How Much Different Are Two Words with Different Shortest Periods

Mai Alzamel[1], Maxime Crochemore[1,2], Costas S. Iliopoulos[1], Tomasz Kociumaka[3], Ritu Kundu[1], Jakub Radoszewski[1,3(✉)], Wojciech Rytter[3], and Tomasz Waleń[3]

[1] Department of Informatics, King's College London, London, UK
{mai.alzamel,maxime.crochemore,costas.iliopoulos,ritu.kundu}@kcl.ac.uk
[2] Université Paris-Est, Marne-la-Vallée, France
[3] Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Warsaw, Poland
{kociumaka,jrad,rytter,walen}@mimuw.edu.pl

Abstract. Sometimes the difference between two distinct words of the same length cannot be smaller than a certain minimal amount. In particular if two distinct words of the same length are both periodic or quasiperiodic, then their Hamming distance is at least 2. We study here how the minimum Hamming distance $\mathit{dist}(x, y)$ between two words x, y of the same length n depends on their periods. Similar problems were considered in [1] in the context of quasiperiodicities. We say that a period p of a word x is *primitive* if x does not have any smaller period p' which divides p. For integers p, n ($p \le n$) we define $\mathcal{P}_p(n)$ as the set of words of length n with primitive period p. We show several results related to the following functions introduced in this paper for $p \ne q$ and $n \ge \max(p, q)$.

$$\mathcal{D}_{p,q}(n) = \min\{\, \mathit{dist}(x, y) \;:\; x \in \mathcal{P}_p(n),\; y \in \mathcal{P}_q(n) \,\},$$
$$N_{p,q}(h) = \max\{\, n \;:\; \mathcal{D}_{p,q}(n) \le h \,\}.$$

1 Introduction

Consider a word x of length $|x| = n$, with its positions numbered 0 through $n-1$. We say that x has a period p if $x_i = x_{i+p}$ for all $0 \le i < n - p$. Our work can be seen as a quest to extend Fine and Wilf's Periodicity Lemma [14], which is a ubiquitous tool of combinatorics on words.

Lemma 1 (Periodicity Lemma [14]). *If a word x has periods p and q and $|x| \ge p + q - \mathrm{GCD}(p, q)$, then x also has a period $\mathrm{GCD}(p, q)$.*

Other known extensions of this lemma include a variant with three [10] and an arbitrary number of specified periods [11,16,17,23], the so-called new periodicity

W. Rytter—Supported by the Polish National Science Center, grant no 2014/13/B/ST6/00770.

L. Iliadis et al. (Eds.): AIAI 2018, IFIP AICT 520, pp. 168–178, 2018.
https://doi.org/10.1007/978-3-319-92016-0_16

lemma [3,13], a periodicity lemma for repetitions that involve morphisms [19], and extensions into periodicity of partial words [4–9,22], into abelian [12] and k-abelian [18] periodicity, into bidimensional words [20], and other variations [15,21].

We say that a word x of length n is *periodic* if it has a period p such that $2p \le n$. For two words x and y of length n, by $dist(x, y)$ we denote their Hamming distance being the number of positions $i = 0, \ldots, n-1$ such that $x_i \neq y_i$. The following folklore fact gives a lower bound on how different are two distinct periodic words. Its proof can be found in [1].

Fact 2. *If x and y are distinct periodic words of the same length, then $dist(x, y) \neq 1$.*

We present several generalizations of this fact.

Results similar to Fact 2 were presented recently in the context of quasiperiodicity [1]. We say that a word x has a *cover* u if each position in x is located inside an occurrence of u in x. The word x is called *quasiperiodic* if it has a cover u other than x. In [1] the following generalization of Fact fct:folklore was shown: $dist(x, y) > 1$ for any two distinct quasiperiodic words x, y of the same length. This type of fact has potential applications; see [2].

There is a quantitative difference between periods and covers. For example, there are words x and y of length 1024 with shortest covers of length 4 and 5, respectively, and $dist(x, y) = 2$:

$$x = (abaa)^{256} \quad \text{and} \quad y = aaba(abaa)^{255}$$

with covers $abaa$ and $aabaa$. However, if x and y are words of length 1024 with shortest periods 4 and 5, respectively, then we must have $dist(x, y) \ge 357$.

