Indonesian Journal of Combinatorics **4** (1) (2020), 1–9

INDONESIAN JOURNAL OF COMBINATORICS

On M-unambiguity of Parikh matrices

Wen Chean Teh

School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM, Malaysia

dasmenteh@usm.my

Abstract

The Parikh matrix mapping was initiated in 2001 by Mateescu, Salomaa, Salomaa, and Yu as a canonical extension of the well established Parikh mapping. The so called injectivity problem in this area, even when restricted to ternary words, has withstanded numerous attempts over two decades by multiple researchers, including Şerbănuţă. Certain M-ambiguous words are crucial in Şerbănuţă's findings about the number of M-unambiguous prints. We will show that these words are in fact strongly M-ambiguous, thus suggesting a possible extension of the results of Şerbănuţă to the context of the newly introduced strong version of M-equivalence. In addition, initial results pertaining to some conjecture by Şerbănuţă will be presented.

Keywords: Parikh mapping, subword occurrence, injectivity problem, print, strongly M-equivalent
Mathematics Subject Classification: 68R15
DOI: 10.19184/ijc.2020.4.1.1

1. Introduction

The well-known Parikh Theorem [7] says that the set of Parikh vectors corresponding to all words of a language is a semilinear set, provided the language is regular. The Parikh matrix [6] of a word arithmetizes more structural information of the word compared to the Parikh vector of the word, the latter which appears as the diagonal above the main diagonal of the Parikh matrix. Parikh matrices are useful tools in studying subword occurrences and the theory of formal languages. The problem of characterizing when two distinct words are having a common Parikh matrix relates to the degree of non-injectivity of the Parikh matrix mapping, thus known as the injectivity problem.

Received: 2 May 2019, Revised: 14 April 2020, Accepted: 18 April 2020.

This problem has thus far defied a satisfactory solution, even for the ternary alphabet, despite significant interest on it [1, 2, 3, 4, 5, 8, 9, 10, 11, 13, 14, 15, 16].

A word is M-unambiguous iff it is uniquely associated to its Parikh matrix. Şerbănuţă [10] showed that over any given ordered alphabet, there exist only finitely many words that are M-unambiguous and being prints, meaning consecutive letters are distinct. This result would be a remarkable advancement towards the injectivity problem if an earlier conjecture [11] by Şerbănuţă is true, which says that if a word is M-unambiguous, then so is its print. In this work, we will show that the so called n-ambiguous words (see Definition 2.2) that play essential roles in [10] turn out to be strongly M-ambiguous, where the strong version of M-equivalence was proposed in [12] to get rid of the dependence on the ordering of the alphabet. Also, we will present our initial findings pertaining to Serbănuţă's conjecture.

Throughout this work, Σ denotes a finite alphabet. The set of all words over Σ is denoted by Σ^*, which includes the empty word λ. Together with some total ordering on the alphabet Σ, it is then called an *ordered alphabet*. For convenience, if $\Sigma = \{a_1, a_2, \dots, a_s\}$ and $a_1 < a_2 < \cdots < a_s$, then we may write $\Sigma = \{a_1 < a_2 < \cdots < a_s\}$. When it is without confusion, we will let Σ to stand for both an ordered alphabet and the correspondinng underlying alphabet. For $v, w \in \Sigma^*$, the concatenation of v and w is denoted by juxaposition vw while $|w|$ denotes the length of w.

A word v is said to be a *(scattered) subword* of $w \in \Sigma^*$ iff there exist $x_1, x_2, \dots, x_n \in \Sigma^*$ and $y_0, y_1, \dots, y_n \in \Sigma^*$, possibly empty, such that $v = x_1x_2\cdots x_n$ and $w = y_0x_1y_1\cdots y_{n-1}x_ny_n$. We say that v is a *factor* of w if the letters in v occur contiguously in w. Two occurrences of v are considered distinct iff at least some letter in v differs in position in the occurrences. Let $|w|_v$ denote the number of distinct occurrences of v as a subword of w. For example, $|aabbab|_{ab} = 7$ and $|baacbc|_{abc} = 2$. By standard convention, $|w|_\lambda = 1$ for every $w \in \Sigma^*$.

