*Fundamenta Informaticae XX (2006) 1–19*
*IOS Press*



# Injectivity of the Parikh matrix mappings revisited

**Virgil Nicolae Şerbănuţă**
*Faculty of Mathematics and Informatics*
*University of Bucharest*
*Academiei 14, sector 1, 010014, Bucharest, Romania*
*serbanutav@gmail.com*

**Traian Florin Şerbănuţă***
*Department of Computer Science*
*University of Illinois at Urbana-Champaign*
*201 N. Goodwin, Urbana, IL 61801, USA*
*tserban2@cs.uiuc.edu*

**Abstract.** We deal with the notion of M-unambiguity [5] in connection with the Parikh matrix mapping introduced by Mateescu and others in [7]. M-unambiguity is studied both in terms of words and matrices and several sufficient criteria for M-unambiguity are provided in both cases, nontrivially generalizing the criteria based on the $\gamma$-property introduced by Salomaa in [15]. Also, the notion of M-unambiguity with respect to a word is defined in connection with the extended Parikh matrix morphism [16] and some of the M-unambiguity criteria are lifted from the classical setting to the extended one.

This paper is an revised and extended version of [17].

**Keywords:** subword, scattered subword, Parikh matrix, ambiguity

## 1. Introduction

The Parikh matrix mapping was introduced by Mateescu and others in [7] as a maping from words to algebraic structures (matrices) in the spirit of the classical Parikh mapping [9] which associates vectors to

Address for correspondence: Traian Florin Şerbănuţă, 201 N Goodwin, Urbana IL 61801, USA
*Also, Faculty of Mathematics and Informatics, University of Bucharest

words. By using matrices instead of vectors more information about the word is preserved and numerical facts such as the number of occurrences of certain subwords in a word can be elegantly computed (by matrix multiplication). Because of the easiness in dealing with subword occurrences some interesting problems were discovered and solved using this tool in fields like combinatorics on words [6, 4, 8, 5, 13, 14, 3] and theory of codes [2, 1].

Also the question of a word being determined by the number of occurrences of some of its subwords has been asked in this framework leading to the notion of M-unambiguity of an word - that is a word being uniquely determined by its corresponding Parikh matrix. Although several articles [2, 4, 5, 15, 3] are dealing with this notion and M-unambiguous words for alphabets with two letters have been completely characterized ([2, 4, 5]), it seems that a complete characterization of M-unambiguous words for general alphabets is still long ahead of us. We add our contribution to this still open question by giving new syntactical (in terms of words) and semantical (in terms of matrices) criteria for M-unambiguity. Although developed independently, our results seem to non-trivially generalize the results obtained by Salomaa in [15] using the so-called $\gamma$-property; yet the way the results from [15] were expressed enabled us to strengthen our results by expressing them in a different manner.

The paper is structured as follows: Section 2 reproduces some known definitions and results from [7, 5, 15, 16] in order to allow a self-contained reading of the paper. Section 3 gives M-unambiguity criteria for words and Parikh matrices. Section 4 lifts some of the results obtained in Section 3 to the case of extended Parikh Matrices [16]. We conclude in Section 5 mentioning some open problems. A characterization of M-unambiguous words for three letter alphabets is given in the appendix in the hope that some of the techniques used there might be generalized at some point in the future.

## 2. Preliminaries

We will assume the reader familiar with the basics of formal languages. Whenever necessary, [12, 10] may be consulted. As customary, we use small letters from the begining of the English alphabet $a, b, c, d$ possibly with indices, to denote letters of our formal alphabet $\Sigma$. Words are usually denoted by small letters from the end of the English alphabet.

### 2.1. Subwords

Let $\Sigma$ be an alphabet. The set of all words over $\Sigma$ is denoted $\Sigma^*$ and the empty word is $\lambda$. If $w \in \Sigma^*$ then $|w|$ denotes the length of $w$.

**Definition 2.1.** Let $\Sigma$ be an alphabet and $u, w \in \Sigma^*$. We say that $u$ is a *scattered subword* (or simply *subword*) of $w$ if $w$, as a sequence of letters, contains $u$ as a subsequence. Formally, this means that there exist words $x_1, \ldots, x_k$ and $y_0, \ldots, y_k$ in $\Sigma^*$, some of them possibly empty such that

$$u = x_1 \ldots x_k \text{ and } w = y_0 x_1 y_1 \ldots x_k y_k.$$

More formally, $a_1 a_2 \ldots a_k$ is a subword of $b_1 b_2 \ldots b_n$ (where $a_i \in \Sigma$ for all $1 \leq i \leq k$ and $b_j \in \Sigma$ for all $1 \leq j \leq n$) if there exists a mapping $f : \{1, \ldots, k\} \to \{1, \ldots, n\}$ so that $f(i) < f(i+1)$ for all $1 \leq i < k$ and $b_{f(i)} = a_i$ for all $1 \leq i \leq k$.

We will denote by $|w|_u$ the *number of occurrences* of word $u$ as a subword in $w$, that is the number of mappings that can be defined with respect to the above definition. For instance,

$$|abba|_{ba} = 2 \text{ and } |aabbc|_{abc} = 4.$$

In some works ([11]), the number $|w|_u$ is denoted as the binomial coefficient. Indeed, if the alphabet $\Sigma$ contains only one letter, the number $|w|_u$ reduces to the number of mappings $f : \{1, \ldots, |u|\} \to \{1, \ldots, |w|\}$ so that $f(i) < f(i+1)$ for all $1 \le i < |u|$, and that is exactly the binomial coefficient.

It is easy to see that if $|w| < |u|$ then $|w|_u = 0$. Also, if $u = \lambda$ then $|w|_u = 1$ because $\{1, \ldots, |u|\} = \emptyset$ and the inclusion $\emptyset \hookrightarrow \{1, \ldots, |w|\}$ is the only possible mapping (it clearly satisfies the definition).

Let $a$, $b$ be two letters in an alphabet $\Sigma$. We denote by $\delta_{a,b}$ be the *Kronecker Symbol* regarding letters, that is

$$\delta_{a,b} = \begin{cases} 1, \text{ if } a = b \\ 0, \text{ if } a \neq b \end{cases}$$

**Fact 2.1.** It is shown in [11] that the equation

$$|vb|_{ua} = |v|_{ua} + \delta_{a,b}|v|_u, a, b \in \Sigma; u, v \in \Sigma^*$$

together with the equations $|w|_\lambda = 1$ and $|w|_u = 0$ for $|w| < |u|$ suffice to compute all values $|w|_u$.

## 2.2. Parikh matrices

The notion of Parikh matrix was introduced in [7]. All definitions and results presented in this subsection can be found in [7, 6, 8].

The definition of the Parikh matrix mapping presented below uses a special type of matrices, called *triangle matrices*. A triangle matrix is a square matrix $M = (m_{i,j})_{1 \le i,j \le k}$, such that $m_{i,j}$ is a nonnegative integer for all $1 \le i, j \le k$, $m_{i,j} = 0$ for all $1 \le j < i \le k$ and $m_{i,i} = 1$ for all $1 \le i \le k$.

The set of all triangle matrices is denoted by $\mathcal{M}$. The set of all triangle matrices of dimension $k \ge 1$ is denoted by $\mathcal{M}_k$. Clearly $(\mathcal{M}_k, \cdot, I_k)$, where $\cdot$ represents the matrix multiplication and $I_k$ is the unit matrix, is a monoid.

An *ordered alphabet* is an alphabet $\Sigma = \{a_1, \ldots, a_k\}$ with a relation of order $<$ on it. If we have $a_1 < a_2 < \ldots < a_k$, then we will use the notation $\Sigma = \{a_1 < a_2 < \ldots < a_k\}$.

**Definition 2.2.** Let $\Sigma = \{a_1 < \ldots < a_k\}$ be an ordered alphabet. The *Parikh matrix mapping*, denoted $\Psi_\Sigma$, is the monoid morphism:

$$\Psi_\Sigma : (\Sigma^*, \cdot, \lambda) \to (\mathcal{M}_{k+1}, \cdot, I_{k+1}),$$

defined by the condition: if $\Psi_\Sigma(a_q) = (m_{i,j})_{1 \le i,j \le (k+1)}$, then for each $1 \le i \le (k+1)$, $m_{i,i} = 1$, $m_{q,q+1} = 1$, and all other elements of the matrix $\Psi_\Sigma(a_q)$ are $0$.

For the ordered alphabet $\Sigma = \{a_1 < \ldots < a_k\}$, we denote by $a_{i,j}$ the word $a_i a_{i+1} \ldots a_j$, where $1 \le i \le j \le k$.

The following theorem characterizes the entries of the Parikh matrix.