We say that a period p of a word x is *primitive* if no proper divisor of p is a period of x, i.e., if $p' \mid p$ and p' is a period of x, then $p' = p$. We define

$$\mathcal{P}_p(n) = \{\, |x| = n,\, p \text{ is a primitive period of } x \,\}.$$

The ultimate goal of this work is a characterization of the function $\mathcal{D}_{p,q}$ defined for $p \neq q$ and $n \ge \max(p, q)$ as:

$$\mathcal{D}_{p,q}(n) = \min \{\, dist(x, y) \;:\; x \in \mathcal{P}_p(n), y \in \mathcal{P}_q(n) \,\}.$$

As $\mathcal{D}_{p,q}$ is non-decreasing for given p, q, it can be described by the following auxiliary function:

$$N_{p,q}(h) = \max \{\, n \;:\; \mathcal{D}_{p,q}(n) \le h \,\}.$$

One can note that Lemma 1 can be equivalently formulated as $N_{p,q}(0) < p+q-\mathrm{GCD}(p, q)$ (Fig. 1). Similarly, an equivalent formulation of Fact 2 is $N_{p,q}(1) < 2q$.

Fine and Wilf [14] also proved that the bound $p+q-\mathrm{GCD}(p, q)$ of Lemma 1 cannot be improved. Consequently, $N_{p,q}(0) = p + q - \mathrm{GCD}(p, q) - 1$. On the other hand, we show that $N_{p,q}(1) = 2q - 1$ only for $p \mid q$. Hence, the bound $N_{p,q}(1) < 2q$ of Fact 2 is not tight in general.

Our Results. In Sect. 2 we consider the case that $p \mid q$. In the remaining sections, we only consider the case of $p < q$ and $p \nmid q$. In Sect. 3 we show exact values of the function $\mathcal{D}_{p,q}$ for $p + q - \mathrm{GCD}(p, q) \le n \le 2q$. In Sect. 4, we show the following bounds valid for abitrary $n \ge q$:

$$\left\lfloor \frac{n-q}{p} \right\rfloor \ \le\ \mathcal{D}_{p,q}(n) \ \le 2 \left\lceil \frac{n-q}{p} \right\rceil .$$

We also prove an alternative bound $\mathcal{D}_{p,q}(n) \ge \left\lfloor \frac{2n}{p+q} \right\rfloor$ valid for $n \ge p + q$.

h	$N_{3,4}(h)$	example
1	6	$a\,a\,b\,a\,a\,\boldsymbol{b}$ $a\,a\,b\,a\,a\,\boldsymbol{a}$
2	8	$a\,a\,b\,a\,a\,\boldsymbol{b}\,\boldsymbol{a}\,a$ $a\,a\,b\,a\,a\,\boldsymbol{a}\,\boldsymbol{b}\,a$
3	10	$a\,a\,b\,a\,a\,\boldsymbol{b}\,\boldsymbol{a}\,a\,\boldsymbol{b}\,a$ $a\,a\,b\,a\,a\,\boldsymbol{a}\,\boldsymbol{b}\,a\,\boldsymbol{a}\,a$
4	11	$a\,a\,b\,a\,a\,\boldsymbol{b}\,\boldsymbol{a}\,a\,\boldsymbol{b}\,a\,\boldsymbol{a}$ $a\,a\,b\,a\,a\,\boldsymbol{a}\,\boldsymbol{b}\,a\,\boldsymbol{a}\,a\,\boldsymbol{b}$
5	17	$a\,a\,b\,a\,a\,\boldsymbol{b}\,\boldsymbol{a}\,a\,\boldsymbol{b}\,a\,\boldsymbol{a}\,\boldsymbol{b}\,a\,a\,b\,a\,a$ $a\,a\,b\,a\,a\,\boldsymbol{a}\,\boldsymbol{b}\,a\,\boldsymbol{a}\,a\,\boldsymbol{b}\,\boldsymbol{a}\,a\,a\,b\,a\,a$

h	$N_{2,3}(h)$	example
0	3	$a\,b\,a$ $a\,b\,a$
1	4	$a\,b\,a\,\boldsymbol{b}$ $a\,b\,a\,\boldsymbol{a}$
2	5	$a\,b\,a\,\boldsymbol{b}\,\boldsymbol{a}$ $a\,b\,a\,\boldsymbol{a}\,\boldsymbol{b}$
3	9	$a\,b\,a\,\boldsymbol{b}\,\boldsymbol{a}\,\boldsymbol{b}\,a\,b\,a$ $a\,b\,a\,\boldsymbol{a}\,\boldsymbol{b}\,\boldsymbol{a}\,a\,b\,a$
4	10	$a\,b\,a\,\boldsymbol{b}\,\boldsymbol{a}\,\boldsymbol{b}\,a\,b\,a\,\boldsymbol{b}$ $a\,b\,a\,\boldsymbol{a}\,\boldsymbol{b}\,\boldsymbol{a}\,a\,b\,a\,\boldsymbol{a}$
5	11	$a\,b\,a\,\boldsymbol{b}\,\boldsymbol{a}\,\boldsymbol{b}\,a\,b\,a\,\boldsymbol{b}\,\boldsymbol{a}$ $a\,b\,a\,\boldsymbol{a}\,\boldsymbol{b}\,\boldsymbol{a}\,a\,b\,a\,\boldsymbol{a}\,\boldsymbol{b}$
6	15	$a\,b\,a\,\boldsymbol{b}\,\boldsymbol{a}\,\boldsymbol{b}\,a\,b\,a\,\boldsymbol{b}\,\boldsymbol{a}\,\boldsymbol{b}\,a\,b\,a$ $a\,b\,a\,\boldsymbol{a}\,\boldsymbol{b}\,\boldsymbol{a}\,a\,b\,a\,\boldsymbol{a}\,\boldsymbol{b}\,\boldsymbol{a}\,a\,b\,a$
7	16	$a\,b\,a\,\boldsymbol{b}\,\boldsymbol{a}\,\boldsymbol{b}\,a\,b\,a\,\boldsymbol{b}\,\boldsymbol{a}\,\boldsymbol{b}\,a\,b\,a\,\boldsymbol{b}$ $a\,b\,a\,\boldsymbol{a}\,\boldsymbol{b}\,\boldsymbol{a}\,a\,b\,a\,\boldsymbol{a}\,\boldsymbol{b}\,\boldsymbol{a}\,a\,b\,a\,\boldsymbol{a}$