Let $\mathbb{N}$ be the set of nonnegative integers. For any positive integer $k \geq 2$, let the multiplicative monoid consisting of $k \times k$ upper triangular matrices with unit diagonal and nonnegative integral entries be denoted by $\mathcal{M}_k$.

Definition 1.1. *Let $\Sigma = \{a_1 < a_2 < \cdots < a_s\}$ be an ordered alphabet. The* Parikh matrix mapping*, denoted by Ψ_Σ, is the monoid morphism $\Psi_\Sigma \colon \Sigma^* \to \mathcal{M}_{s+1}$ defined as follows: $\Psi_\Sigma(\lambda) = I_{s+1}$; for each $1 \leq q \leq s$, say $\Psi_\Sigma(a_q) = (m_{i,j})_{1\leq i,j\leq s+1}$, then $m_{q,q+1} = 1$, $m_{i,i} = 1$ for every $1 \leq i \leq s+1$, and the rest of the entries are zero. The images of Ψ_Σ are called* Parikh matrices.

Theorem 1.1. [6] *Let $\Sigma = \{a_1 < a_2 < \cdots < a_s\}$ be an ordered alphabet and suppose $w \in \Sigma^*$. The entries of the Parikh matrix $\Psi_\Sigma(w) = (m_{i,j})_{1\leq i,j\leq s+1}$ have the following properties:*

- $m_{i,i} = 1$ *whenever* $1 \leq i \leq s+1$;
- $m_{i,j} = 0$ *whenever* $1 \leq j < i \leq s+1$;
- $m_{i,j} = |w|_{a_ia_{i+1}\cdots a_{j-1}}$ *whenever* $1 \leq i < j \leq s+1$.

www.ijc.or.id

Example 1. Let $\Sigma = \{a < b < c\}$ and consider $w = bcabac$. Then

$$\Psi_\Sigma(w) = \Psi_\Sigma(b)\Psi_\Sigma(c)\Psi_\Sigma(a)\Psi_\Sigma(b)\Psi_\Sigma(a)\Psi_\Sigma(c)$$
$$= \begin{pmatrix}1&0&0&0\\0&1&1&0\\0&0&1&0\\0&0&0&1\end{pmatrix}\cdots\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&1\\0&0&0&1\end{pmatrix} = \begin{pmatrix}1&2&1&1\\0&1&2&3\\0&0&1&2\\0&0&0&1\end{pmatrix} = \begin{pmatrix}1&|w|_a&|w|_{ab}&|w|_{abc}\\0&1&|w|_b&|w|_{bc}\\0&0&1&|w|_c\\0&0&0&1\end{pmatrix}.$$

Definition 1.2. *Let Σ be an ordered alphabet and suppose $v, w \in \Sigma^*$. If $\Psi_\Sigma(v) = \Psi_\Sigma(w)$, then we say that v and w are M-*equivalent *and we denote this by $v \equiv_M w$. If w is M-equivalent to another distinct word, then we say that w is M-*ambiguous*. Otherwise, we say that w is M-*unambiguous.

We now state two elementary rules that govern M-equivalence. Let $\Sigma = \{a_1 < a_2 < \cdots < a_s\}$ and suppose $w, w' \in \Sigma^*$.

$E1$. If $w = xa_ka_ly$ and $w' = xa_la_ky$ for some words $x, y \in \Sigma^*$ with $|k-l| \geq 2$, then $w \equiv_M w'$.

$E2$. If $w = xa_ka_{k+1}ya_{k+1}a_kz$ and $w' = xa_{k+1}a_kya_ka_{k+1}z$ for some $1 \leq k \leq s-1$ and some words $x, z \in \Sigma^*$ and $y \in (\Sigma\backslash\{a_{k-1}, a_{k+2}\})^*$, then $w \equiv_M w'$.

Definition 1.3. *Let Σ be an ordered alphabet and suppose $w, w' \in \Sigma^*$. We say that w' is ME-*equivalent *to w iff w' can be obtained from w by applying Rule $E1$ and Rule $E2$ finitely many times. A word $w \in \Sigma^*$ is ME-*ambiguous *iff there is a distinct word ME-equivalent to w. Otherwise, we say that w is ME-*unambiguous*. An M-ambiguous word $w \in \Sigma^*$ is said to be* properly M-ambiguous *iff w is ME-unambiguous.*

As opposed to M-equivalence, the following version of M-equivalence is defined for any alphabet because it is independent from the ordering of the alphabet.