**Theorem 2.1.** Let $\Sigma = \{a_1 < \ldots < a_k\}$ be an ordered alphabet and $w \in \Sigma^*$. The matrix $\Psi_\Sigma(w) = (m_{i,j})_{1 \le i,j \le (k+1)}$, has the following properties:

- $m_{i,j} = 0$, for all $1 \le j < i \le (k+1)$,
- $m_{i,i} = 1$, for all $1 \le i \le (k+1)$,
- $m_{i,j+1} = |w|_{a_{i,j}}$, for all $1 \le i \le j \le k$.

Let $M = (m_{i,j})_{1 \le i,j \le k}$ be a triangle matrix. The *alternate matrix of* $M$, denoted by $\overline{M}$, is the matrix $\overline{M} = (m'_{i,j})_{1 \le i,j \le k}$, where $m'_{i,j} = (-1)^{i+j}(M)_{i,j}$ for all $1 \le i,j \le k$. The *reverse of* $M$, denoted by $M^{(rev)}$, is the matrix $M^{(rev)} = (m''_{i,j})_{1 \le i,j \le k}$, where $m''_{i,j} = m_{k+1-j,k+1-i}$, for all $1 \le i < j \le k$. (The entries below the main diagonal are the same in $M$ and $M^{(rev)}$). Given a word $w = a_1 \ldots a_n$ ($a_i \in \Sigma$ for all $1 \le i \le n$), we denote by *mi*$(w)$ the *mirror image* of word w, that is *mi*$(w) = a_n a_{n-1} \ldots a_1$. Let $(A, <)$ be an ordered set. The *dual order* of the order $<$, denoted $<^\circ$, is defined as:

$$a <^\circ b \text{ iff } b < a.$$

Let $\Sigma = \{a_1 < a_2 < \ldots < a_k\}$ be an ordered alphabet. The *dual ordered alphabet*, denoted $\Sigma_\circ$, is $\Sigma_\circ = \{a_k < a_{k-1} < \ldots < a_1\}$. The following theorem characterizes the inverse of a Parikh matrix.

**Theorem 2.2.** Let $\Sigma = \{a_1 < a_2 < \ldots < a_k\}$ be an ordered alphabet and let $w \in \Sigma^*$ be a word. Then:

$$[\Psi_\Sigma(w)]^{-1} = \overline{\Psi_\Sigma(\textit{mi}(w))} = \overline{\Psi_{\Sigma_\circ}(w)^{(rev)}}$$

## 2.3. Ambiguity

The notion of ambiguity was studied in [4, 2] for two letter alphabets even before it was introduced in [5]. Instead of reproducing the original definition, we prefer to give here a rephrased version of it taken from [3].

**Definition 2.3.** Let $\Sigma = \{a_1 < \cdots < a_k\}$ be an ordered alphabet. Two words $w_1, w_2 \in \Sigma^*$ are termed *M-equivalent*, in symbols $w_1 \equiv_M w_2$, if $\Psi_\Sigma(w_1) = \Psi_\Sigma(w_2)$. A word $w \in \Sigma^*$ is termed *M-unambiguous* if there is no word $w' \neq w$ such that $w \equiv_M w'$. Otherwise, $w$ is termed *M-ambiguous*. If $w \in \Sigma^*$ is M-unambiguous (resp. ambiguous), then also the Parikh matrix $\Psi_\Sigma(w)$ is called unambiguous (resp. ambiguous).

A word being M-unambiguous means that it is uniquely determined by its Parikh matrix. Let us list some basic results about M-unambiguity from [5] (see also [17]). The first result shows that any factor of an M-unambiguous word is also M-unambiguous.

**Proposition 2.1.** If a word $y \in \Sigma$ is M-ambiguous, so is every word $xyz$ where $x, z \in \Sigma^*$.

Next result lists some short M-ambiguous words.

**Proposition 2.2.** Consider the alphabet $\Sigma = \{a_1 < \cdots < a_k\}$. The following words are M-ambiguous:

- $a_i a_j$ with $|i - j| > 1$;

- $a_i a_j^{m+2} a_i$ and $a_j a_i a_j^m a_i a_j$ where $|i-j| = 1$ and $m \geq 0$.

The following corollary says that adjacent letters in a M-unambiguous word must be equal or consecutive in the alphabet.

**Corollary 2.1.** If $w$ is M-unambiguous (over $\Sigma = \{a_1 < \cdots < a_k\}$) and $a_i a_j$ is a factor of $w$ then $|i-j| \leq 1$. That is, the only factors of length two of $w$ are of form:

$$a_i a_i, a_i a_{i+1} \text{ or } a_{i+1} a_i$$

Next result from [5] (see also [4, 2]) gives a complete characterization for M-unambiguous words of length 2.

**Theorem 2.3.** A word in $\{a < b\}^*$ is M-ambiguous if and only if it contains disjoint occurrences of $ab$ and $ba$. A word is M-unambiguous if and only if it belong to the language denoted by the regular expression

$$a^*b^* + b^*a^* + a^*ba^* + b^*ab^* + a^*bab^* + b^*aba^*$$

In [15] another useful property, namely the $\gamma$-*property* is introduced to give M-unambiguity criteria. We reproduce below the definition along with some results concerning the $\gamma$-property presented in [15].

**Definition 2.4.** Let $\gamma : \mathbb{N} \times \mathbb{N} \to 2^{\mathbb{N}}$ be the mapping defined by:

$$\gamma(m,n) = \begin{cases} \{i \mid 0 \leq i \leq mn\} & \text{if } m \leq 1 \text{ or } n \leq 1, \\ \{0, 1, mn, mn-1\} & \text{if } m > 1 \text{ and } n > 1. \end{cases}$$

A $(k+1)$-dimensional Parikh matrix $M$, $k \geq 2$, possesses the $\gamma$-*property* if each entry $m_{i,i+2}$ in the third diagonal is in the set $\gamma(m_{i,i+1}, m_{i+1,i+2})$.

The following result is an alternative characterization of unambiguous Parikh matrices over binary alphabets.

**Theorem 2.4.** A Parikh matrix over a binary alphabet is unambiguous if and only if it possesses the $\gamma$-property.

Also, a M-unambiguity criteria for an arbitrary ordered alphabet $\Sigma = \{a_1 < \cdots < a_k\}$ is given.

**Theorem 2.5.** Assume that $\Psi_\Sigma(w)$ possesses the $\gamma$-property and that every length two factor of $w$ has one of the forms

$$a_i a_i, 1 \leq i \leq k, \text{ or } a_i a_{i+1}, a_{i+1} a_i, 1 \leq i \leq k-1.$$

Then $w$ is $M$-unambiguous (and so is $\Psi_\Sigma(w)$).

For more results and interesting examples of M-unambiguous words, consult [5, 15, 3].

## 2.4. Extended Parikh Matrices

When studying a word $w$ in terms of the number of occurrences of certain subwords in it, one may think of considering a so called *basic word* $u$ (see for example [15]) and count the number of occurrences of each of its factors in $w$.

The Parikh matrix introduced above uses the catenation $a_1 \dots a_k$ of all letters in the alphabet $\Sigma = \{a_1 < \cdots < a_k\}$ in the proper order as the basic word: the matrix giving the values $|w|_v$ for factors $v$ of the basic word. In the extended Parikh matrix mapping [16] any word (also with repeating letters) can be chosen as the basic word. The following definitions and results can be found in [16].

**Definition 2.5.** Let $\Sigma$ be an alphabet and $u = b_1 \dots b_{|u|}$ be a word in $\Sigma^*$ ($b_i \in \Sigma$ for all $1 \le i \le |u|$). The *Parikh matrix mapping induced by the word* $u$ *over the alphabet* $\Sigma$ (shortly, the $u$*-Parikh matrix mapping*), denoted $\Psi_{\Sigma,u}$, is the monoid morphism

$$\Psi_{\Sigma,u} : (\Sigma^*, \cdot, \lambda) \to (\mathcal{M}_{|u|+1}, \cdot, I_{|u|+1}),$$

defined by the condition: if $a \in \Sigma$ and $\Psi_{\Sigma,u}(a) = (m_{i,j})_{1 \le i,j \le (|u|+1)}$, then:

$$m_{i,j} = \begin{cases} 1 & \text{if } j = i \\ \delta_{b_i,a} & \text{if } j = i+1 \\ 0 & \text{otherwise} \end{cases}$$

Since in the sequel we will mainly be concerned with M-unambiguity, we will assume that $\Sigma$ is determined by $u$ (for reasons which will become clear shortly), that is, $u$ contains all letters of $\Sigma$, and use the notation $\Psi_u$ for the $u$-Parikh matrix mapping.

Similarly to the notation $a_{i,j}$ in the case of an ordered alphabet we introduce the following notation: given the word $u = b_1 \dots b_n$, we denote by $u_{i,j}$ the word $b_i b_{i+1} \dots b_j$, where $1 \le i \le j \le n$. Using this notation we can give a similar theorem characterizing the entries of an $u$-Parikh matrix.

