ACADEMIC PRESS

Available at WWW.MATHEMATICSWEB.ORG POWERED BY SCIENCE DIRECT®

Journal of Algebra 259 (2003) 1–42

JOURNAL OF Algebra

www.elsevier.com/locate/jalgebra

Binary equality sets are generated by two words

Štěpán Holub [1]

Matematicko-fyzikální fakulta, Univerzita Karlova, Sokolovská 83, 186 75 Praha 8, Czech Republic

Received 18 April 2000

Communicated by T.E. Hall

Abstract

We show that the equality set $\mathrm{Eq}(g,h)$ of two non-periodic binary morphisms $g,h:A^*\to\Sigma^*$ is generated by at most two words. If the rank of $\mathrm{Eq}(g,h)=\{\alpha,\beta\}^*$ is two, then α and β begin and end with different letters. This in particular implies that any binary language has a test set of cardinality at most two.

1. Introduction

Binary equality language, i.e. the set on which two binary morphisms agree, is the simplest non-trivial example of an equality language, the notion of which was introduced in [9]. Equality languages in general play an important role in formal language theory. For a survey and bibliography see [5, Section 5].

In the binary case the morphisms are defined on a monoid generated by two letters. It was for the first time extensively studied by K. Čulík II and J. Karhumäki in [6]. There the main claim of our work was conjectured, viz. that a binary equality language is generated by at most two words as soon as at least one of the morphisms is non-periodic (or, equivalently, injective). An important step towards the proof of the conjecture was made in [8] where the following partial characterization was obtained.

E-mail address: holub@karlin.mff.cuni.cz.
[1] Supported by Turku Centre for Computer Science.

0021-8693/02/$ – see front matter
PII: S0021-8693(02)00534-3

Theorem 1. *The equality set of two binary morphisms g and h has the following structure*:
(A) *If h and g are periodic, then either $E(h,g)=\{\varepsilon\}$ or*

$$E(h,g)=\{\varepsilon\}\cup\left\{\alpha\in A^+\;\middle|\;\frac{|\alpha|_a}{|\alpha|_b}=k\right\}$$

for some $k\geqslant 0$ or $k=\infty$.
(B) *If exactly one morphism is periodic, then*

$$E(h,g)=\alpha^*$$

for some word $\alpha\in\Sigma^*$.
(C) *If both g and h are non-periodic, then either*

$$E(h,g)=\{\alpha,\beta\}^*$$

for some words $\alpha,\beta\in\Sigma^*$, or

$$E(h,g)=(\alpha\gamma^*\beta)^*$$

for some words $\alpha,\beta,\gamma\in\Sigma^+$.

The question remained open whether the second possibility of case (C), contradicting the conjecture, can actually occur. In the present paper we show that the answer is negative and, moreover, if α and β are both non-empty, they start and end with different letters. This is formulated in

Theorem 2. *Let $g,h:A^*\to\Sigma^*$ be non-periodic binary morphisms.*
(A) *Let α and β, with $\alpha\neq\beta$, be non-empty minimal elements of* $\mathrm{Eq}(g,h)$. *Then*

$$\mathrm{pref}_1(\alpha)\neq\mathrm{pref}_1(\beta)\quad\textit{and}\quad\mathrm{suff}_1(\alpha)\neq\mathrm{suff}_1(\beta).$$

(B) $\mathrm{Eq}(g,h)$ *is generated by at most two words.*

Note that (B) is a trivial consequence of (A). Our proof does not deal directly with (B), but is focused on (A). We are not aware of any way how to prove (B) not using (A).

Remark. The case $g=h$ is trivial. Throughout the paper we shall implicitly suppose $g\neq h$.

Let us mention two problems closely related to the question about the structure of binary equality sets. The first one is the binary case of the famous Post Correspondence Problem, shortly PCP(2). The cohesion of the two problems is obvious, as PCP(2) consists in deciding, given two binary morphisms, whether their equality set is empty. The proof that the question is algorithmically decidable (see [2]) was one of the important moments in the development of theoretical computer science. A survey of recent results concerning PCP

can be found in [3]. It was especially shown in [4] that generalized PCP of arbitrary size is decidable in marked case.

The second problem akin to the structure of binary equality languages is the existence of a test set for binary languages. Indeed, if two morphisms agree on a language, it must be a subset of an equality language. In [8] it is shown that all binary languages have a test set of three elements. Our result allows to cut down this bound to two. Let us remark that this improvement is not a simple consequence of the fact that the equality language is generated by two words—the difference in the first (or last) letter is a necessary ingredient.

This paper has the following structure. In Section 2 some definitions and elementary combinatorial tools are given. In Section 3 we study basic properties of words on which two binary morphisms agree. In Section 4 we describe a typical case of a pair of binary morphisms to which all other cases can be reduced. The findings of Section 3 are applied in Section 5 to marked morphisms. The main result is proved in Section 6 divided into several subsections. Within that section the existence of an equality set not fitting Theorem 2 is gradually shown to be contradictory. Section 7 is dedicated to the test set of binary languages bounding its cardinality by two.

2. Preliminaries

We use the basic notation from [1,5]. By Σ we denote an arbitrary alphabet, by A the two-letter alphabet $\{a, b\}$. Σ^* is the free monoid and Σ^+ the free semigroup generated by Σ. The empty word is denoted by ε.

Expression $|u|$ represents the length of a word u, and $|u|_x$ the number of occurrences of the letter x in u. The set of all letters having at least one occurrence in the word u is denoted by $\mathrm{alph}(u)$.

A *prefix* of u is any word $v \in \Sigma^*$ such that there exists a word $v' \in \Sigma^*$ with $u = vv'$. The set of all prefixes of u is denoted by $\mathrm{pref}(u)$. A prefix v of u is *proper* if $v \neq \varepsilon$ and $v \neq u$. Similarly *suffix* and *proper suffix* are defined. The set of all suffixes of u is denoted by $\mathrm{suff}(u)$. The first (the last respectively) letter of a non-empty word u is also denoted by $\mathrm{pref}_1(u)$ ($\mathrm{suff}_1(u)$ respectively). A word v is called a *factor* of u if there exist words $w, w' \in \Sigma^*$ such that $u = wvw'$.

The positive powers u^n of a word are defined as usually, with $u^0 = \varepsilon$. We shall also use negative powers to simplify notation. If $uv = w$, we write $u = wv^{-1}$ and $v = u^{-1}w$. Obviously, u^{-n} is an abbreviation for $(u^n)^{-1}$.

The notions of prefix, suffix and factor can be extended to languages: a prefix (suffix, factor respectively) of a language is prefix (suffix, factor) of any of its elements. Accordingly,

$$\mathrm{pref}(L) = \bigcup_{u \in L} \mathrm{pref}(u).$$

Similarly for $\mathrm{suff}(L)$.

The language $\{u^i \mid i \in \mathbb{N}_+\}$ is denoted by u^+ and

$$u^* = u^+ \cup \{\varepsilon\}.$$

A word u is called *primitive* if and only if $u = v^n$ implies $u = v$. The *primitive root* of u is the (uniquely given) primitive word r such that $u \in r^+$. Words u and v are called *conjugates* if $u = ww'$ and $v = w'w$.

If we speak about minimality or maximality of some element, the implicit ordering is the prefix one, i.e. $v \leqslant u$ if and only if $v \in \mathrm{pref}(u)$. (While by the *shortest* word we mean the word with the smallest length!) If $v \in \mathrm{pref}(u)$ or $u \in \mathrm{pref}(v)$, we say that they are *comparable*, denoted by u Pref v. The maximal common prefix of words u and v is denoted by $u \wedge v$. It is empty if and only if one of the words is empty or they start with different letters. If u and v are words, the maximal *u-prefix* of v is the maximal element of

$$\mathrm{pref}(v) \cap \mathrm{pref}\left(u^+\right).$$

Let $u \in \Sigma^+$ be a word $u = l_1 l_2 \dots l_d$, with $d = |u|$ and $l_i \in \Sigma$. Then the *mirror image* of the word u, denoted by $\overline{u}$, is obtained by inverting the order of the letters, viz.

$$\overline{u} = l_d l_{d-1} \dots l_1.$$

Let g be an arbitrary morphism. The *mirror image* of g is the morphism denoted by $\overline{g}$ and defined by

$$\overline{g}(x) = \overline{g(x)},$$

for each $x \in \Sigma$. Note that in general $\overline{g}(u)$ need not be equal to $g(\overline{u})$ or $\overline{g(u)}$. Instead,

$$\overline{g}(\overline{u}) = \overline{g(u)}.$$

All concepts and reasonings regarding prefixes are valid dually for suffixes, mirror images considered. We shall often use this fact.

A morphism g defined on Σ is called non-erasing if $g(x)$ is non-empty for all $x \in \Sigma$.

Let S be a subsemigroup of Σ^+ generated by a set M. The *rank* of M is the cardinality of the minimal set generating S. We can write

$$\mathrm{rank}(M) = \mathrm{Card}(S \setminus S \cdot S).$$

By the rank of a monoid M we mean the rank of semigroup $M \setminus \{\epsilon\}$.

It is a well-known fact that for each set $M \subset \Sigma^+$ there exists the smallest free subsemigroup of Σ^+ containing M and called its *free hull*.

Let $g, h : \Sigma^* \to \Sigma^*$ be binary morphisms. Their *equality set* is defined by

$$\mathrm{Eq}(g, h) = \left\{u \in \Sigma^* \;\middle|\; g(u) = h(u)\right\}.$$

It is easy to verify that the set $\mathrm{Eq}(g, h)$ is a free submonoid of Σ^* generated by the set of its minimal elements

$$\mathrm{eq}(g, h) = \mathrm{Eq}(g, h) \setminus \left(\mathrm{Eq}(g, h) \setminus \{\varepsilon\}\right)^2 \setminus \{\varepsilon\}.$$

Note that $\mathrm{eq}(g, h)$ is a biprefix code.

Let $g: A^* \to \Sigma^*$ be a non-periodic binary morphism. By z_g we denote the maximal common prefix of $g(ab)$ and $g(ba)$, i.e.

$$z_g = g(ab) \wedge g(ba).$$

Since g is non-periodic, we have, by Periodicity Lemma (see below), $|z_h| < |g(a)| + |g(b)|$. If $\mathrm{pref}_1(g(a)) \neq \mathrm{pref}_1(g(b))$, i.e. $z_g = \varepsilon$, we say that g is *marked*.

Similarly we define $\underline{z}_g$ as a maximal common suffix of $g(ab)$ and $g(ba)$. Note that

$$\underline{z}_g = \overline{\overline{g(ab)} \wedge \overline{g(ba)}} = \overline{z_{\bar{g}}}$$

and $\underline{z}_g = \varepsilon$ is equivalent to $\bar{g}$ being marked.

Cartesian product $A^* \times A^*$ is the set of ordered pairs (u, v) of words. It can be seen as a monoid with operation of catenation defined by $(u, v)(u', v') = (uu', vv')$, with the unit $(\varepsilon, \varepsilon)$. Such a monoid is obviously not free, it is even not isomorphic to a submonoid of a free monoid.

Let $g, h: A^* \to \Sigma^*$ be binary morphisms. The subset of $A^* \times A^*$ denoted by $\mathbb{C}(g, h)$ and defined by

$$\mathbb{C}(g, h) = \big\{(u, v) \bigm| g(u) = h(v)\big\}$$

will be called the *coincidence set* of morphisms g and h. It is generated by the set

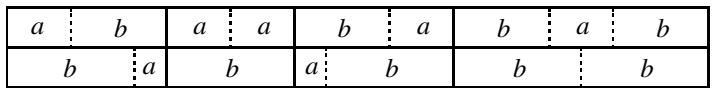
$$\mathbf{c}(g, h) = \mathbb{C}(g, h) \setminus \big(\mathbb{C}(g, h) \setminus \{(\varepsilon, \varepsilon)\}\big)^2 \setminus \{(\varepsilon, \varepsilon)\}.$$

It is not difficult, but quite important to note the following statement.

Lemma 3. *$\mathbb{C}(g, h)$ is, as a submonoid of $A^* \times A^*$, freely generated by $\mathbf{c}(g, h)$. Moreover, if (u_1, v_1), (u_2, v_2), and $(u_1 x u_2, v_1 y v_2)$ are elements of $\mathbb{C}(g, h)$ then also $(x, y) \in \mathbb{C}(g, h)$ (i.e., $\mathbb{C}(g, h)$ is left and right unitary in $A^* \times A^*$).*

This fact is illustrated by the following picture, which represents the unique factorization of the pair $(abaababab, bababbb) \in \mathbb{C}(g, h)$ into elements of the base, namely:

$$(abaababab, bababbb) = (ab, ba)(aa, b)(ba, ab)(bab, bb).$$



Obviously, (u, u) is an element of $\mathbb{C}(g, h)$ for each $u \in \mathrm{Eq}(g, h)$, and $\mathrm{Eq}(g, h)$ is given uniquely by $\mathbb{C}(g, h)$ as

$$\mathrm{Eq}(g, h) = \big\{u \bigm| (u, u) \in \mathbb{C}(g, h)\big\}.$$

We present several combinatorial lemmas for future (often implicit) reference. The following three lemmas are part of the folklore.

Lemma 4. *The words u and v commute if and only if they have the same primitive root.*

Lemma 5 (Periodicity Lemma). *Let u^+ and v^+ have a common factor of the length $|u| + |v|$. Then the words u and v commute.*

Lemma 6. *The following conditions are equivalent*:

(i) *Words u and v are conjugates.*
(ii) *There is a word z such that $uz = zv$.*
(iii) *There are words t_1 and t_2 such that t_2 is non-empty, t_1t_2 is primitive, and*

$$u \in (t_1t_2)^+, \qquad v \in (t_2t_1)^+.$$

Moreover, if t_1 and t_2 are like in (iii) *and z is like in* (ii), *then $z \in (t_1t_2)^*t_1$.*

We shall often use the following lemma. It is based on the well-known fact that a primitive word t cannot satisfy equality $tt = utv$, with u and v non-empty.

