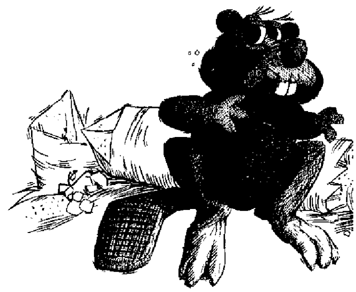
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT



Inverse Morphic Equivalence on Languages

Juhani Karhumäki
Derick Wood

Data Structuring Group
CS-84-06

February, 1984

INVERSE MORPHIC EQUIVALENCE ON LANGUAGES [(1)]

Juhani Karhumäki [(2)]

Derick Wood [(3)]

ABSTRACT

We introduce the notion of inverse morphic equivalence of two morphisms g and h on a language L. Two variants are considered, the universal version, that is $h^{-1}(x) = g^{-1}(x)$, for all x in L, and the existential version, that is $h^{-1}(x) \cap g^{-1}(x) \neq \emptyset$, for all x in L with $h^{-1}(x) \cup g^{-1}(x) \neq \emptyset$.

Keywords: morphism; homomorphism; inverse morphism; equality set; inverse morphic equivalence; test set.

1. INTRODUCTION

In [CS1] the notion of the agreement of two morphisms on a language was abstracted from previous work on the equivalence of iterated morphisms, see [RS] for example. This work heralded the subsequent studies of morphisms in which well known families were characterized morphically [C2], the Ehrenfeucht Conjecture was confirmed in a number of cases, [EKR] and [AW], and the relationship with algebraic systems of equations was explored, [CK2] and [ACK]. More recently, in [CFS], the family of regular languages has been shown to be generable by sequences of morphisms and inverse morphisms of length four applied to the language 0^*1. This has been generalized to arbitrary families in [KL] and [T], while [LL] have studied the generative power of sequences of morphisms and inverse morphisms. In particular they prove that a sequence of length greater than four has an equivalent sequence of length four. It is these two avenues of

---

(1) The work of the first author was supported by the Academy of Finland and by the Magnus Ehrnrooth Foundation, while that of the second author was supported under a Natural Sciences and Engineering Research Council of Canada Grant No. A-5692.

(2) Department of Mathematics, University of Turku, SF-20500 Turku 50, Finland.

(3) Data Structuring Group, Computer Science Department, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada.

research that led us to study inverse morphic equivalence on languages. Inverse morphisms may be viewed as a restricted class of nondeterministic generalized sequential machines (ngsms). The equivalence of ngsms and other transducer models has been studied in [C1] for example; where equivalence means equality of the collections of input-output word pairs. Inverse morphisms correspond to functional ngsms.

Although we only study inverse morphic equivalence in this paper, equivalences may be studied over any class of mappings, for example mappings of the form gh^{-1} , $g^{-1}h$, or even $fg^{-1}h$; where f,g, and h are morphisms. Recently this suggestion has been followed in [KK] and [KM]. Clearly the equivalence problem may also be studied when different types of mappings are to be applied, for example a morphism and an inverse morphism.

2. PRELIMINARIES

Our basic notions in this note are those of a morphism or homomorphism from one finitely generated free monoid into another and the inverse relation of it. We call a homomorphism $h : \Sigma^* \to \Delta^*$ *periodic* if there exists an element u in Δ^* such that $h(\Sigma) \subseteq u^*$, and *binary* if the cardinality of Σ equals two. We say that h has *bounded delay* p (from left to right), for some $p \geq 0$, if the following holds for all words u and v in Σ^* and all letters a and b in Σ:

$$\left.\begin{array}{l} h(au) \prec h(bv) \\ |u| \geq p \end{array}\right\} \text{ implies } a = b$$

where $\prec$ means: "is a prefix of" and $|u|$ denotes the length of the word u.

Let g and h be morphisms from Σ^* into Δ^* and L a language over Δ. We say that the inverse morphisms $g^{-1}, h^{-1} : \Delta^* \to 2^{\Sigma^*}$ *universally* (resp. *existentially*) *agree* on L if $g^{-1}(x) = h^{-1}(x)$ for all x in L (resp. $g^{-1}(x) \cap h^{-1}(x) \neq \emptyset$ for all x in $L \cap (h(\Sigma^*) \cup g(\Sigma^*))$). Letting $L = \Delta^*$ we obtain the notions of the *universal inverse equality language of morphisms* g and h, denoted by $IE_\forall(g,h)$, and the *existential inverse equality language* of g and h, denoted by $IE_\exists(g,h)$, as follows:

$$IE_\forall(g,h) = \{ x \in \Delta^* \mid g^{-1}(x) = h^{-1}(x) \neq \emptyset \}$$

and

$$IE_\exists(g,h) = \{ x \in \Delta^* \mid g^{-1}(x) \cap h^{-1}(x) \neq \emptyset \} .$$