Definition 1.4. *Suppose Σ is an alphabet. We say that two words $w, w' \in \Sigma^*$ are* strongly M-equivalent *iff w and w' are M-equivalent with respect to any ordering on the underlying alphabet Σ. A word $w \in \Sigma^*$ is* strongly M-ambiguous *iff there is a distinct word strongly M-equivalent to w. Otherwise, w is said to be* strongly M-unambiguous.

The following theorem that characterizes strong M-equivalence is immediate by Theorem 1.1.

Theorem 1.2. [12] *Let Σ be an alphabet and suppose $w, w' \in \Sigma^*$. Then w and w' are strongly M-equivalent if and only if $|w|_u = |w'|_u$ whenever $u \in \Sigma^*$ satisfying $|u|_a \leq 1$ for all $a \in \Sigma$.*

2. Strong M-ambiguity of n-ambiguous Words

Definition 2.1. *Let Σ be an alphabet and $w \in \Sigma^*$. If $w = b_1^{n_1}b_2^{n_2}\cdots b_l^{n_l}$ where $b_i \in \Sigma$ and $n_i > 0$ for every $1 \leq i \leq l$, and additionally $b_{i+1} \neq b_i$ for every $1 \leq i < l$, then the* print *of w is $b_1b_2\cdots b_l$. We denote the print of w by $\mathrm{pr}(w)$. If $\mathrm{pr}(w) = w$, then we say that w is a* print.

Şerbănuţă's Conjecture. [11] *Let Σ be an ordered alphabet and $w \in \Sigma^*$. If w is M-unambiguous, then the print of w is also M-unambiguous.*

www.ijc.or.id

Şerbănuţă's conjecture is valid for the binary alphabet and the ternary alphabet. The later is verifiable through a list of M-unambiguous words over the ternary alphabet[1] obtained by Şerbănuţă in [11]. The validity of Şerbănuţă's Conjecture is desirable in view of the next theorem.

Theorem 2.1. [10] *Over any ordered alphabet, the number of M-unambiguous prints is finite.*

The main idea of the proof of Theorem 2.1 is that if a print $w \in \Sigma^*$ is sufficiently long, then either w is trivially M-ambiguous or w contains a certain $(|\Sigma|-1)$-ambiguous word as a factor. Our instinct that a $(|\Sigma|-1)$-ambiguous word is not only M-ambiguous as already shown in [10] but is actually strongly M-ambiguous turns out to be right (see Theorem 2.3).

Definition 2.2. *Let Σ be an alphabet with $|\Sigma| \geq 2$ and $w \in \Sigma^*$.*

- *w is said to be 0-ambiguous iff $w_0 = bab$ for some letters $a, b \in \Sigma$ with $a \neq b$.*[2]
- *For each positive integer n, we say w is n-ambiguous iff $w = w''w'w''$ for some $w', w'' \in \Sigma^*$ such that w'' is $(n-1)$-ambiguous.*

Theorem 2.2. *Suppose Σ is an alphabet with $|\Sigma| \geq 2$ and a, b are distinct letters in Σ. Suppose $y_0, y_1, y_2, \ldots$ are words over Σ. Let $v_0 = bba$, $v'_0 = abb$, and $w_0 = bab$. Recursively, define $v_{n+1} = v_n y_n v'_n$, $v'_{n+1} = v'_n y_n v_n$ and $w_{n+1} = w_n y_n w_n$. Then the following assertions hold.*

1. *For each $n \in \mathbb{N}$, it holds that $|v_n|_u + |v'_n|_u = 2|w_n|_u$ whenever $u \in \Sigma^*$ satisfying $|u|_x \leq 1$ for every $x \in \Sigma$.*
2. *For each $n \in \mathbb{N}$, it holds that $|v_n|_u = |v'_n|_u = |w_n|_u$ whenever $u \in \Sigma^*$ satisfying $|u| \leq n+1$ and $|u|_x \leq 1$ for every $x \in \Sigma$.*

Proof. (Part 1) First, it can be shown by induction that $\pi_{\Sigma\backslash\{a\}}(v_n) = \pi_{\Sigma\backslash\{a\}}(v'_n) = \pi_{\Sigma\backslash\{a\}}(w_n)$ for every $n \in \mathbb{N}$. Hence, it follows that $|v_n|_u = |v'_n|_u = |w_n|_u$ for every $u \in (\Sigma\backslash\{a\})^*$ and $n \in \mathbb{N}$. This simple fact will be used a few times in this proof.