Lemma 7. (A) *Let sw be a factor of w^+. Then s is a suffix of w^+.*
(B) *Let wp be a factor of w^+. Then p is a prefix of w^+.*
(C) *Let uw be a prefix of w^+. Then u and w commute.*
(D) *Let $u_1, u_2, w, w' \in \Sigma^+$ be words such that w' is a conjugate of w, $|u_1| \leqslant |u_2|$, and the words u_1w', u_2w' are prefixes of w^+. Then u_1 is suffix of u_2 and $u_2u_1^{-1}$ commutes with w.*

The last preliminary lemma plays an important rôle in this paper.

Lemma 8. *Let $g : A^* \to A^*$ be a marked morphism and let $u, v \in A^*$. Then there exists a word $w \in A^*$ such that $g(w) = g(u) \wedge g(v)$.*

3. The coincidence set

In this section we study the relation of coincidence sets of non-periodic morphisms to their equality set. We will partially follow the exposition from [1, pp. 347–351]. First notice the following nice lemma.

Lemma 9. *Let $X = \{x, y\} \subseteq \Sigma^+$ be non-periodic set* (*i.e. $xy \neq yx$*). *Let $u \in xX^*$, $v \in yX^*$ be words such that $|u|, |v| \geqslant |xy \wedge yx|$. Then $u \wedge v = xy \wedge yx$.*

The proof is not difficult (see [1, p. 348]).

The lemma immediately implies that for a non-periodic binary morphism h and an arbitrary sufficiently long word $u \in A^+$, the word z_h is a prefix of $h(u)$ and the $(|z_h| + 1)$th letter of $h(u)$ indicates the first letter of u. For any $u, v \in A^*$ we have

$$z_h = h(au)z_h \wedge h(bv)z_h. \tag{1}$$

It is now easy to see that the morphism h_{m}, such that

$$h_{\mathrm{m}}(u) = z_h^{-1} h(u) z_h, \tag{2}$$

$u \in A$, is well defined. Moreover, it is marked, and Eq. (2) holds for any $u \in A^*$. We shall call it the *marked version* of h. Similarly we can define marked version of g. In the following, however, we shall simply suppose that g is marked. This restriction will be justified in Lemma 22.

Let $u, v \in \Sigma^*$ be words such that $g(u)$ Pref $h(v)$. Following lemmas show that the possibility to lengthen the words u, v to words u', v' such that $g(u') = h(v')$ is very restricted.

Lemma 10. *Let g and h be binary morphisms, and let g be marked. Let $u, v \in A^*$ be words such that $g(u)$* Pref *$h(v)$ and let*

$$g(u) \neq h(v) z_h.$$

Let $u_1, u_2, v_1, v_2 \in A^+$ be words such that

$$g(uu_1) = h(vv_1), \qquad g(uu_2) = h(vv_2).$$

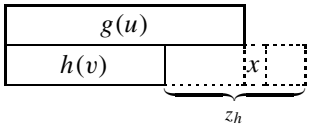
Then $\mathrm{pref}_1(u_1) = \mathrm{pref}_1(u_2)$ *or* $\mathrm{pref}_1(v_1) = \mathrm{pref}_1(v_2)$.

Proof. If u_1, u_2, v_1, and v_2 satisfy the conditions of the lemma, then the same conditions are satisfied also by the words $u_1 u u_1$, $u_2 u u_2$, $v_1 v v_1$, and $v_2 v v_2$ respectively. Hence we can suppose that each of the words u_1, u_2, v_1, v_2 is longer than z_h. Consider three cases.

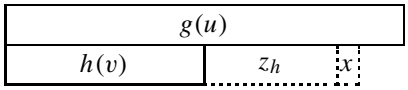
1. First suppose that $|g(u)| < |h(v)| + |z_h|$. By Eq. (1), $h(v)z_h$ is a prefix of both $h(vv_1)$ and $h(vv_2)$, and

   $$\mathrm{pref}_1\big(g(u_1)\big) = \mathrm{pref}_1\big(g(u_2)\big) = \mathrm{pref}_1\big(g(u)^{-1} h(v) z_h\big) = x.$$

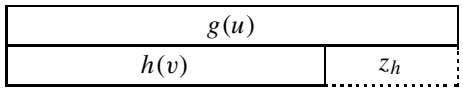
   Since g is a marked morphism, this implies that $\mathrm{pref}_1(u_1) = \mathrm{pref}_1(u_2)$.



2. Suppose, on the other hand, that $|g(u)| > |h(v)| + |z_h|$. Then v_1, v_2 have the common prefix longer than z_h, and $\mathrm{pref}_1(v_1) = \mathrm{pref}_1(v_2)$ is determined by the letter $x = \mathrm{pref}_1((h(v)z_h)^{-1} g(u))$.



3. If $|g(u)| = |h(v)| + |z_h|$, then, clearly, $g(u) = h(v)z_h$.



□

The following immediate corollary of the previous lemma describes the unique case in which u, v can be extended in two different ways.

Corollary 11. *Let g and h be binary morphisms, and let g be marked. Let (c, d) and (c', d') be distinct elements of $\mathbf{c}(g, h)$. Put*

$$u = c \wedge c', \qquad v = d \wedge d'.$$

Then

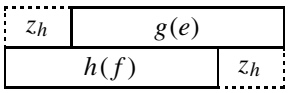
$$g(u) = h(v)z_h.$$

The ground for the characterization of coincidence set is the following lemma.

Lemma 12. *Let g and h be binary morphisms, and let g be marked. Let the words $e, f \in A^+$ satisfy the following conditions*:

(i) $z_h g(e) = h(f)z_h$.
(ii) *The words e, f are minimal, i.e.: if u is a proper prefix of e and v is a proper prefix of f then $z_h g(u) \neq h(v)z_h$.*

Then, given the first letter of e or the first letter of f, the words e and f are determined uniquely.



Proof. Suppose e, f, and e', f' satisfy (i) and (ii), and $\mathrm{pref}_1(e) = \mathrm{pref}_1(e')$. Put $c = e \wedge e'$, $d = f \wedge f'$. Since g is a marked morphism, we have

$$z_h g(e) \wedge z_h g(e') = z_h g(c). \tag{3}$$

From Eq. (1) we deduce

$$h(f)z_h \wedge h(f')z_h = h(d)z_h. \tag{4}$$

Since $z_h g(e) = h(f)z_h$ and $z_h g(e') = h(f')z_h$, Eqs. (3), (4) yield

$$z_h g(c) = h(d)z_h.$$

Since c is non-empty, we deduce from (ii) that $c = e = e'$ and $d = f = f'$. Similarly for $\mathrm{pref}_1(f) = \mathrm{pref}_1(f')$.

Lemma 13. *Let g and h be binary morphisms, and let g be marked.*

(A) *The rank of $\mathbb{C}(g, h_{\mathrm{m}})$ is at most two.*
(B) *If rank of $\mathbb{C}(g, h_{\mathrm{m}})$ is two and $\mathbf{c}(g, h_{\mathrm{m}}) = \{(e, f), (e', f')\}$, then*

$$\mathrm{pref}_1(e) \neq \mathrm{pref}_1(e'), \qquad \mathrm{pref}_1(f) \neq \mathrm{pref}_1(f').$$

Proof. Recall that $h_{\mathrm{m}}(u) = z_h^{-1} h(u) z_h$ to see that

$$\mathbb{C}(g, h_{\mathrm{m}}) = \big\{(u, v) \in A^* \times A^* \;\big|\; z_h g(u) = h(v) z_h\big\}.$$

The rest is a consequence of Lemma 12. □

Note that both g and h_{m} are marked morphisms, and $(e, f) \in \mathbf{c}(g, h_{\mathrm{m}})$ is just an expression of the fact that the pair (e, f) satisfies conditions described in Lemma 12.

The question on the structure of the equality set $\mathrm{Eq}(g, h)$ can be seen as a special case of the above considerations. If conditions

$$u = v, \qquad c = d, \qquad e = f, \qquad u_i = v_i, \qquad c' = d', \qquad e' = f',$$

with $i = 1, 2$, are added, then we get following modification of Lemmas 10, 12, and 13 and Corollary 11.

Lemma 14. *Let g and h be binary morphisms, and let g be marked. Let $u \in A^*$ be a word such that $g(u)$ Pref $h(u)$ and*

$$g(u) \neq h(u) z_h.$$

Let $u_1, u_2 \in A^+$ be words such that

$$g(uu_1) = h(uu_1), \qquad g(uu_2) = h(uu_2).$$

Then $\mathrm{pref}_1(u_1) = \mathrm{pref}_1(u_2)$.

Corollary 15. *Let g and h be binary morphisms, and let g be marked. Let c and c' be distinct elements of* $\mathrm{eq}(g, h)$. *Put $u = c \wedge c'$. Then*

$$g(u) = h(u) z_h.$$

Lemma 16. *Let g and h be binary morphisms, and let g be marked. Let the word $e \in A^+$ satisfy the following conditions*:

(i) $z_h g(e) = h(e) z_h$.

(ii) *The word e is minimal, i.e.: if e_1 is a proper prefix of e then $z_h g(e_1) \neq h(e_1) z_h$.*

Then the word e is determined uniquely by its first letter.

Lemma 17. *Let g and h be binary morphisms, and let g be marked.*

(A) *The rank of* $\mathrm{Eq}(g, h_{\mathrm{m}})$ *is at most two.*
(B) *If rank of* $\mathrm{Eq}(g, h_{\mathrm{m}})$ *is two and* $\mathrm{eq}(g, h_{\mathrm{m}}) = \{e, e'\}$, *then*

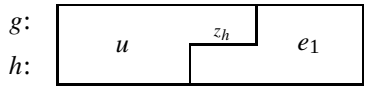
$$\mathrm{pref}_1(e) \neq \mathrm{pref}_1(e').$$

We can now give the following proof.

Proof of Theorem 1(C). By Lemma 22 below, we can assume that g is marked.

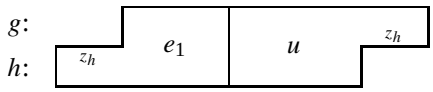
1. If there do not exist a word $u \in A^*$ such that $g(u) = h(u)z_h$, then, by Corollary 15, $\mathrm{Eq}(g, h)$ is generated by at most one word.
2. Suppose that such a word u exists and suppose no non-empty prefix of u is an element of $\mathrm{Eq}(g, h)$. By Corollary 15, the word u is a prefix of any $u' \in \mathrm{Eq}(g, h)$. If $\mathrm{eq}(g, h)$ is not empty, there exists a (minimal) word e_1 such that

$$g(ue_1) = h(ue_1).$$



Note that in such a case

$$z_h g(e_1 u) = h(e_1 u) z_h,$$



and thus $e_1 u$ is an element of $\mathrm{eq}(g, h_{\mathrm{m}})$.
2.1. If $\mathrm{Eq}(g, h_{\mathrm{m}})$ is generated by $e_1 u$ then

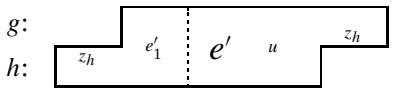
$$\mathrm{eq}(g, h) = \{ue_1\}.$$

2.2. The cardinality of $\mathrm{eq}(g, h_{\mathrm{m}})$ is at most two, by Lemma 17. Suppose that there is another word $e' \in \mathrm{eq}(g, h_{\mathrm{m}})$. We have two possibilities.
2.2.1. Let first exist a prefix e'_1 of e' such that

$$z_h g(e'_1) = h(e'_1).$$

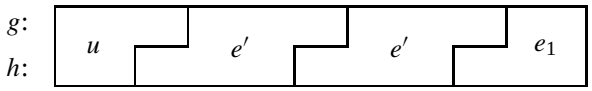
Then $e' = e'_1 u$, and thus $g(ue'_1) = h(ue'_1)$ and

$$\mathrm{eq}(g,h) = \{ue_1, ue_1'\}.$$



2.2.2. If such a prefix does not exist, then

$$\mathrm{eq}(g,h) = ue'^{*}e_1.$$



□

4. Typical morphisms

In this section we introduce some properties of morphisms g, h and show that in the following investigation these properties can be assumed without loss of generality.

Definition 18. We say that an (unordered) pair of binary morphisms $g, h : A^* \to \Sigma^*$ is *principal* if the target alphabet Σ is the base of the free hull of the set $\{g(a), g(b), h(a), h(b)\}$.

Definition 19. An ordered pair (g, h) of morphisms is called *typical* if

(a) both g, h are binary non-periodic morphisms $A^* \to A^*$;
(b) the morphism g is marked;
(c) $|g(a)| > |h(a)|$, $|g(b)| < |h(b)|$.

The following is an important property of the base of the free hull generated by a set.

Lemma 20. *Let X be a finite subset of Σ^* and let Y be the base of the free hull of X. Then for each element $y \in Y$ there is a word $x \in X$ such that y is a prefix (suffix) of x.*

For the proof see [7, Lemma 3.1]. For our purpose note the following immediate consequence.

Corollary 21. *Let X be a finite subset of Σ^* such that Σ is the base of the free hull of X. Then*

$$\Sigma = \{\mathrm{pref}_1(u) \mid u \in X\} = \{\mathrm{suff}_1(u) \mid u \in X\}.$$

We can now prove a statement allowing to restrict our considerations to principal and typical pairs.