Fig. 1. Upper table: values of $N_{3,4}(h)$ for $h = 1, \ldots, 5$ together with pairs of words of length $N_{3,4}(h)$ that have the Hamming distance h. Lower table: values of $N_{2,3}(h)$ for $h = 0, \ldots, 7$.

2 Preliminaries

Let us consider a finite alphabet Σ. If x is a word of length $|x| = n$, then by $x_i \in \Sigma$ for $i = 0, \ldots, n-1$ we denote its ith letter. We say that a word v is a factor of a word x if there exist words u and w such that $x = uvw$. A factor v is called a prefix of x if u is an empty word in some such decomposition and a suffix if w is an empty word in some such decomposition. By $x[i..j]$ we denote the factor $x_i \dots x_j$.

If $x_i = x_{i+p}$ for all $0 \le i < n - p$ for some integer p, then p is called *a period* of x and the prefix of x of length p is called *a string period* of x. If x has period p, then y is called a *periodic extension of x with period p* if y also has period p and has x as a prefix.

We say that a period p is *primitive* if no proper divisor of p is a period of x. Note that the shortest period (denoted $p = \mathbf{per}(x)$) is always primitive.

We say that a word x is *primitive* if there exists no other word u and integer $k > 1$ such that $x = u^k$. Note that p is a primitive period of x if and only if the corresponding string period is a primitive word. Two words x and y are each other's cyclic rotations if there exist words u and v such that $x = uv$ and $y = vu$. In this case we also say that $|u|$ is the shift between x and y.

For a sequence of positive integers $(a_1, \dots, a_m)$, we define a $(a_1, \dots, a_m)$-*decomposition* of a word x as a sequence of consecutive factors of x of lengths $a_1, \dots, a_m, a_1, \dots, a_m, \dots$ The sequence ends at the last *complete* factor that can be cut out of x; see Fig. 2 for an example.

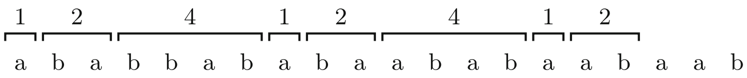

Fig. 2. The $(1, 2, 4)$-decomposition of *ababbababaababaabaab* is *a ba bbab a ba abab a ab*.

If $p \mid q$, we can give a simple complete characterization of functions $N_{p,q}$ and $\mathcal{D}_{p,q}$.

Fact 3. *If $p \mid q$ and $p < q$, then $\mathcal{D}_{p,q}(n) = \left\lfloor \frac{n}{q} \right\rfloor$ and $N_{p,q}(h) = q \cdot (h+1) - 1$.*

Proof. We first show that $\mathcal{D}_{p,q}(n) \ge \left\lfloor \frac{n}{q} \right\rfloor$. Consider a positive integer n, words $x \in \mathcal{P}_p(n)$, $y \in \mathcal{P}_q(n)$, and the (q)-decompositions of x and y: $\alpha_1, \dots, \alpha_k$ and $\beta_1, \dots, \beta_k$. Observe that $\alpha_1 = \dots = \alpha_k$ and $\beta_1 = \dots = \beta_k$ because q is a period of both x and y, but $\alpha_1 \ne \beta_1$ because q is a primitive period of y, but not a primitive period of x. Hence, $dist(x, y) \ge k$.

As for the other inequality on $\mathcal{D}_{p,q}(n)$, let us take $x = (a^{p-1}b)^{\lfloor n/p \rfloor} a^{n \bmod p}$ and let y be the word that is obtained from x by changing the letters at positions $i \equiv q - 1 \pmod{q}$ from b to c. Then $dist(x, y) = \left\lfloor \frac{n}{q} \right\rfloor$.

