Binary morphisms with stable suffix complexity

Štěpán Holub

Abstract

Let g, h be marked morphisms of words. A pair (g_1, h_1) of marked morphisms is called a successor of (g, h) if $g \circ g_1(a) = h \circ h_1(a)$ and $g \circ g_1(b) = h \circ h_1(b)$ and the images of g_1 and h_1 are shortest possible. Successors play an important role in studying the Post Correspondence Problem. Typically, they are simpler than the original morphisms, measured by the number of suffixes of their images (called the *suffix complexity*). In some cases, however, the suffix complexity is stable – it does not decrease. In this paper we study the binary case, that is, morphisms defined on a two-letter alphabet. We give a full characterization of binary morphisms with stable suffix complexity.

It is a surprising fact that solutions w of the equation $g(w) = h(w)$, where $g, h : \Sigma^* \to \Delta^*$ are morphisms of a free monoid, are poorly understood even if the cardinality of Σ is two (*i.e.* in the binary case). The situation is easy only in some special cases, for example when at least one of the morphisms is periodic. The Post Correspondence Problem (PCP), which asks whether the equation has a nonempty solution, is known to be undecidable in general, but decidable in the binary case. The first proof of decidability of the binary case was announced in [1], however, it contains a gap. A complete proof was given in [4], and [5] shows that the decision process is of polynomial time. The structure of the set of all solutions is known for some binary cases. However, even in the binary case, there is still no description of general solutions, in particular, there is no efficient algorithm known that decides whether a given word can be a solution of the above equation. For more details and references see [6, 7, 2].

If a single reason for the complications in the above mentioned research should be pointed out, it could be the existence of so called *successor morphisms* (called "equality collectors" in [1]). Studying a pair of binary morphisms, sooner or later one may discover that the existence of a solution, as well as its structure, depends essentially on the same question for a different pair of morphisms. Turning the attention to the new pair (the successors) the situation may occur again. The crucial question is whether the resulting chain of reductions can be kept under control. This can be done using so called suffix complexity of morphisms, which in most cases decreases; the concept was first introduced in [1]. However, there are some cases in which the suffix complexity is stable. These cases were once again studied in [1] where a characterization is given, which is sufficient for a proof that the binary PCP is decidable. In this paper we give a complete classification of stable instances for both balanced and unbalanced binary morphisms. One of the crucial advantages of the binary alphabet is that it is enough to consider marked morphisms only, which makes the exposition simpler. This makes no real harm to generality, since each binary morphism has its "marked version", which is, for the purposes of the PCP, equivalent to the original one (for more detailed explanation of marked versions, see the literature mentioned above). Moreover, even if the original morphisms are not marked, already the first "generation" of successors is marked always.

1 Preliminaries

We shall mostly use only elementary notation of combinatorics on words, see for example [8].

The empty word is denoted by ε. The first and the last letter of the word u is denoted by $\mathrm{pref}_1(u)$ and $\mathrm{suff}_1(u)$ respectively. We write $u \leq v$ to denote that u is a prefix of the word v. When we say that two words u and v are *comparable*, we refer to the prefix ordering. Analogously, we say that two words are *suffix-comparable*, if one of them is a suffix of the other. In this paper we define simple languages by regular expressions, for instance $(ab)^+a$. We say that u is a prefix of a language L if it is a prefix of at least one word in L. Therefore, if we write for example $u \leq (ab)^+a$, we mean that u is a prefix of $(ab)^i a$ for some $i \geq 1$.

A morphism g is called *marked* if $\mathrm{pref}_1(g(a)) \neq \mathrm{pref}_1(g(b))$ for any two distinct letters a and b. It is length preserving if $|g(u)| = |u|$ for all u.

2 Successors and suffix complexity

Let $g, h : \{a,b\}^* \to \Delta^*$ be two marked binary morphisms. The markedness in particular implies that none of the images is empty. Since each alphabet can be encoded into $\{a,b\}$, it is convenient to suppose that also $\Delta = \{a,b\}$.

A prefix-minimal pair of words (e,f) satisfying

$$g(e) = h(f) \tag{1}$$

is called a *block* of (g,h). (By "prefix-minimal" we mean that if nonempty prefixes e' of e and f' of f satisfy $g(e') = h(f')$, then $e' = e$ and $f' = f$.)

Since the morphisms are marked, it is not difficult to see that for each nonempty word $p \in \Delta^+$ there is at most one prefix-minimal pair of words (u,v) such that

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

Indeed, the continuation of u or of v is always uniquely determined by the overflow of already constructed prefix of (u,v).

