	Gln	Pro	Arg	Leu	G
	His	Pro	Arg	Leu	U
G	Glu	Ala	Gly	Val	A
	Asp	Ala	Gly	Val	C
	Glu	Ala	Gly	Val	G
	Asp	Ala	Gly	Val	U
U	**Stop**	Ser	**Stop**	Leu	A
	Tyr	Ser	Cys	Phe	C
	Stop	Ser	Trp	Leu	G
	Tyr	Ser	Cys	Phe	U

6.3 Coding

We shall now consider coding properties from the point of view of similarity relations. Our objective is to generalise the basic theorem (Sardinas and Patterson, 1953) on variable length codes of words for similarity relations.

The corresponding problem for partial words was proved to be decidable using a domino technique introduced in (Head and Weber, 1995); see (Blanchet-Sadri, 2004a). We give here a solution for the more general problem of deciding whether a given set X is an (R,S)-code or not.

Most of the material of this section is from (Halava *et al.*, 2007a) and (Kärki, 2008).

6.3.1 Some language theory

For languages $L, K \subseteq \Sigma^*$ of words and for a word $u \in \Sigma^*$, we define

$$\begin{aligned}
LK &= \{uv \mid u \in L, v \in K\},\\
L^+ &= \bigcup_{i \geq 1} L^i = \{u_1 u_2 \cdots u_n \mid n \geq 1, u_j \in L\},\\
L^* &= L^+ \cup \{\varepsilon\},\\
u^{-1}K &= \{v \mid uv \in K\},\\
Ku^{-1} &= \{v \mid vu \in K\},\\
L^{-1}K &= \bigcup_{u \in L} u^{-1}K = \{v \mid \exists u \in L\colon uv \in K\}.
\end{aligned}$$

Let 2^X denote the *power set* of X, i.e., the family of all subsets of X. For a relation $R \subseteq X \times X$, let the corresponding function $R : 2^X \to 2^X$ be defined by

$$R(Y) = \{x \in X \mid \exists y \in Y : yRx\}. \tag{6.5}$$

A sequence $x_1, x_2, \ldots, x_m$ of words from $X \subseteq \Sigma^*$ is an X-*factorisation* (or simply a *factorisation*) of a word $w \in \Sigma^*$ if $w = x_1 x_2 \cdots x_n$. Figure 6.2 illustrates a word having two different factorisations, $x_1, x_2, \ldots, x_m$ and $y_1, y_2, \ldots, y_n$.

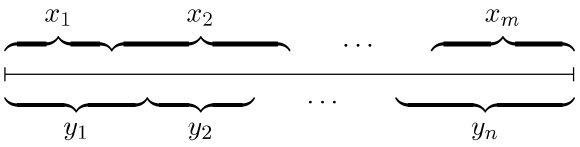

Figure 6.2 Two distinct factorisations of a word.

A subset $X \subseteq \Sigma^*$ is a *code* if every word $w \in X^*$ has a unique X-factorisation, i.e., for all $n, m \geq 1$ and $x_1, x_2, \ldots, x_m \in X$ and $y_1, y_2, \ldots, y_n \in X$,

$$x_1 x_2 \cdots x_m = y_1 y_2 \cdots y_n \implies n = m \text{ and } x_i = y_i \text{ for } i = 1, 2, \ldots, m. \tag{6.6}$$

More precisely, such a code is a *variable length* code as opposed to an *error correcting* code; see (Baylis, 1998). (For more on codes, see also Section 7.4.5.)

6.3.2 Sardinas–Patterson theorem

We begin with a multiplicative property of the function $R : 2^{\Sigma^*} \to 2^{\Sigma^*}$ of a similarity relation R on Σ.

Theorem 6.3.1 *Let R be a similarity relation on Σ^*. Then $R(X)R(Y) = R(XY)$ for all $X, Y \subseteq \Sigma^*$. Especially, $R(X)^* = R(X^*)$ for all $X \subseteq \Sigma^*$.*

Proof Let $w \in R(X)R(Y)$. Then $w = uv$ for some words $u \in R(X)$ and $v \in R(Y)$.

Now, there exist $x \in X$ and $y \in Y$ such that xRu and yRv. By multiplicativity of the relation R, see (6.3), we have $xyRuv$, and thus $w \in R(XY)$ as required.

Conversely, let $w \in R(XY)$, i.e., $xyRw$ for some $x \in X$ and $y \in Y$, where $|w| = |x| + |y|$. Thus, $w = uv$ such that $|u| = |x|$ and $|v| = |y|$. By simplifiability (6.2) of R, it holds xRu and yRv. Hence, $w = uv \in R(X)R(Y)$.

The second claim follows from $R(X)^n = R(X^n)$, for all $n \geq 1$, by induction on n. Also, $R(X)^0 = \varepsilon = R(\{\varepsilon\}) = R(X^0)$. □

Definition 6.3.2 Let R and S be similarity relations over the alphabet Σ. A subset $X \subseteq \Sigma^*$ is called an (R,S)-*code* if for all $n, m \geq 1$ and $x_1x_2,\ldots,x_m, y_1,y_2,\ldots,y_n \in X$, the following condition holds

$$x_1x_2\cdots x_m R y_1y_2\cdots y_n \implies n = m \text{ and } x_i S y_i \text{ for } i = 1,2,\ldots,m. \tag{6.7}$$

The relations R and S are called the *alteration* and *fidelity relations*, respectively.

The alteration relation R may be seen as describing errors or differences in a coded message, and the fidelity relation S can be thought of as describing how well messages can be decoded. If $S = \iota$, the identity relation, then an (R,S)-code is called an *R-code* In an R-code X different elements are necessarily pairwise non-similar.

An (R,R)-code is called a *weak R-code*. An (ι,ι)-code is simply a *code*. Indeed, the definition coincides with the usual definition of a variable length code.

Figure 6.3 illustrates the following results.

Theorem 6.3.3 *(i) Let R_1, R_2 and S be similarity relations on Σ^* with $R_1 \subseteq R_2$. Then each (R_2,S)-code is an (R_1,S)-code.*

(ii) Let R, S_1 and S_2 be similarity relations on Σ^* with $S_1 \subseteq S_2$. Then each (R,S_1)-code is an (R,S_2)-code.

Proof For (i), let X is an (R_2,S)-code, and let xR_1y for some $x = x_1x_2\cdots x_m$ and $y = y_1y_2\cdots y_n$, where $x_i, y_i \in X$ for each i. Now, also xR_2y because $R_1 \subseteq R_2$. Since X is an (R_2,S)-code, $m = n$ and x_iSy_i holds for all $i = 1,2,\ldots,n$. This proves (i).

For (ii), suppose that X is an (R,S_1)-code and let xRy for some $x = x_1x_2\cdots x_m$ and $y = y_1y_2\cdots y_n$ where $x_i, y_i \in X$ for each i. Then $m = n$ and $x_iS_1y_i$ for all $i = 1,2,\ldots,m$. Because $S_1 \subseteq S_2$, this implies that $x_iS_2y_i$ for all $i = 1,2,\ldots,n$ proving the second claim. □

When we consider unions and intersections of similarity relations the previous result implies the following corollary.

Corollary 6.3.4 *Let X be an (R_1,S_1)-code and let R_2 and S_2 be two similarity relations on Σ^*. Then X is an $(R_1 \cap R_2, S_1 \cup S_2)$-code.*

$$\begin{tikzcd}
\Omega & \Omega \\
R_2 \arrow[u, phantom, sloped, "\subseteq"] \arrow[r, dashed, leftrightarrow] \arrow[dr, leftrightarrow] & S_2 \arrow[u, phantom, sloped, "\subseteq"] \\
R_1 \arrow[u, phantom, sloped, "\subseteq"] \arrow[r, dashed, leftrightarrow] & S_1 \arrow[u, phantom, sloped, "\subseteq"] \\
\iota \arrow[u, phantom, sloped, "\subseteq"] \arrow[r, dashed, leftrightarrow] & \iota \arrow[u, phantom, sloped, "\subseteq"]
\end{tikzcd}$$

Figure 6.3 An (R_2,S_1)-code is also an (R_1,S_1)-code, an (R_2,S_2)-code and an (ι,ι)-code.

For sets that are both (R,S_1)-codes and (R,S_2)-codes, the coding property can also be preserved when the fidelity relation is restricted to the intersection of S_1 and S_2 relations.

Corollary 6.3.5 *Let X be both an (R,S_1)-code and an (R,S_2)-code. Then it is an $(R,S_1 \cap S_2)$-code.*

Proof Again, let xRy for some $x = x_1x_2\cdots x_m$ and $y = y_1y_2\cdots y_n$, where $x_i, y_i \in X$ for all i. Then $m = n$ and $x_iS_jy_i$ for $j = 1,2$ and for all $i = 1,2,\ldots,n$. Thus, $x_i\,(S_1 \cap S_2)\,y_i$ for all $i = 1,2,\ldots,n$, and, consequently, X is an $(R,S_1 \cap S_2)$-code. ☐

The next theorem characterises general (R,S)-codes in terms of weak R-codes.

Theorem 6.3.6 *A subset $X \subseteq \Sigma^*$ is an (R,S)-code if, and only if, X is an (R,R)-code and $R_X \subseteq S_X$ for the restrictions of R and S on X.*

Proof Suppose first that X is an (R,S)-code. By Theorem 6.3.3(ii), X is an (R,Ω)-code, where Ω is the universal relation. Therefore if $x_1x_2\cdots x_mRy_1y_2\cdots y_n$ with $x_i, y_j \in X$, then $m = n$ and $|x_i| = |y_i|$ for all i, and by the simplifiability property of the similarity relations, x_iRy_i for all i. Hence, X is an (R,R)-code. By choosing $n = 1 = m$ in the definition of an (R,S)-code we see that $R_X \subseteq S_X$.

Conversely, let X be an (R,R)-code such that $R_X \subseteq S_X$, and consider the words $x = x_1x_2\cdots x_m$ and $y = y_1y_2\cdots y_n \in X$ with xRy, and x_iRy_i for all i and j. Hence, $m = n$ and also x_iSy_i for all $i = 1,2,\ldots,n$ by the assumption $R_X \subseteq S_X$. ☐

As a corollary to Theorem 6.3.6, we show that the (R,S)-codes are always codes in the usual sense.

Corollary 6.3.7 *Every (R,S)-code is a code.*

Proof An (R,S)-code is an (ι,S)-code by Theorem 6.3.3(i). Thus, it is an (ι,ι)-code by Theorem 6.3.6. ☐

We now introduce a modification of the Sardinas–Patterson theorem.

Theorem 6.3.8 (Modified Sardinas–Patterson) *Let R be a similarity relation on Σ^* and let $X \subseteq \Sigma^+$. Set $U_1 = R(X)^{-1}X \setminus \{\varepsilon\}$, and define*

$$U_{n+1} = R(X)^{-1}U_n \cup R(U_n)^{-1}X$$

for $n \geq 1$. The set X is a weak R-code if, and only if, $\varepsilon \notin U_n$ for all n.

We need the following lemma.

Lemma 6.3.9 *Let $X \subseteq \Sigma^+$. For all n and k with $1 \leq k \leq n$, we have $\varepsilon \in U_n$ if, and only if, there exist $u \in U_k$ and $i, j \geq 0$ such that*

$$uX^i \cap R(X^j) \neq \emptyset \quad \textit{and} \quad i + j + k = n. \tag{6.8}$$

Proof We prove the statement by descending induction on k. Assume first that $k = n$. If $\varepsilon \in U_n$, then the condition (6.8) is satisfied with $u = \varepsilon$ and $i = j = 0$. Conversely, if (6.8) holds, then $i = j = 0$ and $\{u\} \cap \{\varepsilon\} \neq \emptyset$. Thus, $u = \varepsilon$ and consequently $\varepsilon \in U_n$.