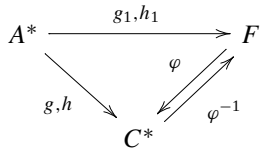
Lemma 22. *Let g_1, h_1 be non-periodic binary morphisms $A^* \to \Sigma^*$ such that the rank of* $\mathrm{Eq}(g_1, h_1)$ *is at least two. Then there exists a pair of morphisms (g, h) which is principal and typical, and*

$$\mathbb{C}(g, h) = \mathbb{C}(g_1, h_1).$$

Moreover, if h_1 ($\overline{g}_1$, $\overline{h}_1$ respectively) is marked then such is also h ($\overline{g}$, $\overline{h}$ respectively).

Proof. Let $F \subset \Sigma^*$ be the free hull of the set $\{g_1(a), g_1(b), h_1(a), h_1(b)\}$ and let C be an alphabet whose cardinality equals the rank of F. Let $\varphi : F \to C^*$ be an isomorphism. Define morphisms $g, h : A^* \to C^*$ by

$$g = \varphi \circ g_1, \qquad h = \varphi \circ h_1. \tag{5}$$



Obviously, (g, h) is a principal pair of non-periodic morphisms, the above diagram commutes, and $\mathbb{C}(g, h) = \mathbb{C}(g_1, h_1)$.

By symmetry of letters a and b, suppose that the condition of Definition 19(c) is satisfied. By symmetry of g and h, we can assume

$$|z_h| \geqslant |z_g|. \tag{6}$$

By Corollary 21,

$$C = \{\mathrm{pref}_1(g(a)), \mathrm{pref}_1(g(b)), \mathrm{pref}_1(h(a)), \mathrm{pref}_1(h(b))\}.$$

1. Suppose g is not marked. Then also h is not marked, by Eq. (6), and

   $$\mathrm{pref}_1(g(a)) = \mathrm{pref}_1(g(b)), \qquad \mathrm{pref}_1(h(a)) = \mathrm{pref}_1(h(b)).$$

   Let x be a first letter of a word $u \in \mathrm{Eq}(g, h)$. Then

   $$\mathrm{pref}_1(g(x)) = \mathrm{pref}_1(h(x))$$

   implies that the cardinality of C is one, a contradiction to non-periodicity of g and h.
2. Thus g is marked. We claim that cardinality of C is two. This completes the proof, since we can then choose $C = A$.
   2.1. To prove the claim suppose first that h is marked. By Lemma 17, the set $\mathrm{Eq}(g, h)$ contains two words starting with different letters. This implies

   $$\mathrm{pref}_1(h(a)) = \mathrm{pref}_1(g(a)), \qquad \mathrm{pref}_1(h(b)) = \mathrm{pref}_1(g(b)),$$

and cardinality of C is two.

2.2. Suppose z_h is non-empty. Let again x be the starting letter of a word in $\mathrm{Eq}(g,h)$. Then

$$\mathrm{pref}_1\big(h(y)\big) = \mathrm{pref}_1\big(h(x)\big) = \mathrm{pref}_1\big(g(x)\big),$$

where $\{x, y\} = \{a, b\}$, and cardinality of C is again two.

We have proved that a pair (g,h) is principal and typical. If h is not marked then neither $h_1 = \varphi^{-1} \circ h$ is. Similarly for $\overline{g}$ and $\overline{h}$. □

Let $\pi : A^* \to A^*$ denote the morphism exchanging letters a and b,

$$\pi(a) = b, \qquad \pi(b) = a.$$

From a typical and principal pair we can derive another one using the mirror image.

Lemma 23. *Let (g,h) be a pair of morphisms which is typical and principal.*

(A) *If $\underline{z}_g = \varepsilon$ then $(\overline{g}, \overline{h})$ is typical and principal.*
(B) *If $\underline{z}_g \neq \varepsilon$ then $(\overline{h} \circ \pi, \overline{g} \circ \pi)$ is typical and principal.*

Proof. (A) The verification is straightforward.

(B) The pair $(\overline{h} \circ \pi, \overline{g} \circ \pi)$, clearly, satisfies conditions (a) and (c) of Definition 19. We have to show that $\overline{h} \circ \pi$ is marked. Since (g,h) is principal,

$$A = \big\{\mathrm{suff}_1\big(g(a)\big), \mathrm{suff}_1\big(g(b)\big), \mathrm{suff}_1\big(h(a)\big), \mathrm{suff}_1\big(h(b)\big)\big\},$$

by Corollary 21. Suppose $\overline{h} \circ \pi$ is not marked. Then both $\underline{z}_g$ and $\underline{z}_h$ are non-empty and from $\mathrm{eq}(g,h) \neq \emptyset$ we can conclude

$$\mathrm{suff}_1\big(g(a)\big) = \mathrm{suff}_1\big(g(b)\big) = \mathrm{suff}_1\big(h(a)\big) = \mathrm{suff}_1\big(h(b)\big),$$

a contradiction to cardinality of A being two.

Principality of both studied pairs follows directly from mirror symmetry. □

5. Marked morphisms

The structure of an equality set is much more transparent if both morphisms are marked as shown in the following lemma.

Lemma 24. *Let $g, h : A^* \to A^*$ be marked morphisms. Then*

$$\mathrm{Eq}(g,h) = \{\alpha, \beta\}^*$$

with $\alpha, \beta \in A^*$. *If rank of* $\mathrm{Eq}(g, h)$ *is two, then*

$$\mathrm{pref}_1(\alpha) \neq \mathrm{pref}_1(\beta).$$

Proof. The claim follows directly from Lemma 17. □

Let us further investigate the relation between the coincidence and equality sets of two marked binary morphisms.

Lemma 25. *Let* $g, h : A^* \to A^*$ *be marked morphisms such that the rank of*

$$\mathrm{Eq}(g, h) = \{v, w\}^*$$

is two. Then

$$\mathbb{C}(g, h) = \big\{(e, f), (e', f')\big\}^*$$

for some non-empty words e, f, e', f' *such that*

$$\mathrm{pref}_1(e) = p_1(f) = a, \qquad \mathrm{pref}_1(e') = p_1(f') = b.$$

Define a pair of morphisms $g_1, h_1 : A^* \to A^*$ *by*

$$\begin{cases} g_1(a) = e, \\ g_1(b) = e', \end{cases} \qquad \begin{cases} h_1(a) = f, \\ h_1(b) = f'. \end{cases}$$

Then g_1, h_1 *are marked non-erasing morphisms and there exist non-empty words* v_1, w_1 *such that*

$$g_1(v_1) = h_1(v_1) = v, \qquad g_1(w_1) = h_1(w_1) = w,$$

and

$$\mathrm{Eq}(g_1, h_1) = \{v_1, w_1\}^*.$$

Proof. Suppose that $\mathrm{pref}_1(v) = a$, $\mathrm{pref}_1(w) = b$. By Lemma 12, the words (e, f) can be defined as the minimal prefix of (v, v) satisfying $g(e) = h(f)$. Similarly, (e', f') is the minimal prefix of (w, w) with the same property. Thus the words e, f, e', f' are non-empty and, still by Lemma 12,

$$\mathbb{C}(g, h) = \big\{(e, f), (e', f')\big\}^*,$$

and g_1, h_1 are marked non-erasing morphisms.

Let u be an element of $\mathrm{Eq}(g, h)$. Since $(u, u) \in \mathbb{C}(g, h)$, there exists a word u' such that

$$g_1(u') = h_1(u') = u.$$

Conversely, if $g_1(u') = h_1(u')$ then also

$$g \circ g_1(u') = h \circ h_1(u').$$

Therefore both g_1 and h_1 are bijections between $E(g_1, h_1)$ and $E(g, h)$. □

In the previous lemma the morphisms g_1, h_1 have the properties assumed for g, h and the construction can be iterated to obtain a sequence of pairs of morphisms. We formulate the fact in the following lemma.

Lemma 26. *Let $g_0, h_0 : A^* \to A^*$ be marked morphisms such that the rank of*

$$\mathrm{Eq}(g_0, h_0) = \{v_0, w_0\}^*$$

is two. Then the following statements hold.

(A) *There exists a sequence of non-erasing marked morphisms $(g_i, h_i)_{i\in\mathbb{N}}$ such that for each $i \in \mathbb{N}$*

$$\mathbb{C}(g_i, h_i) = \big\{(e_i, f_i), \big(e_i', f_i'\big)\big\}^*,$$

with

$$\begin{cases} e_i = g_{i+1}(a), \\ e_i' = g_{i+1}(b), \end{cases} \qquad \begin{cases} f_i = h_{i+1}(a), \\ f_i' = h_{i+1}(b), \end{cases}$$

and

$$\mathrm{pref}_1\big(g_i(a)\big) = \mathrm{pref}_1\big(h_i(a)\big) = a,$$
$$\mathrm{pref}_1\big(g_i(b)\big) = \mathrm{pref}_1\big(h_i(b)\big) = b.$$

(B) *For any $i < j$*

$$E(g_i, h_i) = g_{i+1} \circ g_{i+2} \circ \cdots \circ g_j\big(E(g_j, h_j)\big).$$

(C) *There exists a number m such that $e_m = f_m$, $e_m' = f_m'$, and $e_i = f_i = a$, $e_i' = f_i' = b$ for all $i > m$.*

Proof. The items (A) and (B) follow from Lemma 25 by induction. For item (C) it is enough to note that unless $|e_i| = |f_i| = |e_i| = |f_i| = 1$, the length of the word $v_i w_i$ is strictly decreasing. □

The construction of the sequence $(g_i, h_i)_{i\in\mathbb{N}}$ is similar to an idea used in the proof that Post Correspondence Problem is decidable in the binary case (see [2]). The sequence has also the following interesting property.

Lemma 27. *Let $i, j \geqslant 0$ and let*

$$g_{i+j}(u) = h_{i+j}(v),$$

with $u \neq v$. Then $i + j < m$ and

$$g_i \circ g_{i+1} \circ \cdots \circ g_{i+j}(u) = h_i \circ h_{i+1} \circ \cdots \circ h_{i+j}(v)$$

if and only if j is even.

Proof. By induction, it is enough to show

$$g_i \circ g_{i+1}(u) \neq h_i \circ h_{i+1}(v)$$

and

$$g_i \circ g_{i+1} \circ g_{i+2}(u) = h_i \circ h_{i+1} \circ h_{i+2}(v).$$

By definition, $g_i(u') = h_i(v')$ if and only if $(u', v') \in \{(e_i, f_i), (e_i', f_i')\}^*$, i.e. if and only if there exists a word w such that $u' = g_{i+1}(w)$, $v' = h_{i+1}(w)$. But we assume $u \neq v$. On the other hand, if $j = 2$, put $w = g_{i+2}(u) = h_{i+2}(v)$.

For $k \geqslant m$, we have $e_k = f_k$, $e_k' = f_k'$ and thus $g_k(u) = h_k(v)$ if and only if $u = v$. Therefore, $i + j$ must be less than m. □

Latter, we will need the following technical lemma.

Lemma 28. *Let $g, h : A^* \to A^*$ be two marked morphisms. Let $u, u', v, v' \in A^*$ be words, and s, r, q be positive integers, such that*

$$g(a^s bu) = h(a^s bu'), \qquad g(a^r bv) = h(a^q bv').$$

Then $s = r = q$.

Proof. Recall that we assume $g \neq h$ (for $g = h$ only $r = q$ holds, as is easy to see).

Let g and h be morphisms satisfying assumptions, but $s = r = q$ does not hold. Suppose, moreover, that the length of $a^s bu$ is smallest possible. We show that $a^s bu$ can be shortened, and thus we obtain a contradiction.

We first claim that $g(a)$ and $h(a)$ do not commute. Suppose for a while that $|g(a)| > |h(a)|$ (similarly, if $|g(a)| < |h(a)|$). The claim follows from h being marked and

$$\operatorname{pref}_1\big(h(b)\big) = \operatorname{pref}_1\big(h(a)^{-s} g(a)^s bu\big).$$

(Clearly, $g(a) = h(a)$ implies $g = h$.)

By Corollary 11,

$$g(a^s bu \wedge a^r bv) = h(a^s bu' \wedge a^q bv'). \tag{7}$$

1. If $s \neq r$ and $s \neq q$, then (7) yields

$$g(a^i) = h(a^j),$$

with $i = \min(s, r)$, $j = \min(s, q)$. Therefore, the words $g(a)$ and $h(a)$ commute, a contradiction.
2. Suppose next, by symmetry, $s = r$ and $s \neq q$. Put $m = \min(s, q)$. Equality (7) implies

$$g(a^s bw) = h(a^m), \tag{8}$$

where $w = u \wedge v$.
The set $\mathbb{C}(g, h)$ contains elements $(a^s bu, a^s bu')$ and $(a^s bw, a^m)$, whence it is not difficult to see that the rank of $\mathbb{C}(g, h)$ is two. Let e, f, e', f' be words, and g_1, h_1 morphisms, defined as in Lemma 25.
Equality (8) implies that there is a positive integer p, such that $f = a^p$. From this we deduce $e \notin a^+$ and thus $|e| > s$. Since $a^s bu'$ and $a^q bv'$ are elements of $\{f, f'\}^*$, both s and q are multiples of p. Put

$$s_1 = \frac{s}{p}, \qquad q_1 = \frac{q}{p},$$

and define words u_1 and v_1 by

$$g_1(u_1) = a^s bu, \qquad h_1(u_1) = a^s bu',$$
$$g_1(v_1) = a^s bv, \qquad h_1(v_1) = a^q bv'.$$

Since $h_1(a) = f = a^p$, the words u_1 and v_1 can be factorized as

$$u_1 = a^{s_1} bu_2, \qquad v_1 = a^{q_1} bv_2,$$

with $u_2, v_2 \in A^*$. If $s > q$, from

$$h_1(a^{s_1} bv_2) = h_1(a^{s_1 - q_1} a^{q_1} bv_2) = a^{s-q} a^q bv' = a^s bv'$$

we deduce

$$g \circ g_1(a^{s_1} bu_2) = h \circ h_1(a^{s_1} bu_2), \qquad g \circ g_1(a^{q_1} bv_2) = h \circ h_1(a^{s_1} bv_2).$$

The same equalities are obtained in a similar way if $s < q$.
Inequality $s \neq q$ implies $s_1 \neq q_1$, and $|e| > s$ yields $|a^{s_1} bu_2| < |a^s bu|$. This completes the proof. □

6. The (non-existence of a) counter-example

The consecutive proof of our main claim, Theorem 2, will be essentially made by contradiction. We shall assume that there exist a counter-example to it and gradually show that such an assumption is wrong. Actually, the first step in this direction has been already made in Lemma 24, where we proved that a counter-example cannot consist of two marked morphisms. (Note that our proof does not deal directly with the rank of equality set. It is rather concentrated on the different first letter of generating elements.)