Example 1. As an example, let $p = a$, and

$$\begin{aligned} g(a) &= aa, & g(b) &= bb, \\ h(a) &= ab, & h(b) &= baa. \end{aligned}$$

Obviously, if $ag(u) = h(v)$, then $\mathrm{pref}_1(v) = a$ since the first letter of $h(b)$ is b, not a. Let $v = av'$. We have $g(u) = bh(v)$, which again implies that $\mathrm{pref}_1(v) = b$. After four similar steps we see that the unique minimal pair (u, v) satisfying $ag(u) = h(v)$ is (ba, ab).

Similarly, for each nonempty word $q \in \Delta^+$ there is at most one prefix-minimal pair of words (u, v) such that

$$g(u) = qh(v).$$

This, in particular, implies that there are at most two distinct blocks. Moreover, if two distinct blocks (e, f) and (e', f') exist, then $\mathrm{pref}_1(e) \neq \mathrm{pref}_1(e')$ and $\mathrm{pref}_1(f) \neq \mathrm{pref}_1(f')$. In such a case we define

$$\begin{aligned} g_1(a) &= e, & g_1(b) &= e', \\ h_1(a) &= f, & h_1(b) &= f', \end{aligned}$$

and say that the pair (g_1, h_1) is a *successor* of (g, h). Note that both morphisms g_1 and h_1 are marked.

The *suffix complexity* $\sigma(g)$ of a morphism g is in [1] defined as the number of different non-empty words occurring as suffixes of an image of g. Formally,

$$\sigma(g) = \mathrm{Card}\{x \mid x \neq \varepsilon;\ x \in \mathrm{suff}(g(a)) \cup \mathrm{suff}(g(b))\}.$$

Let $\mathcal{O}_g$ be the set of all suffixes of $g(a)$ and $g(b)$ that occur as *overflows* in at least one block. More precisely, $u \in \mathcal{O}_g$ if and only if u is a nonempty suffix of $g(a)$ or $g(b)$, and there are (possibly empty) words e_1, f_1, e_2 and f_2 such that

$$g(e_1) = h(f_1)u \qquad \text{and} \qquad ug(e_2) = h(f_2). \tag{2}$$

As explained above, each $u \in \mathcal{O}_g$ determines uniquely a word $f_2 \in \mathrm{suff}(f) \cup \mathrm{suff}(f')$ such that $ug(e_2) = h(f_2)$ for some e_2. That defines a mapping

$$\pi_g \colon \mathcal{O}_g \to \mathrm{suff}(f) \cup \mathrm{suff}(f').$$

It should be clear that π_g is surjective: for each non-empty word v from $\mathrm{suff}(f) \cup \mathrm{suff}(f')$ there is a corresponding overflow u, which is in $\mathcal{O}_g$; in fact, this is how the set $\mathcal{O}_g$ was defined. Let, for example, $v \in \mathrm{suff}(f)$. Then

$$u = g(e_1)(h(fv^{-1}))^{-1},$$

where e_1 is the minimal prefix of e satisfying $|g(e_1)| > |h(fv^{-1})|$.

This immediately implies

$$\sigma(h_1) \leq \mathrm{Card}(\mathcal{O}_g) \leq \sigma(g). \tag{3}$$

Moreover, the equality holds if and only if

- each suffix of $g(a)$ and $g(b)$ is in $\mathcal{O}_g$, and

- the mapping π_g is injective.

Remark 1. Later we shall mainly use the first of the two above conditions. Injectivity, however, is not guaranteed, as the following example shows:

$$\begin{aligned} g(a) &= abb \qquad & g(b) &= b \\ h(a) &= a \qquad & h(b) &= bb. \end{aligned}$$

Here $\mathcal{O}_g = \mathrm{suff}(g(a)) \cup \mathrm{suff}(g(b))$, but both overflows b and bb map to $b \in \mathrm{suff}(h_1(b))$.

Similarly we define $\mathcal{O}_h$, and obtain

$$\sigma(g_1) \leq \mathrm{Card}(\mathcal{O}_h) \leq \sigma(h). \tag{4}$$

We say that a pair of morphisms (g, h) has *stable suffix complexity* if (g, h) has a successor (g_1, h_1) such that $\sigma(g) = \sigma(h_1)$ and $\sigma(h) = \sigma(g_1)$.

3 Stable suffix complexity

In this section we give a complete characterization of marked binary morphisms with stable suffix complexity. In the sequel we let (g, h), written also as (g_0, h_0), denote a pair of marked binary morphisms, and (g_{i+1}, h_{i+1}), with $i \geq 0$, will denote the successor of (g_i, h_i).

We let g_a denote either $g(a)$ or $g(b)$ so that $\mathrm{pref}_1(g_a) = a$. Similarly we define g_b, h_a and h_b. The symmetry allows us to adopt the convention that g_a is always equal to $g(a)$, and g_b to $g(b)$. For the morphism h, however, there remain two different possibilities: either $\mathrm{pref}_1(h(a)) = a$ and $\mathrm{pref}_1(h(b)) = b$, or vice versa. We shall often cover both cases by using variables x and y, and shall write $h(x) = h_a$ and $h(y) = h_b$, where $\{x, y\} = \{a, b\}$.

