GENERALIZED PARIKH MAPPINGS AND HOMOMORPHISMS

Juhani Karhumäki
Department of Mathematics
University of Turku
Turku, Finland

Abstract

The notion of the Parikh mapping is generalized by considering numbers of occurrences of segments of a fixed length instead of considering numbers of letters (i.e. segments of length one) only as is done in connection with the Parikh mappings. It is easily seen that the families of regular and context-free languages make difference with respect to these generalized Parikh mappings. On the other hand, properties of the Parikh mappings in connection with λ-free homomorphisms are, in general, preserved in the generalization.

1. Introduction

In this paper we consider a notion which can be regarded as a generalization of a well known notion of the Parikh mapping. This is defined by counting the segments of a fixed length of a word instead of counting only occurrences of letters, i.e. segments of length one, as is done in Parikh mappings. If the length of segments is k, then we refer our mapping as a k-generalized Parikh mapping and denote it by π_k. Two words u and v are called k-equivalent, in symbols $u \equiv_k v$, iff $\pi_k(u) = \pi_k(v)$.

The properties of k-generalized Parikh mappings are studied. Certainly, π_k-images of languages give more information about languages than ordinary Parikh images. For instance it immediately turns out that there exists a context-free language the π_2-image of which is not π_2-image of any regular language, i.e. the theorem of Parikh is not valid for generalized Parikh mappings.

Especially, the k-generalized Parikh mappings are considered in connection with homomorphisms. Let $h: \Sigma^* \to \Delta^*$ be a λ-free homomorphism and $\wedge_k$ a mapping which takes each word to a word which is obtained from the original word by catenating its all segments preserving their order. For example, $\wedge_2(aba) = \neq a\ ab\ ba\ a\neq$, where $\neq$ denotes the endmarker. With these notions we show the existence of a homomorphism $\hat{h}$ (in a suitable alphabet) such that $\wedge_k h = \hat{h} \wedge_k$. This result makes it possible to reduce certain problems concerning k-generalized Parikh properties to problems concerning ordinary Parikh properties (in a larger alphabet, of course).

As an application we show that it is decidable whether the sequences generated by two HDOL systems are k-generalized Parikh equivalent. We also show that two problems related to the Post Correspondence Problem are decidable. Let h and g be two homomorphisms of a free monoid. Define, for each $k \geq 0$,

$$E^k(h,g) = \{ x \in \Sigma^+ \mid h(x) \equiv_k g(x)\},$$

where $\equiv_0$ stands for the length relation, i.e. $u \equiv_0 v$ iff u and v are of the same length. We prove that it is decidable whether $E^k(h,g)$ is empty.

In another problem we consider sets of the form

$$P_k(h,g) = \{ x \in \Sigma^+ \mid \exists\, y \in \Sigma^+ \colon x \equiv_k y,\ h(x) = g(y)\}.$$

The emptiness problem for $P_k(h,g)$ is decidable for each $k \geq 0$, too. The cases $k = 0$ and $k = 1$ are solved by Greibach (1975) and Ibarra and Kim (1976), respectively. For $k \geq 2$ the problem is solved here.

It is instructive to note that above there cannot exist a "universal algorithm" which would solve whether $P_k(h,g)$ is empty for all $k \geq 0$. This is because such an algorithm would imply the decidability of the Post Correspondence Problem. Indeed, the intersection of all $P_k(h,g)$ sets equals to the set of all solutions of an instance of the Post Correspondence Problem.

2. Preliminaries

We fix here the notions and the notations needed in this paper. For unexplained standard notions of formal language theory we refer to any of the text books of the area, e.g. Salomaa (1973) or Harrison (1978).

The free monoid generated by a finite alphabet Σ is denoted by Σ^*. The identity of Σ^*, so-called empty word, is denoted by λ and $\Sigma^+ = \Sigma^* - \{\lambda\}$. The notation $|x|$ is used for the length of a word as well as $|\Sigma|$ for the cardinality of an alphabet Σ. $Pref_k(x)$ and $suff_k(x)$ denote the prefix and the suffix of length k of a word x, respectively. Finally, the notation xy^{-1} (resp. $y^{-1}x$) is used for the right (resp. left) difference of x by y.