To enable an argument by induction, we can also assume that the counter-example is in a sense of minimal length. This leads to the following definitions.

Definition 29. We say that a pair of morphisms (g, h) is *a counter-example* if

(a) (g, h) is a typical pair of morphisms.
(b) $\mathrm{eq}(g, h)$ contains two distinct elements u, v such that $\mathrm{pref}_1(u) = \mathrm{pref}_1(v)$.

We say that a pair of morphisms (g, h) is a *shortest counter-example* if it satisfies the following additional condition.

(c) Let (g', h') be a counter-example. Let d (d' respectively) be the length of the shortest element of $\mathrm{eq}(g, h)$ ($\mathrm{eq}(g', h')$ respectively). Then $d \leqslant d'$.

We say that a pair of morphisms (g, h) is *simple* if $g(e) = h(f)$ implies $e = f$.

The following lemma yields basic information about the structure of the equality set of a counter-example.

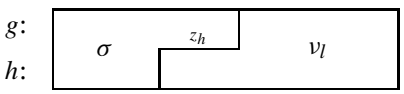
Lemma 30. *Let (g, h) be a counter-example. Then z_h is non-empty and there exist non-empty words σ, v_a, and v_b such that*

$$\mathrm{pref}_1(v_a) = a, \qquad \mathrm{pref}_1(v_b) = b,$$

the words σv_a, σv_b are the two shortest elements of $\mathrm{eq}(g, h)$*, and*

$$g(\sigma) = h(\sigma) z_h, \tag{9}$$

$$z_h g(v_l) = h(v_l), \quad \text{with } l \in A. \tag{10}$$



Moreover,

$$\mathrm{pref}_1\big(g(x)\big) \neq \mathrm{pref}_1\big(g(y)\big) = \mathrm{pref}_1\big(h(x)\big) = \mathrm{pref}_1\big(h(y)\big) \tag{11}$$

holds, with $\{x, y\} = \{a, b\}$ and $y = \mathrm{pref}_1(\sigma)$.

Proof. If the words σv_a and σv_b are elements of $\mathrm{eq}(g,h)$, then Eqs. (9) and (10) follow from Corollary 15.

Corollary 15 and Definition 29(b) imply that z_h is non-empty and $\mathrm{pref}_1(u) = \mathrm{pref}_1(v)$ for any two words $u, v \in \mathrm{Eq}(g,h)$. Therefore, suppose that u, v mentioned in Definition 29(b) are the two shortest elements of $\mathrm{eq}(g,h)$. Put $\sigma = u \wedge v$. Since u and v are minimal, there exist non-empty words u_1 and v_1 such that $\sigma u_1 = u$, $\sigma v_1 = v$, and $\mathrm{pref}_1(u_1) \neq \mathrm{pref}_1(v_1)$. The choice of v_a and v_b is obvious.

Equality (11) follows, since g is marked while h is not. □

From Definition 19(c) and Eqs. (9), (10) we deduce that

$$|\sigma|_a \geqslant 1, \qquad |v_x|_b \geqslant 1, \quad \text{with } x \in A. \tag{12}$$

The following lemma shows the connection between a general counter-example and marked morphisms.

Lemma 31. *Let (g,h) be a counter-example. Then (g, h_{m}) is a typical pair of morphisms,*

$$v_a\sigma,\, v_b\sigma \in \mathrm{Eq}(g, h_{\mathrm{m}}),$$

and the rank of $\mathrm{Eq}(g, h_{\mathrm{m}})$ *is two.*

Proof. The claim is a direct consequence of Lemmas 24 and 30. □

6.1. The case $\underline{z}_g \neq \varepsilon$

In this subsection we shall assume that $\underline{z}_g$ is non-empty, i.e. $\overline{g}$ is not marked. By Lemma 23, the pair $(\overline{h} \circ \pi, \overline{g} \circ \pi)$ is typical. Since $\underline{z}_g \neq \varepsilon$, it is also a counter-example, by Corollary 15. This implies that we can suppose

$$|\underline{z}_g| \geqslant |z_h|, \tag{13}$$

because otherwise we consider $(\overline{h} \circ \pi, \overline{g} \circ \pi)$ instead of (g,h).

Let τ denote the maximal common suffix of two different elements of $\mathrm{Eq}(g,h)$. Then any solution of (g,h) looks like

The mirror variant of (1) implies that $\underline{z}_g$ is a suffix of any $g(u)$, sufficiently long. Especially

$$\underline{z}_g \in \mathrm{suff}\big(g(a)^+\big), \qquad \underline{z}_g \in \mathrm{suff}\big(g(b)^+\big). \tag{14}$$

Thus also

$$z_h \in \mathrm{suff}(\underline{z}_g). \tag{15}$$

The following lemma is the first of several claims investigating the possible structure of z_h.

Lemma 32. *Let (g,h) be a counter-example such that $\underline{z}_g \neq \varepsilon$. Let $\mathrm{pref}(\sigma) = b$. Then $z_h \in g(b)^+$.*

Proof. Let b^l be the maximal b-prefix of σ and let b^k be the maximal b-prefix of $\nu_b\sigma$. From (14) and (15) we deduce that $z_h = sg(b)^i$ for some suffix s of $g(b)$ and $i \in \mathbb{N}$. Thus $z_h g(b) = sg(b)^{i+1}$. Equalities (9) and (10) imply that $g(b)$ is a prefix of $z_h g(b)$ and thus s, $g(b)$, and z_h commute. Let

$$z_h = t^{m_1}, \qquad g(b) = t^{m_2},$$

with t primitive and $m_1, m_2 \in \mathbb{N}_+$. Then (10) yields that $t^{m_1+k\cdot m_2}$ is the maximal t-prefix of $z_h g(\nu_b\sigma)$. Similarly from (9) follows that $t^{l\cdot m_2}$ is the maximal t-prefix of $h(\sigma\nu_a)$.

1. Suppose that $h(b) = t^{m_3}$ for some $m_3 \in \mathbb{N}_+$. Then, by (9), the word $g(b)^l \cdot \mathrm{pref}_1(g(a))$ is a prefix of $t^{l\cdot m_3}$, a contradiction with g being marked.
2. This implies, by Periodicity Lemma, that the maximal t-prefix of $h(b)z_h$ is shorter than $|h(b)t|$. Hence $t^{l\cdot m_2}$ is the maximal t-prefix of any word $h(bu)z_h$, $u \in A^*$. Equality (10) now implies

$$m_1 + k\cdot m_2 = l\cdot m_2.$$

Thus

$$m_1 = (l-k)\cdot m_2 \quad \text{and} \quad z_h = g(b)^{l-k}. \qquad \square$$

Next lemma is similar to Lemma 32.

Lemma 33. *Let (g,h) be a counter-example such that $\underline{z}_g \neq \varepsilon$. Let $\mathrm{pref}(\sigma) = a$. Then $z_h \in h(a)^+$.*

Proof. Equality (9) yields $h(a) \in \mathrm{pref}(g(a))$. Equality (10) implies that $h(a)z_h$ is a prefix of $z_h g(a)$ and thus $z_h h(a) = h(a)z_h$. Hence we have

$$z_h = t^{m_1}, \qquad h(a) = t^{m_2}$$

for a primitive word t and some $m_1, m_2 \in \mathbb{N}_+$.

Let a^l be the maximal a-prefix of $\sigma\nu_b$ and a^k be the maximal a-prefix of $\nu_a\sigma$. Since z_h is the maximal t-prefix of every $h(au)z_h$, the word $t^{l\cdot m_2+m_1}$ is the maximal t-prefix of $g(\sigma\nu_b)$. The maximal t-prefix of $h(\nu_a\sigma)$ is $t^{k\cdot m_2+m_1}$.

1. First suppose that $g(a) = t^{m_3}$ for some $m_3 \in \mathbb{N}_+$. Since g is marked, the word $t^{k\cdot m_3}$ is the maximal t-prefix of $g(v_a\sigma)$ and $t^{k\cdot m_3+m_1}$ is the maximal t-prefix of $z_h g(v_a\sigma)$. Thus, by Eq. (10),

$$k \cdot m_3 + m_1 = k \cdot m_2 + m_1, \tag{16}$$

and $m_2 = m_3$, a contradiction to $|g(a)| > |h(a)|$.
2. From Eqs. (14), (15) we deduce that t is a suffix of $g(a)$. Since $g(a) \notin t^+$, the maximal t-prefix of $g(\sigma v_b)$, i.e. $t^{l\cdot m_2+m_1}$ is also the maximal t-prefix of $g(a)$. Eq. (10) now yields

$$k \cdot m_2 + m_1 = l \cdot m_2 + 2 \cdot m_1 \quad \text{and} \quad z_h = h(a)^{k-l}. \qquad \square$$

Now we can complete the subsection by showing that if $\underline{z}_g$ is non-empty, then (g, h) is not a counter-example.

Lemma 34. *Let (g, h) be a counter-example. Then $\underline{z}_g = \varepsilon$.*

Proof. 1. Suppose first $\mathrm{pref}_1(\sigma) = a$ and $\underline{z}_g \neq \varepsilon$. By Lemma 33, $z_h = h(a^s)$, $s \in \mathbb{N}_+$. From (9) and (10) we have

$$z_h g(\sigma) = h(a^s\sigma)z_h, \qquad z_h g(v_a\sigma) = h(v_a\sigma)z_h.$$

Verify that morphisms h_{m}, g satisfy the assumptions of Lemma 28, a contradiction.

2. Suppose $\mathrm{pref}_1(\sigma) = b$ and $\underline{z}_g \neq \varepsilon$. Let l (k respectively) be the maximal b-prefix of σ ($v_b\sigma$ respectively) and let, by the proof of Lemma 32, $z_h = g(b^s)$, with $s = l - k$. Put $\sigma' = b^{-s}\sigma$. Then

$$z_h g(\sigma') = h(b^s\sigma')z_h, \qquad z_h g(v_b\sigma) = h(v_b\sigma)z_h,$$

and Lemma 28, applied to morphisms $h_{\mathrm{m}} \circ \pi$ and $g \circ \pi$, again yields a contradiction. This completes the proof. $\square$

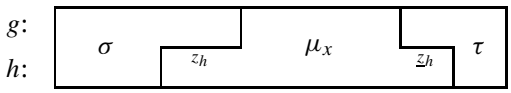
6.2. The case $\underline{z}_h \neq \varepsilon$

In this subsection we show that we can assume $\underline{z}_h = \varepsilon$, i.e. $\overline{h}$ is marked. First we give a more precise description of possible counter-example structure.

Lemma 35. *Let (g, h) be a counter-example.*

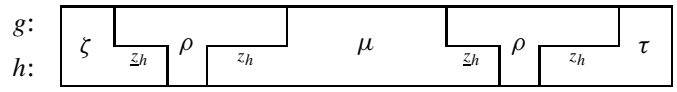
(A) *Let the rank of* $\mathrm{Eq}(g, h)$ *be two. Then there exist words $\sigma, \mu_a, \mu_b \in A^+$ and $\tau \in A^*$ such that*

$$\begin{aligned}
&\mathrm{eq}(g,h)=\{\sigma\mu_a\tau,\sigma\mu_b\tau\}, && z_h g(\mu_a)\underline{z}_h = h(\mu_a), && \mathrm{pref}_1(\mu_a)=a,\\
&g(\sigma)=h(\sigma)z_h, && z_h g(\mu_b)\underline{z}_h = h(\mu_b), && \mathrm{pref}_1(\mu_b)=b,\\
&g(\tau)=\underline{z}_h h(\tau), && \mathrm{suff}_1(\mu_a)\neq\mathrm{suff}_1(\mu_b). &&
\end{aligned}$$



(B) *Let the rank of* $\mathrm{Eq}(g,h)$ *be* ω*. Then there exist words* $\zeta,\mu,\rho,\tau\in A^+$ *such that*

$$\begin{aligned}
&\mathrm{eq}(g,h)=\zeta(\rho\mu)^*\rho\tau=\zeta\rho(\mu\rho)^*\tau,\\
&g(\zeta)\underline{z}_h=h(\zeta), && z_h g(\mu)\underline{z}_h=h(\mu), && \mathrm{pref}_1(\mu)\neq\mathrm{pref}_1(\tau),\\
&g(\rho)=\underline{z}_h h(\rho)z_h, && z_h g(\tau)=h(\tau), && \mathrm{suff}_1(\mu)\neq\mathrm{suff}_1(\zeta).
\end{aligned}$$



Proof. This proof is in fact a refinement of the proof of Theorem 1(C) (we shall refer to it as the Proof).

Since the rank of $\mathrm{Eq}(g,h)$ is at least two, the rank of $\mathrm{Eq}(g,h_{\mathrm{m}})$ is two. Let $\mathrm{eq}(g,h_{\mathrm{m}})=\{e,e'\}$, with $\mathrm{pref}_1(e)\neq\mathrm{pref}_1(e')$.