Now, let $n > k \geq 1$ and suppose that the claim holds for $n, n-1, \ldots, k+1$. If $\varepsilon \in U_n$, then by the induction hypothesis, there exist a word $u \in U_{k+1}$ and integers $i, j \geq 0$ such that $uX^i \cap R(X^j) \neq \emptyset$ and $i + j + (k+1) = n$. Thus, there exist words $x_1, x_2, \ldots x_i, y_1, y_2, \ldots, y_j \in X$ such that $y_1 y_2 \cdots y_j \, R \, u x_1 x_2 \cdots x_i$. Since $u \in U_{k+1}$, there are two cases (1) or (2).

(1) There exists $y \in R(X)$ with $yu \in U_k$. In this case, $y' R y$ for some $y' \in X$ and, by the multiplicativity of R, $y' y_1 y_2 \cdots y_j \, R \, y u x_1 x_2 \cdots x_i$. Consequently, there exists a word $yu \in U_k$ such that $yuX^i \cap R(X^{j+1}) \neq \emptyset$ and $i + (j+1) + k = n$.

(2) There exists $v \in R(U_k)$ with $vu \in X$. Now, there exists $v' \in U_k$ such that $v' R v$ and by the multiplicativity and symmetry of R, $v u x_1 x_2 \cdots x_i \, R \, v' y_1 y_2 \cdots y_j$. Hence, there is a word $v' \in U_k$ such that $v' X^j \cap R(X^{i+1}) \neq \emptyset$ and $j + (i+1) + k = n$ as required.

Conversely, assume that there exist a word $u \in U_k$ and integers $i, j \geq 0$ such that $uX^i \cap R(X^j) \neq \emptyset$ and $i + j + k = n$. Then $y_1 y_2 \cdots y_j \, R \, u x_1 x_2 \cdots x_i$ for some $x_1, x_2, \ldots, x_i, y_1, y_2, \ldots, y_j \in X$. If $j = 0$, then $i = 0$, $k = n$ and we are in the case considered in the beginning of the proof.

Let us assume that $j > 0$. There are two cases to consider.

Case 1. Assume that $|u| \geq |y_1|$. By the simplifiability property of R, we may write $u = y_1' v$, where $y_1 R y_1'$ and $v \in \Sigma^*$. Then $v \in U_{k+1}$ and $y_2 \cdots y_j \, R \, v x_1 x_2 \cdots x_i$. Thus, $vX^i \cap R(X^{j-1}) \neq \emptyset$ and $i + (j-1) + (k+1) = n$. By the induction hypothesis, $\varepsilon \in U_n$.

Case 2. Assume that $|u| < |y_1|$. We write $y_1 = u'v$, where $u' R u$ and $v \in \Sigma^+$. Then, by the symmetry of R, $v \in U_{k+1}$ and $x_1 x_2 \cdots x_i \, R \, v y_2 \cdots y_j$. Thus, $vX^{j-1} \cap R(X^i) \neq \emptyset$ and $(j-1) + i + (k+1) = n$. Again $\varepsilon \in U_n$ by the induction hypothesis. □

We shall now to prove the modified Sardinas–Patterson theorem.

Proof of Theorem 6.3.8. If X is not a weak R-code, then there are positive integers m and n and words $x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_n \in X$ such that

$$x_1 x_2 \cdots x_m \, R \, y_1 y_2 \cdots y_n \quad \text{and} \quad (x_1, y_1) \notin R.$$

By the simplifiability property of R, $|x_1| \neq |y_1|$. By symmetry, we may assume that $|x_1| < |y_1|$. Hence $y_1 = x_1'u$ for some $u \in \Sigma^+$ with $x_1 R x_1'$. Also $x_2 \cdots x_m R u y_2 \cdots y_n$. Thus, $u \in U_1$ and $uX^{n-1} \cap R(X^{m-1}) \neq \emptyset$. By Lemma 6.3.9, $\varepsilon \in U_{m+n-1}$.

Conversely, if $\varepsilon \in U_n$, then choose $k = 1$ in Lemma 6.3.9. Therefore, there exist a word $u \in U_1$ and integers $i, j \geq 0$ such that $i + j = n - 1$ and $uX^i \cap R(X^j) \neq \emptyset$. Hence there exist words $x_1, x_2, \ldots, x_i, y_1, y_2, \ldots, y_j \in X$ such that $y_1 y_2 \cdots y_j R u x_1 x_2 \cdots x_i$. Since $u \in U_1$, necessarily $x = yu$ for some $y \in R(X)$ and $x \in X$. Furthermore, $|x| \neq |y|$, since $u \neq \varepsilon$ by the definition of U_1. Since $y \in R(X)$, there exists $y' \in X$ such that $y' R y$. By the multiplicativity property, $y' y_1 \cdots y_j R y u x_1 \cdots x_i$, which therefore gives $y' y_1 \cdots y_j R x x_1 \cdots x_i$ on X^+. Since $|y| = |y'| \neq |x|$, also $(y', x) \notin R$. This means that X is not a weak R-code. □

Corollary 6.3.10 *Given a similarity relation R and a finite set X it is decidable whether or not X is a weak R-code.*

Proof Note that if X is finite, there exist only finitely many different sets U_n. Indeed, the elements of U_n are suffixes of words in X, and hence the lengths of the elements are less than $\max\{|x| \mid x \in X\}$. Consequently the sequence of the sets U_i is ultimately periodic, i.e., $U_j = U_{j+1}$ for some index j, and therefore it can be effectively determined whether a finite set of words is a relational code or not. □

6.4 Relational periods

We consider in this section variations of the theorem of Fine and Wilf; see Theorem 6.2.1. Halava *et al.* (2008a) considered three types of relational period: global, local and external. Of these, global and local cases generalise the corresponding concepts of partial words. We shall concentrate mostly on these two types of relational periods. In Section 6.4.2 we prove interaction theorems for relational periods. Different interaction types are treated in separate subsections. Typically, we consider words with one 'pure' period and one relational period. The question is whether these two periods induce a third period for long enough words. A summary of the exact length bounds in different interaction cases is given at the end of the section. So called extremal relational Fine and Wilf words are discussed in Section 6.4.2. These are words of maximal length which do not express period interaction behaviour.

6.4.1 Types of relational periods

Recall that an integer $p \geq 1$ is a period of a word $x = x_1 x_2 \cdots x_n \in \Sigma^*$ (where each $x_i \in \Sigma$) if, for all positions i and j with $i \equiv j \pmod{p}$, we have $x_i = x_j$. The minimal period of x is denoted by $\pi(x)$. In the sequel these periods are called *pure* as distinction from *relational periods* defined as follows.

Definition 6.4.1 Let R be a similarity relation on an alphabet Σ. For a word $x = x_1 x_2 \cdots x_n$, where $x_i \in \Sigma$, an integer $p \geq 1$ is

(i) a *global R-period* of x if, for all $1 \le i, j \le n$,

$$i \equiv j \pmod{p} \implies x_i R x_j;$$

(ii) an *external R-period* of x if there exists a word $y = y_1 y_2 \cdots y_p$ such that, for all $1 \le i \le n$ and $1 \le j \le p$,

$$i \equiv j \pmod{p} \implies x_i R y_j$$

(in this case, the word y is called an *external word* of x);

(iii) a *local R-period* of x if, for all $1 \le i \le n - p$, it holds $x_i R x_{i+p}$.

These definitions generalise naturally to infinite words. For a word x, *the minimal global*, *local* and *external R-periods* are denoted by $\pi_{R,g}(x)$, $\pi_{R,l}(x)$, and $\pi_{R,e}(x)$, respectively. In the sequel, we may omit the subscript R or the argument x if these are clear from the context. Similarly, if R is understood from the context, then we talk about global and local periods.

Example 6.4.2 Let $\Sigma = \{a,b,c,d\}$ and choose $R = \langle (a,b),(b,c),(c,d),(d,a) \rangle$. Hence, the graph of R forms a cycle. Consider the word $x = babbbcbd \in \Sigma^*$. The minimal pure period of x is $\pi(x) = |x| = 8$. In this case $\pi_{R,l}(x) = 2$ for the local R-period. (Note that $(x_7, x_8) = (b,d) \notin R$.) The external period of x is $\pi_{R,e}(x) = 3$. Indeed, for the external word we can choose $y = bab$.

Neither 1 nor 2 can do for the external R-period, because in both cases there would be a letter related to every letter of Σ. Finally, the global period of x equals $\pi_{R,g}(x) = 6$. Indeed, since $(b,d) \notin R$, we have $\pi_{R,g}(x) > 5$. But $\pi_{R,g}(x) = 6$ will do because $ba R bd$. Hence, for the word x, we obtain

$$\pi = 8 > \pi_g = 6 > \pi_e = 3 > \pi_l = 2.$$

As another example of relational periods we consider periods of partial words.

Example 6.4.3 In (Berstel and Boasson, 1999) two types of period were defined for partial words. A partial word w has a *(partial) period* p if, for all $i, j \in D(w)$,

$$i \equiv j \pmod{p} \implies w(i) = w(j).$$

A partial word w has a *local (partial) period* p if

$$i, i+p \in D(w) \implies w(i) = w(i+p).$$

Applying the similarity relation $R_\uparrow = \langle \{(\diamond, a) \mid a \in \Sigma\} \rangle$, we see that each period of a partial word w corresponds to the global $R_\uparrow$-period of the companion $w_\diamond$. Similarly, the local partial period of w corresponds to the local $R_\uparrow$-period.

On the other hand, the external period is not meaningful for partial words. Namely, we always have $\pi_e(w) = 1$ after choosing $y = \diamond$ as the external word, since $\diamond^{|w|}$ is compatible with every partial word w.

In the next theorem we show how different types of period are related to each other.

Theorem 6.4.4 *Let $x \in \Sigma^*$ be a word.*

(i) Every pure period of x is a relational (global, external and local) R-period for any similarity relation R on Σ.

(ii) Every global R-period of x is an external R-period and a local R-period of x. Hence $\pi(x) \geq \pi_g(x) \geq \max(\pi_e(x), \pi_l(x))$.

Proof Let R be a similarity relation. Hence $\iota \subseteq R$, and (i) follows from this.

If $x = x_1 x_2 \cdots x_n$ has a global R-period p, then we can choose $x_1 x_2 \cdots x_p$ as an external word of x. Hence a global R-period is an external R-period. Clearly, a global period satisfies the definition of a local period. For the minimal periods, these considerations imply the inequalities of the statement. □

Note that an external period need not be a local period and a local period need not be an external period. For instance, in Example 6.4.2 the minimal local R-period $\pi_l(x)$ is not an external R-period, and $\pi_e(x)$ is not a local R-period. There it was the case $\pi_e(x) > \pi_l(x)$. Next we give an example where $\pi_l(x) > \pi_e(x)$.

Example 6.4.5 Let $\Sigma = \{a,b,c,d\}$. Let $R = \langle (a,b),(b,c),(c,d),(d,a) \rangle$ and choose $x = adcbccccbd$. We determine first the minimal local R-period of x. Since $(x_9, x_{10}) = (x_4, x_2) = (b,d) \notin R$ and 3 is a local R-period, $\pi_l(x) = 3$. Since $x_1 = a, x_4 = b, x_7 = c$ and $x_{10} = d$, there does not exist any external word $y = y_1 y_2 y_3$ of length 3. Otherwise, y_1 would be compatible with all letters of Σ. Hence, 3 is not an external R-period. For the same reason, 1 is not an external R-period, but by choosing $y = bc$, we see that $\pi_e = 2$. As noted above, 2 is not a local period. Since $(a,c) \notin R$, the minimal global R-period satisfies $\pi_g > 7$. Actually, $\pi_g = 8$ since aRb. Clearly, $\pi = 10$, and therefore

$$\pi = 10 > \pi_g = 8 > \pi_l = 3 > \pi_e = 2.$$

We state as an exercise the result that if the similarity relation R is an equivalence relation then the definitions of the relational periods coincide. In particular, we have the following result.