Observe that if g and h are injective, then $IE_\forall(g,h) = IE_\exists(g,h)$. Moreover, the inverse equality languages are connected to ordinary equality languages, denoted by $E(g,h)$ and defined as $\{x \in \Sigma^* : g(x) = h(x)\}$, in the following way:

Lemma 2.1: *For all pairs* (g,h) *of morphisms from* Σ^* *into* Δ^* *we have:*

$$IE_\forall(g,h) \subseteq IE_\exists(g,h) = h(E(g,h)) .$$

Proof: The inclusion is clear by the definitions. To prove the equality we first assume that $x \in IE_\exists(g,h)$. This means that for some word y we have $y \in h^{-1}(x) \cap g^{-1}(x)$ which implies that $h(y) = x = g(y)$, that is $y \in E(g,h)$. Consequently, $x \in h(E(g,h))$. Conversely, if $x \in h(E(g,h))$, then there exists a word y in $E(g,h)$ such that $x = h(y)$, which means that $h(y) = x = g(y)$. Therefore, $x \in IE_{\exists}(g,h)$. □

A well known conjecture of Ehrenfeucht, cf. [CS2], states that for each language L over Σ there exists a finite subset F of L such that to test whether or not two arbitrary morpisms g and $h : \Sigma^* \to \Delta^*$ agree on L it suffices to test whether or not they agree on F. F is called a (morphic equivalence) *test set* for L.

Here we consider a similar problem for inverse morphisms. Accordingly we say that a finite subset F of a language L over Δ is an *inverse morphic equivalence test set* for L in the *universal* (resp. *existential*) sense if for all morphisms g and $h : \Sigma^* \to \Delta^*$ the relation $g^{-1}(x) = h^{-1}(x)$ for all x in $F \cap (h(\Sigma^*) \cup g(\Sigma^*))$ implies the relation $g^{-1}(x) = h^{-1}(x)$ for all x in $F \cap (h(\Sigma^*) \cup g(\Sigma^*))$ (resp. the relation $g^{-1}(x) \cap h^{-1}(x) \neq \emptyset$ for all x in $F \cap (h(\Sigma^*) \cup g(\Sigma^*))$ implies the relation $g^{-1}(x) \cap h^{-1}(x) \neq \emptyset$ for all x in $L \cap (h(\Sigma^*) \cup g(\Sigma^*))$). We simply call such test sets $\forall$*−test sets* and $\exists$*−test sets*, respectively.

Finally, we note that without loss of generality we can assume that the alphabet Δ is binary. This situation is achieved by the standard encodings of arbitrary alphabets into binary ones. Consequently from now on we assume that Δ is binary.

3. TEST SETS FOR INVERSE-MORPHIC EQUIVALENCE

We first point out that the main results of [ChK] and [EKR] can be modified for inverse morphisms as follows.

Theorem 3.1: *Let Σ be a fixed alphabet. For each language L and each integer $p \geq 0$ there exists a finite subset F of L such that for all pairs (g,h) of morphisms from Σ^* having bounded delay p we have:*

$$L \subseteq IE_{\forall}(g,h) \quad \text{iff} \quad F \subseteq IE_{\forall}(g,h) \ .$$

Proof: It was proved in [ChK] that for each integer $p \geq 0$ there exists a regular language R over some alphabet V such that for each pair (g,h) of morphisms having bounded delay p there exists another morpism τ such that $E(g,h) = \tau(R)$. So it follows from Lemma 2.1 (since bounded delay morphisms are injective) that

$$IE_{\forall}(g,h) = h\tau(R) \ .$$

Now, the result follows from the following lemma of [ChK]:

Lemma 3.2: *Let L' and R' be languages over Δ and V, respectively. There exists a finite subset F' of L' such that for each morphism τ from V^* into Δ^* we have*

$$L' \subseteq \tau(R') \quad \text{iff} \quad F' \subseteq \tau(R') \ . \quad \square$$

Theorem 3.3: *For each language L there exists a finite subset F of L such that for any pair (g, h) of binary morphisms we have*

$$L \subseteq IE_{\exists}(g,h) \quad \text{iff} \quad F \subseteq IE_{\exists}(g,h) \ .$$

The theorem follows from Lemma 3.1 and the following characterization result:

Lemma 3.4: *For each pair (g,h) of binary morphisms there exists another morphism τ over the alphabet $\{a,b,c,d\}$ such that*

$$IE_{\exists}(g,h) = \tau((\{a\} \cup \{bc^*d\})^*) \ .$$

Proof: In [EKR] it is proved that for any pair (g,h) of binary morphisms from $\{a,b\}^*$ their equality language is of one of the following three forms: $\{\beta,\gamma\}^*$, $(\beta\gamma^*\delta)^*$ for some words β,γ and δ, or $\{\lambda\} \cup \{x : |x|_a / |x|_b = k\}$, where k is nonnegative rational or $k = \infty$ and $|x|_a$ denotes the number of a's in the word x. Moreover, in the third case both of the morpisms are periodic with the same minimal period u, and therefore $h(E(g,h)) = w^*$ for some w. Indeed if $k = 0$ or $k = \infty$ then $w = b$ or a, respectively, and in all other cases $w = h(a)^t h(b)^q$ where t and q are coprimes and $t/q = k$. So the lemma follows from Lemma 2.1. □

Lemma 3.4 gave a partial characterization for existential inverse equality languages of two binary morphisms. Our next example shows that in the case of universal interpretation there are more possibilities.

Example 3.1: Let $n \geq 1$ and the morphisms g and h be defined as:

$$g: \begin{array}{l} a \to a^n \\ b \to a^{n+1} \end{array} \qquad h: \begin{array}{l} a \to a^{n+1} \\ b \to a^n \end{array} .$$

Clearly, for any x in $\{a,b\}^*$ if $g(x) = h(x)$ then $|x|_a = |x|_b$. Therefore $IE_{\forall}(g,h) \subseteq (a^{2n+1})^*$. We claim that

$$IE_{\forall}(g,h) = \{(a^{2n+1})^i \mid 0 \leq i \leq n\} . \tag{3.1}$$

Assume first that $i > n$. Let $i = n + i_0$, where $i_0 > 0$. Then we have

$$g(a^{2n+1}a^{i_0}b^{i_0}) = a^{(2n+1)n}\, a^{(2n+1)i_0} = a^{(2n+1)(n+i_0)}$$
$$\neq a^{(2n+1)(n+1)}\, a^{(2n+1)i_0} = h(a^{2n+1}a^{i_0}b^{i_0}) \ ,$$

i.e., $a^{(2n+1)i} \notin IE_{\forall}(g,h)$. Second let $i \leq n$. We consider the equation $i(2n+1) = yn + z(n+1)$ and show that it can hold with nonnegative values of y and z only in the case $y = z = i$. Clearly, y and z must be $< 2i$. Solving for z we obtain $z = i + (i-y)n/(n+1)$ which is, by the inequalities $0 \leq y < 2i$ and $i \leq n$, an integer only if $i = y$ (and hence also $z = i$). Consequently, the identity $h(a^i b^i) = g(a^i b^i) = a^{(2n+1)i}$ shows that $a^{(2n+1)i} \in IE_{\forall}(g,h)$. Therefore (3.1) holds.

It is easy to see that the test set conjecture for inverse morphisms with either of the interpretations does not hold if the cardinality of the domain alphabet Σ of the morphisms is not bounded. Indeed there exists an infinite sequence $(g_i, h_i)_{i\geq 0}$ of morphisms such that $E(g_i, h_i) = \alpha_i^*$ for some word α_i, and $h_i(\alpha_i^*) \cap h_j(\alpha_j^*) = \emptyset$ for all $i \neq j$, cf. [CK1]. Now for each $i \geq 0$ morphisms $\bar{g}_i$ and $\bar{h}_i$ can be defined such that $E(\bar{g}_i, \bar{h}_i) = (\bigcup_{j=0}^{i} E(g,h))^*$, hence the conclusion follows.

If we fix the domain alphabet Σ of g and h beforehand, then we do not know whether or not "the test set conjecture" for inverse morphisms with either of the interpretations holds. What we do know is that the size of a $\exists$-test set cannot be bounded by any function dependent only on the cardinality of Σ.