(A) Suppose that $e=e_1u$, $e'=e'_1u$, and $\mathrm{eq}(g,h)=\{ue_1,ue'_1\}$ (cf. Proof 2.2.1). Let v be the maximal common suffix of e_1 and e'_1. Since e_1 and e'_1 are not a suffix one of the other, the word v is a proper suffix of both e_1 and e'_1, and $e_1=cv$, $e'_1=c'v$.
By Corollary 15, applied to morphisms $\overline{g}$ and $\overline{h}$, we have

$$g(v)=\underline{z}_h h(v).$$

Now it suffices to identify σ with u, τ with v, and μ_a, μ_b with c, c', according to the first letter.

(B) Suppose now that $e=e_1u$, u is not a suffix of e', and $\mathrm{eq}(g,h)=ue'^*e_1$ (cf. Proof 2.2.2). Let v be the maximal common suffix of e and e'. The word v is a proper suffix of both words and, by assumption, it is also a proper suffix of u. Let $u=pv$ and $e'=qv$.
The word ve_1 is the maximal common suffix of ue_1 and $ue'e_1$, and thus

$$g(ve_1)=\underline{z}_h h(ve_1).$$

Now identify ζ with p, ρ with v, μ with q, and τ with e_1. □

Note that between Lemma 30 and Lemma 35 there exists the following correspondence. In the case (A) of Lemma 35, the word σ is the same as in Lemma 30, and

$$v_a=\mu_a\tau,\qquad v_b=\mu_b\tau.$$

In the case (B) of Lemma 35,

$$\sigma = \zeta\rho, \qquad \{\nu_a, \nu_b\} = \{\tau, \mu\rho\tau\}.$$

The following lemma shows that we can suppose, without loss of generality, that $\underline{z}_h$ is empty, i.e. $\overline{h}$ is marked.

Lemma 36. *If there exists any shortest counter-example, then there exists also a shortest counter-example* (g, h) *such that* $\underline{z}_h = \varepsilon$.

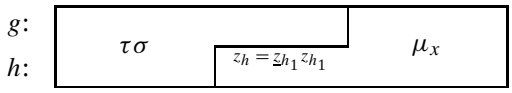
Proof. Let (g_1, h_1) be a shortest counter-example. Suppose $\underline{z}_{h_1} \neq \varepsilon$ and define g and h by

$$g(u) = g_1(u), \qquad h(u) = \underline{z}_{h_1} h_1(u)(\underline{z}_{h_1})^{-1}.$$

It is not difficult to see that morphism h is well defined. The claim is now a consequence of the characterization presented in Lemma 35. Let words ζ, σ, τ, ρ, μ_a, μ_b, and μ be as in that lemma with respect to the pair (g_1, h_1).

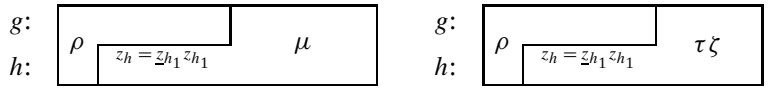
1. If rank of $\mathrm{Eq}(g_1, h_1)$ is two then, by Lemma 35(A),

$$\mathrm{Eq}(g, h) = \{\tau\sigma\mu_a, \tau\sigma\mu_b\}.$$



2. If, on the other hand, rank of $\mathrm{Eq}(g_1, h_1)$ is ω then, by Lemma 35(B),

$$\mathrm{Eq}(g, h) = \{\rho\mu, \rho\tau\zeta\}.$$



By Lemma 22, we can assume that (g, h) is typical. The words $\tau\sigma$ and ρ are non-empty and (g, h) is a counter-example with $\underline{z}_h = \varepsilon$. It is also a shortest counter-example, because the length of words in $\mathrm{eq}(g, h)$ has not changed. □

In the previous lemma the equality set of morphisms g_1 and h_1 is possibly of infinite rank. We have reduced that pair to a pair (g, h) with equality set generated by two words. The claim that rank of $\mathrm{Eq}(g_1, h_1)$ is not ω is now reduced to the claim that σ is empty.

6.3. *The case* $\mathrm{pref}_1(\sigma) = a$

In this subsection we show that we can assume the word σ starts with a letter b.

First note that if both $\overline{g}$ and $\overline{h}$ are marked, then, by Lemma 15, the set $\mathrm{Eq}(g,h)$ contains an element u with $\mathrm{suff}_1(u)=a$. This implies, since $|g(a)|>|h(a)|$,

$$h(a)\in \mathrm{suff}\big(g(a)\big). \tag{17}$$

The next lemma is a parallel to Lemma 33.

Lemma 37. *Let (g,h) be a counter-example such that both $\overline{g}$ and $\overline{h}$ are marked. Let $\mathrm{pref}(\sigma)=a$. Then $z_h\in h(a)^+$.*

Proof. The proof is identical to the proof of Lemma 33, with the only exception that $t\in\mathrm{suff}(g(a))$ (in the beginning of part 2) is deduced from (17). □

We can now prove the claim of this subsection.

Lemma 38. *Let (g,h) be a counter-example such that both $\overline{g}$ and $\overline{h}$ are marked. Then $\mathrm{pref}_1(\sigma)\neq a$.*

Proof. Suppose $\mathrm{pref}_1(\sigma)=a$ and $\underline{z}_g=\underline{z}_h=\varepsilon$. By Lemma 37, $z_h=h(a^s)$, $s\in\mathbb{N}_+$. From (9) and (10) we have

$$z_h g(\sigma)=h(a^s\sigma)z_h,\qquad z_h g(\nu_a\sigma)=h(\nu_a\sigma)z_h.$$

Verify that morphisms h_{m}, g satisfy the assumptions of Lemma 28; a contradiction. □

The results of Sections 6.1–6.3 are summarized by the following lemma.

Lemma 39. *If there exists a counter-example, then there exists a shortest counter-example (g,h) such that*

(A) *$\overline{g}$ and $\overline{h}$ are marked* (*i.e.* $\underline{z}_g=\underline{z}_h=\varepsilon$),
(B) *$\mathrm{pref}_1(u)=b$ for each $u\in\mathrm{eq}(g,h)$,*
(C) *$g(b)$ and $h(b)$ do not commute,*
(D) *$\mathrm{suff}_1(\sigma)=b$.*

Proof. (A) The claim follows from Lemmas 34 and 36.

(B) Follows directly from Lemma 38.

(C) Proof by contradiction. Let t be the common primitive root of $g(b)$ and $h(b)$. Then from $|g(b)|<|h(b)|$ and from (9) we deduce that $\mathrm{pref}_1(t)=\mathrm{pref}_1(g(a))$, a contradiction with g being marked.

(D) By Lemmas 23 and 31, the pair of morphisms $(\overline{g},\overline{h_{\mathrm{m}}})$ is typical. The set $\mathrm{Eq}(\overline{g},\overline{h_{\mathrm{m}}})$ contains two distinct elements $\overline{\nu_a\sigma}=\overline{\sigma}\,\overline{\nu_a}$ and $\overline{\nu_b\sigma}=\overline{\sigma}\,\overline{\nu_b}$ with a common prefix $\overline{\sigma}$ and distinct last letters. Thus $\mathrm{pref}_1(\overline{\sigma})=\mathrm{suff}_1(\sigma)=b$, by Lemma 38. □

To rule out the remaining possibility described in Lemma 39 we shall deal separately with cases $|g(ba)|<|h(b)|$ and $|g(ba)|\geqslant|h(b)|$.

6.4. Relative position

In this section we define some important concepts.

Let u be an element of $\mathrm{Eq}(g,h)$ and let $q_u = g(u) = h(u)$.

The *position* p in $q = q_u$ is given by the factorization $q = q_1q_2$ with $|q_1| = p$. By $q[i,j]$, with $i \leqslant j$, we shall denote the factor of q spreading between positions i and j, i.e.

$$q = vq[i,j]v', \quad \text{with } |v| = i, \qquad \big|vq[i,j]\big| = j.$$

Clearly, $|q[i,j]| = j - i$.

By "ith occurrence of $g(b)$ ($h(b)$ respectively) in q" we mean the occurrence of the factor $g(b)$ ($h(b)$ respectively) in q which is the image of the ith occurrence of the letter b in u by the morphism g (h respectively).

We define integers $c_i \in \{0,1,2,\ldots,|u|\}$, $i = 1,2,\ldots,|u|_b$, as follows. Let u_i' and u_i be prefixes of u such that

$$u_i = u_i'b \quad \text{and} \quad \big|g(u_i)\big|_b = i.$$

Then

$$c_i = \big|g(u_i)\big| - \big|g(b)\big| = \big|g\big(u_i'\big)\big|.$$

The integer c_i is the *starting position* of ith occurrence of $g(b)$ in q. Similarly, we define the starting position of ith occurrence of $h(b)$ in q by

$$d_i = \big|h(u_i)\big| - \big|h(b)\big| = \big|h\big(u_i'\big)\big|.$$

Note that

$$c_{i+1} - c_i \geqslant \big|g(b)\big| \quad \text{and} \quad d_{i+1} - d_i \geqslant \big|h(b)\big|,$$

for each $i = 1,\ldots,|u|_b - 1$. Note also that

$$q\big[c_i + \big|g(b)\big|, c_{i+1}\big] \in g(a)^* \quad \text{and} \quad q\big[d_i + \big|h(b)\big|, d_{i+1}\big] \in h(a)^*,$$

for each $i = 1,\ldots,|u|_b - 1$.

The relation between the occurrences of $g(b)$ and $h(b)$ in q is given by the mappings

$$\Phi = \Phi_u,\ \Psi = \Psi_u : \{1,\ldots,|u|_b\} \to \{0,1,\ldots,|u|_b\}$$

defined as follows:

$$\Phi(i) = \begin{cases} j, & \text{if } d_j \leqslant c_i < d_j + |h(b)| \text{ for some } 1 \leqslant j \leqslant |u|_b, \\ 0, & \text{otherwise;} \end{cases}$$

$$\Psi(i) = \begin{cases} j, & \text{if } d_j < c_i + |g(b)| \leqslant d_j + |h(b)| \text{ for some } 1 \leqslant j \leqslant |u|_b, \\ 0, & \text{otherwise.} \end{cases}$$

The value of $\Phi(i)$ is 0 if the ith occurrence of $g(b)$ in q begins within $h(a)$ in the factorization of q induced by h, and $\Phi(i) = j$ if the ith occurrence of $g(b)$ in q begins within the jth occurrence of $h(b)$ in q.

The map Ψ has similar values for the positions in which the occurrences of $g(b)$ end.

The following lemma shows the way mappings Φ and Ψ will be used.

Lemma 40. *Let (g, h) be a typical pair of morphisms, and let $u \in bA^*b$ be an element of* $\mathrm{Eq}(g, h)$.

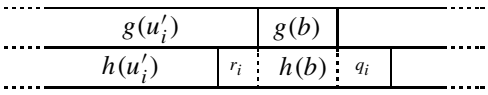
(A) *Let an integer $j \in \{1, \ldots, |u|_b\}$ be not in the range of Φ. Then $h(b)$ is either a factor of $g(a)^+$ or a prefix of $sg(a)^+$ for some proper suffix s of $g(b)$.*
(B) *Let an integer $j \in \{1, \ldots, |u|_b\}$ be not in the range of Ψ. Then $h(b)$ is either a factor of $g(a)^+$ or a suffix of $g(a)^+p$ for some proper prefix p of $g(b)$.*
(C) *Let*

$$\mathrm{Range}(\Psi) = \mathrm{Range}(\Phi) = \{1, \ldots, |u|_b\}.$$

Then

$$h(b) = r_i g(b) q_i, \qquad g\big(u_i'\big) = h\big(u_i'\big) r_i, \qquad g(u_i) q_i = h(u_i), \tag{18}$$

with $i = 1, \ldots, |u|_b$, and $r_i \in \mathrm{suff}(g(a)^+)$, $q_i \in \mathrm{pref}(g(a)^+)$.



Proof. Put $q = g(u) = h(u)$.

(A) Let $j \notin \mathrm{Range}(\Phi)$. By assumption, we have $\Phi(1) = 1$ and $\Phi(|u|_b) = |u|_b$. Therefore, there exists an integer i such that

$$c_i < d_j < d_j + \big|h(b)\big| \leqslant c_{i+1}.$$

This implies, looking at the jth occurrence of $h(b)$ in q, that $h(b)$ is a factor of $q[c_i + 1, c_{i+1}]$, which is a proper suffix of $g(b)g(a)^*$. The claim follows.

(B) Similarly as (A).

(C) Clearly,

$$d_i \leqslant c_i < c_i + \big|g(b)\big| \leqslant d_i + \big|h(b)\big|, \tag{19}$$

with $i = 1$. By assumption, within each occurrence of $h(b)$ in q, some occurrences of $g(b)$ start and some end. One can easily see, by induction, that occurrences of $g(b)$ starting and ending within one occurrence of $h(b)$ coincide. Therefore, Eq. (19) holds for each $i = 1, \ldots, |u|_b$. Thus $h(b) = r_i g(b) q_i$ for some $r_i, q_i \in A^*$. From injectivity of Φ we also deduce that r_i is a suffix and q_i a prefix of $g(a)^+$. □

6.5. The case $|g(ba)| < |h(b)|$

First adopt the following definitions.

Convention 41.

- Henceforward, if we speak about a counter-example, we implicitly suppose that it has properties described in Lemma 39.
- Let ξ denote the word σv_b or σv_a so that $\mathrm{pref}_1(\xi) = \mathrm{suff}_1(\xi) = b$ (see Lemma 39). In the rest of this section variables k, l, l' will have the following meaning:
  - b^l is the maximal b-prefix of σ,
  - b^k is the maximal b-prefix of $v_b\sigma$,
  - $b^{l'}$ is the maximal b-prefix of $\overline{\xi}$ (i.e. the maximal b-suffix of ξ).

Note that by (12), the word b^l (b^k respectively) is the proper prefix of σ ($v_b\sigma$ respectively). Also $b^{l'}$ is the proper suffix of ξ. Since σ is the common prefix of all elements in $\mathrm{Eq}(g,h)$, the word b^l is also the maximal b-prefix of ξ.