It is natural to identify cases that only differ by exchanging morphisms g and h and/or letters a and b. Let therefore $\mu : \{a, b\}^* \to \{a, b\}^*$ denote the morphism defined by $a \mapsto b$ and $b \mapsto a$. We denote the identity morphism on $\{a, b\}^*$ by id. We say that two pairs of morphisms (g, h) and (g', h') are *symmetric* if $\{g, h\} = \{\mu_1 \circ g' \circ \mu_2, \mu_1 \circ h' \circ \mu_2\}$ with $\mu_1, \mu_2 \in \{\mu, \mathrm{id}\}$. Note also that $\{\mu, \mathrm{id}\}$ are the only length preserving morphisms.

We start with a simple lemma, which follows from the fact that each suffix is an overflow.

Lemma 1. *Assume that the pair (g, h) has stable suffix complexity. Let u be a suffix of g_a or g_b. If c is the first letter of u, then u is comparable with h_c. Moreover, there is a prefix v of g_a or g_b such that $vu \in \{g_a, g_b\}$ and v is suffix-comparable with h_a or h_b.*

The lemma holds also if we exchange g and h.

Next lemma is a slightly more advanced tool for studying the stable suffix complexity. Once more, it has variants given by symmetries of morphisms g and h, as well as letters a and b.

Lemma 2. *Let (g,h) have stable suffix complexity. If*

$$\mathrm{suff}_1(h_a) = \mathrm{suff}_1(h_b) = a, \tag{5}$$

then $g_a \in a^+$ and $g_b \in ba^*$.

Proof. Suppose $\mathrm{suff}_1(h_a) = \mathrm{suff}_1(h_b) = a$. Let k be the largest positive integer such that ba^k is a suffix of g_a or g_b. Since $a^k \in \mathcal{O}_g$, there is a (possibly empty) word v such that $va^k \in \{g_a, g_b\}$ and v is suffix-comparable with h_a or h_b, by Lemma 1. If v is nonempty, the assumption (5) yields that a is its suffix, a contradiction with the maximality of k. Hence v is empty, and $g_a \in a^+$.

Suppose that $g_b = bu$ and $u \notin a^*$. Since $u \in \mathcal{O}_g$ and u is not a suffix of g_a, we deduce that there are words e_1, f_1 such that $g(e_1 b) = h(f_1)u$, which implies $g(e_1)b = h(f_1)$. This is a contradiction with the assumption about the last letters of h_a and h_b. □

We can now state our main result.

Theorem 1. *Let (g,h) be a pair of marked binary morphisms. The suffix complexity of (g,h) is stable if and only if (at least) one of the following cases takes place (up to symmetry):*

(Case 1) *At least one of the morphisms is length preserving.*

(Case 2) *There are integers $j,k,\ell,m \geq 1$ such that*

$$g_a = a^j, \qquad g_b = b^k,$$
$$h_a = a^\ell, \qquad h_b = b^m,$$

and $\gcd(j,\ell) = \gcd(k,m) = 1.$

(Case 3) *There are integers $j,\ell \geq 0$ and integers $k,m \geq 1$ such that*

$$g_a = ab^j, \qquad g_b = b^k,$$
$$h_a = ab^\ell, \qquad h_b = b^m,$$

and $\gcd(k,m) = 1,$ $j \leq k$ *and* $\ell \leq m.$

(Case 4) *There are integers $j,k \geq 1$ and integers $\ell,m \geq 0$ such that*

$$g_a = (ab)^j, \qquad g_b = (ba)^k,$$
$$h_a = (ab)^\ell a, \qquad h_b = (ba)^m b,$$

with $\gcd(j,\ell+m+1) = 1$ *and* $\gcd(k,\ell+m+1) = 1.$

(Case 5) *There are integers $j,k,\ell,m \geq 0$ such that*

$$g_a = (ab)^j a, \qquad g_b = (ba)^k b,$$
$$h_a = (ab)^\ell a, \qquad h_b = (ba)^m b,$$

and *a)* *either* $\gcd(j+k+1,\ell+m+1) = 1$
b) *or* $\gcd(j+k+1,\ell+m+1) = 2$ *and* $j-\ell$ *is odd.*

(Case 6) *There are integers* $t \geq 2$ *and* $k, m \geq 0$ *such that*

$$\begin{aligned} g_a &= a, & g_b &= (ba^t)^k b, \\ h_a &= a, & h_b &= (ba^t)^m b, \end{aligned}$$

and $\gcd(k+1, m+1) = 1.$

(Case 7) *There are integers* $t, k, \ell, m \geq 1$ *such that*

$$\begin{aligned} g_a &= a, & g_b &= b^k, \\ h_a &= a^t b^\ell, & h_b &= b^m, \end{aligned}$$