Let Σ be a finite alphabet and $k \geq 1$. A new alphabet $\hat{\Sigma}$, so-called k-generalization of Σ, is defined as

$$\hat{\Sigma} = \Sigma^k \cup \bigcup_{i=0}^{k-2} (\neq\Sigma^{k-1} \cup \neq\Sigma^i \neq \cup \Sigma^{k-1}\neq),$$

where $\neq$ is a new symbol not in Σ. A mapping $\wedge_k\colon \Sigma^* \to \hat{\Sigma}^*$ is now defined by

$$\wedge_k(x) = \begin{cases} \neq x \neq & \text{if } |x| < k-1, \\ \neq x_1 \ldots x_{k-1} | x_1 \ldots x_k | \ldots | x_{t-k+2} \ldots x_t \neq & \\ \qquad \text{if } x = x_1 \ldots x_t,\ t \geq k-1 \text{ and } x_i \in \Sigma,\ i=1,\ldots,t, & \end{cases}$$

where $|$ is used for clarity as the operation of $\hat{\Sigma}^*$. For convenience we may

write $\wedge$ instead of $\wedge_k$ as well as $\hat{x}$ instead of $\wedge(x)$.

Clearly, $\wedge$ is a generalized sequential mapping, and hence

$$R = \wedge(\Sigma^+)$$

is a regular subset of $\hat{\Sigma}^*$. The mapping $\wedge$ is also injective. To find its inverse let us define a homomorphism $c: \hat{\Sigma}^* \to \Sigma^*$ by

$$c(y) = \begin{cases} \neq^{-1} y \neq^{-1} & \text{if } y \in \hat{\Sigma} - (\neq\Sigma^{k-1} \cup \Sigma^k \cup \Sigma^{k-1}\neq), \\ \neq^{-1} y & \text{if } y \in \neq\Sigma^{k-1}, \\ \lambda & \text{if } y \in \Sigma^{k-1}\neq, \\ \mathrm{suff}_1(y) & \text{if } y \in \Sigma^k. \end{cases}$$

Now c restricted to $R \cup \{\lambda\}$ gives the inverse of $\wedge$, i.e.

$$c(\wedge(x)) = x \quad \text{for all} \quad x \in \Sigma^*,$$
$$\wedge(c(y)) = y \quad \text{for all} \quad y \in R \cup \{\lambda\}.$$

Next our central notion, a k-generalized Parikh mapping, is defined. Let Σ be an alphabet and $k \geq 0$. A k-generalized Parikh mapping $\pi_k: \Sigma^* \to \mathbb{N}^m$, where $m = (|\Sigma|^{k+1}-1)/(|\Sigma|-1) + |\Sigma|^{k-1}$ is defined by

$$\pi_0(x) = |x| \quad \text{and}$$
$$\pi_k(x) = \pi_1(\wedge_k(x)) \quad \text{for } k \geq 1.$$

Two words u and v are called k-equivalent iff $\pi_k(u) = \pi_k(v)$. Similarly languages are called k-equivalent iff their π_k-images coincide.

The notion of a k-equivalence is similar to that one used when defined k-testable sets, cf. Brzozowski and Simon (1973). The only difference is that now we take care of multiplicities, too. Observe that the k-equivalence of u and v implies that they have the same prefixes and suffixes of length $k-1$, respectively. This follows since we used the endmarkers, which, in turn, was done to guarantee the following property.

Lemma 1. For words u and v in Σ^* and $k \geq 0$, the following holds true

$$u \equiv_{k+1} v \quad \Rightarrow \quad u \equiv_k v.$$

The proof of the lemma is immediate.

The notion of an equality set of two homomorphisms h and $g: \Sigma^* \to \Delta^*$ was introduced in Salomaa (1978) by

$$E(h,g) = \{ x \in \Sigma^+ \mid h(x) = g(x)\}.$$

For our purposes we define, for each $k \geq 0$, somewhat similar sets as follows

$$E^k(h,g) = \{ x \in \Sigma^+ \mid h(x) \equiv_k g(x)\}$$

and

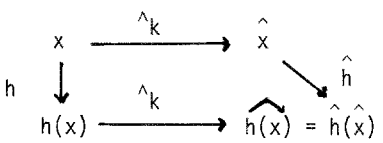
$$P_k(h,g) = \{ x \in \Sigma^+ \mid \exists\, y \in \Sigma^+ \colon x \equiv_k y,\ h(x) = g(y)\}.$$

We call $P_k(h,g)$ sets as k-generalized Parikh equality sets.