**Theorem 2.6.** Consider $u, w \in \Sigma^*$. The matrix $\Psi_u(w) = (m_{i,j})_{1 \le i,j \le (|u|+1)}$, has the following properties:

$(i)$ $m_{i,j} = 0$, for all $1 \le j < i \le (|u| + 1)$,

$(ii)$ $m_{i,i} = 1$, for all $1 \le i \le (|u| + 1)$,

$(iii)$ $m_{i,j+1} = |w|_{u_{i,j}}$, for all $1 \le i \le j \le |u|$.

A result similar to Theorem 2.2 cannot be given for Parikh matrices induced by any word. However, it can be given for all words $u$ not having consecutive equal letters.

**Theorem 2.7.** Let $u \in \Sigma^*$ be a word such that $aa$ is not a factor of $u$ for any $a \in \Sigma$. Then:

$$[\Psi_u(w)]^{-1} = \overline{\Psi_u(\mathit{mi}(w))}.$$

Related to the inverse of a Parikh matrix the following holds for arbitrary $u$.

**Theorem 2.8.**

$$\Psi_u(\mathit{mi}(w)) = \Psi_{\mathit{mi}(u)}(w)^{(rev)}.$$

The following result shows that any $u$-Parikh matrix can be obtained as a Parink matrix over a (different) ordered alphabet. To make the presentation clearer we will use a different style than in [16].

Fix a word $u = a_1 \ldots a_k \in \Sigma^*$. Associate to $u$ the ordered alphabet $\Sigma_k = \{1 < 2 < \cdots < k\}$ and for each letter $a \in \Sigma$, let $\mathit{trace}_u(a)$ be the ordered sequence $i_1 i_2 \ldots i_{|u|_a} \in \Sigma_k^*$ of positions in $u$ on which $a$ occurs, that is, $a_{i_j} = a$ for all $1 \leq j \leq |u|_a$. For example, $\mathit{trace}_{baraba}(b) = 1\,5$ and $\mathit{trace}_{baraba}(a) = 2\,4\,6$.

**Theorem 2.9.** Let $\varphi : \Sigma^* \to \Sigma_k^*$ be the morphism given by $\varphi(a) = \mathit{mi}(\mathit{trace}_u(a))$. Then for each $w \in \Sigma^*$,

$$\Psi_u(w) = \Psi_{\Sigma_k}(\varphi(w))$$

## 3. New M-unambiguity results

Let $\Sigma = \{a_1 < \ldots < a_k\}$ be an ordered alphabet. Let $\varphi^\circ : \Sigma^* \to \Sigma^*$ denote the only morphism given by $\varphi^\circ(a_i) = a_{k-i+1}$ for any $1 \leq i \leq k$. It is easy to see that $\Psi_\Sigma(\varphi^\circ(w)) = \Psi_{\Sigma,\circ}(w)$.

**Proposition 3.1.** For any word $w$ and any ordered alphabet $\Sigma$, the following are equivalent:

1. $w$ is M-unambiguous;
2. $\mathit{mi}(w)$ is M-unambiguous;
3. $\varphi^\circ(w)$ is M-unambiguous;
4. $\mathit{mi}(\varphi^\circ(w))$ is M-unambiguous;

**Proof:**

**"$1 \iff 2$":** $\Psi_\Sigma(\mathit{mi}(w)) = \Psi_\Sigma(\mathit{mi}(w'))$ iff $\overline{\Psi_\Sigma}(\mathit{mi}(w)) = \overline{\Psi_\Sigma}(\mathit{mi}(w'))$ iff $[\Psi_\Sigma(w)]^{-1} = [\Psi_\Sigma(w')]^{-1}$ iff $\Psi_\Sigma(w) = \Psi_\Sigma(w')$

**"$1 \iff 3$":** $\Psi_\Sigma(\varphi^\circ(w)) = \Psi_\Sigma(\varphi^\circ(w'))$ iff $\Psi_{\Sigma,\circ}(w) = \Psi_{\Sigma,\circ}(w')$ iff $\overline{\Psi_{\Sigma,\circ}}(w) = \overline{\Psi_{\Sigma,\circ}}(w')$ iff $[\overline{\Psi_{\Sigma,\circ}}(w)]^{(rev)} = [\overline{\Psi_{\Sigma,\circ}}(w')]^{(rev)}$ iff $[\Psi_\Sigma(w)]^{-1} = [\Psi_\Sigma(w')]^{-1}$ iff $\Psi_\Sigma(w) = \Psi_\Sigma(w')$

**"$1 \iff 4$":** $\Psi_\Sigma(w) = \Psi_\Sigma(w')$ iff $\Psi_\Sigma(\varphi^\circ(w)) = \Psi_\Sigma(\varphi^\circ(w'))$ iff $\Psi_\Sigma(\mathit{mi}(\varphi^\circ(w))) = \Psi_\Sigma(\mathit{mi}(\varphi^\circ(w')))$ ∎

□

Let $\Sigma$ be an alphabet and $\Sigma' \subseteq \Sigma$ be a subalphabet of $\Sigma$. *The projection of* $\Sigma^*$ *to* $\Sigma'^*$ is the only morphism mapping the letters of $\Sigma'$ to themselves and the remaining letters of $\Sigma$ to the empty word (see [15], for example). In the sequel, given an ordered alphabet $\Sigma = \{a_1 < \cdots < a_k\}$, for any $1 \leq i \leq j \leq k$, let $\pi_{i,j}$ denote the projection of $\Sigma^*$ to $\{a_i < \cdots < a_j\}$. Also, we will use $\Psi_{i,j}$ as a short notation for $\Psi_{\{a_i < \cdots < a_j\}}$. Given a matrix $A \in \mathcal{M}_k$ and $1 \leq p \leq q \leq k$, let $A_{p,q}$ denote the submatrix of $A$ at the intersection of lines and columns between $p$ and $q+1$. This notation is not so intuitive in terms of matrices, but as next result shows, it is closely related to the projection on a restricted alphabet.

**Theorem 3.1.** Let $\Sigma = \{a_1 < \cdots < a_k\}$ and let $1 \leq p \leq q \leq k$. Then for any word $w$ we have that

$$[\Psi_\Sigma(w)]_{p,q} = \Psi_{p,q}(\pi_{p,q}(w))$$

**Proof:**
It clearly holds due to the Theorem 2.1 and to the fact that for each $p \leq i \leq j \leq q$ we have that $|w|_{a_{i,j}} = |\pi_{p,q}(w)|_{a_{i,j}}$. The latter is true since the projection neither deletes nor changes the order of letters between $a_p$ and $a_q$ □

As a corollary we get the following characterization of M-equivalence for projections.

**Corollary 3.1.** Let $\Sigma = \{a_1 < \cdots < a_k\}$ and $u, w \in \Sigma^*$ such that $u \equiv_M w$. Then for each $1 \leq p \leq q \leq k$,

- $\pi_{p,q}(u) \equiv_M \pi_{p,q}(w)$;
- if $\pi_{p,q}(w)$ is M-unambiguous then $\pi_{p,q}(u) = \pi_{p,q}(w)$.

**Proof:**
The first is a direct consequence of the theorem. The second holds from the first since M-equivalence for M-unambiguous words reduces to equality. □

The following result shows that one may prove a word $w \in \Sigma^*$ to be M-unambiguous if it manages to prove that its projection on selected subalphabets of $\Sigma$ is M-unambiguous.

**Theorem 3.2.** Let $\Sigma = \{a_1 < \cdots < a_k\}$ be an ordered alphabet. Let $w \in \Sigma^*$ such that all its length two factors are of the form

$$a_i a_i, 1 \leq i \leq k, \text{ or } a_i a_{i+1}, a_{i+1} a_i, 1 \leq i \leq k-1.$$

If there exist $p, q$ such that $1 < p \leq q < k$ and both $\pi_{1,q}(w)$ and $\pi_{p,k}(w)$ are M-unambiguous then $w$ is also M-unambiguous.

**Proof:**

If $\pi_{1,q}(w)$ or $\pi_{p,k}(w)$ are $\lambda$ then our proof is done.

Else, suppose by contradiction there exists $u \neq w$ such that $\Psi_\Sigma(w) = \Psi_\Sigma(u)$. Then, by Corollary 3.1 we have that $\pi_{1,q}(u) = \pi_{1,q}(w)$ and $\pi_{p,k}(u) = \pi_{p,k}(w)$. Suppose now that $w = b_1 \ldots b_n$ and $u = c_1 \ldots c_n$ and let $1 \leq i \leq n$ be the smallest integer such that $b_i \neq c_i$. Also, suppose $\pi_{1,q}(w) = d_1 \ldots d_m (= \pi_{1,q}(u))$ and let $j$ be such that $d_1 \ldots d_{j-1} = \pi_{1,q}(b_1 \ldots b_{i-1}) = \pi_{1,q}(c_1 \ldots c_{i-1})$.