We proceed by induction on n. For the base step, $|v_0|_u + |v'_0|_u = 2|w_0|_u$ can be verified for $u \in \{\lambda, a, b, ab, ba\}$. For the other u, both sides are clearly zero. Hence, we complete the base step.

Now, for the induction step, let $u \in \Sigma^*$ such that $|u|_x \leq 1$ for all $x \in \Sigma$. If $u \in (\Sigma\backslash\{a\})^*$, then

[1]The M-unambiguous words $a^m bab^n cb^p$, $b^m ab^n cbc^p$, $b^m cb^n aba^p$, and $c^m bcb^n ab^p$ for $m, n, p > 0$ are mistakenly omitted from the final list but accounted for in the proof.

[2]The original definition of 0-ambiguity by Serbănuţă stipulates that Σ is ordered and a, b are the first two letters in Σ.

www.ijc.or.id

by our simple fact we are done. Suppose $|u|_a = 1$. Hence,

$$
\begin{aligned}
&|v_{n+1}|_u + |v'_{n+1}|_u \\
={}& \sum_{\substack{u_1,u_2,u_3\in\Sigma^* \\ u_1u_2u_3=u}} |v_n|_{u_1}|y_n|_{u_2}|v'_n|_{u_3} + \sum_{\substack{u_1,u_2,u_3\in\Sigma^* \\ u_1u_2u_3=u}} |v'_n|_{u_1}|y_n|_{u_2}|v_n|_{u_3} \\
={}& \sum_{\substack{u_1,u_2,u_3\in\Sigma^* \\ u_1u_2u_3=u \\ |u_1|_a=1}} |v_n|_{u_1}|y_n|_{u_2}|v'_n|_{u_3} + \sum_{\substack{u_1,u_2,u_3\in\Sigma^* \\ u_1u_2u_3=u \\ |u_2|_a=1}} |v_n|_{u_1}|y_n|_{u_2}|v'_n|_{u_3} + \sum_{\substack{u_1,u_2,u_3\in\Sigma^* \\ u_1u_2u_3=u \\ |u_3|_a=1}} |v_n|_{u_1}|y_n|_{u_2}|v'_n|_{u_3} \\
&+ \sum_{\substack{u_1,u_2,u_3\in\Sigma^* \\ u_1u_2u_3=u \\ |u_1|_a=1}} |v'_n|_{u_1}|y_n|_{u_2}|v_n|_{u_3} + \sum_{\substack{u_1,u_2,u_3\in\Sigma^* \\ u_1u_2u_3=u \\ |u_2|_a=1}} |v'_n|_{u_1}|y_n|_{u_2}|v_n|_{u_3} + \sum_{\substack{u_1,u_2,u_3\in\Sigma^* \\ u_1u_2u_3=u \\ |u_3|_a=1}} |v'_n|_{u_1}|y_n|_{u_2}|v_n|_{u_3}
\end{aligned}
$$

Now, we have the following sequence of equalities, where the first equality is due to our simple fact and the third equality is by the induction hypothesis:

$$
\begin{aligned}
&\sum_{\substack{u_1,u_2,u_3\in\Sigma^* \\ u_1u_2u_3=u \\ |u_1|_a=1}} |v_n|_{u_1}|y_n|_{u_2}|v'_n|_{u_3} + \sum_{\substack{u_1,u_2,u_3\in\Sigma^* \\ u_1u_2u_3=u \\ |u_1|_a=1}} |v'_n|_{u_1}|y_n|_{u_2}|v_n|_{u_3} \\
={}& \sum_{\substack{u_1,u_2,u_3\in\Sigma^* \\ u_1u_2u_3=u \\ |u_1|_a=1}} |v_n|_{u_1}|y_n|_{u_2}|w_n|_{u_3} + \sum_{\substack{u_1,u_2,u_3\in\Sigma^* \\ u_1u_2u_3=u \\ |u_1|_a=1}} |v'_n|_{u_1}|y_n|_{u_2}|w_n|_{u_3} \\
={}& \sum_{\substack{u_1,u_2,u_3\in\Sigma^* \\ u_1u_2u_3=u \\ |u_1|_a=1}} \left(|v_n|_{u_1} + |v'_n|_{u_1}\right)|y_n|_{u_2}|w_n|_{u_3} = 2\sum_{\substack{u_1,u_2,u_3\in\Sigma^* \\ u_1u_2u_3=u \\ |u_1|_a=1}} |w_n|_{u_1}|y_n|_{u_2}|w_n|_{u_3}.
\end{aligned}
$$

Similarly, it can be shown that

$$
\sum_{\substack{u_1,u_2,u_3\in\Sigma^* \\ u_1u_2u_3=u \\ |u_3|_a=1}} |v_n|_{u_1}|y_n|_{u_2}|v'_n|_{u_3} + \sum_{\substack{u_1,u_2,u_3\in\Sigma^* \\ u_1u_2u_3=u \\ |u_3|_a=1}} |v'_n|_{u_1}|y_n|_{u_2}|v_n|_{u_3} = 2\sum_{\substack{u_1,u_2,u_3\in\Sigma^* \\ u_1u_2u_3=u \\ |u_3|_a=1}} |w_n|_{u_1}|y_n|_{u_2}|w'_n|_{u_3}.
$$

Furthermore, by our observation

$$
\sum_{\substack{u_1,u_2,u_3\in\Sigma^* \\ u_1u_2u_3=u \\ |u_2|_a=1}} |v_n|_{u_1}|y_n|_{u_2}|v'_n|_{u_3} + \sum_{\substack{u_1,u_2,u_3\in\Sigma^* \\ u_1u_2u_3=u \\ |u_2|_a=1}} |v'_n|_{u_1}|y_n|_{u_2}|v_n|_{u_3} = 2\sum_{\substack{u_1,u_2,u_3\in\Sigma^* \\ u_1u_2u_3=u \\ |u_2|_a=1}} |w_n|_{u_1}|y_n|_{u_2}|w_n|_{u_3}.
$$

Therefore, $|v_{n+1}|_u + |v'_{n+1}|_u = 2\sum_{\substack{u_1,u_2,u_3\in\Sigma^* \\ u_1u_2u_3=u}} |w_n|_{u_1}|y_n|_{u_2}|w_n|_{u_3} = 2|w_{n+1}|_u$.

(Part 2) Again we proceed by induction on n. We omit the base step as it is trivial. Suppose

www.ijc.or.id

$u \in \Sigma^*$, $|u| \leq n+2$, and $|u|_x \leq 1$ for all $x \in \Sigma$ for the induction step. Then

$$\begin{aligned} |w_{n+1}|_u &= |w_n|_u + |y_n|_u + |w_n|_u + \sum_{\substack{u_1,u_2,u_3 \in \Sigma^* \\ u_1 \neq u, u_2 \neq u, u_3 \neq u \\ u_1 u_2 u_3 = u}} |w_n|_{u_1} |y_n|_{u_2} |w_n|_{u_3} \\ &= |v_n|_u + |y_n|_u + |v'_n|_u + \sum_{\substack{u_1,u_2,u_3 \in \Sigma^* \\ u_1 \neq u, u_2 \neq u, u_3 \neq u \\ u_1 u_2 u_3 = u}} |v_n|_{u_1} |y_n|_{u_2} |v'_n|_{u_3} = |v_{n+1}|_u. \end{aligned}$$

Using our induction hypothesis, it follows that the two last summations equal each other because $|u_1| \leq n+1$ and $|u_3| \leq n+1$. Meanwhile, $2|w_n|_u = |v_n|_u + |v'_n|_u$ by Part 1. We can show similarly that $|w_{n+1}|_u = |v'_{n+1}|_u$. $\square$