Theorem 6.4.6 *If R is a transitive similarity relation on Σ then*

$$\pi_g(x) = \pi_e(x) = \pi_l(x).$$

Proof See Exercise 6.6.4. □

If R is not transitive, local R-periods differ from global and relational periods by the following property.

Lemma 6.4.7 *Let $x \in \Sigma^*$ be a word and R a similarity relation on Σ. If p is a global R-period (an external R-period, respectively) of x, then any multiple of p is a global R-period (an external R-period, respectively) of x. A multiple of a local R-period of x need not be a local R-period.*

Proof If p is a global R-period of x and $i \equiv j \pmod{kp}$ for a non-negative integer

k then $i \equiv j \pmod{p}$ and, by the assumption, $x_i R x_j$. Hence kp is a global R-period. The proof is similar for the external R-periods.

Consider then $x = abc \in \{a,b,c\}^*$, and choose $R = \langle (a,b),(b,c) \rangle$. The word x has 1 as a local R-period, but 2 is not a local R-period. This proves the last claim. □

6.4.2 Variants of the theorem of Fine and Wilf

The theorem of Fine and Wilf, Theorem 6.2.1, is one of the cornerstones in combinatorics on words. In this theorem the derived period is the greatest common divisor of the original periods. This phenomenon was also the starting point of the study of partial words in the seminal paper (Berstel and Boasson, 1999). They proved the following variant for partial words with one hole.

Theorem 6.4.8 *Let $\Sigma_\diamond$ be an alphabet with a hole symbol.*

(i) Let $w \in \Sigma_\diamond^*$ be a partial word of length n with at most one hole. If w has local periods p and q such that $n \geq p+q$ then w is purely $\gcd(p,q)$*-periodic.*

(ii) The bound $p+q$ on the length of the word is sharp, i.e., there are partial words with one hole with local periods p and q and $n = p+q-1$ such that w is not purely $\gcd(p,q)$*-periodic.*

Example 6.4.9 For the sharpness in (ii), consider the word $w_{k+4} = aba\diamond a^k$ of length $k+4$ with $k \geq 1$. Then w_{k+4} has local periods 2 and $|w_{k+4}| - 1 = k+3$. However, if k is even, $\gcd(k+3,2) = 1$ is not a period of w_{k+4}.

Generalisations of Theorem 6.4.8 for several holes has been studied in (Blanchet-Sadri, 2004b); see also (Blanchet-Sadri and Hegstrom, 2002), where it was shown that local partial periods p and q force a sufficiently long word to have a (global) partial period $\gcd(p,q)$ when certain unavoidable special cases are excluded. The bound on the length depends on the number of holes in the word. On the other hand, in (Shur and Gamzova, 2004) there are bounds for the length of a word with k holes such that (global) partial periods p and q imply a (global) partial period $\gcd(p,q)$. These results indicate that finding simple formulations for the interaction of periods in general relational periods is not feasible except for equivalence relations.

The following example shows that without any additional assumption we cannot find a general bound for the interaction of relational periods.

Example 6.4.10 Let $R = \langle (a,b),(b,c) \rangle$ and let $i_1, i_2, \ldots$ be a sequence of integers. Define an infinite word w as follows:

$$w = w_1 w_2 w_3 \cdots = acb^{6i_1-2}acb^{6i_2-2}\cdots \qquad \text{with } w_i \in \{a,b,c\}.$$

The word w has global R-periods 2 and 3, since

$$w_1, w_3, w_5, \ldots \in \{a,b\} \text{ and } w_2, w_4, w_6, \ldots \in \{b,c\},$$

and

$$w_1, w_4, w_7, \ldots \in \{a,b\},\ w_2, w_5, w_8, \ldots \in \{b,c\} \text{ and } w_3, w_6, w_9, \ldots \in \{b\}.$$

However, $\gcd(2,3) = 1$ is not a global R-period of w.

Example 6.4.11 For the ultimately periodic word $w = acb^{\omega} = acbb\cdots$, all integers $p \geq 2$ are local and global R-periods with respect to $R = \langle (a,b)(b,c) \rangle$, but 1 is not.

Nonetheless, some interaction results can be obtained. If the relation R is an equivalence relation, the situation reduces to Theorem 6.2.1. Recall that the global, local and external periods coincide for the equivalence relations; see Exercise 6.6.5.

Theorem 6.4.12 *Let R be an equivalence relation. If a word x has R-periods p and q and $|x| \geq p + q - \gcd(p,q)$, then $\gcd(p,q)$ is an R-period of x.*

The following example shows that there are infinite words having a pure period q and a local R-period p but that do not have a local R-period $\gcd(p,q)$.

Example 6.4.13 Let $R = \langle (a,b),(b,c) \rangle$. Every non-transitive similarity relation must have a subrelation isomorphic to R such that a and c are not compatible. Consider an infinite periodic word $x = (bbcab)^{\omega}$. Clearly, w has a pure period $q = 5$. It also has a local R-period $p = 3$, since the distance of the letters a and c in x is different from 3. Since $(x_3, x_4) = (a,c) \notin R$, $\gcd(p,q) = 1$ is neither a local R-period nor a global R-period.

Bounds on interaction

We concentrate on global and local rational periods. For the external period, we refer to (Halava *et al.*, 2008a).

In Example 6.4.13 the local relational period p is too weak to imply the desired interaction result. However, in the sequel we obtain several results that depend on the type of the relational period p.

In the following we adopt the short-hand notation

$$t \in \{g, l\}$$

for the *types of relational periods*: global and local, respectively.

Definition 6.4.14 Let $P \geq 2$ and $Q \geq 3$ be positive integers, and let t_1 and t_2 be two types of relational periods. A positive integer

$$B = B_{t_1,t_2}(P,Q)$$

is called the *bound of t_1-t_2 interaction for P and Q*, if it satisfies (i) and (ii).

(i) The bound B is *sufficient*, i.e., if R is a similarity relation and w is a word with $|w| \geq B$ that has a pure period Q and a t_1-type R-period P, then w has a t_2-type R-period $\gcd(P,Q)$.

(ii) The bound is *strict*, i.e., there exist a similarity relation R and a word w with

$|w| = B - 1$ having a pure period Q and a t_1-type R-period P such that $\gcd(P, Q)$ is not a t_2-type R-period of w.

In the definition we excluded the special cases where $P = 1$ or $Q \leq 2$. Indeed, if $Q \leq 2$ is a pure period, then the word contains at most two letters, and this case is covered by Theorem 6.4.12. By the following lemma, it suffices to consider cases where $\gcd(P, Q) = 1$.

Lemma 6.4.15 *Let P and Q be positive integers with* $\gcd(P, Q) = d$*, and let* $B = B_{t_1,t_2}(P/d, Q/d)$*. Then* $Bd = B_{t_1,t_2}(P, Q)$*.*

Proof Let $q = Q/d$ and $p = P/d$. Hence $\gcd(p, q) = 1$. Suppose that $w = w_1 w_2 \cdots w_n$ has a pure period Q and a relational t_1-type period P. Assume that $n \geq dB$. Consider the word

$$w^{(i)} = w_i w_{i+d} \cdots w_{i+k_i d} \qquad \text{where } 1 \leq i \leq d \text{ and } k_i = \lfloor (n-i)/d \rfloor .$$

Then $w^{(i)}$ has a pure period q and a t_1-type relational period p. Since $|w^{(i)}| \geq B$ and $1 = \gcd(p, q)$, $w^{(i)}$ has a t_2-type period 1 by the definition of $B = B_{t_1,t_2}(p, q)$. Since this is true for all $i = 1, 2, \ldots, d$, we conclude that d is a t_2-type relational period of w.

In order to prove that the bound Bd is strict, we give an example of a word u of length $Bd - 1$ such that it has a period Q and an R-period P but no R-period d. By the definition of B, there exists a word $v = v_1 v_2 \cdots v_{B-1}$ that has a pure period q and a t_1-type period p, but $\gcd(p, q) = 1$ is not a t_2-type relational period of v. Let a be a letter and define a word u as follows:

$$u = a^{d-1} v_1 a^{d-1} v_2 \cdots a^{d-1} v_{B-1} a^{d-1} .$$

Now u has a pure period $Q = qd$ and a t_1-type period $P = pd$, but by the property of v, $\gcd(P, Q) = d$ is not a t_2-type R-period of u. □

Global–global interaction

We consider the case where one pure period and one global relational period imply a derived global relational period. The bounds of global–global interaction $B_{g,g}(p, q)$ for coprime integers p and q are given in Table 6.3.

Theorem 6.4.16 *Let p and q be positive integers with* $\gcd(p, q) = 1$*. The bound of the global–global interaction for p and q is* $B_{g,g}(p, q)$ *given in Table 6.3.*

We divide the proof into two parts. In the proof $[n]_q$ denotes the *least positive residue of n* $\pmod{q}$.

Lemma 6.4.17 *The bound* $B_{g,g}(p, q)$ *defined in Theorem 6.4.16 is sufficient.*

Proof Denote $B = B_{g,g}(p, g)$. Let R be a similarity relation, and w be a word with a pure period q and a global R-period p such that $|w| \geq B$. We show that $\gcd(p, q) = 1$ is a global R-period of w. Since w has pure period q, it has at most q different letters.

Table 6.3 *Table of the bounds* $B_{g,g}(p,q)$*, where* $\gcd(p,q) = 1$.

$B_{g,g}(p,q)$	$p < q$	$p > q$
p, q odd	$(p+1)q/2$	$q+(q-1)p/2$
p odd, q even	$(p+1)q/2$	$(p+1)q/2$
p even, q odd	$q+(q-1)p/2$	$q+(q-1)p/2$

Hence, it suffices to show that a letter w_n in an arbitrary position $1 \leq n \leq q$ is R-compatible with the other letters of w. For a position $1 \leq n \leq q$, let

$$\tau(n) = \max\{m \mid 1 \leq m \leq |w|,\ m \equiv n \pmod{q}\}.$$

Note that if w has exactly q different letters, then $\tau(n)$ is the last occurrence of the letter w_n in w.

Since w has the global relational period p, it follows that w_n is related to all letters in the positions

$$S(n) = \{n + ip \mid i = 0, 1, \ldots, \lfloor(|w| - n)/p\rfloor\}$$

and

$$T(n) = \{\tau(n) - ip \mid i = 1, 2, \ldots, \lfloor(\tau(n) - 1)/p\rfloor\}.$$

We prove that $S(n) \cup T(n)$ contains a complete residue system modulo q. It follows then that each letter w_n is R-compatible with all letters w_i for $i = 1, 2, \ldots, q$. Then, by the q-periodicity, 1 is an external R-period of w.

Assume first that $\tau(n) \equiv n \pmod{p}$, and hence also $\tau(n) \equiv n \pmod{pq}$, since $\tau(n) \equiv n \pmod{q}$ and $\gcd(p,q) = 1$. Now, $\tau(n) \geq n + pq$ and hence also $|w| \geq n + pq$. Therefore $S(n)$ contains the complete residue system $\{n + ip \mid i = 0, 1, \ldots, q-1\}$ modulo q.

Suppose then that $\tau(n) \not\equiv n \pmod{p}$. In this case, $S(n) \cap T(n) = \emptyset$.

Let

$$M(n) = \tau(n) - (q - \lfloor(B-n)/p\rfloor - 1)p.$$

Claim 1. One has $M(n) > 0$.