Case 1: $i > 1$

If $b_i \in \{a_1 \ldots a_{p-1}\}$ then $b_{i-1} \in \{a_1 \ldots a_p\}$. But $b_{i-1} = c_{i-1}$, so $c_i \in \{a_1 \ldots a_{p+1}\}$. Since $b_i \in \{a_1 \ldots a_{p-1}\}$,it must be that $d_j = b_i$. If $p < q$ or [$p = q$ and $c_i \neq a_{p+1}$] then, since $c_i \in \{a_1 \ldots a_q\}$ we must have that $d_j = c_i$. But this leads to $b_i = c_i$, a contradiction. If $p = q$ and $c_i = a_{p+1}$ then $b_{i-1} = c_{i-1} = a_p$ whence the first letter in $u$ whose index is greater than $i$ and belongs to $\{a_1 \ldots a_p\}$ must be $a_p$, implying that $d_j = a_p$, contradiction with $d_j = b_i \in \{a_1 \ldots a_{p-1}\}$

By a similar argument, but using $\pi_{p,k}$, $b_i \notin \{a_{q+1} \dots a_k\}$.

If $b_i \in a_p \dots a_q$ then if $c_i \in \{a_1 \dots a_q\}$ then at position $j$ we can observe that $\pi_{1,q}(b_i) = \pi_{1,q}(c_i)$, contradiction. The same way $c_i \in \{a_{q+1} \dots a_k\}$ leads to a contradiction using $\pi_{p,k}$.

Case 2: $i = 1$

If $b_1 \in \{a_1 \dots a_{p-1}\}$ then the first letter in $w$ whose index is greater than 1 and does not belong to $\{a_1 \dots a_{p-1}\}$ is $a_p$, so $\pi_{p,k}(w)$ starts with $a_p$; thus $u$ cannot start with a letter from $\{a_{q+1} \dots a_k\}$. This means that (using $\pi_{1,q}$) $b_1 = d_1 = c_1$, a contradiction.

The same way, $b_1 \notin \{a_{q+1} \dots a_k\}$

If $b_1 \in \{a_p \dots a_q\}$ then $\pi_{1,q}(w)$ and $\pi_{p,k}(w)$ start with $b_1$, so $\pi_{1,q}(u)$ and $\pi_{p,k}(u)$ start with $b_1$, which means that $b_1 = c_1$, contradiction.

□

By iteratively applying the above theorem, the following result is immediate.

**Corollary 3.2.** Let $w$ be as in Theorem 3.2. If there exists a sequence of $n$ pairs $(p_i, q_i)$, $1 \leq i \leq n$ such that

- $p_0 = 1$ and $q_n = k$,
- $p_i < q_i$ and
- $p_{i+1} \leq q_i$

and $\pi_{p_i,q_i}(w)$ is M-unambiguous for each $1 \leq i \leq n$, then $w$ is also M-unambiguous.

Analyzing the $\gamma$-property more carefully, one can see it as an instance of the above Corollary for the sequence $(i, i+1)$ where $1 \leq i < k$. Indeed, applying the Corollary we get that $w$ is M-unambiguous if $\pi_{i,i+1}(w)$ is M-unambiguous for $1 \leq i < k$. But this is equivalent with $\Psi_\Sigma(w)$ having the $\gamma$-property, since Theorems 2.4 and 2.5 basically say that $A = \Psi_\Sigma(w)$ has the $\gamma$-property if and only if each submatrix $A_{i,i+2} = \Psi_{i,i+1}(\pi_{i,i+1}(w))$ of $A$, $1 \leq i < k$ is unambiguous.

To show that the above theorem is more powerful than Theorem 2.5, consider the word $abcdcbcde$ over the alphabet $a < b < c < d < e$. The projections of $w$ on alphabets $\{a < b\}$, $\{b < c\}$, $\{c < d\}$ and $\{d < e\}$ are:

- $\pi_{a<b}(abcdcbcde) = abb$,
- $\pi_{b<c}(abcdcbcde) = bccbc$,
- $\pi_{c<d}(abcdcbcde) = cdccd$ and
- $\pi_{d<e}(abcdcbcde) = dde$.

It is clear that we cannot apply Theorem 2.5 to prove its M-unambiguity, since only the first and the last projections yield M-unambiguous words. On the other hand, we can use the decomposition $\{a < b < c < d\}$ and $\{b < c < d < e\}$ to obtain:

- $\pi_{a<b<c<d}(abcdcbcde) = abcdcbcd$ and
- $\pi_{b<c<d<e}(abcdcbcde) = bcdcdcde$ and

First notice that $bcdcdcd$ is M-unambiguous (see appendix, Theorem A.1), so the order of $b$, $c$ and $d$ is fixed. Now, since the number of $ab$s in $abcdcbcd$ is 2, $a$ can only occur as the first letter in the word, thus the word is completely determined. A similar argument holds for $bcdcdcde$, thus $abcdcbcde$ is M-unambiguous

Next theorem gives a similar criteria for M-unambiguity, this time without imposing special conditions on the factors of $w$.

**Theorem 3.3.** Let $\Sigma = \{a_1 < \cdots < a_k\}$ be an ordered alphabet, let $p, q \in \mathbb{N}$ such that $1 < p < q < k$ and let $w \in \Sigma^*$ such that $\pi_{1,q}(w)$ and $\pi_{p,k}(w)$ are M-unambiguous and $\pi_{p,q}(w) \neq \lambda$. Then $w$ is also M-unambiguous.

Since one can easily check whether the two length factors word $w$ satisfy the conditions of Theorem 3.2, this is a strictly less useful result than Theorem 3.2. However, it can be rephrased as the following completely algebraic (not referring to words) equivalent criteria for unambiguity of Parikh matrices giving a test for unambiguity for matrices known to be Parikh but with unknown generating word.

**Theorem 3.4.** Let $A$ be a Parikh matrix. If we can find $p$ and $q$, $1 \leq p < q \leq k$ such that $A_{1,q}$ and $A_{p,k}$ are unambiguous and $A_{p,q} \neq I$, then $A$ is also unambiguous

**Proof:**
We will show that we can apply Theorem 3.2. Suppose by contradiction there exists $a_i a_j$ a factor of $w$ such that $|j - i| > 1$. Since $\pi_{1,q}(w)$ and $\pi_{p,k}(w)$ are both M-unambiguous, it must be that either $i < p$ and $j > q$ or $i > q$ and $j < p$. Without loss of generality, let us assume that $i < p$ and $j > q$. Also, since $\pi_{p,q}(w) \neq \lambda$ it must be that there exist an occurrence of a letter $a_r$ in $w$ such that $p \leq r \leq q$.

If $a_r$ occurs at the right of $a_i a_j$ then since $\pi_{1,q}(w)$ is M-unambiguous all letters in $\Sigma$ having indexes between $i$ and $r$ must occur in $w$ between $a_i a_j$ and $a_r$. Moreover, $a_p$ must be the first letter occuring at right of $a_i a_j$ having index greater than $p - 1$. But this precisely means that $a_j a_p$ is a factor of the M-unambiguous word $\pi_{p,k}(w)$, a contradiction, since $j - p > q - p = 1$.

If $a_r$ occurs at the left of $a_i a_j$ the same argument as above holds interchanging the roles of $\pi_{1,q}$ and $\pi_{p,k}$.

□

The following is a converse of Theorem 3.3.

**Theorem 3.5.** Let $\Sigma = \{a_1 < \cdots < a_k\}$ be an ordered alphabet, let $p, q \in \mathbb{N}$ such that $1 < p < q < k$ and let $w \in \Sigma^*$ such that $w$ and $\pi_{p,q}(w)$ are M-unambiguous. Then both $\pi_{1,q}(w)$ and $\pi_{p,k}(w)$ are also M-unambiguous.

With its equivalent matrix formulation.

**Theorem 3.6.** Let $A$ be an unambiguous Parikh matrix. If we can find $p$ and $q$, $1 \leq p < q \leq k+1$ such that $A_{p,q}$ is unambiguous, then both $A_{1,q}$ and $A_{p,k}$ are unambiguous.

**Proof:**

First notice that inside an M-unambiguous word, the first and the last letters in the alphabet may have powers greater than 1 only at the beginning or at the end of the word (it is almost obvious from Proposition 2.2). Let $w' = \pi_{p,q}(w)$. Since $w'$ is M-unambiguous it follows that $a_p$ and $a_q$ can have powers greater than 1 only at the beginning or the end of $w'$.