Finally, the formula for $N_{p,q}(h)$ follows directly from the other one. $\square$

Henceforth, we will always assume that $p \nmid q$ and $q \nmid p$.

3 Exact Values for Small n

Let us start with the following useful lemma.

Lemma 4. *Let x be a word of length n and let y by its cyclic rotation by s characters. If $x \neq y$, then $dist(x, y) \geq 2$. Moreover, there are two mismatches between x and y located at least $\mathrm{GCD}(n, s)$ positions apart.*

Proof. Note that $y_i = x_{(i+s) \bmod n}$ for $0 \leq i < n$. Since $x \neq y$, we have $x_a \neq y_a = x_{(a+s) \bmod n}$ for some position a. Let k be the smallest positive integer such that $x_a = x_{(a+ks) \bmod n}$. Due to $x_{(a+s) \bmod n} \neq x_a$ and $x_{(a+ns) \bmod n} = x_a$, we have $1 < k \leq n$. Let $b = (a + (k-1)s) \bmod n$. Note that $x_b \neq x_a = x_{(b+s) \bmod n} = y_b$. Hence, a and b are positions of two distinct mismatches between x and y. Moreover, $b \equiv (a + (k-1)s) \bmod n \equiv a \pmod{\mathrm{GCD}(n, s)}$. Consequently, these two mismatches are indeed located at least $\mathrm{GCD}(n, s)$ positions apart. □

For an illustration of the following Lemma 5, see Fig. 3.

Lemma 5. *Consider positive integers p, q satisfying $p < q$ and $p \nmid q$. Let x and y be words of length n such that $p + q - \mathrm{GCD}(p, q) \leq n \leq q + p\lceil \frac{q}{p} \rceil - 1$, p is a period of x, and q is a period of y but not a period of x. Then*

$$dist(x, y) \geq \left\lfloor \tfrac{n-q}{p} \right\rfloor + \left\lfloor \tfrac{n-q+\mathrm{GCD}(p,q)}{p} \right\rfloor .$$

Proof. Let $u = x[0..p-1]$ and let v be the cyclic rotation of u by q characters. Note that u is a string period of x, so $u \neq v$; otherwise, q would be a period of x. Consequently, Lemma 4 provides two distinct indices a, b such that $u_a \neq v_a$, $u_b \neq v_b$, and $a \leq b - \mathrm{GCD}(p, q) < p - \mathrm{GCD}(p, q)$. Let us define

$$A = \left\{ kp + a : 0 \leq k < \left\lfloor \tfrac{n-q+\mathrm{GCD}(p,q)}{p} \right\rfloor \right\},$$
$$B = \left\{ kp + b : 0 \leq k < \left\lfloor \tfrac{n-q}{p} \right\rfloor \right\}.$$

Observe that

$$\max A = \left\lfloor \tfrac{n-q+\mathrm{GCD}(p,q)}{p} \right\rfloor p - p + a \leq n - q + \mathrm{GCD}(p, q) - p + a < n - q$$

and
$$\max B = \left\lfloor \tfrac{n-q}{p} \right\rfloor p - p + b \leq n - q - p + b < n - q.$$

Moreover,

$$\max A \leq \left\lfloor \tfrac{p\lceil q/p\rceil - 1 + \mathrm{GCD}(p,q)}{p} \right\rfloor p - p + a = \left\lceil \tfrac{q}{p} \right\rceil p - p + a < q + a \leq q + \min(A \cup B),$$

and

$$\max B \leq \left\lfloor \tfrac{p\lceil q/p\rceil - 1}{p} \right\rfloor p - p + b = \left\lceil \tfrac{q}{p} \right\rceil p - p < q \leq q + \min(A \cup B).$$

Consequently, for each $i \in A \cup B$, there are positions x_i and x_{i+q}, and all these $2(|A| + |B|)$ positions are distinct. Moreover, observe that for $i \in A$, we have $x_i = u_a \neq v_a = x_{i+q}$, while for $i \in B$, $x_i = u_b \neq v_b = x_{i+q}$. Thus, for $i \in A \cup B$, we have $x_i \neq x_{i+q}$, but $y_i = y_{i+q}$; hence $x_i \neq y_i$ or $x_{i+q} \neq y_{i+q}$. The positions we consider are distinct, so $dist(x, y) \geq |A \cup B| = |A| + |B| = \left\lfloor \frac{n-q}{p} \right\rfloor + \left\lfloor \frac{n-q+\mathrm{GCD}(p,q)}{p} \right\rfloor$, as claimed. □

For an illustration of the following Lemma 6, see Figs. 4 and 5.