Lemma 42. *Let (g,h) be a counter-example. Then $g(b)^l$ is a proper prefix of $h(b)$ and $g(b)^{l'}$ is a proper suffix of $h(b)$.*

Proof. By (9), the words $h(b)$ and $g(b)^l$ are comparable. Since $g(b)$ is a suffix of $h(b)$, there exist a non-empty word u such that $h(b) = ug(b)$. If $h(b)$ were a prefix of $g(b)^l$, the words u and $g(b)$ would commute, by Lemma 7(C). This is a contradiction to Lemma 39(C). □

The proof of the second part of the statement is symmetric.

Lemma 43. *Let (g,h) be a counter-example and let $|g(ba)| < |h(b)|$. Then*

$$g\big(b^l a\big) \in \mathrm{pref}\big(h(b)\big) \quad \textit{and} \quad g\big(ab^{l'}\big) \in \mathrm{suff}\big(h(b)\big).$$

Proof. With $g(\xi) = h(\xi)$, it is enough to prove $|g(b^l a)| \leqslant |h(b)|$ and $|g(ab^{l'})| \leqslant |h(b)|$. Consideration for the two cases is mirror symmetric.

Proceed by contradiction and suppose $|g(b^l a)| > |h(b)|$. Since $|g(ba)| < |h(b)|$, $l \geqslant 2$ and the word $g(b^l a)$ is a prefix of $h(b)g(b)^{l-1}$. By Lemma 42, there are words u, q_1, and r_1 such that

$$g(b) = q_1 r_1, \qquad h(b) = g(b)^l u, \qquad g(a) = ug(b)^i q_1,$$

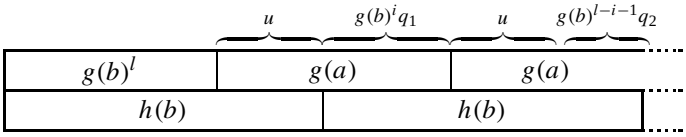
with $0 \leqslant i \leqslant l-2$.

1. Suppose that $b^l ab$ is a prefix of σ. Then $g(b)^i q_1 g(b)$ is a prefix of $g(b)^l$ and q_1 commutes with $g(b)$. This is a contradiction to $\overline{g}$ being marked. Similarly if $bab^{l'}$ is a suffix of ξ.

2. Thus $b^l aa$ is a prefix, and $aab^{l'}$ is a suffix of ξ.
   2.1. First suppose $|g(b^l aa)| > |h(bb)|$. Then

$$g(b) = q_2 r_2, \qquad h(b) = g(b)^i q_1 u g(b)^{l-i-1} q_2,$$

where $|q_2| = |r_1|$.



Since $l - i - 1 \geqslant 1$ and $g(b)$ is a suffix of $h(b)$, the word q_2 commutes with $g(b)$. Thus $r_1 = q_2$ and also q_1 commutes with $g(b)$, a contradiction to $\bar{g}$ being marked. Similarly we obtain contradiction if $|g(aab^{l'})| > |h(bb)|$.

   2.2. Suppose now

$$\left|g\left(b^l aa\right)\right| \leqslant \left|h(bb)\right|, \qquad \left|g(aab)^{l'}\right| \leqslant \left|h(bb)\right|.$$

Put $\Phi = \Phi_\xi$. Since $l \geqslant 2$, the range of Φ does not contain some $j \in \{1, 2, \ldots, |\xi|_b\}$. By Lemma 40, either $h(b)$ is a factor of $g(a)^+$ or a prefix of $sg(a)^+$ for some proper suffix s of $g(b)$.

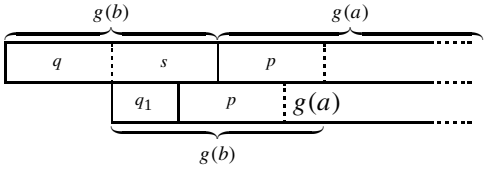
   2.2.1. Suppose $h(b)$ is a prefix of $sg(a)^+$.

      2.2.1.1. If $|s| \leqslant |g(b)^i q_1|$, then $sg(a)$ is a factor of $g(a)^+$. This implies that the word s is a suffix $g(a)^+$, a contradiction to $\bar{g}$ being marked.

      2.2.1.2. If, on the other hand, $|s| > |g(b)^i q_1|$, then clearly $i = 0$ and $q_1 g(a)$ is a prefix of $sg(a)$. Let

$$qs = q_1 r_1 = g(b).$$

Since $sg(a)$ is a prefix of $h(b)$, there is a prefix p of $g(a)$ such that $sp = g(b)$.



From $qs = sp$ follows the existence of a primitive word $t = t_1 t_2$ such that t_2 is non-empty,

$$s = (t_1 t_2)^{i_1} t_1, \qquad q = (t_1 t_2)^{j_1}, \qquad p = (t_2 t_1)^{j_1},$$

with $i_1 \in \mathbb{N}$, $j_1 \in \mathbb{N}_+$. Since $q_1 g(a)$ is a prefix of $h(b)$, the word $q_1 p$ is a prefix of $g(b)$. From

$$g(b) = (t_1 t_2)^{i_1+j_1} t_1, \qquad p = (t_2 t_1)^{j_1}$$

we deduce $q_1 \in (t_1 t_2)^* t_1$, and q_1 is a suffix of $g(b)$. This is a contradiction to $\overline{g}$ being marked, since q_1 is a suffix of $g(a)$.

2.2.2. Suppose $h(b)$ is a factor of $g(a)^+$. Let t_a be the primitive root of $g(a)$, and let $v_1 \in \mathrm{suff}(t_a)$ and $v_2 \in \mathrm{pref}(t_a)$ be words such that

$$h(b) \in \left(v_1 t_a^* v_2\right).$$

Since $g(b)^i q_1 t_a$ is a prefix of $h(b)$, it is also a prefix of $v_1 t_a^+$ and we conclude that

$$g(b)^i q_1 \in v_1 t_a^*.$$

Therefore, $h(bb)$ is a prefix of $g(b)^l t_a^+$. Similarly we deduce that $h(bb)$ is a suffix of $t_a^+ g(b)^{l'}$. Hence, by primitivity of t_a,

$$h(bbb) = g(b)^l t_a^m g(b)^{l'}$$

for some $m \in \mathbb{N}_+$. From

$$|t_a| + \left|g(b)\right| \leqslant \left|g(a)\right| + \left|g(b)\right| < \left|h(b)\right|,$$
$$3 \cdot \left|h(b)\right| = (l + l') \cdot \left|g(b)\right| + m \cdot |t_a|,$$

it is not difficult to deduce that either

$$(l + l') \cdot \left|g(b)\right| > \left|g(b)\right| + \left|h(b)\right| \quad \text{or} \quad m \cdot |t_a| > |t_a| + \left|h(b)\right|.$$

This implies, by Periodicity Lemma, that either of $g(b)$ or t_a commutes with $h(b)$. We thus obtain a contradiction to Lemma 39(C) or to $\mathrm{pref}_1(h(b)) \neq \mathrm{pref}_1(g(a))$ (see Eq. (11)). □

Lemma 44. *Let (g, h) be a counter-example such that $|g(ba)| < |h(b)|$. Let u be an element of* $\mathrm{Eq}(g, h)$. *Then*

$$\mathrm{Range}(\Phi_u) = \mathrm{Range}(\Psi_u) = \{1, 2, \ldots, |u|_b\}.$$

Proof. Suppose, for a contradiction, that $1 \leqslant j \leqslant |u|_b$ is not in the range of Φ_u. By Lemma 40, the word $h(b)$ is either a factor of $g(a)^+$ or a prefix of $s g(a)^+$ for some proper suffix s of $g(b)$.

1. If $h(b)$ is a factor of $g(a)^+$ then, by (9) and Lemma 43, the word $g(b)^l g(a)$ is a factor of $g(a)^+$. This implies, by Lemma 7(A), that $g(b)$ is a suffix of $g(a)^+$, a contradiction to $\overline{g}$ being marked.
2. Consider now the latter possibility. Let r be the word such that $rs = g(b)$. Observe that

$$(rs)^l g(a) \in \operatorname{pref}\big(sg(a)^+\big). \tag{20}$$

Define s' by $ss' = rs$. By (20), the word $s'(rs)^{l-1}g(a)$ is a prefix of $g(a)^+$. By Lemma 7(C), the words $s'(rs)^{l-1}$ and $g(a)$ commute. This is a contradiction to $\overline{g}$ being marked, since $s'(rs)^{l-1}$ is a suffix of $g(b)^l$.

We have proved $\operatorname{Range}(\Phi_u) = \{1, 2, \ldots, |u|_b\}$. The rest follows from mirror considerations. □

Lemma 45. *Let (g, h) be a counter-example such that $|g(ba)| < |h(b)|$. Let $u = x_1 w x_2$, with $x_1, x_2 \in A$ and $u \in A^+$, be an element of* $\operatorname{eq}(g, h)$. *Then $w \in a^*$.*

Proof. In this proof p_i (s_i respectively) will always denote a proper prefix (a proper suffix) of $g(a)$, and r_i, q_i, u_i, u'_i are like in (18).

Lemmas 43 and 44 imply

$$h(b) = g(b)q_1, \qquad h(b) = r_n g(b).$$

Suppose $|w|_b \geqslant 1$. Then

$$h(b) = r_2 g(b) q_2.$$

1. First suppose that both r_2 and q_2 are non-empty. Then we have

$$h(b) = g(b)g(a)^{m_1} p_1 = s_2 g(a)^{m_2} g(b) = s_3 g(a)^{m_3} g(b) g(a)^{m_4} p_4,$$

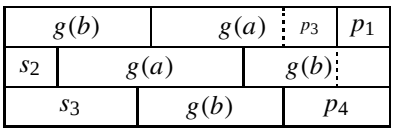
with $m_1, m_2, m_3, m_4 \in \mathbb{N}$. Since $g(a)^{m_3} r$ is a factor of $s_2 g(a)^{m_2}$ for a non-empty prefix r of $g(b)$, Lemma 7(B) and g being marked imply that $m_3 = 0$. The mirrored consideration yields $m_4 = 0$.

Hence $|h(b)| < |g(b)| + 2 \cdot |g(a)|$, and therefore $m_1 = m_2 = 1$. We can write

$$h(b) = g(b)g(a)p_1, \tag{21}$$

$$h(b) = s_2 g(a) g(b), \tag{22}$$

$$h(b) = s_3 g(b) p_4, \tag{23}$$



where $|s_2| < |s_3|$ and $|p_1| < |p_4|$. From (21) and (23) we deduce $p_4 = p_3p_1$ and $g(b)g(a) = s_3g(b)p_3$, with $p_3s_3 = g(a)$. Hence $g(b)p_3s_3 = s_3g(b)p_3$, and words $g(b)p_3$ and s_3 have a common primitive root, say t. Let $t = t_1t_2$ be a factorization of t such that

$$g(b) = (t_1t_2)^{i_1}t_1, \qquad p_3 = t_2(t_1t_2)^{i_2}, \qquad s_3 = (t_1t_2)^j,$$

with $i_1, i_2, j \in \mathbb{N}$, $j \geqslant 1$. Then also

$$g(a) = p_3s_3 = (t_2t_1)^{i_2+j}t_2, \qquad g(b)g(a) = (t_1t_2)^{i_1+i_2+j+1},$$
$$g(a)g(b) = (t_2t_1)^{i_1+i_2+j+1}.$$

From (22) and (21) it follows that $s_2(t_2t_1)$ is a prefix of $g(b)g(a)$ and thus

$$s_2 = (t_1t_2)^{i_3}t_1, \qquad h(b) = s_2g(a)g(b) = (t_1t_2)^{i_1+i_2+i_3+j+1}t_1,$$

with $i_3 \geqslant 0$. Equality (23) gives

$$p_4 = (t_1t_2)^{i_2+i_3+1}$$

and, since p_4 is a prefix of $g(a)$, the words t_1 and t_2 commute. Therefore, also $g(a)$ and $g(b)$ commute; a contradiction.

2. If, on the other hand, either of r_2 or q_2 is empty, then

$$g\big(u_2'\big) = h\big(u_2'\big) \quad \text{or} \quad g(u_2) = h(u_2).$$

This contradicts the minimality of x_1wx_2. □

Lemma 46. *Let (g, h) be a counter-example. Then $|g(ba)| \geqslant |h(b)|$.*

Proof. Suppose $|g(ba)| < |h(b)|$. By Lemma 39, (B) and (D),

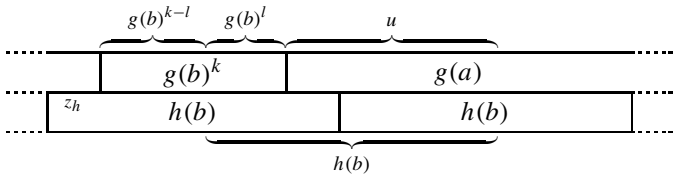
$$\mathrm{pref}_1(\sigma) = \mathrm{suff}_1(\sigma) = b.$$

Lemma 45 applied to σv_a implies $\sigma = b$, a contradiction with $|g(\sigma)| = |h(\sigma)z_h| > |h(\sigma)|$. □

6.6. The case $|g(ba)| \geqslant |h(b)|$

Recall Convention 41. Following two lemmas, a more complicated parallel of Lemma 32, claim that in the given case the word $h(b)$ commutes with the word $z_hg(b)^{k-l}$. The two lemmas correspond to different signs of $k - l$.