Since $B - n < pq$, we have $q - \lfloor(B-n)/p\rfloor - 1 \geq q - (q-1) - 1 = 0$. For the proof of Claim 1, we observe first that, since $B \geq q$,

$$\tau(n) \geq \begin{cases} B - [B]_q + n, & \text{if } 1 \leq n \leq [B]_q; \\ B - [B]_q - q + n, & \text{if } [B]_q + 1 \leq n \leq q. \end{cases}$$

We divide our considerations of Claim 1 into two cases according to the form of B.

Case 1. Let $B=(p+1)q/2$. In this case p is odd, and hence $[B]_q=q$, and

$$\begin{aligned} M(n) &> B-[B]_q+n-(q-1-((B-n)/p-1))\,p \\ &= B-q+n-qp+B-n=2B-(p+1)q \\ &= (p+1)q-(p+1)q=0. \end{aligned}$$

Case 2. Let $B=q+(q-1)p/2$. We note that

$$\lfloor (B-n)/p\rfloor=(q-1)/2+\lfloor (q-n)/p\rfloor \geq (q-1)/2,$$

since q is odd and $q\geq n$. If $1\leq n\leq [B]_q\}$, then

$$\begin{aligned} M(n) &\geq B-[B]_q+n-(q-(q-1)/2-1)p \\ &= q+(q-1)p/2-[B]_q+n-qp+(q-1)p/2+p \\ &= q-[B]_q+n\geq n>0. \end{aligned}$$

On the other hand, if $[B]_q+1\leq n\leq q$, then

$$\begin{aligned} M(n) &\geq B-[B]_q-q+n-(q-(q-1)/2-1)p \\ &= q+(q-1)p/2-[B]_q-q+n-qp+(q-1)p/2+p \\ &= n-[B]_q>0. \end{aligned}$$

Claim 2. One has $|S(n)\cup T(n)|\geq q$.

By Claim 1 we have that $\tau(n)-1\geq (q-\lfloor (B-n)/p\rfloor-1)p$, since $M(n)$ is an integer. Therefore also $\lfloor (\tau(n)-1)/p\rfloor \geq q-\lfloor (B-n)/p\rfloor-1$. Now, since $|w|\geq B$,

$$|S(n)\cup T(n)|=1+\lfloor (|w|-n)/p\rfloor+\lfloor (\tau(n)-1)/p\rfloor\geq q. \tag{6.9}$$

This shows Claim 2.

To conclude the proof of the lemma, we observe that, by Claim 2, there exist an integer $k\in\{0,q-2\}$ and sets

$$S'(n)=\{n+ip\mid i=0,1,\dots,k\}$$

and

$$T'(n)=\{\tau(n)-jp\mid j=1,2,\dots,q-k-1\}$$

such that $S'(n)\cup T'(n)$ is a complete residue system modulo q. Indeed, note that the elements of $S'(n)$ are pairwise incongruent modulo q, since $\gcd(p,q)=1$ and $k<q$. The same holds for $T'(n)$. Assume next that for some $0\leq i\leq k$ and $1\leq j\leq q-k-1$,

$$n+ip\equiv \tau(n)-jp \pmod{q}. \tag{6.10}$$

This holds if, and only if, $(i+j)p\equiv \tau(n)-n\equiv 0 \pmod{q}$, where the second congruence follows from the definition of $\tau(n)$. Since $\gcd(p,q)=1$, we conclude that (6.10) holds if, and only if, $i+j\equiv 0 \pmod{q}$. However, this is a contradiction, since $0 < i + j \le k + (q - k - 1) = q - 1 < q$. Therefore, the set $S'(n) \cup T'(n)$ contains exactly $(k+1) + (q-k-1) = q$ pairwise incongruent elements, which proves the lemma. □

We prove then that the bound is strict.

Lemma 6.4.18 *The bound $B_{g,g}(p,q)$ defined in Theorem 6.4.16 is strict.*

Proof We follow the notation of the previous proof. In particular, $B = B_{g,g}(p,q)$. Fix $n = [B]_q$, and define the *critical positions* $m = m(p,q) \in \{1,2,\ldots,q\}$ according to Table 6.4. We show that there exists a word v of length $B-1$ with a pure period q and a global R-period p such that the letter in the critical position is not related to the letter in position n. In the sequel we denote critical positions succinctly by m.

Table 6.4 *Table of critical positions $m = m(p,q)$.*

$m(p,q)$	$p < q$	$p > q$
p, q odd	$(q-p)/2$	q
p odd, q even	$q/2$	$q/2$
p even, q odd	q	q

Consider solutions (i,j) in non-negative integers of the equation

$$m + iq \equiv n + jq \pmod{p}. \tag{6.11}$$

By a *minimal solution* we mean a solution where $\max(n+iq, m+jq)$ is as small as possible. Note that if $i > j$ in a solution, then $m + (i-j)q \equiv n \pmod{p}$ is a smaller solution. Similarly, if $j > i$, then $m \equiv n + (j-i)q \pmod{p}$ is a smaller solution. Thus, in a minimal solution either $i = 0$ or $j = 0$. Also, such a solution is unique. Namely, let (i,j) and (i',j') be distinct minimal solutions, say with $j = 0$, $i' = 0$ and $m + iq = n + j'q$. Then $m \equiv n \pmod{q}$; a contradiction, since $1 \le m \le q$ and $n \ne m$ as can be verified from Tables 6.3 and 6.4.

Since $\gcd(p,q) = 1$, the sets

$$\{m + iq \mid i = 0,1,\ldots,p-1\} \text{ and } \{n + jq \mid j = 0,1,\ldots,p-1\}$$

are complete residue systems modulo p. Hence there exist exactly one integer j with $0 \le j \le p-1$ that satisfies $m \equiv n + jq \pmod{p}$, and exactly one integer i with $0 \le i \le p-1$ such that $m + iq \equiv n \pmod{p}$. Furthermore, for $1 \le j \le p-1$, we have

$$m \equiv n + jq \pmod{p} \implies m + (p-j)q = m + pq - jq \equiv n \pmod{p},$$

and $1 \leq p-j \leq p-1$. Hence, the minimal solution of (6.11) is either of the form $(0,j)$ or $(p-j,0)$.

We show that regardless of the parity of p and q and which of them is greater, the minimal solution is $i=0$ and $j=(B-n)/q$.

Since $B < pq$, it is $1 \leq (B-n)/q \leq p-1$ in all cases. Consider first those cases of Table 6.3 where $B=(p+1)q/2$ and, consequently, $n=[B]_q=q$. Let $j=(B-n)/q$.

Case 1. Let p and q be both odd and $p<q$. In Table 6.4, we have $m=(q-p)/2$. Now $n+jq=B$ and, since q is odd, it follows that

$$(n+jq)-m=(p+1)q/2-(q-p)/2=(q+1)p/2 \equiv 0 \pmod{p}.$$

Hence, $(0,(B-n)/q)$ is a solution. Also, $jq=B-n=(p+1)q/2-q=(p-1)q/2$ and

$$m+(p-j)q=m+pq-(p-1)q/2=m+(p+1)q/2=m+B>B.$$

Hence, in the solution $(p-(B-n)/q,0)$ it is the case $\max(m+iq,n+jq)>B$, whereas in the solution $(0,(B-n)/q)$ we have $\max(m+iq,n+jq)=B$. Thus, $(0,(B-n)/q)$ is the minimal solution.

Case 2. Suppose that p is odd and q is even. By the parity of q, $m=q/2$ is an integer and $(n+jq)-m=(p+1)q/2-q/2=pq/2\equiv 0 \pmod{p}$. Hence, $(0,(B-n)/2)$ is a solution. As in Case 1, we have $m+(p-j)q=m+B>B$, and therefore $(0,(B-n)/q)$ is the minimal solution also in this case.

Consider then the cases where $B=q+(q-1)p/2$. According to Table 6.3 and Table 6.4, we have $m=q$ and q is odd. Clearly, $(i,j)=(0,(B-n)/q)$ is a solution, since

$$(n+jq)-m=q+(q-1)p/2-q=(q-1)p/2\equiv 0 \pmod{p}.$$

As in the above, $m+(p-j)q=m+pq-B+n$. By substituting m and B we get

$$m+(p-j)q=q+(2\cdot(q-1)p/2+p)-(q+(q-1)p/2)+n=B+(p-q)+n.$$

Case 3. Assume that $p>q$. Then $p-q$ is positive and $m+(p-j)q>B$. Thus, $(0,(B-n)/q)$ is the smallest solution.

Case 4. Assume that p is even, q is odd and $p<q$. Then $n=q-p/2$, and

$$m+(p-j)q=B+(p-q)+q-p/2=B+p/2>B.$$

Hence $(0,(B-n)/q)$ is the smallest solution also in this final case.

Define now a word w over the $\{a,b,c\}$ by the rule

$$w=\begin{cases} (b^{m-1}ab^{q-m-1}c)^{\frac{p+1}{2}} & \text{if } B=(p+1)q/2, \\ (b^{n-1}cb^{q-n-1}a)^{\frac{B-n}{q}}b^{n-1}c & \text{if } B=q+(q-1)p/2, \end{cases} \tag{6.12}$$

where $m=m(p,q)$ is given by Table 6.4 and $n=[B]_q$. Also, let $w=vc$ and choose $R=\langle(a,b),(b,c)\rangle$. Now q is a pure period of v and p is a global R-period. Namely,

b is related to each letter in v, and the first occasion where the distance between the letters a and c in w is a multiple of p is the case where a is in the position m and c is in the position B. This does not happen in v, since v is one letter shorter. Since $(a,c) \notin R$, 1 is not a global R-period of v. □

Example 6.4.19 For $p = 5$ and $q = 7$, the bound for global–global interaction is $B_g(p,q) = (p+1)q/2 = 21$. Hence, any word w with a global R-period 5 and a pure period 7 and no relational period $\gcd(p,q)$ satisfies $|w| \leq 20$. The situation is illustrated in Figure 6.4. The table with 20 entries represents a word

$$w = (w_1 w_2 \cdots w_7)^2 w_1 w_2 \cdots w_6$$

that is a rational power of the word $w_1 w_2 \cdots w_q$, written into p columns. Letters in each column are R-related, since p is a global period. The graph with vertices $w_1, w_2, \ldots, w_q$ represents all necessary relations in the word. If two vertices occur in the same column of the table, then there is an edge between those vertices. We notice that the graph is almost complete, only the edges $\{w_1, w_7\}$ and $\{w_6, w_7\}$ are missing. Hence, we conclude that $w = (abbbbbc)^2 abbbbb$ and $w' = (abbbbac)^2 abbbba$ with relation $R = \langle (a,b),(b,c) \rangle$ are words such that they have global R-period 5 and pure period 7, but 1 is not a global R-period. Note that w satisfies the formula given by (6.12). Namely, $m(5,7) = 1$ by Table 6.4. Moreover, from the figure we see that increasing the length of the word w by 1 is not possible. The next letter is indicated in the table by $\cancel{w_7}$. We notice that increasing the length causes w_1, w_6 and w_7 to occur in the same column. Hence, the graph would become perfect (dashed lines) implying that all the letters would be related to each other and, therefore, 1 would be a global R-period.

w_1	w_2	w_3	w_4	w_5
w_6	w_7	w_1	w_2	w_3
w_4	w_5	w_6	w_7	w_1
w_2	w_3	w_4	w_5	w_6
$\cancel{w_7}$				

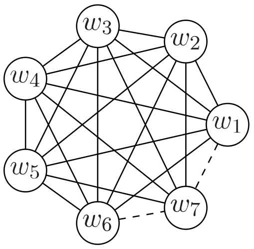

Figure 6.4 Global–global interaction for $p = 5$ and $q = 7$.

Global–local interaction

Instead of attaining a global period $\gcd(p,q)$ we loosen our requirements and consider the case where $\gcd(p,q)$ becomes a local relational period.