$\ell \leq m$ *and* $\gcd(k, m) = 1.$

Before starting the proof, let us make the following remark. The proof is quite concise, leaving many claims to be verified by the reader. However, each such verification should be straightforward, even if possibly laborious. In most cases, such a claim can be verified by a direct construction similar to the one in Example 1. The leading and ubiquitous idea, hidden behind references to Lemma 1 but also used implicitly, is the fact that every suffix has to be an overflow. For example, if *bba* is a suffix of $g(a)$ or $g(b)$, then the word h_b has to start with *bb* (if its length is not one), otherwise *bba* cannot be an overflow, and the suffix complexity cannot be stable. Also claims counting suffix complexities are laconic since their verification is direct. An important part of the proof validity is the completeness of the case division. Even in this regard, the reader has to check carefully that the claims invoking the symmetry are correct. However, the author tried hard to make the classification as tabular as possible.

Proof. The proof is a case analysis. The main criterion of the division in subcases is the number of words of length one among g_a, g_b, h_a, h_b.

If one of the morphisms is length preserving, it is easy to see that the suffix complexity is stable (see Example 2). Let therefore at least one of the images of each of the two morphisms have length at least two.

1. Suppose, first, that all four words g_a, g_b, h_a, h_b have length at least two.

1.1. Let $aa \leq g_a$ and $bb \leq g_b$. Repeated application of Lemma 1 yields that $g_a, h_a \in a^+$ and $g_b, h_b \in b^+$. Let $h_a = h(x)$ and $h_b = h(y)$. Let

$$\begin{aligned} g(a) &= a^j, & g(b) &= b^k, \\ h(x) &= a^\ell, & h(y) &= b^m. \end{aligned}$$

The suffix complexities are $\sigma(g) = j + k$ and $\sigma(h) = \ell + m$. It easy to verify that

$$\begin{aligned} g_1(a) &= a^{\frac{\ell}{\gcd(j,\ell)}}, & g_1(b) &= b^{\frac{m}{\gcd(k,m)}}, \\ h_1(a) &= x^{\frac{j}{\gcd(j,\ell)}}, & h_1(b) &= y^{\frac{k}{\gcd(k,m)}}. \end{aligned}$$

Therefore the suffix complexity of (g,h) is stable if and only if $\gcd(j,\ell) = \gcd(k,m) = 1$.

1.2. Let $ab \leq g_a$ and $bb \leq g_b$. From Lemma 1 we deduce $g_b, h_b \in b^+$, and $g_a, h_a \in ab^+$. Let again $h_a = h(x)$ and $h_b = h(y)$, and let

$$g(a) = ab^j, \qquad g(b) = b^k,$$
$$h(x) = ab^\ell, \qquad h(y) = b^m.$$

Since the suffix complexity of (g,h) is stable, we have $b \in \mathcal{O}_g$. It is not difficult to see that words u and v satisfying $bg(u) = h(v)$ exist if and only if k and m are coprime.

A direct construction of the successor then yields

$$g_1(a) = ab^{j'}, \qquad g_1(b) = b^m,$$
$$h_1(a) = xy^{\ell'}, \qquad h_1(b) = y^k,$$

where j', ℓ' are the smallest nonnegative integers satisfying

$$j + j'k = \ell + \ell' m.$$

Let $j > k$. Then $\sigma(g) = j+1$; whence $\sigma(h_1) = \max\{\ell', k\} + 1$ implies $\ell' = j$. From the equality $j + j'k = \ell + jm$ we deduce $j' \geq m$. However, now also nonnegative integers $j' - m$ and $j - k$ satisfy

$$j + (j' - m) = \ell + (j - k)m,$$

a contradiction with the minimality of j' and ℓ'. Therefore $j \leq k$. Similarly we conclude that $\ell \leq m$, which yields the conditions of Case 3. The suffix complexity is stable, since $\sigma(g) = \sigma(h_1) = k+1$ and $\sigma(h) = \sigma(g_1) = m+1$.

1.3. Let $ab \leq g_a$ and $ba \leq g_b$. Lemma 1 implies $g_a, h_a \leq (ab)^+$ and $g_b, h_b \leq (ba)^+$. Lemma 2 yields $\mathrm{suff}_1(h_a) \neq \mathrm{suff}_1(h_b)$ and $\mathrm{suff}_1(g_a) \neq \mathrm{suff}_1(g_b)$. Morphisms of the form

$$g(a) = (ab)^j, \qquad g(b) = (ba)^k,$$
$$h(x) = (ab)^\ell, \qquad h(y) = (ba)^m$$

do not have stable suffix complexity. This can be seen from the fact that suffixes of odd length cannot be overflows. Therefore, up to symmetry, we have in this subcase two possibilities.