Theorem 2.3. *Let Σ be an alphabet with $|\Sigma| \geq 2$ and suppose $w \in \Sigma^*$. If w is $(|\Sigma|-1)$-ambiguous, then w is strongly M-ambiguous.*

Proof. Say that $|\Sigma| = N$. Since w is $(N-1)$-ambiguous, there exists $w_0, w_1, \ldots, w_{N-1} \in \Sigma^*$ and $y_0, y_1, \ldots, y_{N-2} \in \Sigma^*$ such that $w_0 = bab$ for some distinct $a, b \in \Sigma$, $w_{n+1} = w_n y_n w_n$ for each $0 \leq n \leq N-2$ and $w_{N-1} = w$. Let $v_0 = bba$, $v'_0 = abb$, $v_{n+1} = v_n y_n v'_n$, and $v'_{n+1} = v'_n y_n v_n$ for each $0 \leq n \leq N-2$. By Theorem 2.2, $|v_{N-1}|_u = |w_{N-1}|_u$ whenever $u \in \Sigma^*$ satisfying $|u|_x \leq 1$ for every $x \in \Sigma$ and $|u| \leq N$. It follows that v_{N-1} and w_{N-1} are strongly M-equivalent by Theorem 1.2. Furthermore, the words v_{N-1} and w_{N-1} are clearly distinct. Therefore, we have shown that $w = w_{N-1}$ is a strongly M-ambiguous word. $\square$

Example 2. The 2-ambiguous word $babcbabcbabcbab$ is strongly M-ambiguous over the ternary alphabet $\{a, b, c\}$.

Because of Theorem 2.3, we find it interesting and not surprising if Şerbănuţă's work in [10] can be extended to the context of strong M-equivalence. Precisely, we conjecture that in fact over any alphabet there are finitely many strongly M-unambiguous words that are prints. However, one cannot simply reproduce the proof of Theorem 2.1. The word $abcabcabcabcabcabc$ is strongly M-equivalent to $cababab cabccabccab$ and thus strongly M-ambiguous but does not contain a 2-ambiguous word as a factor. This conjecture will be addressed by our forthcoming contribution.

3. Various Results about Şerbănuţă's Conjecture

First of all, assuming Şerbănuţă's Conjecture, we show that M-unambiguous words are sparse.

Theorem 3.1. *Let Σ be an ordered alphabet. Assume Şerbănuţă's Conjecture holds. Then*

$$\lim_{n\to\infty} \frac{\big|\{\, w \in \Sigma^* \mid |w| \leq n \text{ and } w \text{ is } M\text{-unambiguous} \,\}\big|}{\big|\{\, w \in \Sigma^* \mid |w| \leq n \,\}\big|} = 0.$$

Proof. Suppose $\Sigma = \{a_1 < a_2 < \cdots < a_s\}$. Let $\mathcal{L}$ consist of all M-unambiguous words over Σ, and let $\mathcal{L}_{pr}$ consist of all M-unambiguous prints over Σ. By Theorem 2.1, $\mathcal{L}_{pr}$ is finite and thus there is an upper bound on the length of words in $\mathcal{L}_{pr}$, say N. Suppose n is a (sufficiently large) positive

www.ijc.or.id

integer. Clearly, there are s^n many words over Σ of length n. Hence, $\left|\{\, w \in \Sigma^* \colon |w| \le n \,\}\right| \ge s^n$. Suppose $a_{t_1}a_{t_2}\cdots a_{t_l} \in \mathcal{L}_{pr}$ and $w = a_{t_1}^{n_1}a_{t_2}^{n_2}\cdots a_{t_l}^{n_l}$ for some $n_i > 0$ ($1 \le i \le l$). Note that w is uniquely determined by $(n_1, n_1+n_2, \ldots, n_1+n_2+\cdots n_l)$. Since the number of strictly increasing l-tuple of positive integers with the last and largest entry at most n is $\binom{n}{l}$, it follows that there are at most $\binom{n}{l}$ words of length at most n with the same given print of length l. Because $l \le N$, if n is sufficiently large, then $\binom{n}{l} \le \binom{n}{N}$. Hence, for each $v \in \mathcal{L}_{pr}$, there are at most $\binom{n}{N}$ M-unambiguous w of length at most n such that $\mathrm{pr}(w) = v$. Therefore, assuming Şerbănuţă's Conjecture, $\left|\{\, w \in \mathcal{L} \colon |w| \le n \,\}\right| \le |\mathcal{L}_{pr}| \binom{n}{N} \le |\mathcal{L}_{pr}|\, n^N$. Hence, $\lim_{n\to\infty} \frac{\left|\{\, w\in\mathcal{L} \colon |w|\le n \,\}\right|}{\left|\{\, w\in\Sigma^* \colon |w|\le n \,\}\right|} \le \lim_{n\to\infty} \frac{|\mathcal{L}_{pr}|n^N}{s^n} = 0$. □