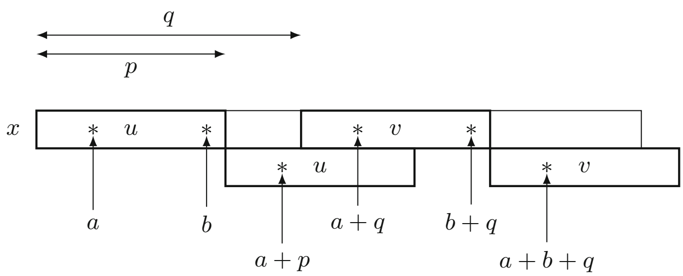

Fig. 3. Illustration of the equalities in the bound in Lemma 5 for $\lceil \frac{q}{p} \rceil = 1$.

Lemma 6. *Consider coprime integers p, q satisfying $1 < p < q$. Let w be a word of length $p + q - 2$ with periods p and q, but without period 1. Moreover, let n be an integer such that $p + q - 1 \leq n \leq q + \lceil \frac{q}{p} \rceil p - 1$, and let x and y be periodic extensions of w of length n preserving periods p and q, respectively. Then $\mathbf{per}(x) = p$, $\mathbf{per}(y) = q$, and*

$$dist(x, y) \leq \left\lfloor \frac{n-q}{p} \right\rfloor + \left\lfloor \frac{n-q+1}{p} \right\rfloor,$$

Proof. Claim. If a position i satisfies $i < q$ or $(i-q) \bmod p < p-2$, then $x_i = y_i$.

Proof. The claim is clear for $i < q+p-2$ since due to the common prefix of x and y. Thus, we consider a position $i = q+kp+r$ with $1 \leq k < \lceil \frac{q}{p} \rceil$ and $0 \leq r < p-2$. We have $x_{q+kp+r} = x_{q+r} = y_{q+r} = y_r = x_r = x_{kp+r} = y_{kp+r} = y_{q+kp+r}$. This is because positions $r < kp+r < q+r$ are within the common prefix of x and y. □

Consequently,

$$\begin{aligned} dist(x, y) \leq \{i : q \leq i < n \wedge (i-q) \bmod p \geq p-2\} = \\ \{j : 0 \leq j < n-q \wedge j \bmod p = p-1\} + \{j : 0 \leq j < n-q \wedge j \bmod p = p-2\} = \\ \left\lfloor \frac{n-q}{p} \right\rfloor + \left\lfloor \frac{n-q+1}{p} \right\rfloor, \end{aligned}$$

as claimed. Next, we prove that $p' := \mathbf{per}(x)$ is equal to p. Note that $p' \leq p$ by definition of x. For a proof by contradiction, suppose that $p' < p$. Note that w has periods p' and q. Moreover, $|w| = p + q - 2 \geq p' + q - 1$, so $\mathrm{GCD}(p', q)$ is a period of w. Moreover, $n \geq p+q-1 \geq p+\mathrm{GCD}(p', q)-1$, so $\mathrm{GCD}(\mathrm{GCD}(p', q), p)$

is a period of x. However, $\mathrm{GCD}(\mathrm{GCD}(p', q), p) = \mathrm{GCD}(p', \mathrm{GCD}(q, p)) = 1$ is not a period of w, which is a prefix of x.

Similarly, suppose that $q' := \mathbf{per}(y) < q$. We observe that $|w| = p + q - 2 \geq p + q' - 1$, so $\mathrm{GCD}(p, q')$ is a period of w. Moreover, $n \geq p + q - 1 \geq \mathrm{GCD}(p, q') + q - 1$, so $\mathrm{GCD}(\mathrm{GCD}(p, q'), q)$ is a period of y. However, $\mathrm{GCD}(\mathrm{GCD}(p, q'), q) = \mathrm{GCD}(q', \mathrm{GCD}(p, q)) = 1$ is not a period of w, which is a prefix of y. □

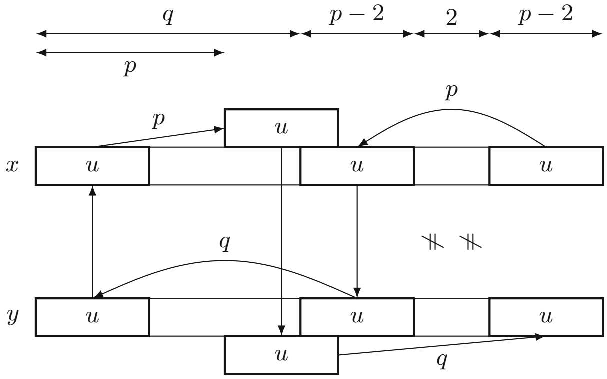

Fig. 4. Illustration of the equalities in the lower bound in Lemma 6 for $n = q + 2p - 2$.