1.3.1. Suppose, first, that there are integers j, k, ℓ, m such that

$$g(a) = (ab)^j, \qquad g(b) = (ba)^k,$$
$$h(x) = (ab)^\ell a, \qquad h(y) = (ba)^m b.$$

Then we obtain, by a direct inspection of blocks,

$$g_1(a) = a^{\frac{\ell+m+1}{\gcd(j,\ell+m+1)}}, \qquad g_1(b) = b^{\frac{\ell+m+1}{\gcd(k,\ell+m+1)}},$$
$$h_1(a) = (xy)^{\frac{j}{\gcd(j,\ell+m+1)}}, \qquad h_1(b) = (yx)^{\frac{k}{\gcd(k,\ell+m+1)}}.$$

Since $\sigma(g) = 2(j+k)$ and $\sigma(h) = 2(\ell+m+1)$ while

$$\sigma(g_1) = \frac{\ell+m+1}{\gcd(j,\ell+m+1)} + \frac{\ell+m+1}{\gcd(k,\ell+m+1)},$$

and

$$\sigma(h_1) = \frac{2j}{\gcd(j,\ell+m+1)} + \frac{2k}{\gcd(k,\ell+m+1)},$$

it follows that the suffix complexity is stable if and only if $\gcd(j,\ell+m+1) = \gcd(k,\ell+m+1) = 1$.

1.3.2. The second possibility is

$$\begin{aligned} g(a) &= (ab)^j a, & g(b) &= (ba)^k b, \\ h(x) &= (ab)^\ell a, & h(y) &= (ba)^m b. \end{aligned}$$

for some j,k,ℓ,m. Put

$$r = \frac{\ell+m+1}{\gcd(j+k+1,\ell+m+1)}, \qquad s = \frac{j+k+1}{\gcd(j+k+1,\ell+m+1)}.$$

It is not difficult to verify that r and s are the smallest positive integers such that $g((ab)^r) = h((xy)^s)$. Also, they are the smallest positive integers satisfying $g((ba)^r) = h((yx)^s)$.

Recall that there is at most one block (e,f) with $\mathrm{pref}_1(e) = a$ and at most one block (e',f') with $\mathrm{pref}_1(e') = b$. We deduce that the pairs $((ab)^r,(xy)^s)$ and $((ba)^r,(yx)^s)$ are either blocks, or they both split into blocks. The first possibility yields

$$\begin{aligned} g_1(a) &= (ab)^r, & g_1(b) &= (ba)^r, \\ h_1(a) &= (xy)^s, & h_1(b) &= (yx)^s. \end{aligned} \tag{6}$$

The second possibility implies that there are nonnegative integers j',k',ℓ',m' such that

$$\begin{aligned} g_1(a) &= (ab)^{j'} a, & g_1(b) &= (ba)^{k'} b, \\ h_1(a) &= (xy)^{\ell'} x, & h_1(b) &= (yx)^{m'} y. \end{aligned} \tag{7}$$

Since

$$g \circ g_1(ab) = g((ab)^{j'} ab(ab)^{k'}) = h((xy)^{\ell'} xy(xy)^{m'}) = h \circ h_1(ab),$$

the minimality of r and s implies $j'+k'+1 = r$ and $\ell'+m'+1 = s$. It is straightforward to verify that integers j' and ℓ' are (minimal) solutions of the equation

$$j'(j+k+1) + j = \ell'(\ell+m+1) + \ell. \tag{8}$$

The suffix complexity of g is $2(j+k+1)$. The suffix complexity of h_1 is either $4s$ (if the possibility (6) holds), or $2s$ (if the case (7) takes place). Since we are interested in the stable suffix complexity, we can suppose $\gcd(j+k+1, \ell+m+1) \leq 2$.

Let, first, the numbers $j+k+1$ and $\ell+m+1$ be coprime. Basic modular arithmetic implies that equality (8) has nonnegative solutions $j' < \ell+m+1$ and $\ell' < j+k+1$, and we have the case (7). It is straightforward to verify that this already yields stable suffix complexity and the first option of Case 5.

Let now $\gcd(j+k+1, \ell+m+1) = 2$. In this case, the stable suffix complexity is achieved if and only if (8) has no solution, which occurs if and only if $j-\ell$ is odd. This is the second option of Case 5.

Up to symmetry we have checked all cases in which all four words have length at least two.

2. Let now $g_a = a$, and the remaining three words have length at least two.
2.1. Suppose that $bb \leq g_b$. Then Lemma 1 implies that

$$\begin{aligned} g_a &= a, & g_b &= b^k, \\ h_a &= a^j b^\ell, & h_b &= b^m, \end{aligned}$$

for some $j \geq 1$, $k, m \geq 2$ and $j+\ell \geq 2$. If $\ell = 0$, then we obtain a situation leading to Case 2. If $\ell > 0$, then we can use similar argumentation as in **1.2.** The suffix complexity is stable if and only if $\gcd(k,m) = 1$ and $\ell \leq m$, which is Case 7.