Theorem 6.4.20 *Let p and q be positive integers with* $\gcd(p,q) = 1$. *Let k be the* *smallest integer satisfying* $kp \equiv \pm 1 \pmod{q}$. *The bound of global–local interaction for* p *and* q *is*

$$B_{g,l}(p,q) = \begin{cases} q + kp - 1, & \text{if } q \equiv 2 \pmod{p} \text{ and } kp \equiv +1 \pmod{q}; \\ q + kp, & \text{otherwise.} \end{cases}$$

We again divide the proof into two parts.

Lemma 6.4.21 *The bound* $B_{g,l}(p,q)$ *defined in Theorem 6.4.20 is sufficient.*

Proof Denote $B = B_{g,l}(p,q)$, and let w have a pure period q and a global R-period p. We show that 1 is a local R-period of w if $|w| \geq B$. As in the proof of Lemma 6.4.17 we conclude that there are at most q different letters in w. Hence, w has a local R-period 1 if, and only if, for all $n = 1, 2, \dots, q$,

$$w_{[n]_q} R\, w_{[n+1]_q}. \tag{6.13}$$

We show that, for each $1 \leq n \leq q$, there are non-negative i_n and j_n such that

$$[n]_q + i_n q \equiv [n+1]_q + j_n q \pmod{p} \tag{6.14}$$

and both sides of the congruence belong to $\{1, 2, \dots, B\}$. From this it follows together with the global period p that (6.13) holds if $|w| \geq B$.

Case 1. Assume first that $kp \equiv 1 \pmod{q}$. For each $1 \leq n \leq q-1$, choose $j_n = (kp-1)/q$ and $i_n = 0$. Here j_n is an integer by the definition of k. Then

$$(n+1) + j_n q = n + 1 + kp - 1 = n + kp \equiv n \pmod{p}.$$

Clearly, both sides of the congruence belong to $\{1, 2, \dots, B\}$. Also, let $i_q = 0$ and $j_q = (kp-1)/q + 1$. Now

$$1 + j_q q = 1 + kp - 1 + q = q + kp \equiv q \pmod{p}.$$

The left-hand side is less than or equal to B only if $q \not\equiv 2 \pmod{p}$. However, in the special case where $1 \equiv q - 1 \pmod{p}$, we can choose $i_q = (kp-1)/q$ and $j_q = 0$ so that

$$q + i_q q = q + kp - 1 \equiv q - 1 \equiv 1 \pmod{p}.$$

Now the left-hand side is exactly $B = q + kp - 1$.

Case 2. Assume that $kp \equiv -1 \pmod{q}$. For $1 \leq n \leq q-1$, let $i_n = (kp+1)/q$ and $j_n = 0$. Note that i_n is an integer by the definition of k. Hence,

$$n + i_n q = n + kp + 1 \equiv n + 1 \pmod{p}.$$

Choose furthermore $i_q = (kp+1)/q - 1$ and $j_q = 0$. Then

$$q + i_q q = q + kp + 1 - q \equiv 1 \pmod{p}.$$

Note that both sides of both congruences belong to the set $\{1, 2, \dots, B\}$.

Hence, we have shown that (6.13) is satisfied for all $n = 1, 2, \dots, q$, if $|w| \geq B$. Therefore, w has $\gcd(p,q) = 1$ as its local relational period. □

Lemma 6.4.22 *The bound $B_{g,l}(p,q)$ defined in Theorem 6.4.20 is strict.*

Proof Again, let $B = B_{g,l}(p,q)$, where $\gcd(p,q) = 1$. We show that there is a word w of length $B-1$ with a global period p and a pure period q but that w does not have local relational period $\gcd(p,q) = 1$. We show that, at least for one position n with $1 \leq n \leq q$, there is no solution i_n, j_n of (6.14) where both sides of the equation belong to the set $\{1,2,\ldots,B-1\}$. Without contradicting the assumption that p is a global period of w we may then assume that $(w_{[n]_q}, w_{[n+1]_q}) \notin R$ and therefore $\gcd(p,q) = 1$ is not a local R-period of w.

Let again k be the smallest integer satisfying $kp \equiv \pm 1 \pmod{q}$.

If $k > q/2$, then also $(q-k)p \equiv -kp \equiv \mp 1 \pmod{q}$. Now $q-k < k$, and this is a contradiction. On the other hand, if $k = q/2$, then, by the definition of k, $0 \equiv qp = 2kp \equiv \pm 2 \pmod{q}$; a contradiction, since $q \geq 3$ by the definition of the interaction bound. Hence $k < q/2$.

Next we will consider minimal solutions of (6.14). As in the proof of Lemma 6.4.18, by a minimal solution we mean a solution (i_n, j_n) where

$$M(i_n, j_n) = \max([n]_q + i_n q, [n+1]_q + j_n q)$$

is as small as possible. Recall that the minimal solution is unique, and if $1 \leq j \leq p-1$, then the minimal solution has the form $(0, j)$ or $(p-j, 0)$.

We divide the considerations into three cases. In each case there exists $1 \leq n \leq q$ such that the minimal solution (i_n, j_n) of (6.14) satisfies $M(i_n, j_n) \geq B$.

Case 1. Assume first that $kp \equiv 1 \pmod{q}$ and $q \not\equiv 2 \pmod{p}$. We consider (6.14) with $n = q$, i.e., the congruence $q + i_q q \equiv 1 + j_q q \pmod{p}$. Note that in the solution $(i_q, j_q) = (0, (kp-1)/q + 1)$, it is $1 + j_q q = q + kp = B$. Now assume that the solution $(p - (kp-1)/q - 1, 0)$ is smaller, i.e., $q + (p - (kp-1)/q - 1)q < B$. Thus,

$$0 < q + kp - (q + (p - (kp-1)/q - 1)q) = 2kp - qp + q - 1 < q - 1, \qquad (6.15)$$

where the last inequality follows, since by the above $k < q/2$. Since $kp \equiv 1 \pmod{q}$, we have

$$2kp - qp + q - 1 \equiv 1 \pmod{q}. \qquad (6.16)$$

Combining (6.15) and (6.16), we obtain $2kp - qp + q - 1 = 1$. On the other hand,

$$1 = 2kp - qp + q - 1 \equiv q - 1 \pmod{p},$$

which contradicts our assumption. Therefore, for the minimal solution,

$$\max(q + i_q q, 1 + j_q q) \geq B,$$

and thus $M(i_q, j_q) \geq B$. Define then a rational power in the ternary alphabet $\{a,b,c\}$ with length $B-1$:

$$w = (ab^{q-2}c)^{(B-1)/q}.$$

Let $R = \langle (a,b), (b,c) \rangle$. By the above, w has a period q and a global R-period p. However, 1 is not a local R-period of w, since a and c are unrelated.

Case 2. Assume next that $kp \equiv 1 \pmod{q}$ and $q \equiv 2 \pmod{p}$. Consider the congruence

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

In the solution $(i_{q-1}, j_{q-1}) = (0, (kp-1)/q)$, we have $q + j_{q-1}q = q + kp - 1 = B$. Also, in the solution $(p - (kp-1)/q, 0)$,

$$q - 1 + (p - (kp-1)/q)q = q - 1 + qp - kp + 1 > q + kp > B,$$

where the second to last inequality is due to $k < q/2$. Hence, the minimal solution satisfies $\max(q - 1 + i_{q-1}q, q + j_{q-1}q) \geq B$, and hence $M(i_{q-1}, j_{q-1}) \geq B$.

In this case, the rational power

$$w = (b^{q-2}ac)^{(B-1)/q}$$

and the relation $R = \langle (a,b),(b,c) \rangle$ together with the above calculations show that the bound B is strict.

Case 3. Finally, assume that $kp \equiv -1 \pmod{q}$. Consider the same congruence as in Case 2. However, note that now $B = q + kp$. Now $(i_{q-1}, j_{q-1}) = ((kp+1)/q, 0)$ is one solution, where $q - 1 + ((kp+1)/q)q = q + kp = B$. In the other solution $(0, p - (kp+1)/q)$, we have

$$q + (p - (kp+1)/q)q = q + (q-k)p + 1 > q + kp + 1 > B$$

by using the fact $k < q/2$. Hence, the word

$$w = (b^{q-2}ac)^{(B-1)/q}$$

with the relation $R = \langle (a,b),(b,c) \rangle$ shows that the bound B is strict also in this case.

□

Theorem 6.4.20 follows from Lemma 6.4.21 and Lemma 6.4.22. Note that the value of k can be calculated easily using an elementary theorem by Fermat and Euler. Namely, the smallest solution k' of the equation $k'p \equiv 1 \pmod{q}$ is called the *reciprocal* of p modulo q and, by the theorem,

$$k' = [p^{\varphi(q)-1}]_q,$$

where φ is Euler's totient function.[1] Thus $k = \min(k', q-k')$, since $(q-k')p \equiv -1 \pmod{q}$.

Example 6.4.23 Let $p = 5$ and $q = 7$. Then $k' = [5^{\varphi(7)-1}]_7 = 3 = k$ and $kp = 15 \equiv 1 \pmod{7}$. Since $q = 7 \equiv 2 \pmod{5}$, the bound of global–local interaction is equal to $B_{g,l}(p,q) = q + kp - 1 = 21$. Since $B_{g,l}(5,7) = B_{g,g}(5,7)$, a word w of length $B_{g,l}(p,q) - 1 = 20$ with local R-period p and pure period q but not having 1 as a local R-period can be represented in a table form exactly as in Figure 6.4.

---

[1] Recall that $\varphi(n)$ is the number of non-negative integers in the set $\{m \leq n \mid \gcd(m,n) = 1\}$.

Local interactions

Despite the negative result in Example 6.4.13 there exist interaction bounds for some integers p and q also in the case where p is a local period. In the following, if no bound $B(p,q)$ of interaction for p and q exists, then we set $B(p,q) = \infty$.

Theorem 6.4.24 *Let p and q be positive integers with* $\gcd(p,q) = 1$. *Then the bound of local–local interaction for p and q is*

$$B_{l,l}(p,q) = \begin{cases} p+q, & \text{if } p-1 \equiv 0 \pmod{q} \text{ or } p+1 \equiv 0 \pmod{q}; \\ \infty, & \text{otherwise.} \end{cases}$$

Proof Let w be a word of length $B_{l,l} = B_{l,l}(p,q)$ with a pure period q and a local R-period p. Suppose that $\gcd(p,q) = 1$. Assume first that $p+1 \equiv 0 \pmod{q}$. By the periodicity assumption, we then have

$$w_i R w_{i+p} = w_{i-1}$$

for all $i = 2,3,\ldots,q$, and $w_1 R w_{1+p} = w_q$. Since q is a period of w, 1 is a local R-period of w. On the other hand, if we set $R = \langle (a,c),(b,c) \rangle$, the word

$$w = (c^{q-2}ab)^{(p+q-1)/q}$$

has a pure period q and a local R-period p. However, $\gcd(p,q) = 1$ is not a local R-period of w, since $(w_{q-1}, w_q) \notin R$. Note that in order to check that w has a local period p, it suffices to ensure that the distance from any occurrence of a to any occurrence of b is not p. By the length of w, this holds. Namely, the only position i such that $w_i \in \{a,b\}$ and $i+p \leq |w|$ is the first occurrence of the letter a. We have $a = w_{q-1} R w_{q-1+p} = w_{q-2} = c$. Moreover, let $kq = p+1$ for some positive k. Then the only position i such that $w_i \in \{a,b\}$ and $i-p > 0$ is the last occurrence of b. Hence, this position is kq and $b = w_{kq} R w_{kq-p} = w_1 = c$.