Example 3.2: For any integer $t \geq 1$ let g_t and h_t be morphisms defined by

$$g_t : \begin{array}{l} a \to (0101)^t \\ b \to 0 \\ c \to (10)^t \end{array} \qquad h_t : \begin{array}{l} a \to (01)^t \\ b \to 0 \\ c \to (1010)^t \end{array}$$

Clearly, $E(g_t, h_t) = \{a^n b c^n \mid n \geq 0\}^*$ and, consequently,

$$IE_{\exists}(g,h) = \{((01)^{3nt})^* 0 \mid n \geq 0\}^* . \tag{3.2}$$

Let $\{p_1, ..., p_k\}$ be a set of k primes which does not include 3. Let $p = p_1 \cdots p_k$ and $x_i = (01)^{3p_i'}0$ where $p_i' = 3(p/p_i)$ for $i = 1, ..., t$. Letting $L = \{x_1, ..., x_t\}$ we claim that for each proper subset L' of L there exists a pair of morphisms such that their inverses existentially agree on L' but do not agree on L. Indeed, for $L' = L - \{x_i\}$, for $i = 1, ..., k$, a suitable pair is (g_{p_i}, h_{p_i}) by (3.2).

We have

Theorem 3.5: *For each natural number k there exists a language L_k such that any set testing the existential inverse morphism equivalence on L for morphisms from a three letter alphabet contains at least k words.* □

4. INVERSE MORPHISM EQUIVALENCE ON LANGUAGES

In this section we discuss the problem of whether or not two given inverse morphisms agree existentially or universally on a given language. The study of such problems in connection with morphisms rather than inverse morphisms was initiated in [CS1]. Our first result reduces the universal inverse morphism problem to the corresponding morphism problem.

Theorem 4.1: *Let $\mathcal{L}$ be a family of languages satisfying the following two conditions:*

(i) *$\mathcal{L}$ is effectively closed under inverse morphisms,*

(ii) *morphism equivalence is decidable in $\mathcal{L}$, that is, given morphisms g and h and a language L in $\mathcal{L}$ it is decidable whether or not the equation $g(x) = h(x)$ holds for all x in L.*

Then universal inverse morphism equivalence is decidable for $\mathcal{L}$.

Proof: Let g and h be homomorphisms and $L \in \mathcal{L}$. Obviously g^{-1} and h^{-1} universally agree on L if and only if g and h agree on $h^{-1}(L)$ and $h^{-1}(L) = g^{-1}(L)$. But these conditions hold if and only if g and h agree on $h^{-1}(L)$ and $g^{-1}(L)$. □

In the case of existential agreement no general result like Theorem 4.1 is known. However, if the class of morphisms is heavily restricted as in Theorems 3.1 and 3.3 then we have such results.

Theorem 4.2: *Let $\mathcal{L}$ be a family of languages satisfying the following two conditions:*

(i) *each L in $\mathcal{L}$ is recursively enumerable,*

(ii) *for a given L in $\mathcal{L}$ and a given regular language R it is decidable whether or not $L \cap R$ is empty.*

Then it is decidable whether or not the inverses of two bounded delay morphisms agree existentially on a given language L in $\mathcal{L}$.

Proof: Let g and h be morphisms from Σ^* into Δ^*. First we recall that the bounded delay p of g and h can be found effectively. Second, it follows from [ChK] that a finite subset F of L can be effectively found for a given L under our current assumptions. Consequently, it is enough to check whether g^{-1} and h^{-1} agree existentially on a finite language which is, of course, effective. □

Exactly in the same way as Theorem 4.2 is obtained from Theorem 3.1 we can deduce from Theorem 3.3:

Theorem 4.3: *Let $\mathcal{L}$ be a family of languages over a binary alphabet satisfying conditions (i) and (ii) of Theorem 4.2. Then it is decidable whether or not the inverses of two binary morphisms existentially agree on a given language of $\mathcal{L}$.* □

Now, we turn to consider the problem of whether or not the inverses of periodic morphisms agree on a given language.

Lemma 4.4: *Let g and h be periodic morphisms from Σ^* into Δ^*. Then there exists, effectively, a finite set K such that $IE_{\exists}(g,h) = K^*$.*

Proof: If g and h have different minimal periods, then $K = \{\lambda\}$. Otherwise we have $g(x) = h(x)$ if and only if

$$\sum_{i=1}^{|\Sigma|} |g(a_i)|\ |x|_{a_i} = \sum_{i=1}^{|\Sigma|} |h(a_i)|\ |x|_{a_i} \text{ if and only if}$$

$$\sum_{i=1}^{|\Sigma|} (|g(a_i)| - |h(a_i)|)\ |x|_{a_i} = 0$$

which is equivalent, by a result in [ES], to the fact that

$$\psi(x) \in \left\{ \sum_{i=1}^{t} \alpha_i e_i \mid \alpha_i \geq 0 \right\}$$

for some vectors $e_1, \ldots, e_t$ in $N^{|\Sigma|}$. (Here $\psi(x)$ denotes the Parikh vector of a word x). Moreover, the vectors $e_1, \ldots, e_t$ can be found effectively. So it follows from Lemma 2.1 that $IE_{\exists}(g,h) = \{h(p_1), \ldots, h(p_t)\}^*$, where $p_i \in \psi^{-1}(e_i)$ for $i = 1, \ldots, t$. □