2.2. Let $ba \leq g_b$ (and thus also $ba \leq h_b$).
2.2.1. If $aa \leq h_a$, then $g_b \in ba^+$, by Lemma 1, and Lemma 2 implies $h_a \in a^+$ and $h_b \in ba^+$. We have a situation symmetric to Case 3.
2.2.2. Let $ab \leq h_a$. Lemma 1 implies that $g_b \leq (ba)^+$, and Lemma 2 yields that $g_b \in (ba)^+b$. In particular, $bab \leq g_b$, which implies that neither h_a, nor h_b contain baa as a factor. Similarly they do not contain the factor bb. Therefore $g_a, h_a \leq (ab)^+$ and $g_b, h_b \leq (ba)^+$, which leaves us with Case 4 or Case 5, as shown above.

3. Suppose that $g_a = h_a = a$. This implies that g_b and h_b have length at least two, since we already suppose that the morphisms are not length preserving.
3.1. If $bb \leq g_b$ or $bb \leq h_b$, then Lemma 1 implies Case 2. Moreover, if a is the last letter of g_b or h_b, then, up to symmetry, we have Case 3, by Lemma 2.
3.2. Let therefore $ba \leq g_b$ and $ba \leq h_b$, and let b be the last letter of both g_b and h_b. Lemma 1 now implies that bb is not a factor of g_b or h_b and we have

$$g_b = ba^{k_1}ba^{k_2}\dots ba^{k_m}b, \quad \text{and} \quad h_b = ba^{\ell_1}ba^{\ell_2}\dots ba^{\ell_{m'}}b.$$

Lemma 1 further implies that $k_i = \ell_1$ for each $i = 1, \dots, m$, and $\ell_j = k_1$ for each $j = 1, \dots, m'$. Therefore we obtain

$$\begin{aligned} g(a) &= a, & g(b) &= (ba^t)^k b, \\ h(x) &= a, & h(y) &= (ba^t)^m b. \end{aligned}$$

An analysis analogous to the one carried on in **1.3.2.** yields that the suffix complexity is stable if and only if $\gcd(k+1, m+1) = 1$, and we have Case 6.
4. It remains to consider the case $g_a = a$ and $h_b = b$.
4.1. Let $bb \leq g_b$. Then $h_a = a^j b^\ell$, by Lemma 1.
4.1.1. If $\ell = 0$, then $aa \leq h_a$, and $g_b = b^t a^m$. We have a situation symmetric to Case 7 treated in **2.1.**
4.1.2. If $\ell \geq 1$, then Lemma 2 implies $g_b \in b^+$, which yields the Case 7 again.
4.2. By symmetry, we can now suppose $ba \leq g_b$ and $ab \leq h_a$. If $g_b = ba$ or $h_a = ab$, then we have a contradiction with Lemma 2. Therefore $bab \leq g_b$ and $aba \leq h_a$. Lemma 1 and Lemma 2 now yield that we have Case 5.
The proof is complete. □

The content of the previous theorem is concisely given in Table 1. Note that in some cases the successor morphisms have again stable suffix complexity. The following example shows a simple situation in which there exists an infinite sequence of successors, all of them having the same suffix complexity.

Example 2. Consider the Case 1 with $g = \mathrm{id}$. There are two possibilities according to the choice of x and y. Suppose the more complex situation, namely, let $h(a) = h_b$ and $h(b) = h_a$. Let (g_i, h_i), $i = 1, 2, \ldots$ be the sequence of successors of (g, h). It is straightforward to verify that

$$\begin{aligned}
(g_1, h_1) &= (h \circ \mu, \mu),\\
(g_2, h_2) &= (\mathrm{id}, \mu \circ h \circ \mu),\\
(g_3, h_3) &= (\mu \circ h, \mu), \text{ and}\\
(g_4, h_4) &= (g, h).
\end{aligned}$$

It is obvious that all pairs have stable suffix complexity and the sequence is infinite with period four.

4 Balanced cases

In this section we focus on a special case of balanced morphisms.