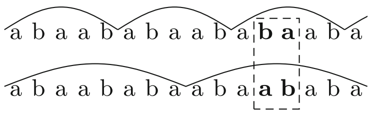

Fig. 5. A periodic prefix of a Fibonacci word and a power of a Fibonacci word that differ only at two positions.

Theorem 7. *If* $p < q$, $p \nmid q$, *and* $p + q - \mathrm{GCD}(p, q) \leq n \leq q + \lceil \frac{q}{p} \rceil p - 1$, *then*

$$\mathcal{D}_{p,q}(n) = \left\lfloor \frac{n - q}{p} \right\rfloor + \left\lfloor \frac{n + \mathrm{GCD}(p, q) - q}{p} \right\rfloor. \tag{1}$$

Proof. Lemma 5 gives a lower bound of $\mathcal{D}_{p,q}(n)$. Our upper bound is based on Lemma 6. Let $d = \mathrm{GCD}(p, q)$, $p' = \frac{p}{d}$, $q' = \frac{q}{d}$, and $n' = \left\lfloor \frac{n}{d} \right\rfloor$. Observe that $1 < p' < q'$ and $p' + q' - 1 \leq n' \leq q' + \lceil \frac{q'}{p'} \rceil p' - 1$. Hence, Lemma 6 results in strings x', y' with of length n' with shortest periods p' and q' respectively, and with $dist(x', y') \leq \left\lfloor \frac{n'-q'}{p'} \right\rfloor + \left\lfloor \frac{n'-q'+1}{p'} \right\rfloor = \left\lfloor \frac{n-q}{p} \right\rfloor + \left\lfloor \frac{n-q+\mathrm{GCD}(p,q)}{p} \right\rfloor$.

Let c be a character occurring neither in x' nor in y'. Let us define x and y so that $x_{id+d-1} = x'_i$ and $y =_{id+d-1}= y'_i$, and $x_j = y_j = c$ if $j \bmod d \neq d-1$. Note that $dist(x, y) = dist(x', y')$ and $|x| = |y| = n$. Also, observe that due to the choice of the character c, all periods of x and y are larger than dn' or multiples of d. Consequently, $\mathbf{per}(x) = d\mathbf{per}(x') = dp' = p$ and $\mathbf{per}(y) = d\mathbf{per}(y') = dq' = q$. This completes the construction. $\square$

Corollary 8. *The formula* (1) *of Theorem 7 applies for* $p + q - \mathrm{GCD}(p, q) \leq n \leq 2q$.

Fact 9. *The function* $\mathcal{D}_{p,q}(n)$ *is non-decreasing for* $n \geq p + q - \mathrm{GCD}(p, q)$. *Moreover:*

$$\mathcal{D}_{p,q}(n) = h \iff N_{p,q}(h-1) < n \leq N_{p,q}(h)$$
$$N_{p,q}(h) = n \iff \mathcal{D}_{p,q}(n) = h < \mathcal{D}_{p,q}(n+1)$$

4 Bounds for $\mathcal{D}_{p,q}(n)$ for Arbitrary n

Lemma 10. *Let* p, q *be integers such that* $p < q$ *and* $p \nmid q$. *Moreover, let* x *and* y *be words of length* $n \geq q$ *such that* p *is a period of* x, *and* q *is a period of* y *but not of* x. *Then* $dist(x, y) \geq \left\lfloor \frac{n-q}{p} \right\rfloor$.

Proof. Since q is not a period of x, we have $x_i \neq x_{i+q}$ for some position i, $0 \leq i < n - q$. Consider a set $J = \{j : 0 \leq j < n - q \wedge j \equiv i \pmod{p}\}$. Since p is a period of x, we have $x_j \neq x_{j+q}$ for each $j \in J$; on the other hand, $y_j = y_{j+q}$, so $x_j \neq y_j$ or $x_{j+q} \neq y_{j+q}$. Moreover, $p \nmid q$ implies that the positions j, $j + q$ across $j \in J$ are pairwise distinct. Consequently, $dist(x, y) \geq |J| \geq \left\lfloor \frac{n-q}{p} \right\rfloor$. $\square$

Theorem 11. *If* $p < q$, $p \nmid q$, *and* $n \geq p + q$, *then* $\left\lfloor \frac{n-q}{p} \right\rfloor \leq \mathcal{D}_{p,q}(n) \leq 2\left\lceil \frac{n-q}{p} \right\rceil$.