3. The Basic Lemma

In this section we establish a result which can be used to reduce problems concerning λ-free homomorphisms and k-generalized Parikh properties to problems concerning λ-free homomorphisms and usual Parikh properties. So a non-commutativity involved when dealing with k-generalized Parikh mappings can be avoided in connection with λ-free homomorphisms.

Basic Lemma. Let $h\colon \Sigma^* \to \Delta^*$ be a λ-free homomorphism and $k \geq 1$. Then there exists a homomorphism $\hat{h}\colon \hat{\Sigma}^* \to \hat{\Delta}^*$ such that $\wedge_k h = \hat{h} \wedge_k$, i.e. the following diagram holds true for all x in Σ^*



Proof. The homomorphism $\hat{h}$ is defined as follows:

(i) For words $\neq x \neq \in \neq\Sigma^i\neq$ in $\hat{\Sigma}$ if $h(x) = y_1 \ldots y_t$, then

$$\hat{h}(\neq x \neq) = \begin{cases} \neq h(x) \neq & \text{if } t < k\text{-}1 \\ \neq y_1 \ldots y_{k-1} \mid y_1 \ldots y_k \mid \ldots \mid y_{t-k+2} \ldots y_t \neq & \text{if } t \geq k\text{-}1. \end{cases}$$

(ii) For words $\neq x \in \neq\Sigma^{k-1}$ in $\hat{\Sigma}$ if $h(x) = y_1 \ldots y_t$, then

$$\hat{h}(\neq x) = \begin{cases} \neq h(x) & \text{if } t = k\text{-}1 \\ \neq y_1 \ldots y_{k-1} \mid y_1 \ldots y_k \mid \ldots \mid y_{t-k+1} \ldots y_t & \text{if } t > k\text{-}1. \end{cases}$$

(iii) For words $x \in \Sigma^{k-1}\neq$ in $\hat{\Sigma}$

$$\hat{h}(x\neq) = \mathrm{suff}_{k-1}(h(x))\neq.$$

(iv) For words $x \in \Sigma^{k}$ in $\hat{\Sigma}$ if $y_1 \ldots y_s = \mathrm{suff}_{k-1}(h(x'))h(a)$, where $x = x'a$ and $a \in \Sigma$, then

$$\hat{h}(x) = y_1 \ldots y_k | \ldots | y_{s-k+1} \ldots y_s .$$

Above all y_i's mean letters in Δ and $|$ is used to denote the operation of $\hat{\Delta}^*$.

It is straightforward to see that $\hat{h}$ satisfies the property of the lemma.

Recalling the notations of the previous section we conclude that $\hat{h}$ restricted to R, i.e. $\hat{h}_{|R}$, is obtained as the composition

$$\hat{h}_{|R} = \wedge_k \, h \, c_{|R} .$$

Observe, however, that the restriction to R is essential. Indeed, $\wedge_k hc: \hat{\Sigma}^* \to \hat{\Delta}^*$ is not even a homomorphism.

In the above lemma it is necessary to assume that h is λ-free. Otherwise the definitions of (iii) and (iv) do not work. In fact, the following example shows that the Basic Lemma is not even true for erasing homomorphisms.

Example. Let $h: \{a,b\}^* \to \{a,b\}^*$ be the homomorphism defined by $h(a) = ab$, $h(b) = \lambda$. Then there does not exist any homomorphism $\hat{h}$ from $\{\neq a, \neq b, aa, ab, ba, bb, a\neq, b\neq\}^*$ into itself such that $\wedge_2 h = \hat{h} \wedge_2$.

To show this assume the contrary that such an $\hat{h}$ exists. Since $h(aa) = abab$ and $h(aaa) = ababab$, then necessarily $|\hat{h}(aa)| = 2$ and $|\hat{h}(\neq a)| + |\hat{h}(a\neq)| = 3$. Now we consider the words aa and $baab$ which are mapped into $abab$ under h. Clearly, $\hat{h}(\neq b)$ and $\hat{h}(b\neq)$ must be nonempty. So $|\hat{h}(ab)| + |\hat{h}(ba)| \leq 1$. This gives a contradiction when we consider words aaa and $ababa$. Indeed,

$$|\hat{h}(\widehat{aaa})| = 7 \quad \text{and} \quad |\hat{h}(\widehat{ababa})| \leq 6$$

although $h(aaa) = h(ababa)$.

4. Applications of the Basic Lemma

In this section we apply our observations to some problems, for detailed proofs we refer to Karhumäki (to appear). First we note that k-generalized Parikh mappings, contrary to ordinary Parikh mappings, make difference between regular and context-free languages.