Since $w'$ is a scattered subword of $w$, $w$ can be obtained by inserting in $w'$ some letters from $\{a_1, \ldots, a_{p-1}, a_{q+1}, \ldots, a_k\}$. From the observation above and from the fact that for all factors $a_i a_j$ of $w$, $|i - j| \leq 1$ must hold, it can easily be seen that the letters may be inserted in $w$ only inside the $a_p$ or $a_q$ groups at the beginning and the end of the word (if such groups exist) or before and after the word, if the first and the last letter allows us to. Because of this fact, it follows that if $w'$ neither begins nor ends with $a_p$ or $a_q$ then $w = w'$, since no letters from $\{a_1, \ldots, a_{p-1}, a_{q+1}, \ldots, a_k\}$ can be inserted in $w'$

If $w'$ begins and ends with $a_p$, then none of the letters $a_{q+1} \ldots a_k$ may be added to $w'$, thus $\pi_{1,q}(w) = w$ and $\pi_{p,k}(w) = w'$.

If $w'$ begins with $a_p$ and ends with $a_q$, then the letters $a_1 \ldots a_{p-1}$ may be added only inside or before the $a_p$ group at the beginning of $w'$, and letters $a_{q+1} \ldots a_k$ may be added only inside or after the $a_q$ group at the end of $w'$. Suppose by contradiction that $\pi_{1,q}(w)$ is M-ambiguous and let $u \neq \pi_{1,q}(w)$ such that $u \equiv_M \pi_{1,q}(w)$. Then $\pi_{p,q}(u) = w'$ (by Corollary 3.1). Now construct $w''$ from $u$ by inserting the letters $a_{q+1} \ldots a_k$ in the same relative positions they have in $w$ with respect to the letters in $w'$ (there may be more than one way to add them). One can see that that $w \equiv_M w''$, a contradiction. Indeed, it is obvious that $|w|_{a_{r,s}} = |w''|_{a_{r,s}}$ if $r, s \leq q$ or $r, s \geq p$. Let's see what happens if $r < p$ and $q < s$. A scattered occurrence of $a_{r,s}$ in $w''$ is given by an occurrence of $a_{r,q}$ and an occurrence of $a_{q+1,s}$ after that. Note that $a_{q+1} \ldots a_s$ may be found only after the last $a_{q-1}$, and $q - 1 \geq p$. For any scattered occurrence of $a_{r,q-1}$ in $w$ we have an occurrence of $a_{r,q-1}$ in $w''$, and for any occurrence of $a_{r,q}$ in $w$ we have one in $w''$. We can't have any occurrence of $a_{r,q}$ after the last $a_{q-1}$, so the number of occurrences of $a_{r,s}$ in $w$ is the same with the number of occurrences in $w''$.

The other cases for the beginning and the end of $w'$ are treated in a similar manner. $\square$

To show that the M-unambiguity condition for $\pi_{p,q}(w)$ is indeed needed by the theorem, consider the word $abcdcba$ which can easily be proven M-unambiguous. However, since its projection on $\{b < c\}$, $bccb$ is M-ambiguous, one cannot guarantee the M-unambiguity of its projections on $\{a < b < c\}$ and $\{b < c < d\}$. Indeed, its projection on $\{a < b < c\}$, $abcc$ is M-ambiguous.

## 4. M-unambiguity on extended Parikh matrices

Although Theorem 2.9 says that any extended Parikh matrix is in fact a Parikh matrix according to the original definition, the unambiguity results cannot be carried on this way, because the image of the $u$-Parikh matrix mapping is a strict subset of all Parikh matrices over $\Sigma_{|u|}$. For example, $\Psi_{aba}(a) = \Psi_{\{1<2<3\}}(31)$ is ambiguous; hence, the image of all words containing $a$ through $\Psi_{aba}$ would be ambiguous. However, it is intuitively clear that $\Psi_{aba}$ gives more information than $\Psi_{ab}$ and there is no word $w$ for which $\Psi_{aba}(a) = \Psi_{aba}(w)$. With this intuition in mind, we refine the notions of M-equivalence and M-(un)ambiguity parametric on the given basic word.

**Definition 4.1.** Let $u \in \Sigma^*$ be an ordered alphabet. Two words $w_1, w_2 \in \Sigma^*$ are termed *M-equivalent w.r.t.* $u$, in symbols $w_1 \equiv_{M(u)} w_2$, if $\Psi_u(w_1) = \Psi_u(w_2)$. A word $w \in \Sigma^*$ is termed *M-unambiguous w.r.t.* $u$ if there is no word $w' \neq w$ such that $w \equiv_{M(u)} w'$. Otherwise, $w$ is termed *M-ambiguous w.r.t.*

$u$. If $w \in \Sigma^*$ is M-unambiguous (resp. M-ambiguous) w.r.t. $u$, then also the (extended) Parikh matrix $\Psi_u(w)$ is called unambiguous (resp. ambiguous) w.r.t. $u$.

Given the fact that the extended Parikh matrix mapping has similar properties as the original Parikh matrix mapping, we will try next to prove some of the already presented M-unambiguity results in the more general context of M-unambiguity w.r.t. a word.

First, a corollary of Theorem 2.9 should be mentioned. Assume $u = a_1 \dots a_k \in \Sigma^*$ and let $\Sigma$ only contain the letters occuring in $u$.

**Corollary 4.1.** In the framework of Theorem 2.9, if $\Psi_u(w)$ is unambiguous as a Parikh matrix over the alphabet $\Sigma_k$ then it is also unambiguous w.r.t. $u$ (and $w$ is M-unambiguous w.r.t. $u$).

**Proof:**
Remember that $\varphi : \Sigma^* \to \Sigma_k^*$ is given by $\varphi(a) = \textit{mi}(\textit{trace}_u(a))$ where $\textit{trace}_u(a)$ is the ordered sequence of occurring positions of $a$ in $u$. Let $w'$ be such that $\Psi_u(w') = \Psi_u(w)$. Then, Theorem 2.9 assures us that, since $\Psi_u(w')$ is unambiguous as a Parikh matrix, $\varphi(w') = \varphi(w)$. The conclusion follows by noticing that $\varphi$ is injective. This is indeed true, since $\varphi(a)$ is not $\lambda$ by the choice of $\Sigma$ and also if $a \neq b$ then $\varphi(a)$ does not have letters in common with $\varphi(b)$ (due to the way *trace* is defined). $\square$

It is easy to see (using the same argument as for Proposition 2.1) that if $w$ is M-unambiguous w.r.t. $u$ than any of its factor has the same property.

Let us now associate to each basic word $u \in \Sigma^*$ a graph $G_u = (V, E)$ where $V = \Sigma$ and $E = \{(a, b) \mid ab \text{ factor in } u\}$. For any two letters $a, b \in \Sigma$, let $d_u(a, b)$ denote *the distance between* $a$ *and* $b$ *in* $u$, that is, the length of the minimum path from $a$ to $b$ in $G_u$. Next result is a generalization of Corollary 2.1.

**Proposition 4.1.** If $w$ is M-unambiguous w.r.t. $u$ and $ab$ is a factor of $w$ then $d_u(a, b) \leq 1$, that is, either $ab$ is a factor of $u$ or $a = b$.

It is interesting to notice that the property of $G_u$ having no self-loops exactly characterizes the words with no consecutive repeating letters which we encountered in the results characterizing the inverse of an extended Parikh matrix. Next results generalize Proposition 3.1.

**Proposition 4.2.** Let $u \in \Sigma^*$ such that $G_u$ has no self-loops. Then $w$ is M-unambiguous w.r.t $u$ if and only if $\textit{mi}(w)$ is M-unambiguous w.r.t $u$;

**Proof:**
Using Theorem 2.7 we obtain: $\Psi_u(\textit{mi}(w)) = \Psi_u(\textit{mi}(w'))$ iff $\overline{\Psi_u}(\textit{mi}(w)) = \overline{\Psi_u}(\textit{mi}(w'))$ iff $[\Psi_u(w)]^{-1} = [\Psi_u(w')]^{-1}$ iff $\Psi_u(w) = \Psi_u(w')$ $\square$

Also, as a consequence of Theorem 2.8, for arbitrary $u$ we have.

**Proposition 4.3.** Let $u, w \in \Sigma^*$. Then $\textit{mi}(w)$ is M-unambiguous w.r.t. $u$ if and only if $w$ is M-unambiguous w.r.t $\textit{mi}(u)$.

**Proof:**
$\Psi_u(\textit{mi}(w)) = \Psi_u(\textit{mi}(w'))$ iff $\Psi_{\textit{mi}(u)}(w)^{(rev)} = \Psi_{\textit{mi}(u)}(w')^{(rev)}$ iff $\Psi_{\textit{mi}(u)}(w) = \Psi_{\textit{mi}(u)}(w')$. $\square$

**Corollary 4.2.** Let $u \in \Sigma^*$ such that $G_u$ has no self-loops. The following are equivalent:

- $w$ is M-unambiguous w.r.t $u$;
- $mi(w)$ is M-unambiguous w.r.t $u$;
- $w$ is M-unambiguous w.r.t $mi(u)$;
- $mi(w)$ is M-unambiguous w.r.t $mi(u)$.