Assume next that $p-1 \equiv 0 \pmod{q}$. Now $w_i R w_{i+p} = w_{i+1}$ for all $i = 1,2,\ldots,q$. As above, this means that w has a local R-period 1. Our bound is strict, since setting again $R = \langle (a,c),(b,c) \rangle$, the word

$$w = (ac^{q-2}b)^{(p+q-1)/q}$$

has a pure period q and a local R-period p. However, $(w_q, w_{q+1}) \notin R$ and 1 is not a local R-period. Again the length of w ensures that a and b do not have to be related. As above, we need to check the first occurrence of a and the last occurrence of b. Assume that $p-1 = kq$ for some positive k. Then $a = w_1 R w_{1+p} = w_2 = c$ and $b = w_{(k+1)q} R w_{(k+1)q-p} = w_{q-1} = c$.

Finally, assume that q does not divide $p-1$ nor $p+1$. Then $i+p \not\equiv i+1 \pmod{q}$ and $i+p \not\equiv i-1 \pmod{q}$. Thus, if $R = \langle (a,c),(b,c) \rangle$, then the infinite word

$$w = (abc^{q-2})^{\omega}$$

has a pure period q and a local R-period p, but clearly 1 is not a local R-period of w. □

The next theorem shows that local R-periods are weak when related to global interaction.

Theorem 6.4.25 *Let p and q be such that* $\gcd(p,q)=1$. *The bound $B_{l,g}(p,q)$ of local–global interaction exists only for $q=3$, in which case $B_{l,g}(p,q)=p+3$.*

Proof Let $B=B_{l,g}(p,q)$, and consider first the case $q=3$. Assume that a word w has a pure period 3 and a local R-period p. If $|w|\geq p+2$, then

$$w_1 R w_{[1+p]_3} \text{ and } w_2 R w_{[2+p]_3}. \tag{6.17}$$

Since $\gcd(p,q)=1$, we have $w_{[1+p]_3}=w_2$ and $w_{[2+p]_3}=w_3$, or $w_{[1+p]_3}=w_3$ and $w_{[2+p]_3}=w_1$. If $|w|\geq p+3$, then in addition to (6.17), it holds $w_3 R w_{[3+p]_3}$ where $w_{[3+p]_3}$ is equal to either w_1 or w_2. Hence, all letters are R-related, and 1 is a global R-period of w. On the other hand, $v=(abc)^{(p+2)/3}$, that has length $p+2$, with $S=\langle(a,w_{[1+p]_3}),(b,w_{[2+p]_3})\rangle$ show that the bound $B_{l,g}$ is strict for $q=3$.

Suppose then that $q\geq 4$, and consider the four-letter alphabet $\{a,b,c,d\}$. Choose $R=\langle(a,b),\ (b,c),(c,d),(d,a)\rangle$. Define an infinite word $w=(w_1w_2\cdots w_q)^{\omega}$ as follows:

$$w_1=a,\ w_{[1+p]_q}=b,\ w_{[1+2p]_q}=c \text{ and } w_{[1+ip]_q}=d$$

for $i=3,4,\ldots,q-1$. Now, $w_i R w_{i+p}$ for all $i=1,2,\ldots,q$, and hence p is a local R-period of w. However, 1 is not a global R-period, since no letter is compatible with all other letters. Hence, $B=\infty$. □

Extremal words

We shall now investigate words that demonstrate that the bound of global–global interaction is strict as given in (Halava *et al.*, 2007b). The study of these words originates from the standard Fine and Wilf case, where for coprime periods p and q, the non-constant words of maximal length $p+q-2$ have very interesting properties. Such words are called *extremal Fine and Wilf words*. In 1994 the following result was proved in (de Luca and Mignosi, 1994).

Theorem 6.4.26 (de Luca and Mignosi) *The extremal Fine and Wilf words are palindromes, and the factors of the extremal Fine and Wilf words are exactly the factors of the Sturmian words.*

Definition 6.4.27 Let $p\geq 2$ and $q\geq 3$ be integers satisfying $\gcd(p,q)=1$. A word is an *extremal relational Fine and Wilf word* if $|w|=B_{g,g}(p,q)-1$ and there exists a similarity relation R such that w has a global R-period p and a pure period q but $\gcd(p,q)=1$ is not an R-period of w.
Let $FW(p,q)$ denote the set of all extremal relational Fine and Wilf words.

Denote by R_w the similarity relation with a minimum number of pairs of letters such that $w\in FW(p,q)$ has an R_w-period p. We leave it as an exercise to show that R_w is well defined; see Exercise 6.6.8.

Lemma 6.4.28 *The number of different letters N occurring in $w \in FW(p,q)$ satisfies* $3 \leq N \leq q$.

Proof The inequality $N \leq q$ follows directly from the requirement of q-periodicity.

Suppose then that $N = 2$, say w is over the binary alphabet $\{a,b\}$ together with a similarity relation R such that w has global period p and a pure period q and $|w| = B_{g,g}(p,q) - 1$. If aRb then the letters are R-similar and $\gcd(p,q) = 1$ is a global R-period of w. In $(a,b) \notin R$, then $R = \iota$ and the theorem of Fine and Wilf implies that $\gcd(p,q) = 1$ is a period of w, since both $B_{g,g}(p,q) = (p+1)q/2$ and $B_{g,g}(p,q) = q + (q-1)p/2$ of Table 6.3 are greater than $p + q - 1$. □

Example 6.4.29 Consider the set $FW(3,7)$. Then

$$B_{g,g}(3,7) = (p+1)q/2 = 14.$$

Hence $|w| = 13$ for each $w \in FW(3,7)$. Choose $\Sigma = \{a,b,c\}$ and $R = \langle (a,b),(b,c) \rangle$. Then $u = babbabcbabbab \in FW(3,7)$. But also, $w = babbbbcbabbbb \in FW(3,7)$, and therefore, even if we restrict our considerations to words having the smallest possible number of different letters, we do not have uniqueness.

Also, let $\Delta = \{a,b,c,d\}$ and $R_v = \langle (a,b),(a,c),(a,d),(b,c),(c,d) \rangle$. Then $v = abcacadabcaca \in FW(3,7)$.

However, we can show that the words in $FW(p,q)$ do share a unique structure in the sense described below.

Definition 6.4.30 Let R be a similarity relation on Σ^*. We say that two letters a and b are *R-isomorphic* if, for all $x \in \Sigma$, we have

$$aRx \iff bRx.$$

Also, a letter a is said to be *relationally universal*, or more precisely, *R-universal* if aRx for all $x \in \Sigma$.

In the sequel we consider words in $FW(p,q)$ that do not have any distinct R-isomorphic letters and the number of occurrences of an R-universal letter is minimal.

This restriction is justified, since all words in $FW(p,q)$ can be obtained, up to renaming of letters, from the word w described in the next theorem by two operations: (1) changing some symbols to universal symbols and (2) replacing a letter with a new letter R_w-isomorphic to the original one. In this respect, $w \in FW(p,q)$ with no distinct R_w-isomorphic letters and with minimal number of occurrences of an R_w-universal letter can be called *minimal*. As above, we use the notation $[n]_q$ for the least positive residue of an integer $n \pmod{q}$. For simplicity, denote $B = B_{g,g}(p,q)$.

The proof of the following result follows the techniques of Lemma 6.4.17. For a detailed proof, we refer to (Halava *et al.*, 2007b).

Theorem 6.4.31 *Let $w \in FW(p,q)$ with no distinct R_w-isomorphic letters and with minimal number of occurrences of an R_w-universal letter. This word is unique up to renaming of letters. Furthermore, w is of the form uc^{-1}, where*

$$u = \begin{cases} \left(\left(b^{[B]_p-1}ab^{p-[B]_p}\right)^{\lfloor \frac{q}{p} \rfloor} b^{q-1-\lfloor \frac{q}{p} \rfloor p} c\right)^{\frac{p+1}{2}} & \text{if } B = \frac{p+1}{2}q \text{ and } p < q, \\ \left(b^{[B]_p-1}ab^{q-1-[B]_p}c\right)^{\frac{p+1}{2}} & \text{if } B = \frac{p+1}{2}q \text{ and } p > q, \\ (b^{[B]_q-1}cb^{q-[B]_q-1}a)^{\frac{B-[B]_q}{q}} b^{[B]_q-1}c & \text{otherwise,} \end{cases}$$

and the relation is $R_w = \langle (a,b),(b,c) \rangle$.

Note that the relation $R_w = \langle (a,b),(b,c) \rangle$ in Theorem 6.4.31 which was used in defining the minimal extremal words in $FW(p,q)$ corresponds to the compatibility relation of partial words.

As in the case of normal extremal Fine and Wilf words (de Luca and Mignosi, 1994; Tijdeman and Zamboni, 2003), the minimal extremal relational Fine and Wilf words given in Theorem 6.4.31 have nice palindromic properties. Recall that a word $w = w_1w_2\cdots w_n$ is a *palindrome* if $w = \widetilde{w}$, where $\widetilde{w} = w_nw_{n-1}\cdots w_1$. A generalisation of palindromic words are so called pseudo-palindromic words.

Definition 6.4.32 Let $\sigma : \Sigma \to \Sigma$ be a morphism satisfying $\sigma^2 = \iota$. A word $w = w_1w_2\cdots w_n$ is a *σ-pseudo-palindrome* if $w = \sigma(\widetilde{w})$ for $\widetilde{w} = w_nw_{n-1}\cdots w_1$.

For more information on palindromes and pseudo-palindromes, see (Anne *et al.*, 2005; de Luca and De Luca, 2006). As a final result of extremal Fine and Wilf words we prove the following palindromic properties.

Theorem 6.4.33 *Let $w \in \Sigma^*$, where $\Sigma = \{a,b,c\}$, belong to $FW(p,q)$ with no distinct R_w-isomorphic letters and with minimal number of occurrences of an R_w-universal letter. Let $R_w = \langle (a,b),(b,c) \rangle$. If $B_{g,g}(p,q) = (p+1)q/2$, then w is a palindrome. Otherwise, it is a σ-pseudo-palindrome, where $\sigma : \Sigma \to \Sigma$ is defined by $\sigma(a) = c$ and $\sigma(b) = b$.*

Proof Let $B_g = B_{g,g}(p,q)$. The word w is given by the formula of Theorem 6.4.31. Consider first $w \in FW(p,q)$ such that $B = (p+1)q/2$. Suppose that $w_m = a$. By Theorem 6.4.31, we have $m = n + iq$ for some i and $1 \le n < q$ satisfying $n \equiv B \pmod{p}$. Since $B \equiv 0 \pmod{q}$, $w_{B-n-iq} = w_{q-n}$ by the period q. Now $q - n \equiv q - B + pq = (p+1)q/2 = B \pmod{p}$ since $n \equiv B \pmod{p}$, and thus $w_{B-m} = w_{q-n} = a$.

Consider then the occurrences of c in w. Suppose that $w_m = c$. By Theorem 6.4.31, we have $m \equiv 0 \pmod{q}$. Since $B = (p+1)q/2$, also $B - m \equiv 0 \pmod{q}$. This implies that $w_{B-m} = c$ and therefore $w_m = w_{B-m} = w_{|w|+1-m}$ if $w_m = a$ or $w_m = c$. Hence, this is true also for $w_m = b$ and so the word w is a palindrome.

Next consider $w \in FW(p,q)$ such that $B(p,q) = q + (q-1)p/2$. We apply again Theorem 6.4.31 to obtain: $w_m = a$ if, and only if, $m \equiv 0 \pmod{q}$ and $w_m = c$ if, and only if, $m \equiv B \pmod{q}$. Hence $B - m \equiv B \pmod{q}$, for $w_m = a$, and therefore

$w_{B-m} = c$. On the other hand, if $w_m = c$, then $B - m \equiv 0 \pmod q$ and $w_{B-m} = a$. Thus, $w_m = \sigma(w_{B-m}) = \sigma(w_{|w|+1-m})$, i.e., w is a σ-pseudo-palindrome. ☐

We close this section by giving some examples of relational extremal Fine and Wilf words demonstrating also the palindromic properties discussed above.