As a corollary of Lemma 4.4 we immediately obtain

Theorem 4.5: *Let $\mathcal{L}$ be family of languages over a one letter alphabet satisfying: given an L in $\mathcal{L}$ and a regular language R it is decidable whether or not $L \subseteq R$. Then it is decidable whether or not two inverses of periodic morphisms agree existentially on a given language of $\mathcal{L}$.*

We finish this section with some comments. By Theorem 4.1 it is decidable whether or not two inverse morphisms agree universally on a given regular or even context-free language, cf. [ACK]). Very recently the same problem with existential interpretation has been shown in [KM] - contrary to our initial intuition - to be undecidable. We also want to mention here the recent results of [KK] where mappings of the form hg^{-1} and $g^{-1}h$ are studied. It is shown that for mappings of the form hg^{-1} it is undecidable whether or not two such mappings agree universally on a given regular language. Furthermore, the same problem for mappings of the form $g^{-1}h$ is shown to be decidable.

Acknowledgement:

The authors are grateful to Y. Maon for pointing out an error in the first version of this note.

REFERENCES

[ACK] Albert, J., Culik II, K., and Karhumäki, J., Test Sets for Context Free Languages and Algebraic Systems of Equations Over a Free Monoid, *Information and Control 52* (1983), 172-186.

[AW] Albert, J., and Wood, D., Checking Sets, Test Sets, Rich Languages and Commutatively Closed Languages, *Journal of Computer and System Sciences 26,* (1983), 82-91.

[ChK] Choffrut, C., and Karhumäki, J., Test Sets for Bounded Delay Morphisms, *Proceedings of ICALP 83, Springer-Verlag Lecture Notes in Computer Science 154* (1983), 118-127.

[C1] Culik II, K., Some Decidability Results about Regular and Pushdown Translations, *Information Processing Letters 8* (1979), 5-8.

[C2] Culik II, K., Homomorphisms: Decidability, Equality and Test Sets, in *Formal Language Theory: Perspectives and Open Problems*, (R.V. Book, ed.), Academic Press, New York (1980), 167-194.

[CFS] Culik II, K., Fitch, F., and Salomaa, A., A Homomorphic Characterization of Regular Languages, *Discrete Applied Mathematics 4* (1982), 149-152.

[CK1] Culik II, K., and Karhumäki, J., On the Equality Sets for Homomorphisms on Free Monoids with Two Generators, *R.A.I.R.O., Informatique Theorique 14* (1980), 349-369.

[CK2] Culik II, K., and Karhumäki, J., Systems of Equations Over a Free Monoid and Ehrenfeucht's Conjecture, *Discrete Mathematics 43* (1983), 109-153.

[CS1] Culik II, K., and Salomaa, A., On the Decidability of Homomorphism Equivalence for Languages, *Journal of Computer and System Sciences 17* (1978), 163-175.

[CS2] Culik II, K., and Salomaa, A., Test Sets and Checking Words for Homomorphism Equivalence, *Journal of Computer and System Sciences 20* (1980), 279-295.

[EKR] Ehrenfeucht, A., Karhumäki, J., and Rozenberg , G., On Binary Equality Sets and a Solution to the Test Set Conjecture in the Binary Case, *Journal of Algebra 85* (1983), 76-85.

[ES] Eilenberg , S., and Schützenberger, J., Rational Sets in Commutative Monoids, *Journal of Algebra 13* (1969), 173-191.

[KK] Karhumäki, J., and Kleijn, H.C.M., On the Equivalence of Compositions of Morphisms and Inverse Morphisms on Regular Languages, manuscript (1983).

[KL] Karhumäki, J., and Linna, M., A Note on Morphic Characterization of Languages, *Discrete Applied Mathematics 5,* (1983), 243-246.

[KM] Karhumäki, J., and Maon, Y., A Simple Undecidable Problem: Existential Inverse Morphic Equivalence on Regular Languages, manuscript (1983).

[LL] Latteux, M., and Leguy, J., On the Composition of Morphisms and Inverse Morphisms, *Proceedings of ICALP 83, Springer-Verlag Lecture Notes in Computer Science 154* (1983), 420-432.

[T] Turakainen, P., On Homomorphic Characterization of Principal SemiAFL's Without Using Intersection with Regular Sets, *Information Sciences 27* (1982), 141-149.