Example. Let $L = \{ a^n b^n \mid n \geq 1 \}$. Then

$$\pi_2(L) = \{ (1,0,n,1,0,n,0,1) \mid n \geq 0 \},$$

where $\widehat{\{a,b\}}$ is ordered as $\neq a, \neq b, aa, ab, ba, bb, a\neq, b\neq$. Moreover, for each $n \geq 0$,

$$\pi_2^{-1}(1,0,n,1,0,n,0,1) = a^n b^n.$$

Hence there cannot exist any regular language which would be 2-equivalent to L.

Our first application of the Basic Lemma is to the theory of DOL systems. For detailed definitions we refer to Rozenberg and Salomaa (1980). We only recall that a DOL system consists of an alphabet Σ, an endomorphism h of Σ^* and a so-called axiom which is an element of Σ^+. When applied iteratively h to the axiom w a sequence of words, a so-called DOL sequence, is obtained: $w, h(w), h^2(w), \ldots$ If this sequence is mapped by another homomorphism, say f, it yields a so-called HDOL sequence. Hence HDOL systems can be regarded as quadruples (Σ, h, w, f).

Now we can show

Theorem 1. Given $k \geq 0$. It is decidable whether the sequences generated by two HDOL systems $(\Sigma_i, h_i, w_i, f_i)$, $i = 1,2$, are k-equivalent, i.e. whether the following holds true

$$f_1(h_1^n(w_1)) \equiv_k f_2(h_2^n(w_2)) \quad \text{for all } n \geq 0.$$

We want to remind here that the decidability of the sequence equivalence problem for HDOL systems, i.e. whether two HDOL systems generate the same sequence of words, is still open. For DOL systems it was solved in Culik and Fris (1977), see also Rozenberg and Salomaa (1980). Our above result shows that equivalence problems related to the sequence equivalence are decidable also for HDOL systems. Certainly, the algorithm in Theorem 1 depends on k. If this would not be the case, then we would have an algorithm for HDOL sequence equivalence.

As another application of the Basic Lemma we consider the sets

$$E^k(h,g) = \{ x \in \Sigma^+ \mid h(x) \equiv_k g(x) \}$$

introduced in Section 2. For these we have the following representation result in the case of λ-free homomorphisms.

Theorem 2. For an integer $k \geq 1$ and λ-free homomorphisms h and $g: \Sigma^* \to \Delta^*$, there exist homomorphisms $\hat{h}$ and $\hat{g}: \hat{\Sigma}^* \to \hat{\Delta}^*$, a regular subset R of $\hat{\Sigma}^*$ and a homomorphism $c: \hat{\Sigma}^* \to \Sigma^*$ such that

$$E^k(h,g) = c(E^1(\hat{h},\hat{g}) \cap R) .$$

From Theorem 2 we easily obtain

Corollary. Given $k \geq 0$. It is decidable whether $E^k(h,g)$ is empty for two λ-free homomorphisms h and g: $\Sigma^* \to \Delta^*$.

We turn to consider the sets of the form

$$P_k(h,g) = \{ x \in \Sigma^+ \mid \exists\, y \in \Sigma^+ : x \equiv_k y,\ h(x) = g(y)\},$$

where h and g are λ-free homomorphisms from Σ^* into Δ^* and $k \geq 0$. Especially, we are interested in the decidability of the emptiness of the set $P_k(h,g)$ and the related sets.

If in the definition of $P_k(h,g)$ it is required that $x = y$, then the emptiness of the set is an undecidable property, since the problem becomes the Post Correspondence Problem. If, on the other hand, no restriction on y is introduced, then the problem is trivially decidable. Indeed, it is the question of the emptiness of the regular set $g^{-1}(h(\Sigma^+))$.

So it is interesting to analyse some cases in between. Greibach (1975) showed that the problem is decidable if it is required that $|x| = |y|$, i.e. she solved the decidability of the emptiness of $P_0(h,g)$. Ibarra and Kim (1976) generalized this for the case $k = 1$, i.e. for the case where x and y are required to be Parikh equivalent. Our purpose is to show that the problem is decidable for any fixed $k \geq 2$.