Let $\Sigma = \{a < b\}$. It is a standard fact that a word $w \in \Sigma^*$ is M-ambiguous if and only if the words ab and ba occur in w as factors in non-overlapping positions [1, 4]. Using this, it can be shown that there is a total of $6n - 10$ M-unambiguous words over Σ of length n, provided $n \ge 4$. It is conceivable that the conclusion of Theorem 3.1 holds regardless of the validity of Şerbănuţă's Conjecture. However, this is left as an open problem.

Definition 3.1. *Suppose Σ is an ordered alphabet and let $w \in \Sigma^*$ be a print. We say that w is a* uniformly M-ambiguous print *iff w' is M-ambiguous whenever $w' \in \Sigma^*$ and $\mathrm{pr}(w') = w$.*

Equivalently, Şerbănuţă's conjecture amounts to saying that over any ordered alphabet, every M-ambiguous print is uniformly M-ambiguous. However, if a print is ME-ambiguous, then it must be uniformly M-ambiguous by the next theorem.

Theorem 3.2. *Let Σ be an ordered alphabet and suppose $w \in \Sigma^*$. If the print of w is ME-ambiguous, then w is ME-ambiguous.*

Proof. There exists $w_0, w_1, \ldots, w_n \in \Sigma^*$ satisfying $w_0 = \mathrm{pr}(w)$, $w_n = w$, and for each $0 \le i \le n-1$, we have $w_i = uav$ and $w_{i+1} = uaav$ for some $a \in \Sigma$ and $u, v \in \Sigma^*$. Hence, it suffices to prove that if uav is ME-ambiguous, then $uaav$ is ME-ambiguous for $a \in \Sigma$ and $u, v \in \Sigma^*$. Suppose uav is ME-ambiguous. Then Rule $E1$ or Rule $E2$ can be applied on uav. It is easy to check case by case that if one such rule can be applied on uav, then the same rule can be applied on $uaav$ accordingly. (Duplicating the letter a does not affect the applicability of Rule $E2$.) Therefore, $uaav$ is ME-ambiguous. □

The following shows that the analogous M-unambiguity version of Definition 3.1 is completely understood.

Theorem 3.3. *Let $\Sigma = \{a_1 < a_2 < \cdots < a_s\}$ be any ordered alphabet and suppose $\lambda \ne w \in \Sigma^*$ is a print. The following are equivalent.*

1. *Every word w' such that $\mathrm{pr}(w') = w$ is M-unambiguous.*
2. *w is either $a_i a_{i+1} \cdots a_j$ or $a_j a_{j-1} \cdots a_i$ for some $1 \le i \le j \le s$.*

Proof. ($1 \Rightarrow 2$) Because w is M-unambiguous, consecutive letters in w must be adjacent in Σ. Hence, the conclusion holds easily if w has length at most two. Assume $|w| \ge 3$. If xyx is a factor of w where the letters x and y are adjacent in Σ, then because $xyyx$ is M-equivalent to

www.ijc.or.id

$yxxy$, it follows that w is not uniformly M-unambiguous. Hence, w must be either $a_i a_{i+1} \cdots a_j$ or $a_j a_{j-1} \cdots a_i$ for some $1 \leq i \leq j \leq s$.