Proof. The lower bound follows directly from Lemma 10. The upper bound is obtained using words $(x, y) \in \mathcal{S}_{p,q,n}$ with string periods $a^{p-1}b$ and $(a^{p-1}b)^k a^r$ where $q = kp + r$. Indeed, x and y agree on the first q positions. After that, inside each pair of corresponding fragments of length at most p they have at most 2 mismatches. $\square$

Remark 12. In general, it is not true that $\mathcal{D}_{p,q}(n) \geq \lceil \frac{n-q}{p} \rceil$. For example, words $x = cacbcacbcacbcacbcac$ and $y = cacbcacacbcacacbcac$ of length 19, with shortest periods $p = 4$ and $q = 6$, respectively, satisfy $dist(x, y) = 3 < \lceil \frac{19-6}{4} \rceil$.

Theorem 13. *If* $p < q$ *and* $p \nmid q$ *and* $n \geq p + q$, *then* $\mathcal{D}_{p,q}(n) \geq \left\lfloor \frac{2n}{p+q} \right\rfloor$.

Proof. We use the following claim:

Claim. If $p < q$ and $p \nmid q$, then

(a) $N_{p,q}(1) = q + p - 1$,
(b) $N_{p,q}(2) = q + 2p - \text{GCD}(p,q) - 1$.

Proof. Observe that $q + p \leq q + 2p - \text{GCD}(p,q) \leq q + \lceil \frac{q}{p} \rceil p - 1$, so the values below are within the scope of Theorem 7. We have:

$$\mathcal{D}_{p,q}(q+p-1) = \left\lfloor \tfrac{p-1}{p} \right\rfloor + \left\lfloor \tfrac{p+\text{GCD}(p,q)-1}{p} \right\rfloor = 0 + 1 = 1$$
$$\mathcal{D}_{p,q}(q+p) = \left\lfloor \tfrac{p}{p} \right\rfloor + \left\lfloor \tfrac{p+\text{GCD}(p,q)}{p} \right\rfloor = 1 + 1 = 2.$$

This concludes the proof of part (a).

$$\mathcal{D}_{p,q}(q+2p-\text{GCD}(p,q)-1) = \left\lfloor \tfrac{2p-\text{GCD}(p,q)-1}{p} \right\rfloor + \left\lfloor \tfrac{2p-1}{p} \right\rfloor = 1 + 1 = 2$$
$$\mathcal{D}_{p,q}(q+2p-\text{GCD}(p,q)) = \left\lfloor \tfrac{2p-\text{GCD}(p,q)}{p} \right\rfloor + \left\lfloor \tfrac{2p}{p} \right\rfloor = 1 + 2 = 3.$$

This concludes the proof of part (b). □

Consider words $x \in \mathcal{P}_p(n)$, $y \in \mathcal{P}_q(n)$ and their $(p+q)$-decompositions: $\alpha_1, \ldots, \alpha_k$ and $\beta_0, \ldots, \beta_{k-1}$. If $\alpha_i = \beta_i$ for some $1 \leq i \leq k$, then, by the Periodicity Lemma, both α_i, β_i have period $\text{GCD}(p,q)$; consequently, both x and y have period $\text{GCD}(p,q)$, a contradiction. Hence, part (a) of the claim implies $\mathit{dist}(\alpha_i, \beta_i) \geq 2$ for each $i = 1, \ldots, k$.

Let α_{k+1} and β_{k+1} be the suffixes of x and y starting immediately after the last factors of the corresponding decompositions. If $|\alpha_{k+1}| < \frac{p+q}{2}$, then we already have that

$$\mathit{dist}(x,y) \geq 2 \left\lfloor \tfrac{n}{p+q} \right\rfloor = \left\lfloor \tfrac{2n}{p+q} \right\rfloor.$$

Otherwise, by part (b) of the claim applied for the words $\alpha_k\alpha_{k+1}$ and $\beta_k\beta_{k+1}$, we have

$$\mathit{dist}(x,y) \geq 2 \left(\left\lfloor \tfrac{n}{p+q} \right\rfloor - 1 \right) + 3 = 2 \left\lfloor \tfrac{n}{p+q} \right\rfloor + 1 = \left\lfloor \tfrac{2n}{p+q} \right\rfloor.$$

In both cases we obtain the desired inequality. □

5 Conclusions

The paper studies the following general type of question:

> How much dissimilar in a whole should be two objects which are different in some specific aspect?

The answer to this type of question heavily depends on the studied type of the objects. Thus sometimes the answer is completely trivial; for example, two different strings of the same length may differ at only a single position. In this work we show that if we consider different strings of the same length that are additionally periodic, then the implied number of positions where the two strings must differ can be large. The exact number depends on the length of the strings and on their periods.