Example 6.4.34 We showed in Example 6.4.19 that the word

$$w = (abbbbac)^2abbbba$$

together with the relation $R = \langle (a,b),(b,c) \rangle$ is a word of maximal length such that w has a global R-period 5 and a pure period 7, but 1 is not a global R-period. Note that in $w_1w_2 \cdots w_7$ the letter c occurs in the position $[B]_7 = 7$ and the letter a occurs exactly in positions 1 and 6, which are the positions congruent to $B = 21$ modulo 5. Hence, w is of the form uc^{-1} given in Theorem 6.4.31. Moreover, this word is clearly a palindrome.

The word w is minimal and therefore acts as a template for other words in $FW(5,7)$. For example, replace the letter w_2 by d and w_6 by e. From Figure 6.4 we clearly see that $w_2 = d$ must be R_w-isomorphic to b. In other words, it must be R-universal. Similarly, $w_6 = e$ must be R_w-isomorphic to a. Hence, $v = (adbbbec)^2adbbbe \in FW(5,7)$ with the relation $R_v = \langle \Omega_\Sigma \setminus \{(a,c),(c,a),(e,c),(c,e)\} \rangle$. Note that this word is neither a palindrome nor a pseudo-palindrome.

As an example of a pseudo-palindrome, consider a minimal word $w' \in FW(7,5)$. Now, $B(7,5) = q + (q-1)p/2 = 19$, $[B(7,5)]_5 = 4$ and $(B(7,5) - [B_g(7,5)]_5)/q = 3$. By the formula of Theorem 6.4.31, $w' = (bbbca)^3bbb$, which is a σ-pseudo-palindrome for the morphism $\sigma : \{a,b,c\}^* \to \{a,b,c\}^*$ such that $\sigma(a) = c$ and $\sigma(b) = b$.

6.5 Repetitions in relational words

Repetitions and repetition freeness have been one of the main subjects in combinatorics on words since the seminal papers of Axel Thue (Thue, 1906a, 1912). Let us recall that he showed that there exist infinite words over a ternary alphabet that do not have any squares as factors (see Theorem 4.2.2). Thue also constructed an infinite binary word t that avoids all overlaps $xyxyx$ for any word y and non-empty x (see Theorem 4.2.8). This celebrated word is called the Thue–Morse word and it has many remarkable properties; see (Allouche and Shallit, 1999) or Example 1.3.1.

A generalisation of repetition freeness has been studied in connection with partial words. In a partial word w a factor uv is a square if the words u and v are compatible. Obviously squares cannot be avoided in partial words since every word containing a hole contains a square $a\diamond$ or $\diamond a$ for some letter a. However, we can avoid larger squares. Namely, over a ternary alphabet there exist uncountably many partial words with an infinite number of holes such that the only square factors are the trivial ones; see (Halava *et al.*, 2008b) and (Blanchet-Sadri *et al.*, 2009). Overlapfreeness of partial words was considered in (Halava *et al.*, 2009). They showed that an infinite overlapfree binary partial word is either full or of the form $\diamond w$ or $a \diamond w$, where w is an infinite full word and a is a letter. There are infinitely many overlapfree words of each type; see also (Blanchet-Sadri *et al.*, 2009).

Here we have chosen to investigate squarefree and overlapfreeness with respect to relational words. The section relies on the article (Harju and Kärki, 2014).

6.5.1 Relations, graphs and squares

Denote by

$$\Gamma = \{0,1,2\}$$

a special ternary alphabet in this section.

Definition 6.5.1 Let R be a similarity relation on Σ.

(i) A non-empty word $w \in \Sigma^*$ has an *R-square* uu' if uRu' and uu' is a factor of w. If w has no R-squares then it is *R-squarefree*. A word w that is ι-squarefree, for the identity relation ι, is simply *squarefree*.

(ii) We say that words x and y *R-overlap* if $x = uv$ and $y = v'u'$ such that vRv' or uRu' for some non-empty words u, v, u', v'.

Clearly every R-squarefree word is squarefree. The following result is a criterion of squarefreeness preserving morphisms for ordinary words; see (Crochemore, 1982b).

Theorem 6.5.2 *Let $\alpha : \Gamma^* \to \Sigma^*$ be a morphism for which every image $\alpha(w)$ is squarefree for $|w| \leq 5$. Then α is a squarefree morphism, i.e., $\alpha(w)$ is squarefree for all squarefree words $w \in \Gamma^*$.*

Recall that a similarity relation $R \subseteq \Sigma^* \times \Sigma^*$ can be represented as a graph $G_R = (\Sigma, E)$ such that the edges correspond to the pairs $(a,b) \in R$ with $a \neq b$. Also, the converse is true, i.e., if G is a finite graph there is a unique similarity relation R such that $G = G_R$.

In the following if $G = (\Sigma, E)$ is a graph, its *order* $n(G)$ is the cardinality of the vertex set Σ and its *size* $e(G)$ is the number of its edges. For a vertex $a \in \Sigma$, let $N_G(a) = \{b \mid \{a,b\} \in E\}$ denote its *neighbourhood*.

We denote by P_n a graph that is isomorphic to a *path* on n vertices, i.e., P_n consists of edges $\{v_i, v_{i+1}\}$, $i = 1,2,\ldots,n-1$, for different vertices $v_1, v_2, \ldots, v_n$. Similarly, C_n denotes a *cycle* on n different vertices.

Definition 6.5.3 We say that a connected graph G on Σ *admits a squarefree word*, if for the corresponding similarity relation $R \subseteq \Sigma \times \Sigma$ there exists an infinite R-squarefree word $w \in \Sigma^\omega$ where all letters $a \in \Sigma$ occur infinitely many times. In this case we also say that the relation R *admits a squarefree word*.

Remark 6.5.4 The requirements of connectivity and infinity of occurrences of letters are essential to our considerations. Indeed, e.g., if $R = \iota$ is on $\Gamma = \{0,1,2\}$, then G_R has no edges, and by Thue's result there exists an infinite squarefree word on Γ (see Theorem 4.2.2).

On the other hand, let $R = \langle (a,b),(b,c) \rangle$. By Thue's result, there exists an infinite overlapfree word over the binary alphabet $\{a,c\}$ (see Theorem 4.2.8). However, if we require that there are infinitely many occurrences of b, then no such word exists; see (Halava *et al.*, 2008b).

Next we prove three general results on the number of edges in the graphs for the similarity relations. Recall that a subgraph H of a graph G is a *spanning subgraph* if H has the same vertex set (but a subset of edges).

Lemma 6.5.5 *If $G = G_R$ admits a squarefree word, then so do its connected spanning subgraphs.*

Proof If H is a spanning subgraph of G, then the corresponding similarity relation R' of H is a subrelation of R. Therefore, if w is an R-squarefree word, then w also is R'-squarefree. □

Lemma 6.5.6 *Let $G = (\Sigma, E)$ be a graph that admits a squarefree word, and let $a \in \Sigma$ be a vertex with the neighbourhood $N_G(a)$, and let $b \notin \Sigma$ be a new vertex, and let $A \subseteq N_G(a) \cup \{a\}$ be a non-empty subset. Then, for $E_A = \{\{b,c\} \mid c \in A\}$, the graph $G' = (\Sigma \cup \{b\}, E \cup E_A)$ of order $n(G') = n(G) + 1$ admits a squarefree word.*

Proof We denote R' the similarity relation according to G'. Let w be an infinite R-squarefree word admitted by $G = G_R$, and let w' be an infinite word that is obtained from w by replacing infinitely many occurrences of a by the new letter b so that there remains infinitely many occurrences of a. Assume that u' and v' are factors of w' such that $u'R'v'$, and let u and v be obtained from u' and v', respectively, by replacing all occurrences of b by a. Then also uRv, since $N_{G'}(b) \subseteq N_G(a) \cup \{a\}$. It follows that $u'v'$ cannot be factor in w'. □

As a corollary to Lemma 6.5.5 and Lemma 6.5.6, we obtain the following result.

Corollary 6.5.7 *Suppose there exists a graph G of order n that admits a squarefree word, and let $m \geq n$. Then there is a tree T, any spanning tree of G, of order m that admits a squarefree word.*

It remains to find the lower bound for the order of graphs admitting squarefree words. Below we shall show that the bound is six.

6.5.2 The lower bound

Theorem 6.5.8 *There are no graphs of order 5 that admit squarefree words.*

Proof By the previous corollary, we need to show only that no such trees exist.

First we list all trees of order 5 up to isomorphism. There are only three of these (see Figure 6.5). We let the vertex set be $\Sigma = \{0,1,2,3,4\}$.

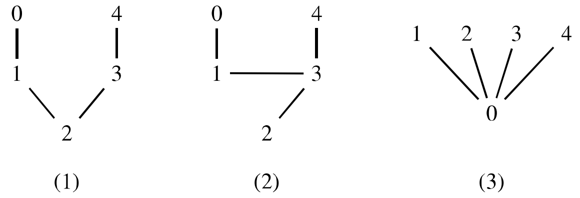

Figure 6.5 The trees of order 5.

In each of the cases, suppose the tree T admits a squarefree word w. Also, let R be the similarity relation corresponding to T.

Consider first the path P_5 in (1). Here the factor 20, if it exists in w, can be followed only by 2, 402 and 42. For instance, in $203a$ the letter a cannot be related to 0 or 3. But this exhausts all possibilities. Therefore between two occurrences of 2, there does not occur any letters 1 or 3. The case for the factor 24 is symmetric, and hence 1 and 3 cannot occur in w infinitely many times. This contradicts the assumption that w should contain infinitely many of each letter.

Let T be the tree in (2). Now, an occurrence of 3 is preceded and followed by 0, since 3 is related to the other letters. Hence 030 is necessarily followed by 2 or 4, but $03R02$ and $03R04$ which is a contradiction.

Finally, if T is the tree in (3), the letter 0 is related to the rest of the letters, and hence 0 cannot occur in w at all. This is again a contradiction. □

Example 6.5.9 Any graph where one of the vertices is adjacent to all other vertices cannot admit a squarefree word. There are also trees of order $n \geq 6$ that do not admit squarefree words. We only consider one of these of order 6. Let T be the first tree in Figure 6.6. Suppose again that w is a squarefree word admitted by T. The letter 3 is necessarily preceded and followed by 0 or 1, but the factors 03 and 13 cannot be followed in w by any word of length two. After Theorem 6.5.10 we learn that the three trees of Figure 6.6 are the only trees of order six that do not admit squarefree words.

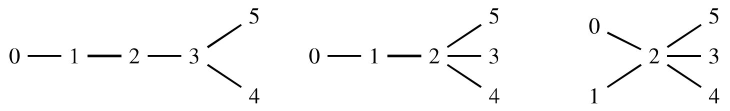

Figure 6.6 Trees of order 6 that do not admit a squarefree word.

The following theorem gives an optimal bound for squarefreeness from (Kärki, 2012) where it was shown that the path P_6 of six vertices admits a squarefree word.

The proof in (Kärki, 2012) relies on the Leech morphism of words which is modified for the similarity relation corresponding to P_6. Below we give another proof of the result using a different graph.

Theorem 6.5.10 *Let* $n \geq 6$. *There exists a graph* G *of order* n *that admits a squarefree word.*

Proof The claim follows from Lemma 6.5.6 after we prove that the graph G of Figure 6.7 on the vertex set $\Sigma = \{0, 1, \ldots, 5\}$ admits a squarefree word. To this end, let R be the similarity relation with $G = G_R$.

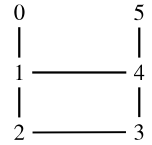

Figure 6.7 A graph G of order 6.