Given $x \in \Sigma^*$ we can define the projection $\pi_x$ to be the projection of $\Sigma^*$ to the alphabet containing only the letters of $x$. Using Theorem 2.6 we can prove a result similar to Theorem 3.1 for extended Parikh matrices.

**Theorem 4.1.** Consider $u \in \Sigma^*$. Then for any $1 \leq p \leq q \leq |u|$ and any word $w \in \Sigma^*$ we have that

$$[\Psi_u(w)]_{p,q} = \Psi_{u_{p,q}}(\pi_{u_{p,q}}(w))$$

And, of course the corresponding corollary.

**Corollary 4.3.** If $w \equiv_{M(u)} w'$ then for any factor $x$ of $u$, $\pi_x(w) \equiv_{M(x)} \pi_x(w')$. Also, if $\pi_x(w)$ is M-unambiguous w.r.t. $x$ then $\pi_x(w') = \pi_x(w)$.

Interesting enough, in the case of extended Parikh matrices we obtain another useful corollary which didn't make sense in the original setting.

**Corollary 4.4.** If $u \in \Sigma^*$ contains a factor $u'$ such that $u'$ contains all letters occuring in $u$ and $w$ is M-unambiguous w.r.t. $u'$ then $w$ is also M-unambiguous w.r.t. $u$.

**Proof:**
Directly from Corollary 4.3, since $\pi_{u'}(w) = w$. □

Next theorem generalizes Theorem 3.2

**Theorem 4.2.** Let $u \in \Sigma^*$ be a basic word and let $w \in \Sigma^*$ be a word such that each two letter factor of $w$ is either a factor of $u$ or of the form $aa$ with $a \in \Sigma$. Let $x, y, z \in \Sigma^*$ be such that $u = xyz$, $|x| > 0$ and $x$ and $z$ don't share any letters besides those in $y$. If $\pi_{xy}(w)$ is M-unambiguous w.r.t. $xy$ and $\pi_{yz}(w)$ is M-unambiguous w.r.t. $yz$ then $w$ is M-unambiguous w.r.t. $u$.

To see that the above theorem indeed generalizes Theorem 3.2, it is enough to take $x = a_1 \cdots a_{p-1}$, $y = a_p \cdots a_q$ and $z = a_{q+1} \cdots a_k$. This is a decomposition satisfying the conditions above since the unambiguity conditons map exactly to the ones in Theorem 3.2. The proof follows the same technique as for Theorem 3.2

**Proof:**
We can assume, without any loss of generality, that the letters adjacent to $y$ (that is last letter of $x$ and first letter of $z$) don't occur in $y$. Indeed by expanding $y$ to $y'$ to satisfy the above property, for the new decomposition $x'y'z'$ we get that $xy$ is a factor of $x'y'$ containing all the letters occuring in $x'y'$

and $yz$ is a factor of $y'z'$ containing all letters occuring in $y'z'$. Applying Corollary 4.4 we get that $\pi_{x'y'}(w) = \pi_{xy}(w)$ is M-unambiguous w.r.t. $x'y'$ and $\pi_{y'z'}(w) = \pi_{yz}(w)$ is M-unambiguous w.r.t. $y'z'$.

If $\pi_{xy}(w)$ or $\pi_{yz}(w)$ are $\lambda$ then our proof is done.

Else, suppose by contradiction there exists $w' \neq w$ such that $\Psi_u(w) = \Psi_w(w')$. Then, by Corollary 4.3 we have that $\pi_{xy}(w') = \pi_{xy}(w)$ and $\pi_{yz}(w') = \pi_{yz}(w)$. Suppose now that $w = b_1 \dots b_n$ and $w' = c_1 \dots c_n$ and let $1 \leq i \leq n$ be the smallest integer such that $b_i \neq c_i$. Also, suppose $\pi_{xy}(w) = d_1 \dots d_m (= \pi_{xy}(w'))$ and let $j$ be such that $d_1 \dots d_{j-1} = \pi_{xy}(b_1 \dots b_{i-1}) = \pi_{xy}(c_1 \dots c_{i-1})$.

Case 1: $i > 1$

If $\pi_{yz}(b_i) = \lambda$ then $d_j = b_i$. If $\pi_{xy}(c_i) = c_i$ then also $d_j = c_i$, contradiction. Else, it must be that $b_{i-1} = c_{i-1}$ occurs in $y$ and $c_i$ occurs in $z$ but not in $xy$. By the hypothesis, one starting with $c_i$ should pass through $y$ before getting to a letter in $x$, whence $d_j$ occurs in $y$ implying $b_i$ occurs in $y$, contradiction with $\pi_{yz}(b_i) = \lambda$. Thus, $\pi_{yz}(b_i) = b_i$

By a similar argument, but using $\pi_{xy}$, $\pi_{xy}(b_i) = b_i$.

Now, if $\pi_{xy}(c_i) = c_i$ then at position $j$ we can observe that $\pi_{xy}(b_i) = \pi_{xy}(c_i)$, contradiction. The same way, $\pi_{yz}(c_i) = c_i$ leads to a contradiction using $\pi_{yz}$.

Case 2: $i = 1$

If $\pi_{yz}(b_1) = \lambda$ then the first letter in $w$ whose index is greater than 1 and does not occur in $x$ must occur in $y$, so $\pi_{yz}(w)$ starts with a letter occuring in $y$; thus $w'$ must also start with a letter occuring in $xy$. This means that (using $\pi_{xy}$) $b_1 = d_1 = c_1$, a contradiction. Thus $\pi_{yz}(b_1) = b_1$

The same way, $\pi_{xy}(b_1) = x_1$. Using the above facts, it follows that both $\pi_{xy}(w)$ and $\pi_{yz}(w)$ start with $b_1$ whence both $\pi_{xy}(w')$ and $\pi_{yz}(w')$ must start with $b_1$, leading to $b_1 = c_1$, contradiction. □

## 5. Conclusion. Open problems

The problem of characterizing the M-unambiguity for arbitrary alphabets or basic words still remains open. However the results presented here give general and practical criteria for M-ambiguity, and hopefully are solid steps towards the higher goal.

We would not want to conclude without pointing an interesting problem related to M-unambiguity. Given a word $w = b_1^{p_1} b_2^{p_2} \dots b_n^{p_n}$ such that for all $1 \leq i \leq n$, $p_i > 0$ and for each $1 \leq i < n$, $b_i \neq b_{i+1}$ (it is clear that each word admits a unique such decomposition), we define *the print of* $w$ to be the word $b_1 b_2 \dots b_n$. We have found out that for alphabets of size two (see Theorem 2.3) and three (see Appendix, Theorem A.1) the M-unambiguity of a word implies the M-unambiguity of its print. Several questions naturally arise in this setting:

1. Does the M-unambiguity of a word imply the M-unambiguity of its print for arbitrary alphabets?

2. Is the maximum length of a M-unambiguous print bounded for a given alphabet, and if so, can it be computed?

3. Given a print, can one characterize all M-unambiguous words having the same print?

One can for example see that if the following conjecture holds, first question would be favorable answered.

**Conjecture 5.1.** Let $\Sigma = \{a_1 < \cdots < a_k\}$ be an ordered alphabet Then for any $u, v \in \Sigma^*$ and $a \in \Sigma$,
- if $uaav$ is M-unambiguous then $uav$ is also M-unambiguous, or, equivalently,
- if $uav$ is M-ambiguous, then so is $uaav$.

## Acknowledgments

We are grateful to our professor Alexandru Mateescu for helping us make our first steps into research.

## References

[1] Atanasiu, A., Martín-Vide, C., Mateescu, A.: Codifiable Languages and the Parikh Matrix Mapping., *J. UCS*, **7**(8), 2001, 783–793.

[2] Atanasiu, A., Martín-Vide, C., Mateescu, A.: On the Injectivity of the Parikh Matrix Mapping., *Fundam. Inform.*, **49**(4), 2002, 289–299.

[3] Ding, C., Salomaa, A.: *On some problems of Mateescu concerning subword occurrences*, Technical Report 701, TUCS, August 2005.

[4] Fossé, S., Richomme, G.: Some characterizations of Parikh matrix equivalent binary words., *Inf. Process. Lett.*, **92**(2), 2004, 77–82.

[5] Mateescu, A., Salomaa, A.: Matrix Indicators For Subword Occurrences And Ambiguity., *Int. J. Found. Comput. Sci.*, **15**(2), 2004, 277–292.

[6] Mateescu, A., Salomaa, A., Salomaa, K., Yu, S.: *On an extension of the Parikh mapping*, Technical Report 364, TUCS, 2000.