Lemma 47. *Let (g,h) be a counter-example and let $k>l$. Then $h(b)$ commutes with the word $z_h g(b)^{k-l}$.*



Proof. The assumption implies $k \geqslant 2$. From (9), Lemmas 42 and 46 we deduce that $h(b) = g(b)^l u$ for some prefix u of $g(a)$. Since $|h(b)| > |g(b)|$, we have

$$\left|h(b)^k z_h\right| > \left|z_h g(b)^{k-l} h(b)\right|.$$

Equality (10) now implies that the word $z_h g(b)^{k-l} g(b)^l u = z_h g(b)^{k-l} h(b)$ is a prefix of $h(b)^+$ and thus $z_h g(b)^{k-l}$ commutes with $h(b)$. □

Lemma 48. *Let (g,h) be a counter-example and let $k \leqslant l$. Then*

$$z_h = s g(b)^{l-k}$$

for some word $s \in A^*$, which commutes with $h(b)$.

Proof. Let u be a prefix of $g(a)$ such that $g(b)^l u = h(b)$. Thus

$$\left|g(b)^l u g(b)^{l-k}\right| < \left|g(b^l a)\right|$$

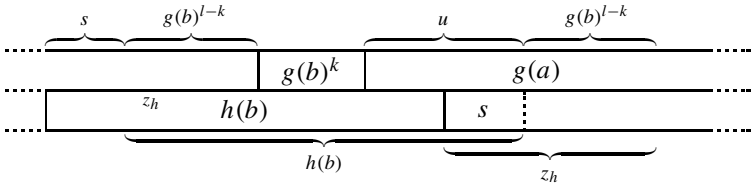
and, by (9), $u g(b)^{l-k}$ is a prefix of $g(a)$. From (10) we deduce

$$h(b) z_h = z_h g(b)^k u g(b)^{l-k}. \tag{24}$$

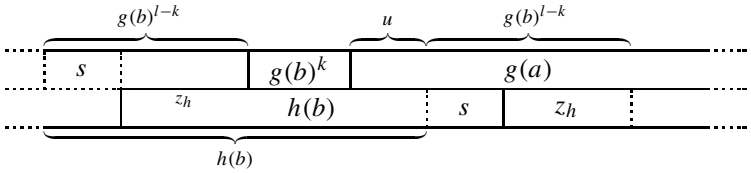
1. First suppose $|z_h| \geqslant |g(b)^{l-k}|$. Equality (24) yields $z_h = s g(b)^{l-k}$ for some $s \in A^*$ and it reads

$$h(b) s g(b)^{l-k} = s g(b)^{l-k} g(b)^k u g(b)^{l-k} = s h(b) g(b)^{l-k}.$$

Thus the words $h(b)$ and s commute and we are through.
Note that the previous considerations pass smoothly even if $k = l$. The case $l = 1$ (and thus $k = 1$) deserves special attention.



2. Suppose now $|z_h| < |g(b)^{l-k}|$ and, consequently, $k < l$, $l \geqslant 2$. Equality (24) implies the existence of a non-empty word $s \in A^+$ such that $sz_h = g(b)^{l-k}$ and $h(b) = z_h g(b)^k us$.



Therefore,

$$sh(b) = sz_h g(b)^k us = g(b)^{l-k} g(b)^k us = h(b)s.$$

Thus, the words $h(b)$, s, and $z_h g(b)^k u$ have the same primitive root, say t. From $sz_h = g(b)^{l-k}$, we have $|t| < |g(b)^{l-k}|$. Since $g(b)^l = sz_h g(b)^k$ is a prefix of t^+, by Periodicity Lemma t is the primitive root of $g(b)$, a contradiction to Lemma 39(C). □

A consequence of Lemmas 47 and 48 appears as two lemmas.

Lemma 49. *Let (g, h) be a counter-example such that $k \geqslant l$. Then either*

$$z_h g\big(b^{k-l}\big) = h\big(b^{k-l}\big) \tag{25}$$

or there exist words $v \in A^+$ and $w \in b^+$ such that

$$z_h g(v) = h(w) z_h. \tag{26}$$

Proof. By Lemma 47, words $h(b)$ and $z_h g(b^{k-l})$ have a common primitive root, say t. Let

$$h(b) = t^{k_1}, \qquad z_h g\big(b^{k-l}\big) = t^{k_2}.$$

The word $h(b)^k z_h = t^{k \cdot k_1} z_h$ is the maximal t-prefix of $h(v_b \sigma)$, and therefore also of $z_h g(v_b \sigma)$. Since b^{k-l} is a prefix of $v_b \sigma$,

$$\text{the word} \quad t^{k \cdot k_1 - k_2} z_h \quad \text{is the maximal } t\text{-prefix of} \quad g\big(b^{-(k-l)} v_b \sigma\big). \tag{27}$$

From (9) one can similarly deduce that

$$\text{the word} \quad t^{l \cdot k_1} z_h \quad \text{is the maximal } t\text{-prefix of} \quad g(\sigma v_a). \tag{28}$$

1. First suppose $k \cdot k_1 - k_2 = l \cdot k_1$. Then one easily verifies $z_h g(b^{k-l}) = h(b)^{(k-l)}$.
2. Suppose then $k \cdot k_1 - k_2 \neq l \cdot k_1$ and put

$$m = \min\{k \cdot k_1 - k_2, l \cdot k_1\}.$$

From (27) and (28) we deduce, by Lemma 8, that $t^m z_h = g(u)$ for some $u \in A^+$, and

$$z_h g\big(b^{k-l}u\big) = t^{m+k_2} z_h.$$

Then also

$$z_h g\big(\big(b^{k-l}u\big)^{k_1}\big) = t^{k_1(m+k_2)} z_h = h(b)^{m+k_2} z_h,$$

and we are through. □

If $k < l$, the first possibility is excluded.

Lemma 50. *Let (g, h) be a counter-example such that $k < l$. Then*

$$z_h g(v) = h(w) z_h$$

for some $v \in A^+$ and $w \in b^+$.

Proof. This proof is essentially the same as the proof of Lemma 49, with $(k - l)$ negative.

By Lemma 48, the word $h(b)$ commutes with $s = z_h g(b)^{-(l-k)}$. Let t be the primitive word and

$$h(b) = t^{k_1}, \qquad s = t^{k_2}.$$

Inequality $k < l$ yields $k \cdot k_1 - k_2 < l \cdot k_1$. Verify that statements (27) and (28) hold. Therefore, by Lemma 8, there is a word $u \in A^+$ such that $g(u) = t^{k\cdot k_1 - k_2} z_h$, and

$$sg(u) = t^{k\cdot k_1} z_h = h(b)^k z_h.$$

Since $sg(b)^{l-k} = z_h$ is a prefix of $t^{k\cdot k_1} z_h$, the word b^{l-k} is a prefix of u. Thus we can write

$$z_h g\big(b^{-(l-k)}u\big) = h(b)^k z_h. \qquad \square$$

6.7. Shortest counter-examples

In this subsection we shall exploit the fact the counter-example can be supposed to be a shortest one. Obviously, if any counter-example exists, there is also a shortest one. A contradiction will be obtained by showing that every counter-example can be shortened.

Next lemma deals with possibilities suggested by Lemmas 49 and 50.

Lemma 51. *Let (g, h) be a shortest counter-example.*

(A) *If $(e, f) \in \mathbf{c}(g, h_{\mathrm{m}})$ then $f \notin b^+$.*
(B) *If $k > l$ and $z_h g(b^{k-l}) = h(b^{k-l})$, then the pair (g, h_m) is simple (i.e. $g(e) = h_m(f) \Rightarrow e = f$).*

Proof.

1. If the pair (g, h_m) is simple then both claims hold, as is easy to see.
2. By Lemma 31, we can assume that there exist words $e \neq f$ and $e' \neq f'$ such that

$$\mathbf{c}(g, h_{\mathrm{m}}) = \big\{(e, f), (e', f')\big\}, \tag{29}$$

with $(e, f) \neq (e', f')$.
Define marked morphisms $g_1, h_1 : A^* \to A^*$ by

$$\begin{cases} g_1(a) = e, \\ g_1(b) = e', \end{cases} \qquad \begin{cases} h_1(a) = f, \\ h_1(b) = f'. \end{cases}$$

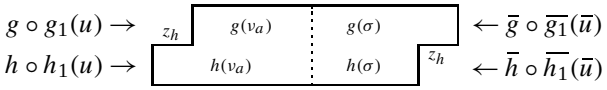
By Lemma 25, there are words u, v such that

$$g_1(u) = h_1(u) = v_a\sigma, \qquad g_1(v) = h_1(v) = v_b\sigma.$$

Then

$$\overline{g_1}(\bar{u}) = \overline{h_1}(\bar{u}) = \overline{v_a\sigma} = \bar{\sigma}\,\overline{v_a}, \qquad \overline{g_1}(\bar{v}) = \overline{h_1}(\bar{v}) = \overline{v_b\sigma} = \bar{\sigma}\,\overline{v_b},$$

and $\bar{u}$, $\bar{v}$ are distinct elements of $\mathrm{Eq}(\overline{g_1}, \overline{h_1})$. The length of $\bar{u}$ and $\bar{v}$ is at least two, because g and h are not simple.



By Lemma 22, there exists a typical pair of morphisms (g', h') such that

$$\bar{u}, \bar{v} \in \mathrm{Eq}(g', h') = \mathrm{Eq}\big(\overline{g_1}, \overline{h_1}\big).$$

Since (g, h) is a shortest counter-example, from

$$|\bar{u}| = |u| < |\sigma v_a| \quad \text{and} \quad |\bar{v}| = |v| < |\sigma v_b|$$

we deduce $\mathrm{pref}_1(\bar{u}) \neq \mathrm{pref}_1(\bar{v})$. By construction of $\overline{g_1}$ and $\overline{h_1}$, either the words $\overline{h_1}(a) = \overline{f}$ and $\overline{h_1}(b) = \overline{f'}$ are comparable, or $\bar{\sigma}$ is a proper prefix of both $\overline{f}$ and $\overline{f'}$.
2.1. Consider the first possibility. By (29) ,

$$\bar{g}(\bar{e}) = \overline{h_{\mathrm{m}}}\big(\overline{f}\big), \qquad \bar{g}\big(\overline{e'}\big) = \overline{h_{\mathrm{m}}}\big(\overline{f'}\big),$$

and the pairs $(\bar{e}, \overline{f})$ and $(\overline{e'}, \overline{f'})$ are minimal elements of $\mathbb{C}(\bar{g}, \overline{h_{\mathrm{m}}})$. Suppose, by symmetry, that $\overline{f}$ is a prefix of $\overline{f'}$. Since $\bar{g}$ is marked, we conclude that also $\bar{e}$ is a prefix of $\overline{e'}$, a contradiction to minimality of $(\overline{e'}, \overline{f'})$.
2.2. Thus $\bar{\sigma}$ is a proper prefix of both $\overline{f}$ and $\overline{f'}$.
The claim (A) now follows from $|\sigma|_a \geqslant 1$.

The assumptions of (B) imply $v_b = b^{k-l}$, whence

$$\overline{h_1}(\bar{v}) = \bar{\sigma} b^{k-l}.$$

Put $x = \mathrm{pref}_1(\bar{v})$. The present assumption (2.2) implies that $\bar{\sigma}$ is a prefix of $\overline{h_1}(x)$. Moreover, $|\bar{v}| \geqslant 2$ and thus

$$\overline{h_1}\big(x^{-1}\bar{v}\big) \in b^+.$$

This yields that either $\overline{f}$ or $\overline{f'}$ is in b^+, a contradiction to $|\sigma|_a \geqslant 1$. □

We are left with the final case, described in the following lemma.

Lemma 52. *Let (g, h) be a shortest counter-example. Then*

(A) *(g, h_m) is simple,*
(B) *$k > l$,*
(C) *$v_b = b^{k-l}$,*
(D) *$|z_h| \geqslant |h(b)| - |g(b)^{l'-1}| > |g(b)^l|$.*

Proof. (B), (C). The possibility $k < l$ is excluded by Lemmas 50 and 51(A). The possibility (26) of Lemma 49 is also in contradiction with Lemma 51(A). Therefore, the possibility (25) remains. The minimality of σv_b yields $v_b = b^{k-l}$ and $k > l$.

(A) Follows from Lemma 51(B).

(D) From the facts that $g(b)^l$ is a prefix of $h(b)$, $g(b)^{l'}$ is a suffix of $h(b)$, and the words $h(b)$ and $g(b)$ do not commute, we deduce, by Periodicity Lemma,

$$\big|h(b)\big| + \big|g(b)\big| > \big|g(b)^l\big| + \big|g(b)^{l'}\big|.$$

This implies the second inequality.

Note that the word ξ from Convention 41 is equal to σb^{k-l}. Moreover, $\mathrm{suff}_1(\sigma) = b$ and thus $l' - 1 \geqslant k - l \geqslant 1$. The first inequality now follows directly from

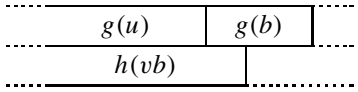
$$|z_h| = \big|h(b)^{k-l}\big| - \big|g(b)^{k-l}\big|. \qquad \square$$

We present two more combinatorial lemmas.

Lemma 53. *Let g, h be binary morphisms and let w be an element of* $\mathrm{Eq}(g, h)$. *Let*

$$ub, vb \in \mathrm{pref}(w)$$

be words such that $g(u)$ is a proper prefix of $h(vb)$ and $h(vb)$ is a proper prefix of $g(ub)$. Then (g, h) is not a shortest counter-example.