Consider the morphism $\alpha : \Gamma^* \to \Sigma^*$ defined by

$$\alpha(0) = 130502420535,$$
$$\alpha(1) = 135305020535,$$
$$\alpha(2) = 153050240535.$$

We show that if $w \in \Gamma^*$ is a squarefree word, then $\alpha(w)$ does not contain R-squares. Assume on the contrary that $w \in \Gamma^*$ is a squarefree word such that $\alpha(w)$ has a factor uv with uRv. A straightforward computer check shows that $|w| > 5$, and hence $\alpha(w)$ is squarefree as an ordinary word by Theorem 6.5.2. Also, it follows that $u = x_2\alpha(x)y_1$ and $v = y_2\alpha(y)z_1$ for some non-empty words $x, y \in \Gamma^*$ such that $\alpha(a) = x_1x_2$, $\alpha(b) = y_1y_2$ and $\alpha(c) = z_1z_2$ for some $a, b, c \in \Gamma$; see Figure 6.8.

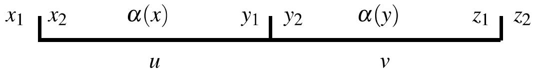

Figure 6.8 The factor ww' in a squarefree word.

We can easily check that none of the images $\alpha(d)$ with $d \in \Gamma$ is similar to a factor of $\alpha(s)$, where $|s| = 2$, except when s has d as a prefix or a suffix. By the form of the images $\alpha(i)$ and the relation R, necessarily $x_2 = y_2$ and $y_1 = z_1$, and thus also $x = y$. Hence $\alpha(b) = y_1y_2 = z_1x_2$, where z_1 is a prefix of $\alpha(c)$ and x_2 is a suffix of $\alpha(a)$. However, the length of the longest R-similar prefixes of the images of different letters is two ($\alpha(0)$ and $\alpha(1)$) and the length of the longest R-similar suffixes is five ($\alpha(0)$ and $\alpha(1)$). Since the images have length 12, this gives a contradiction and proves the claim. □

The graph of Figure 6.7 has three spanning trees as given in Figure 6.9. These trees are obtained by removing one of the edges from the cycle $1-2-3-4$. Hence, by Lemma 6.5.5, the three trees of Figure 6.9 admit a squarefree word. Since, up to isomorphism, there are six trees of order six and the trees in Figure 6.6 do not admit squarefree words, Figure 6.9 lists all the trees admitting a squarefree word.

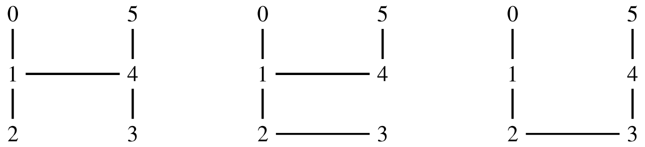

Figure 6.9 The trees of order 6 admitting a squarefree word.

By Lemma 6.5.6, we have the following corollary.

Corollary 6.5.11 *Let $n \geq 6$. There exists a tree T of order n that admits a squarefree word.*

Each tree T of order n has exactly $n-1$ edges. Therefore if a graph of order n admits a squarefree word, then $e(G) \geq n-1$. A trivial upper bound for the size is $e(G) \leq n(n-1)/2$, but this bound is clearly too high; see Exercise 6.6.12.

The following theorem is due to (Kärki, 2012).

Theorem 6.5.12 *The (cordless) cycle C_6 of six vertices does not admit a squarefree word.*

6.5.3 Local and global overlapfreeness

With respect to overlapfreeness in conjunction with similarity relations, there are again two types of case.

Definition 6.5.13 Let $R \subseteq \Sigma^* \times \Sigma^*$ be a similarity relation.

(i) A *local R-overlap* is a word of the form $uu'vv'w$, where uRv, $u'Rv'$ and vRw. In this case, the R-similar words $uu'v$ and $vv'w$ are R-overlapping.

(ii) A *global R-overlap* is a local R-overlap $uu'vv'w$ where also uRw.

Definition 6.5.14 We say that a similarity relation $R \subseteq \Sigma \times \Sigma$ *admits a local (global) overlapfree word*, if its graph G_R is connected and there exists an infinite local (global, resp.) R-overlapfree word $w \in \Sigma^\omega$ where all letters $a \in \Sigma$ occur infinitely many times. In this case we also say that the graph G_R *admits local (global, resp.) overlapfree word*.

Example 6.5.15 Let R correspond to the path P_5 of five edges; see Figure 6.5(1). The word $w = 411132202 = (4111)(3220)2$ is a local R-overlap, but it is not a global R-overlap, since $u = 4$ is not R-compatible $w = 2$.

We show that the smallest order of a graph G_R admitting a local and a global R-overlapfree word is 4.

Theorem 6.5.16 *The similarity relation R corresponding to the path P_4 on four vertices admits a local, and thus also global, R-overlapfree word. No similarity relation on three letters admits a global, and thus local, R-overlapfree word.*

Proof We first prove the negative case on Γ. Let R be the similarity relation on Γ corresponding to the path P_3 of three vertices such that $0R1$ and $2R1$. Consider any infinite word w over Γ where 1 occurs infinitely many times. Let $a1b$ be a factor in w, where $a,b \in \Gamma$. Since $aR1$ and $1Rb$ hold, the factor $a1b$ is a local R-overlap.

Assume then that w is globally R-overlapfree. Hence in a factor $a1b$, always $(a,b) \notin R$, and so $\{a,b\} = \{0,2\}$. Suppose $a1b$ occurs infinitely many times in w. It is necessarily part of $v = ba1ba$, which cannot be followed by 1, since $ba1ba1a$, $ba1ba11$ and $ba1ba1b$ all have global R-overlaps. Also, v cannot be followed by a, since $ba1baaa$, $ba1baa1$ and $ba1baab$ all have an R-overlap. Hence, v is followed by b. Now, $ba1baba$ contains a global R-overlap $1baba$, and $ba1bab1$ contains a global R-overlap $1bab1$ and $ba1babb$ is a global R-overlap. In conclusion, w contains global P_3-overlaps.

For the first claim, let R correspond to P_4 on $\Sigma = \{0,1,2,3\}$. Now $(0,3) \notin R$. Consider the Thue–Morse word $t = \tau^{\omega}(0)$ over $\{0,3\}$, where $\tau : \{0,3\}^* \to \{0,3\}^*$ is defined by

$$
\begin{aligned}
0 &\mapsto 03,\\
3 &\mapsto 30.
\end{aligned}
$$

The word $t = 033030033003\cdots$ is a fixed point of τ, and it is overlapfree; see Theorem 4.2.8 and (Lothaire, 1983). Let t' be a word obtained from t by arbitrarily replacing infinitely many occurrences of 303 by 313 and infinitely many occurrences of 030 by 020. We claim that t' is locally and, consequently, also globally R-overlapfree.

Suppose on the contrary that t' has a local R-overlap. We can assume that this factor is of the form $axbyc$ where $a,b,c \in \Sigma$ and $axbRbyc$. We conclude that there is an index i in $axb = u_1u_2\cdots u_n$ and in $byc = v_1v_2\cdots v_n$ such that either $u_i = 1$ and $v_i = 2$ or vice versa. Indeed, otherwise, we could change all modified letters 1 and 2 in $axbyc$ back to the original letters of t and obtain a global P_4-overlap in t contradicting overlapfreeness of t.

By symmetry, we can choose $u_i = 1$ and $v_i = 2$. If $i > 1$, then $u_{i-1}u_i = 31$ and $v_{i-1}v_h = 02$. These factors are not R-compatible; a contradiction. If $i < |u|$, then $u_iu_{i+1} = 13$ and $v_iv_{i+1} = 20$. As above, we obtain a contradiction. Hence, there are no local R-overlaps in t' and the claim follows. □

The following result on cycles was proven in (Kärki, 2012). We leave the proof as an exercise.

Theorem 6.5.17 *The (cordless) cycle C_5 of five vertices admits both globally and locally an overlapfree word.*

6.6 Exercises and problems

Section 6.3

Exercise 6.6.1 Prove Corollary 6.3.4: if X is an (R_1, S_1)-code and R_2 and S_2 are similarity relations on Σ^*, then X is an $(R_1 \cap R_2, S_1 \cup S_2)$-code.

Exercise 6.6.2 Let R_1 and R_2 be similarity relations on Σ^*, and let $X \subseteq \Sigma^*$. Show that X is not necessarily an $(R_1 \cup R_2, S)$-code even when if it is both an (R_1, S)-code and an (R_2, S)-code.

Exercise 6.6.3 Show that the converse of Corollary 6.3.7 does not hold in general.

Section 6.4

Exercise 6.6.4 Prove Theorem 6.4.6: if a similarity relation R is an equivalence relation, then for the tree types of periods we have $\pi_g(x) = \pi_e(x) = \pi_l(x)$.

Exercise 6.6.5 Consider an equivalence relation R that is a similarity relation. Let a word x have two R-periods p and q and suppose the length of x is at least $p + q - \gcd(p, q)$. Show that $\gcd(p, q)$ is an R-period of x.

Exercise 6.6.6 Note that the roles of the relations R and S are not symmetric in Theorem 6.3.6, i.e., show that not all (R, S)-codes are (S, S)-codes.

Exercise 6.6.7 Show that, in general, it is not the case that $B_{g,g}(p, q) = B_{g,l}(p, q)$.

Exercise 6.6.8 Let R_w be the similarity relation with a minimum number of pairs of letters such that $w \in FW(p, q)$ has an R_w-period p. Show that the relation R_w is well defined.

Problem 6.6.9 How do the periodicity properties of words change if we consider, instead of similarity relations, the relations $R \subseteq \Sigma^* \times \Sigma^*$ that need not be symmetric, but still induced by a relation on the set of the letters Σ, i.e., if $(u, v) \in R$ then $|u| = |v|$? An integer p is then a *global R-period* of a word $w = w_1 w_2 \cdots w_n \in \Sigma^*$, if for all indices i and j,

$$i \equiv j \text{ and } i < j \implies (w_i, w_j) \in R,$$

and p is a *local R-period*, if $(w_i, w_{i+p}) \in R$ for all indices i and j.

Section 6.5

Exercise 6.6.10 Prove Theorem 6.5.12: the (cordless) cycle C_6 of six vertices does not admit squarefree words.

Exercise 6.6.11 Prove Theorem 6.5.17: the (cordless) cycle C_5 of five vertices admits both globally and locally an overlapfree word.

Exercise 6.6.12 Show that for each $n \geq 6$, there exists a graph G of order $n(G) = n$ and size

$$n - 1 \leq e(G) \leq \frac{(n+1)(n-6)}{2} + 6. \qquad (6.18)$$

Problem 6.6.13 Give an optimal upper bound for the number of edges in graphs of order n that admit an overlapfree word.

Problem 6.6.14 Characterise the trees (or graphs in general) of order n that admit a squarefree word.

Problem 6.6.15 Give an optimal upper bound for the number of edges in graphs of order n that admit a squarefree word.

Problem 6.6.16 As in Problem 6.6.9 consider the generalisations of the relations where $R \subseteq \Sigma^* \times \Sigma^*$ need not be symmetric, and say that a factor uv of a word is a *R-square* if $(u,v) \in R$ holds. Now, the graph corresponding to R is directed, i.e., the edges are ordered pairs of letters (vertices). Determine the minimum order of Σ for a relation R for which there exists an infinite word avoiding directed R-squares, and containing all letters infinitely often.

Problem 6.6.17 Theorem 6.5.2 gives an efficient criterion for a morphism to be squarefree. Is there such a criterion for morphisms that preserve R-squarefree words? Does there exists an integer N for which any morphism $\alpha : \Gamma^* \to \Sigma^*$ preserves R-squarefree words if it does so for R-squarefree words of length N?