By $|w|_\ell$, with $\ell \in \{a, b\}$, we shall denote the number of occurrences of letter ℓ in w. We say that a pair of morphisms (g, h) is *balanced* if for each $\rho(.) \in \{|.|, |.|_a, |.|_b\}$ either

$$\begin{aligned}
&\rho(g(a)) < \rho(h(a)), && \rho(g(b)) > \rho(h(b)), \text{ or}\\
&\rho(g(a)) > \rho(h(a)), && \rho(g(b)) < \rho(h(b)), \text{ or}\\
&\rho(g(a)) = \rho(h(a)), && \rho(g(b)) = \rho(h(b)).
\end{aligned}$$

Otherwise we say that the instance is *unbalanced.* We study balanced cases because they constitute the core of the difficulty in PCP. It is obvious that the equation $g(w) = h(w)$ has no nonempty solution in the unbalanced case. Let

Table 1: Morphisms with stable suffix complexity

	$g(a)$	$g(b)$	$g_1(a)$	$g_1(b)$
	$h(x)$	$h(y)$	$h_1(a)$	$h_1(b)$
Case 1	a	b	au	bv
	au	bv	x	y
Case 2	a^j	b^k	a^ℓ	b^m
	a^ℓ	b^m	x^j	y^k
Case 3	ab^j	b^k	$ab^{j'}$	b^m
	ab^ℓ	b^m	$xy^{\ell'}$	y^k
Case 4	$(ab)^j$	$(ba)^k$	$a^{\ell+m+1}$	$b^{\ell+m+1}$
	$(ab)^\ell a$	$(ba)^m b$	$(xy)^j$	$(yx)^k$
Case 5 a)	$(ab)^j a$	$(ba)^k b$	$(ab)^{j'} a$	$(ba)^{k'} b$
	$(ab)^\ell a$	$(ba)^m b$	$(xy)^{\ell'} x$	$(yx)^{m'} y$
Case 5 b)	$(ab)^j a$	$(ba)^k b$	$(ab)^{(\ell+m+1)/2}$	$(ba)^{(\ell+m+1)/2}$
	$(ab)^\ell a$	$(ba)^m b$	$(xy)^{(j+k+1)/2}$	$(yx)^{(j+k+1)/2}$
Case 6	a	$(ba^t)^k b$	a	$(ba^t)^m b$
	a	$(ba^t)^m b$	x	$(yx^t)^k y$
Case 7	a	b^k	$a^t b^{j'}$	b^m
	$a^t b^\ell$	b^m	$xy^{\ell'}$	y^k

us remark that the situation is slightly more complicated in so called *Generalized* PCP, but even in that problem the unbalanced cases can be solved quite straightforwardly (for more details on the Generalized PCP see for example [3])

The following result says that there is practically just one case in which a chain of successors with stable suffix complexity contains only balanced pairs.

Theorem 2. *Let (g_i, h_i), $i = 0, 1, 2$ be pairs of balanced morphisms such that (g_1, h_1) is a successor of (g_0, h_0), and (g_2, h_2) is a successor of (g_1, h_1). Suppose that all three morphisms have the same suffix complexity. Then either both g_1 and h_1 are length preserving, or*

$$\begin{aligned} g_0(a) &= a^j, & g_0(b) &= b^k, \\ h_0(a) &= b^m, & h_0(b) &= a^\ell, \end{aligned}$$

where j, k, ℓ, m are positive integers satisfying

$$\gcd(j, m) = \gcd(k, \ell) = \gcd(m, k) = \gcd(\ell, j) = 1\,.$$

Suppose, by symmetry, that $m = \min\{j, k, l, m\}$. Then $m < k < \ell$ and $m < j < \ell$. The sequence of successors $\{(g_i, h_i)\}_{i \in \mathbb{N}_0}$, where (g_{i+1}, h_{i+1}) denotes a successor of (g_i, h_i) for each $i \geq 0$, is well defined and has period four.

Proof. Theorem 1 yields a list of candidates for the pair (g_0, h_0). It remains to verify that the possibilities of the present theorem are the only ones in which the additional assumptions are satisfied. Namely, the successor (g_1, h_1) has to have stable suffix complexity, and all three pairs (g_i, h_i), $i = 0, 1, 2$ have to be balanced.

Consider the situation when g_1 and h_1 are both length preserving; in other words, we have $\sigma(g_1) = \sigma(h_1) = 2$. Therefore also $\sigma(g_0) = \sigma(h_0) = 2$. It is interesting to note that this does not imply that g_0 and h_0 are themselves length preserving; the morphism f defined by $f(a) = ab$ and $f(b) = b$ represents another possibility satisfying $\sigma(f) = 2$. We can therefore have for example $g_0 = f$ and $h_0 = f \circ \mu$. If g_1 and h_1 are both length preserving, then we have a situation similar to Example 2, the infinite sequence of successors is well defined, and has a period four.

Further we shall suppose that g_1 is not length preserving, which implies that nor h_1 is, since the pair (g_1, h_1) is balanced. We gradually investigate situations in which the pair (g_0, h_0) has one of the forms of Theorem 1. In each case we shall respect the notation used in Theorem 1, in its proof and in Table 1.

Case 1 This case is excluded by the above discussion.