Our work is based on the paper of Ibarra and Kim. We, however, need the following auxilary notion. Let k, h and g be as above and A a regular subset of Σ^+. We define

$$P_k(h,g;A) = \{ x \in A \mid \exists\, y \in A : x \equiv_k y,\ h(x) = g(y)\}.$$

So we are considering P_k-sets with respect to a given regular set. Using ideas from Ibarra and Kim (1976) we obtain

Lemma 2. For two λ-free homomorphisms h and g: $\Sigma^* \to \Delta^*$ and a regular set A, it is decidable whether $P_1(h,g;A)$ is empty.

So we are ready for

Theorem 3. Given an integer $k \geq 2$. It is decidable whether for two λ-free homomorphisms h and g: $\Sigma^* \to \Delta^*$ $P_k(h,g)$ is empty.

Proof. We have

$$\begin{aligned} P_k(h,g) &= \{ x \in \Sigma^+ \mid \exists\, y \in \Sigma^+ \colon x \equiv_k y,\ h(x) = g(y)\} \\ &= \{ x \in \Sigma^+ \mid \exists\, y \in \Sigma^+ \colon \hat{x} \equiv_1 \hat{y},\ \widehat{h(x)} = \widehat{g(y)}\} \\ &= \{ \hat{x} \in R \mid \exists\, \hat{y} \in R : \hat{x} \equiv_1 \hat{y},\ \hat{h}(\hat{x}) = \hat{g}(\hat{y})\} \\ &= P_1(\hat{h},\hat{g};R) , \end{aligned}$$

where the notations of Section 2 are employed.

5. Discussion

We have generalized the notion of the Parikh mapping in a natural way. This generalization takes into an account, in some extent, the order of the letters, too. Hence the properties of generalized Parikh mappings are not quite the same as those of ordinary Parikh mappings. In fact, it turned out that the famous theorem of Parikh is not true for generalized Parikh mappings.

However, in connection with λ-free homomorphisms many problems about generalized Parikh mappings could be reduced to problems (or related problems) about ordinary Parikh mappings. Especially, we introduced an "upper approximation sequence" for an equality set of two homomorphisms in such a way that the emptiness was decidable in all elements of this sequence. Indeed, for sets

$$P_k(h,g) = \{ x \in \Sigma^+ \mid \exists\, y \in \Sigma^+ \colon x \equiv_k y,\ h(x) = g(y)\}$$

we have, by Lemma 1,

$$(1) \qquad E(h,g) \subseteq \ldots \subseteq P_k(h,g) \subseteq \ldots \subseteq P_1(h,g) \subseteq P_0(h,g)$$

and

$$E(h,g) = \bigcap_{k=0}^{\infty} P_k(h,g) .$$

Another way to obtain such an "upper approximation sequence" is to use the sets $E^k(h,g)$.

On the other hand, it is known that so-called k-bounded equality sets $E_k(h,g)$, cf. Rozenberg and Salomaa (1980) form an "lower approximation sequence" for $E(h,g)$, i.e.

$$(2) \qquad E(h,g) \supseteq \ldots \supseteq E_k(h,g) \supseteq \ldots \supseteq E_1(h,g) \supseteq E_0(h,g)$$

and

$$E(h,g) = \bigcup_{k=0}^{\infty} E_k(h,g) .$$

If one could find a class of homomorphisms for which both (1) and (2) would be finite, then the Post Correspondence Problem for this class would be decidable.

Acknowledgement. This research was supported by Finnish Academy which is gratefully acknowledged.

References

Brzozowski, J. and Simon, I., Characterization of locally testable events, Discrete Mathematics 4 (1973), 243-272.

Culik, K. II and Fris, I., The decidability of the equivalence problem for DOL systems, Inform. and Control 35 (1977), 20-39.

Ginsburg, S., The Mathematical Theory of Context-Free Languages, McGraw Hill, New York (1966).

Greibach, S., A remark on code sets and context-free languages, IEEE Trans. on Computers C-24 (1975), 741-742.

Harrison, M., Introduction to Formal Language Theory, Addison Wesley, London (1978).

Ibarra, O. and Kim, C., A useful device for showing the solvability of some decision problems, Proc. of eighth annual ACM symposium on theory of computing (1976), 135-140.

Karhumäki, J., Generalized Parikh mappings and homomorphisms, Inform. and Control (to appear).

Rozenberg, G. and Salomaa, A., The Mathematical Theory of L Systems, Academic Press, New York (1980).

Salomaa, A., Formal Languages, Academic Press, New York (1973).

Salomaa, A., Equality sets for homomorphisms of free monoids, Acta Cybernetica 4 (1978), 127-139.