References

1. Amir, A., Iliopoulos, C.S., Radoszewski, J.: Two strings at Hamming distance 1 cannot be both quasiperiodic. Inf. Process. Lett. **128**, 54–57 (2017)
2. Amir, A., Levy, A., Lubin, R., Porat, E.: Approximate cover of strings. In: Kärkkäinen, J., Radoszewski, J., Rytter, W. (eds.) 28th Annual Symposium on Combinatorial Pattern Matching, CPM 2017, 4–6 July 2017, Warsaw, Poland, vol. 78 of LIPIcs, pp. 26:1–26:14. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik (2017)
3. Bai, H., Franek, F., Smyth, W.F.: The new periodicity lemma revisited. Discrete Appl. Math. **212**, 30–36 (2016)
4. Berstel, J., Boasson, L.: Partial words and a theorem of Fine and Wilf. Theor. Comput. Sci. **218**(1), 135–141 (1999)
5. Blanchet-Sadri, F., Bal, D., Sisodia, G.: Graph connectivity, partial words, and a theorem of Fine and Wilf. Inf. Comput. **206**(5), 676–693 (2008)
6. Blanchet-Sadri, F., Corcoran, K., Nyberg, J.: Fine and Wilf's periodicity result on partial words and consequences. In: Loos, R., Fazekas, S.Z., Martín-Vide, C., (eds.) Proceedings of the 1st International Conference on Language and Automata Theory and Applications, LATA 2007, Report 35/07, pp. 115–126. Research Group on Mathematical Linguistics, Universitat Rovira i Virgili, Tarragona (2007)
7. Blanchet-Sadri, F., Hegstrom, R.A.: Partial words and a theorem of Fine and Wilf revisited. Theor. Comput. Sci. **270**(1–2), 401–419 (2002)
8. Blanchet-Sadri, F., Oey, T., Rankin, T.D.: Fine and Wilf's theorem for partial words with arbitrarily many weak periods. Int. J. Found. Comput. Sci. **21**(5), 705–722 (2010)
9. Blanchet-Sadri, F., Simmons, S., Tebbe, A., Veprauskas, A.: Abelian periods, partial words, and an extension of a theorem of Fine and Wilf. RAIRO - Theor. Inf. Appl. **47**(3), 215–234 (2013)
10. Castelli, M.G., Mignosi, F., Restivo, A.: Fine and Wilf's theorem for three periods and a generalization of Sturmian words. Theor. Comput. Sci. **218**(1), 83–94 (1999)
11. Constantinescu, S., Ilie, L.: Generalised Fine and Wilf's theorem for arbitrary number of periods. Theor. Comput. Sci. **339**(1), 49–60 (2005)
12. Constantinescu, S., Ilie, L.: Fine and Wilf's theorem for Abelian periods. Bull. EATCS **89**, 167–170 (2006)
13. Fan, K., Puglisi, S.J., Smyth, W.F., Turpin, A.: A new periodicity lemma. SIAM J. Discrete Math. **20**(3), 656–668 (2006)
14. Fine, N.J., Wilf, H.S.: Uniqueness theorems for periodic functions. Proc. Am. Math. Soc. **16**(1), 109–114 (1965)
15. Giancarlo, R., Mignosi, F.: Generalizations of the periodicity theorem of Fine and Wilf. In: Tison, S. (ed.) CAAP 1994. LNCS, vol. 787, pp. 130–141. Springer, Heidelberg (1994). https://doi.org/10.1007/BFb0017478
16. Holub, S.: On multiperiodic words. ITA **40**(4), 583–591 (2006)
17. Justin, J.: On a paper by Castelli, Mignosi, Restivo. ITA **34**(5), 373–377 (2000)
18. Karhumäki, J., Puzynina, S., Saarela, A.: Fine and Wilf's theorem for k-Abelian periods. Int. J. Found. Comput. Sci. **24**(7), 1135–1152 (2013)
19. Manea, F., Mercaş, R., Nowotka, D.: Fine and Wilf's theorem and pseudo-repetitions. In: Rovan, B., Sassone, V., Widmayer, P. (eds.) MFCS 2012. LNCS, vol. 7464, pp. 668–680. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-32589-2_58

20. Mignosi, F., Restivo, A., Silva, P.V.: On Fine and Wilf's theorem for bidimensional words. Theor. Comput. Sci. **292**(1), 245–262 (2003)
21. Mignosi, F., Shallit, J., Wang, M.: Variations on a theorem of Fine & Wilf. In: Sgall, J., Pultr, A., Kolman, P. (eds.) MFCS 2001. LNCS, vol. 2136, pp. 512–523. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-44683-4_45
22. Smyth, W.F., Wang, S.: A new approach to the periodicity lemma on strings with holes. Theor. Comput. Sci. **410**(43), 4295–4302 (2009)
23. Tijdeman, R., Zamboni, L.Q.: Fine and Wilf words for any periods II. Theor. Comput. Sci. **410**(30–32), 3027–3034 (2009)