Proof. Suppose, for a contradiction, that (g, h) is a shortest counter-example. Let u' and v' be sufficiently long words, such that

$$g(ubu') = h(vbv').$$

Let s_1 be a proper suffix of $h(b)$ such that

$$g(u)s_1 = h(vb), \tag{30}$$

and p_1 be a proper prefix of z_h such that

$$g(ub) = h(vb)p_1.$$

Since $g(b)$ is a prefix of z_h and a suffix of $h(b)$, we obtain

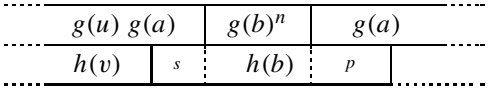
$$g(b) = s_1p_1 = p_1s_1,$$

and the words $g(b)$, s_1, and p_1 have the same primitive root, say r. The maximal r-suffix of $h(vb)$ is equal to the maximal r-suffix of $h(b)$, i.e. to $g(b)^{l'}$. By (30), the word $g(b)^{l'}$ is also the maximal r-suffix of $g(u)s_1$. This is in contradiction with $\overline{g}$ and $\overline{h}$ being marked. □

Lemma 54. *Let g and h be binary morphisms and let w be an element of* $\mathrm{Eq}(g, h)$. *Let n be a positive integer and u, v words such that*

$$vb, uab^na \in \mathrm{pref}(w),$$

where $h(v)$ is a proper prefix of $g(ua)$ and $g(uab^n)$ is a proper prefix of $h(vb)$. Then (g, h) is not a shortest counter-example.



Proof. Proceed by contradiction, and suppose that (g, h) is a shortest counter-example. Let u', v' be sufficiently long words such that

$$g(uab^nau') = h(vbv').$$

The assumptions imply that $g(b)^n$ is a proper factor of $h(b)$,

$$sg(b^nau') = h(bv'), \tag{31}$$

and

$$h(b) = sg(b)^np, \tag{32}$$

for a proper suffix s and a proper prefix p of $g(a)$. From the facts that g and $\overline{g}$ are marked, $g(b)^l \in \mathrm{pref}(h(b))$, $g(b)^{l'} \in \mathrm{suff}(h(b))$, Lemma 7, and Eq. (32) we deduce the following inequalities:

$$|s| > \left|g(b)^{l-1}\right|, \qquad |p| > \left|g(b)^{l'-1}\right|.$$

Thus, by Lemma 52(D),

$$\left|sg(b)^n\right| \leqslant |z_h|. \tag{33}$$

Recall that $|g(b)g(a)| > |h(b)|$. Let w be the prefix of $g(b)g(a)$ of length $h(b)$. From (9) and from $|g(b)^l| < |z_h|$ (see Lemma 52(D)) we deduce

$$g(b)^{l-1}w \in \mathrm{pref}\big(h(b)z_h\big). \tag{34}$$

Since

$$sg(b^n a) \text{ Pref } h(b)z_h,$$

the inequality (33) implies

$$sg(b)^{n-1}w \in \mathrm{pref}\big(h(b)z_h\big). \tag{35}$$

Lemma 7(D), (34) and (35) now yield that $g(b)^{l-1}$ is a suffix of $sg(b)^{n-1}$ and $sg(b)^{n-l}$ commutes with $h(b)$. Hence $n \geqslant l$, because s is a suffix of $g(a)$. Denote by r the common primitive root of $h(b)$ and $sg(b)^{n-l}$ and let

$$h(b) = r^{k_1}, \qquad sg(b)^{n-l} = r^{k_2}, \qquad \text{with } k_1 > k_2 > 0.$$

Equality (9) implies that

$$\text{the maximal } r\text{-prefix of } g(\sigma v_a) = h(\sigma v_a) \text{ is } r^{k_1 \cdot l} z_h. \tag{36}$$

Let b^m be the maximal b-prefix of bv'. It follows from (31) that

$$\text{the maximal } r\text{-prefix of } g\big(b^l a u'\big) \text{ is } r^{k_1 \cdot m - k_2} z_h. \tag{37}$$

From (36) and (37) we deduce, by Lemma 8, that there is a word u_1 such that

$$g(u_1) = r^{k_3} z_h,$$

with $k_3 = \min\{k_1 \cdot l, k_1 \cdot m - k_2\}$. Therefore,

$$z_h g\big(b^{k-l} u_1\big) = h(b)^{k-l} r^{k_3} z_h = r^{(k-l)\cdot k_1 + k_3} z_h$$

and

$$z_h g\big(\big(b^{k-l}u_1\big)^{k_1}\big) = h\big(b^{(k-l)\cdot k_2+k_3}\big)z_h,$$

in contradiction with Lemma 51(A). □

The whole section is concluded by the following lemma. It shows that the possibility excluded by Lemma 54 has to take place in a shortest counter-example. That yields the final contradiction.

Lemma 55. *Let (g,h) be a shortest counter-example. Then there exist words w, u, and v, and a positive integer n such that w is an element of* $\mathrm{eq}(g,h)$ *and*

$$vb, uab^n a \in \mathrm{pref}\big(\mathrm{eq}(g,h)\big),$$

where $h(v)$ is a proper prefix of $g(ua)$ and $g(uab^n)$ is a proper prefix of $h(vb)$.

Proof. Consider the word σv_b. Since $\underline{z_h}$ is empty and since

$$\mathrm{suff}_1(v_b) = \mathrm{suff}_1\left(b^{k-l}\right) = b,$$

we conclude that

$$\mathrm{pref}_1(v_a) = \mathrm{suff}_1(v_a) = a. \tag{38}$$

First we want to show that if w_1 and w_2 are proper prefixes of v_a, then

$$g(\sigma w_1) \neq h(\sigma w_2).$$

Suppose the contrary. The minimality of σv_a implies $w_1 \neq w_2$. From

$$z_h g(w_1\sigma) = h(w_2\sigma)z_h$$

we deduce that (g,h) is not simple, a contradiction to Lemma 52(A).

Put $m = |v_a|_b$. From $|g(v_a)| < |h(v_a)|$ we deduce $m \geqslant 1$. Define words u_i, v_i, $i = 1,\ldots,m$, by

$$u_i v_i = \sigma v_a, \qquad \mathrm{pref}_1(v_i) = b, \qquad |v_i|_b = m - i + 1.$$

Inequalities

$$\big|g(\sigma b)\big| \geqslant \big|h(\sigma b)\big| \quad \text{and} \quad \big|g(a)\big| > \big|h(a)\big|$$

imply

$$\big|g(u_1 b)\big| > \big|h(u_1 b)\big|, \qquad \big|g(u_m b)\big| < \big|h(u_m b)\big|.$$

Let $j \in \{2, 3, \ldots, m\}$ be the smallest integer such that

$$\big|g(u_{j-1}b)\big| > \big|h(u_{j-1}b)\big|, \qquad \big|g(u_jb)\big| < \big|h(u_jb)\big|.$$

$|g(u_j)| > |h(u_j)|$, since $u_j \in u_{j-1}ba^*$ and $|g(a)| > |h(a)|$. Thus

$$\big|h(u_j)\big| < \big|g(u_j)\big| < \big|g(u_jb)\big| < \big|h(u_jb)\big|. \tag{39}$$

Let b^{m_1} be the maximal b-suffix of u_j and b^{m_2} be the maximal b-prefix of v_j. By Lemma 53,

$$\big|g\big(u_jb^{m_2}\big)\big| < \big|h(u_jb)\big|.$$

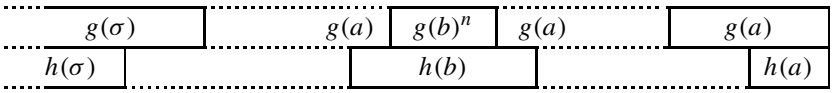
If $m_1 > 0$ then, again by Lemma 53,

$$\big|g\big(u_jb^{-m_1}\big)\big| > \big|h(u_j)\big|.$$

If $m_1 = 0$, then the last inequality is contained in (39). Put

$$u = u_ja{b^{m_1}}^{-1}, \qquad v = u_j, \qquad n = m_1 + m_2,$$

and verify that they satisfy the assumptions. This completes the proof.



□

We can summarize this section by proving Theorem 2 from the introduction.

Proof of Theorem 2. Let g and h be distinct non-periodic binary morphisms. By Lemma 22, we can assume that (g, h) is typical (Definition 19).

For a contradiction suppose that $\mathrm{eq}(g, h)$ contains distinct words u and v such that $\mathrm{pref}_1(u) = \mathrm{pref}_1(v)$, and suppose that (g, h) is shortest possible (Definition 29). The present section shows that the assumption is contradictory.

To prove that $\mathrm{eq}(g, h)$ does not contain two distinct words u' and v' with $\mathrm{suff}_1(u') = \mathrm{suff}_1(v')$, consider mirror morphisms $\overline{g}$ and $\overline{h}$. □

7. Test sets

In this section we show that each binary language has a test set of cardinality at most two. We follow the exposition from [8], where a three element test set is constructed. Our improvement is a direct consequence of Theorem 2.

Test set of a language $L \subset \Sigma^*$ is a subset T of L such that the agreement of two morphisms on the language T guarantees their agreement on L. Formally, for any two morphisms g and h defined on Σ^*,

$$(\forall u \in T)\ \big(g(u) = h(u)\big) \quad \Rightarrow \quad (\forall v \in L)\ \big(g(v) = h(v)\big).$$

Let $L \subset A^*$ be a binary language. The *ratio* of a non-empty word $u \in L$ is denoted by $r(u)$ and defined by

$$r(u) = \frac{|u|_a}{|u|_b}.$$

If $|u|_b = 0$, then $r(u) = \infty$. A word u is said to be *ratio-primitive* if no proper prefix of u has the same ratio as u. Note that each word has a unique factorization into ratio-primitive words (or shortly, ratio-primitive factorization).

Theorem 56. *Let $L \subset A^*$ be a language. Then L possesses a test set of cardinality at most two.*

Proof. Let g and h be binary morphisms. We can assume $|g(a)| \neq |h(a)|$ and $|g(b)| \neq |h(b)|$ (the discussion of the remaining cases is trivial).

Clearly, morphisms g and h can agree on a word u only if they agree lengthwise on it and one easily sees that it is equivalent to

$$r(u) = \frac{|h(b)| - |g(b)|}{|g(a)| - |h(a)|}.$$

This also implies that if $u = u_1 u_2 \cdots u_n$ is the ratio-primitive factorization of u, then $g(u) = h(u)$ if and only if $g(u_i) = h(u_i)$, $i = 1, \ldots, n$. Therefore, g and h agree on L if and only if they agree on language L_r consisting of all ratio-primitive words occurring in ratio-primitive factorization of all elements in L. Moreover, any test set of L_r can be transformed into a test set of L of the same or smaller cardinality: it is enough to assign to each word $u \in L_r$ a word $v \in L$ such that u is contained in the ratio-primitive factorization of v.

The above considerations allow to restrict ourselves to languages consisting of ratio-primitive words. The proof is based on the observation that in such a case, if g and h agree on L, each element of L is in $\mathrm{eq}(g, h)$.

1. If L contains at most two words, we are trivially through.
2. If L contains two words with different ratio, then only morphisms $g = h$ can agree on L and the two words constitute a test set.
3. Suppose that cardinality of L is at least three and all words have the same ratio. Let $T = \{u, v\}$ with $u, v \in L$, $u \neq v$, and $\mathrm{pref}_1(u) = \mathrm{pref}_1(v)$. We claim that T is a test set of L.

   3.1. If both morphisms are periodic, then any single word constitutes a test set.

3.2. If just one morphism is periodic then g and h do not agree on L, by Theorem 1(B), and any two words constitute a test set.
3.3. If both morphisms are non-periodic, they agree on L just in case $g = h$, by Theorem 2. Again by Theorem 2, the two words in T constitute a test set, since $\mathrm{pref}_1(u) = \mathrm{pref}_1(v)$. □

Remark 57. The only known equality languages generated by two words are of the form

$$L = \{a^i b, ba^i\},$$

with $i \in \mathbb{N}_+$ (see [6]). Some partial results of this paper indicate that no other such languages exist. This suggests a direction of further research.

Acknowledgments

A nucleus of this paper was a part of author's PhD thesis supervised by Aleš Drápal [10]. Hunting up all emerging cases, however, took another year and a half. The proof was completed during the postdoctoral stay in Turku granted by Turku Centre for Computer Science (TUCS). The author is grateful especially to Juhani Karhumäki for making this stay possible.

The author wants to thank Vesa Halava, Tero Harju, Juhani Karhumäki, and Juha Kortelainen for useful comments.

References

[1] C. Choffrut, J. Karhumäki, Combinatorics on words, in: G. Rozenberg, A. Salomaa (Eds.), Handbook of Formal Languages, Vol. 1, Springer-Verlag, Berlin, 1997, pp. 329–438.
[2] A. Ehrenfeucht, J. Karhumäki, G. Rozenberg, The (generalized) Post correspondence problem with lists consisting of two words is decidable, Theoret. Comput. Sci. 21 (1982) 119–144.
[3] V. Halava, T. Harju, Some new results on Post correspondence problem and its modifications, Bull. Eur. Assoc. Theor. Comput. Sci. 73 (2001) 131–141.
[4] V. Halava, T. Harju, M. Hirvensalo, Generalized Post correspondence problem for marked morphisms, Internat. J. Algebra Comput. 10 (2000) 757–772.
[5] T. Harju, J. Karhumäki, Morphisms, in: G. Rozenberg, A. Salomaa (Eds.), Handbook of Formal Languages, Vol. 1, Springer-Verlag, Berlin, 1997, pp. 439–510.
[6] K. Čulík II, J. Karhumäki, On the equality sets for homomorphisms on free monoids with two generators, RAIRO Theor. Inform. 14 (1980) 349–369.
[7] J. Berstel, D. Perrin, J.F. Perrot, A. Restivo, Sur le théorème du défaut, J. Algebra 60 (1979) 169–180.
[8] A. Ehrenfeucht, J. Karhumäki, G. Rozenberg, On binary equality sets and a solution to the test set conjecture in the binary case, J. Algebra 85 (1983) 76–85.
[9] A. Salomaa, Equality sets for homomorphisms of free monoids, Acta Cybernet. 4 (1978) 127–139.
[10] Š. Holub, Equations in free monoids, PhD thesis, Charles University, Prague, 2000.