Case 2 This case yields the possibility described in the theorem. Since (g_0, h_0) is balanced, we have $h_0(a) = b^\ell$ and $h_0(b) = a^m$. Moreover, $\gcd(j, \ell) = \gcd(k, m) = 1$ by the stability of the suffix complexity. A direct construction yields the following pairs (g_i, h_i), $i = 0, 1, 2, 3$:

	a	b
g_0	a^j	b^k
h_0	b^m	a^ℓ

$\mapsto$

	a	b
g_1	a^ℓ	b^m
h_1	b^j	a^k

$\mapsto$

	a	b
g_2	a^k	b^j
h_2	b^ℓ	a^m

$\mapsto$

	a	b
g_3	a^m	b^ℓ
h_3	b^k	a^j

and the pair (g_4, h_4) is again equal to (g_0, h_0). The stable suffix complexity of (g_1, h_1) implies $\gcd(j, m) = \gcd(k, \ell) = 1$. The inequalities $m < j < \ell$ and $m < k < \ell$ follow from the fact that the morphisms are balanced.

Case 3 If $j = \ell$, then also $k = m$, otherwise (g_1, h_1) is not balanced in $|.|_b$ or in $|.|_a$; then both g_1 and h_1 are length preserving. Suppose, therefore, by symmetry, that $j < \ell$, which implies $m < k$, and $j' > 0$.

If $x = b$, then the pair (g_1, h_1) is not balanced in $|.|_b$. Suppose therefore $x = a$. Since (g_1, h_1) is balanced, we have also $\ell' < j'$. Recall that j', ℓ' are the smallest nonnegative integers satisfying

$$j + j'k = \ell + \ell' m.$$

From $j' > \ell'$ we have

$$\ell = j + j'k - \ell' m \geq j + k + \ell'(k - m),$$

which implies $\ell > k$ and $\ell > m$ since we have $k > m$; a contradiction with conditions for Case 3 in Theorem 1.

Case 4 Since (g_0, h_0) is balanced, we have $\ell + m > 0$. It is obvious that (g_1, h_1) do not have successor.

Case 5 a) Without loss of generality (by symmetry) we can suppose $j < \ell$ and $k > m$ since (g_0, h_0) is balanced (and since $j = \ell$ and $k = m$ imply that (g_0, h_0) has not stable suffix complexity unless both morphisms are length preserving). Still without loss of generality we can suppose $j + k > \ell + m$, where $j + k = \ell + m > 0$ would again imply that the suffix complexity of (g_0, h_0) is not stable. We have

$$j'(j + k + 1) + j = \ell'(\ell + m + 1) + \ell, \tag{9}$$

$$k'(j + k + 1) + k = m'(\ell + m + 1) + m. \tag{10}$$

As above, we can suppose $m' > k'$ and $\ell' < j'$ since (g_1, h_1) is balanced and has stable suffix complexity (the case $m' < k'$ and $\ell' > j'$ is analogous). The equality (9) can be rewritten as

$$\ell - j = j'(j + k + 1) - \ell'(\ell + m + 1),$$

which is in contradiction with inequalities $j' > \ell'$ and $j + k > \ell + m$, as it is straightforward to see.

Case 5 b) It is easy to see that the pair (g_1, h_1) has not stable suffix complexity.

Case 6 The pair (g_0, h_0) is balanced only if $k = m$. Therefore, either both morphisms are length preserving, or the suffix complexity is not stable.

Case 7 If $t > 1$, then (g_0, h_0) is not balanced in $|.|_a$. If $t = 1$, then we have Case 3, already discussed above. □

References

[1] Andrzej Ehrenfeucht, Juhani Karhumäki, and Grzegorz Rozenberg. The (generalized) Post correspondence problem with lists consisting of two words is decidable. *Theoret. Comput. Sci.*, 21(2):119–144, 1982.

[2] Jana Hadravová and Štěpán Holub. Large simple binary equality words. In *DLT '08: Proceedings of the 12th international conference on Developments in Language Theory*, pages 396–407, Berlin, Heidelberg, 2008. Springer.

[3] Vesa Halava, Tero Harju, and Mika Hirvensalo. Generalized Post correspondence problem for marked morphisms. *Internat. J. Algebra Comput.*, 10(6):757–772, 2000.

[4] Vesa Halava, Tero Harju, and Mika Hirvensalo. Binary (generalized) Post correspondence problem. *Theoret. Comput. Sci.*, 276(1-2):183–204, 2002.

[5] Vesa Halava and Štěpán Holub. Reduction tree of the binary Generalized Post Correspondence Problem, journal = Int. J. Found. Comput. Sci., fjournal = International Journal of Foundations of Computer Science, volume = to appear,.

[6] Štěpán Holub. Binary equality sets are generated by two words. *J. Algebra*, 259(1):1–42, 2003.

[7] Štěpán Holub. A unique structure of two-generated binary equality sets. In *Developments in language theory*, volume 2450 of *Lecture Notes in Comput. Sci.*, pages 245–257. Springer, Berlin, 2003.

[8] M. Lothaire. *Combinatorics on words.* Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1997.