[7] Mateescu, A., Salomaa, A., Salomaa, K., Yu, S.: A sharpening of the Parikh mapping., *ITA*, **35**(6), 2001, 551–564.

[8] Mateescu, A., Salomaa, A., Yu, S.: Subword histories and Parikh matrices., *J. Comput. Syst. Sci.*, **68**(1), 2004, 1–21.

[9] Parikh, R.: On Context-Free Languages, *J. ACM*, **13**(4), 1966, 570–581.

[10] Rozenberg, G., Salomaa, A.: *Handbook of Formal Languages*, Springer, Berlin, 1997.

[11] Sakarovitch, J., Simon, I.: Subwords, in: *Combinatorics on Words* (M. Lothaire, Ed.), Addison-Wesley, Reading, 1983, 105–144.

[12] Salomaa, A.: *Formal Languages*, Academic Press, New York, 1973.

[13] Salomaa, A.: Connections between subwords and certain matrix mappings, *Theor. Comput. Sci.*, **340**(1), 2005, 188–203.

[14] Salomaa, A.: *On languages defined by numerical parameters*, Technical Report 663, TUCS, 2005.

[15] Salomaa, A.: On the Injectivity of Parikh Matrix Mappings, *Fundamenta Informaticae*, **64**, 2005, 188–203.

[16] Şerbănuţă, T. F.: Extending Parikh matrices., *Theor. Comput. Sci.*, **310**(1-3), 2004, 233–246.

[17] Şerbănuţă, V. N.: *Matrice Parikh injective*, Master Thesis, Faculty of Mathematics, University of Bucharest, December 2002, In Romanian.

# A. M-unambiguity on a three-letter alphabet - case study

The results in this section were proved in [17]. We give the same presentation as there, only changing the notation and the language in order to fit this paper more properly.

First, let us give a criteria for the M-ambiguity of a word over a two-letter alphabet which will be extensively used in the sequel:

**Algorithm A.1.** (*The Moving Algorithm.* ) Let $\Sigma = \{a < b\}$ be the alphabet. If $w = a_1 \cdots a_n$ is a word, and $\Psi_\Sigma(w)$ is its Parikh matrix, then any word having the same Parikh matrix as $w$ can be obtained by applying the following rules a finite number of times.

Let $1 \le i < |w|$ and $2 \le j \le |w|$ be two indices such as $i \neq j$, $i + 1 \neq j - 1$, $a_i = a_j = a$ and $a_{i+1} = a_{j-1} = b$. The *(i,j)-rule* consists of swapping $a_i$ with $a_{i+1}$ and $a_j$ with $a_{j-1}$

**Proof:**

Moving the letters does not change the number of $a$ or $b$. Also, if we move letters using only the above rules the number of $ab$ in the word stays the same. We will prove that the rules above are enough by induction on the length of $w$. If $|w| = 1$ then the above is obviously true. Let's suppose that $|w| > 1$ and we have $u$ with $\Psi_\Sigma(w) = \Psi_\Sigma(u)$.

If $w_1 = u_1$, then let $w' = w_2 w_3 \dots w_{|w|}$ and $u' = u_2 u_3 \dots u_{|u|}$. $\Psi_\Sigma(w') = \Psi_\Sigma(u')$, so $w'$ can be transformed to $u'$ with the above algorithm. But this means that we can get $u$ from $w$ by applying the same changes.

If $w_1 \neq u_1$: let's suppose $w_1 = a$ and $u_1 = b$. Since $\Psi_\Sigma(w) = \Psi_\Sigma(u)$ this means that $w$ contains at least one $b$ and $u$ contains at least an $a$. Then we take the leftmost $a$ of $u$ and another $a$ from the same word that has a $b$ to its right (we will prove that this is possible) and apply the algorithm rule. We do this until we obtain a word starting with $a$.

Let's suppose that $u$ does not have another $a$ with a $b$ letter to its right. Then $u$ must be of the form $b^m a b^n a^*$ with $m$ and $n$ being positive integers, $m > 0$. The number of $ab$ in this word is $n$. But $w$ also has $m + n$ $b$ letters and it has an $a$ to the left, so it must have at least $m + n$ occurrences of $ab$ as a subword. This contradicts the fact that $w$ and $u$ have the same Parikh matrix. $\square$

Fix the alphabet $\{a < b < c\}$. Denote $\Psi_{\{a<b<c\}}$ simply by $\Psi$.

**Fact A.1.** The word $ac$ is M-ambiguous.

**Fact A.2.** The M-unambiguous words with at most two distinct letters are: $\lambda$, $a^+$, $a^+b^+$, $a^+ba^+$, $a^+bab^+$, $b^+$, $b^+a^+$, $b^+ab^+$, $b^+aba^+$ $c^+$, $c^+b^+$, $c^+bc^+$, $c^+bcb^+$, $b^+$, $b^+c^+$, $b^+cb^+$ and $b^+cbc^+$.

In order to find all the three letters M-unambiguous words we will take all the M-unambiguous words in the above fact and add letters according to the restrictions from Propositions 2.1 and 2.2.

We can see that if we generate all the M-unambiguous words beginning with $a$, by using $\varphi^\circ$ we will generate all the M-unambiguous words beginning with $c$.

Every $(i, j)$-rule in Algorithm A.1 change the number of $abc$ in the word with the number of $c$ between $i$ and $j$. If $i < j$ then the number is decreased, else it is increased. The same is true if we change the rules by replacing $a$ with $c$ and count the $a$'s between $i$ and $j$.

Since every configuration with the same matrix except the upper-right corner can be reached with the above rules (and swapping consecutive $a$ and $c$ letters freely), if we want to have the same upper-right corner too we must use the same number of rules that decrease and increase it.

Also, it is obvious that a non-empty $a$ or $c$ group inside an M-unambiguous word can only have a $b$ near it, so the group must have exactly one letter (see Proposition 2.2).

It can also easily be seen that an M-unambiguous word over a three letters alphabet belongs to the language obtained from the concatenation of the following languages:

$$\{\lambda,\ ab,\ b^+,\ cb\}$$

$$\{(ab^+cb^+)^*,\ ab^+,\ cb^+\}$$

$$\{\lambda,\ a^+,\ c^+\}$$

This means that the word has a body of $ab^+cb^+ab^+cb^+\dots$ (it cannot have subwords like $b^+ab^+ab^+$, $b^+cb^+cb^+$, $cbcbc$ or $ababa$), a prefix from the language $\{\lambda, ab, b^+, cb\}$ and a suffix from the language $\{\lambda, a^+, c^+\}$.

If we take the mirror of every M-unambiguous word $w$ beginning with $a$ or $c$ and ending with $b$ we will have all the M-unambiguous words beginning with $b$ and ending with $a$ or $c$. The M-unambiguous words beginning with $b$ and ending with $b$ are generated by adding $b$ to the and of a word beginning with $b$ and ending with $a$ or $c$.

In the following $m, n, p, q, r, s, t, u, v, w, x$ are nonzero positive integers.

We can easily see that the word $a^m b^n c^p$ is M-unambiguous, because no rule of Algorithm A.1 can be used. To get another M-unambiguous word, we can add only $b$ letters. However, if $p$ is greater than 1, the word we get is not M-unambiguous (see Proposition 2.2). If $p = 1$, then the word $a^m b^n cb^q$ is M-unambiguous (we can't apply the algorithm rule). If $q = 1$ then we can try to add $c$ letters: $a^m b^n cbc^r$ is M-unambiguous for the same reason. We cannot add more letters to this word, so let's return to $a^m b^n cb^q$.

If we add $a$ letters, we get $a^m b^n cb^q a^r$. We cannot apply any rule of the Algorithm A.1. Indeed, the only change we can make to this word is to move letters from the $a^m$ group to the right and letters from the $a^n$ group to the left. But these moves decrease the number of $abc$ in the word.

If we have $r = 1$ then we can try to add some $b$ letters. The word $a^m b^n cb^q ab^s$ is M-unambiguous, by the same argument as above.

For $s = 1$ we add $a$ letters and get $a^m b^n cb^q aba^t$, which is M-unambiguous for the same reason.

By adding $c$ letters we get $a^m b^n cb^q ab^s c^t$. One can see that moving a pair of $a$ or $c$ decreases the number of $abc$ in the word, so this one is M-unambiguous, too. If $n > 1$ and we try to add a $b$ to this word, we get a ambiguous Parikh matrix, even if $t = 1$:

$$\Psi(a^m b^n cb^q ab^s cb^u) = \Psi(a^{m-1} bab^{n-2} cb^q ab^{s+2} cb^{u-1}).$$

However, for $n = 1$ and $t = 1$ the word $a^m bcb^q ab^s cb^u$ is M-unambiguous. Indeed, if we move the $c$ letters and get $a^m cb^{q+1} ab^{s+1} cb^{u-1}$, we have decreased the number of $abc$ in our word by 1. We cannot move further the $c$ letters without changing the number of $bc$ in the word, and we cannot increase the number of $abc$ by moving the $a$ letters. Moving the $c$ letters the other way, we decreases the number of $abc$ in the word. Further moves that keep the number of $ab$ in the word $a^m b^2 cb^{q-1} ab^{s-1} cb^{u+1}$ (other than moving the $c$ back) will decrease or leave unchanged the number of $abc$. Moving only the $a$ letters will decrease the number of $abc$.