$(2 \Rightarrow 1)$ Suppose $w' = a_i^{n_i} a_{i+1}^{n_{i+1}} \cdots a_j^{n_j}$ for some $1 \leq i \leq j \leq s$ and $n_k > 0$ $(i \leq k \leq j)$ and $w'' \equiv_M w'$. For each $i \leq k \leq j-1$, the M-unambiguity of $a_k^{n_k} a_{k+1}^{n_{k+1}}$ over $\{a_k < a_{k+1}\}$ implies that the projection of w'' into $\{a_k, a_{k+1}\}^*$ is $a_k^{n_k} a_{k+1}^{n_{k+1}}$. It follows that $w'' = w'$ and thus w' is M-unambiguous. □

Due to Theorem 3.2, if w is a counterexample to Şerbănuţă's conjecture, then $\mathrm{pr}(w)$ is a properly M-ambiguous print. We discovered that there are exactly four properly M-ambiguous prints over $\{a < b < c\}$ with the shortest possible length: *ababcbabcb*, *cbcbabcbab*, *bcbabcbaba*, and *babcbabcbc*. However, *abbcbabcb*, *babcbabbc*, *bcbabcbba*, and *cbbabcbab* are shorter properly M-ambiguous words.

Acknowledgement

This work is supported by the grant No. 304/PMATHS/6313077 of Universiti Sains Malaysia. It has been presented in the Asian Mathematical Conference 2016 at Bali, Indonesia. Some time after that, Şerbănuţă's conjecture was disproved for the quaternary alphabet by Teh et al. [17].

References

[1] A. Atanasiu, Binary amiable words, *Internat. J. Found. Comput. Sci.* **18** (2007), 387–400.

[2] A. Atanasiu, Parikh matrices, amiability and Istrail morphism, *Internat. J. Found. Comput. Sci.* **21** (2010), 1021–1033.

[3] A. Atanasiu, R. Atanasiu, and I. Petre, Parikh matrices and amiable words, *Theoret. Comput. Sci.* **390** (2008), 102–109.

[4] S. Fossé and G. Richomme, Some characterizations of Parikh matrix equivalent binary words, *Inform. Process. Lett.* **92** (2004), 77–82.

[5] K. Mahalingam and K. G. Subramanian, Product of Parikh matrices and commutativity, *Internat. J. Found. Comput. Sci.* **23** (2012), 207–223.

[6] A. Mateescu, A. Salomaa, K. Salomaa, and S. Yu, A sharpening of the Parikh mapping, *Theor. Inform. Appl.* **35** (2001), 551–564.

[7] R. J. Parikh, On context-free languages, *J. Assoc. Comput. Mach.* **13** (1966), 570–581.

[8] A. Salomaa, On the injectivity of Parikh matrix mappings, *Fund. Inform.* **64** (2005), 391–404.

[9] A. Salomaa, Criteria for the matrix equivalence of words, *Theoret. Comput. Sci.* **411** (2010), 1818–1827.

[10] V. N. Şerbănuţă, On Parikh matrices, ambiguity, and prints, *Internat. J. Found. Comput. Sci.* **20** (2009), 151–165.

www.ijc.or.id

[11] V. N. Şerbănuţă and T.F. Şerbănuţă, Injectivity of the Parikh matrix mappings revisited, *Fund. Inform.* **73** (2006), 265–283.

[12] W. C. Teh, Parikh matrices and strong M-equivalence, *Internat. J. Found. Comput. Sci.* **27** (2016), 545–556.

[13] W. C. Teh, On core words and the Parikh matrix mapping, *Internat. J. Found. Comput. Sci.* **26** (2015), 123–142.

[14] W. C. Teh, Parikh matrices and Parikh rewriting systems, *Fund. Inform.* **146** (2016), 305–320.

[15] W. C. Teh, Separability of M-equivalent words by morphisms, *Internat. J. Found. Comput. Sci.* **27** (2016), 39–52.

[16] W. C. Teh and A. Atanasiu, On a conjecture about Parikh matrices, *Theoret. Comput. Sci.* **628** (2016), 30–39.

[17] W. C. Teh, A. Atanasiu, and G. Poovanandran, On strongly M-unambiguous prints and Şerbănuţăs conjecture for Parikh matrices, *Theoret. Comput. Sci.* **719** (2018), 86–93.

www.ijc.or.id