For $u = 1$ we can add $c$ letters, and we get $a^m bcb^q ab^s cbc^v$ which is M-unambiguous for the same reason. If $u > 1$ and we add a letters, we get a M-ambiguous word: $a^m bcb^q ab^s cb^u a^v$. Indeed,

$$\Psi(a^m bcb^q ab^s cb^u a^v) = \Psi(a^m cb^{q+2} ab^s cb^{u-2} aba^{v-1}).$$

However, if $u = 1$ then $a^m bcb^q ab^s cba^v$ is M-unambiguous by an argument similar to the above one.

If $q > 1$ and we try to add $b$ letters to this word, we get a M-ambiguous one. For $v > 1$, this is obvious. For $v = 1$, we can see that

$$\Psi(a^m bcb^q ab^s cbab) = \Psi(a^m bbcb^{q-2} ab^s cbbba).$$

If $q = 1$ and $v = 1$ and we add $b$ letters, we get an M-unambiguous word: $a^m bcbab^s cbab^w$. Indeed, let's see that if we move the $c$ letters such as we get $a^m cbbab^{s+1} cab^w$ and we have increased the number of $abc$ in the word; we cannot decrease it by moving the $a$ letters. If we move the $c$ letters the other way we get $a^m bbcab^{s-1} cbbab^w$, and we have decreased by $1$ the number of $abc$ in the word. The only chance to increase it is to move the $a$ letters such as we get $a^m babcb^{s-1} cbbbab^{w-1}$, but this increases the number of $abc$ by $2$, and we have no way to decrease it again.

If we add an $a$ letter to this word, we get $a^m bcbab^s cbab^w a$, which is M-ambiguous:

$$\Psi(a^m bcbab^s cbab^w a) = \Psi(a^m cbbbab^s cab^{w-1} ab)$$

If $w > 1$ and we add $c$ letters, we get $a^m bcbab^s cbab^w c$, which is M-ambiguous:

$$\Psi(a^m bcbab^s cbab^w c) = \Psi(a^m bcab^s cbbbab^{w-2} cb).$$

For $w = 1$ the we get an M-unambiguous word: $w = a^m bcbab^s cbabc^x$. To see why, let's try to move the letters: we can move the $c$ letter from the right and the one in the middle such as we get $a^m cbbab^{s+1} cabc^x$. The number of $abc$ has increased by $1$. Further moves for the $c$ letters (other than moving them back) do not change the number of $abc$. Moving the $a$ letters can only increase the number of $abc$. But we can move the $c$ letters from above such as we get $a^m bbcab^{s-1} cbbabc^x$. The number of $abc$ has decreased by $1$. We can decrease it again by moving (once or twice) the $c$ in the middle to the left and one $c$ from the group in the left to the right. Each of this moves decreases the number of $abc$ by $1$ We get the words:

$$a^m bbcab^s cbacbc^{x-1}$$

and

$$a^m bbcab^{s+1} caccbc^{x-2}$$

For the first word, moving the $a$ can increase the number of $abc$ only by $3$, and we cannot decrease it anymore. For the second word, the number of $abc$ increases by $4$ and cannot be decreased. Let's return to

$$a^m bbcab^{s-1} cbbabc^x.$$

If we try to move the $a$ letters, we can only increase the number of $abc$ by $2$, and we can't decrease it again. Moving the $c$ in the right to the left and one $c$ from the left to the right will take us to one of the words discussed above. We can move the $c$ in the middle and one $c$ from the left and we get $a^m bcbab^{s+1} cacbc^{x-1}$ The number of $abc$ has decreased by $1$. Any further move will take us to a word discussed above, will leave the number of $abc$ unchanged or will increase it by $2$, with no hope of decreasing it again. Moving only the $a$ letters in $w$ will give us words with different numbers of $abc$.

If we add $b$ letters to this word, we get M-ambiguous words. Indeed,

$$\Psi(a^m bcbab^s cbabc^x b) = \Psi(a^{m-1} bacab^s cbbabc^{x-1} bc).$$

For the next M-unambiguous words, the injectivity arguments are the same as for the words above.

Let's start with $a^m bab^n$ and add letters. we can add only $c$ letters, and $a^m bab^n c^p$ is M-unambiguous (it is $\varphi^\circ(mi(a^p b^n cbc^m))$). For $p = 1$ we can add $b$ letters and get $a^m bab^n cb^q$, which is M-unambiguous. if $q = 1$ we can add $c$ letters, an get an M-unambiguous word: $a^m bab^n cbc^r$. We can also add $a$ and get $a^m bab^n cb^q a^r$, which is M-unambiguous. For $r = 1$ we add some $b$ letters and get $a^m bab^n cb^q ab^s$, which is M-unambiguous. for $s = 1$ we add $a$ and get $a^m bab^n cb^q aba^t$, which is M-unambiguous. $a^m bab^n cb^q ab^s c^t$ is M-unambiguous only for $s = 1$ ($a^m bab^n cb^q ab^s c^t = \varphi^\circ(mi(a^t b^s cb^q ab^n cbc^m))$). If we add a $b$ we get a M-ambiguous word even if $t = 1$: $a^m bab^n cb^p abc^q b$.

$$\Psi(a^m bab^n cb^p abc^q b) = \Psi(a^{m-1} baab^{n-1} cb^p abbc^{q-1} bc)$$

Let's see the words starting and ending with $b$ and containing all the three letters (the others can be obtained from the words above with $mi$ and $\varphi^\circ$): $b^m ab^n cb^q$, $b^m ab^n cb^p ab^q$, $b^m abcb^n abcb^p$ and $b^m abcbabcbab^n$ $\blacksquare$ are M-unambiguous, all the others are M-ambiguous.

To summarize, we have:

**Theorem A.1.** The only M-unambiguous words over $\{a, b, c\}$ are:
$a^m bab^n c^p$, $a^m bab^n cb^p a^q$, $a^m bab^n cb^p ab^q$, $a^m bab^n cbc^p$,
$a^m bab^n cb^p aba^q$, $a^m bab^n cb^p abc^q$,
$a^m b^n c^p$, $a^m b^n cb^p$, $a^m b^n cb^p a^q$, $a^m b^n cbc^p$,
$a^m b^n cb^p ab^q$, $a^m b^n cb^p aba^q$, $a^m b^n cb^p ab^q c^r$, $a^m bcb^n ab^p cb^q$,
$a^m bcb^n ab^p cba^q$, $a^m bcb^n ab^p cbc^q$, $a^m bcbab^n cbab^p$, $a^m bcbab^n cbabc^p$,
$b^m ab^n c^p$, $b^m ab^n cb^p$, $b^m ab^n cb^p a^q$, $b^m ab^n cb^p ab^q$,
$b^m ab^n cb^p aba^q$, $b^m ab^n cb^p abc^q$, $b^m abcb^n abcb^p$, $b^m abcb^n abcba^p$,
$b^m abcbabcbab^n$,
$b^m cb^n a^p$, $b^m cb^n ab^p$, $b^m cb^n ab^p c^q$, $b^m cb^n ab^p cbc^q$,
$b^m cb^n ab^p cba^q$, $b^m cbab^n cbab^p$, $b^m cbab^n cbaba^p$,
$c^m b^n a^p$, $c^m b^n ab^p$, $c^m b^n aba^p$, $c^m b^n ab^p c^q$,
$c^m b^n ab^p cb^q$, $c^m b^n ab^p cb^q a^r$, $c^m b^n ab^p cbc^q$, $c^m bab^n cb^p ab^q$,
$c^m bab^n cb^p aba^q$, $c^m bab^n cb^p abc^q$, $c^m babcb^n abcb^p$, $c^m babcb^n abcba^p$,
$c^m bcb^n a^p$, $c^m bcb^n aba^p$, $c^m bcb^n ab^p c^q$, $c^m bcb^n ab^p cb^q$,
$c^m bcb^n ab^p cba^q$, $c^m bcb^n ab^p cbc^q$, $a^m$, $a^m b^n$, $a^m ba^n$, $a^m bab^n$, $b^m$, $b^m a^n$, $b^m ab^n$, $b^m aba^n$ $c^m$, $c^m b^n$, $c^m bc^n$, $c^m bcb^n$, $b^m$, $b^m c^n$, $b^m cb^n$, $b^m cbc^n$ with $m, n, p, q > 0$