PARIKH MATRICES

by

NG YIN YIN

Dissertation submitted in partial fulfillment
of the requirements for the degree
of Master of Science in Mathematics

May 2008

ACKNOWLEDGEMENT

I would like to express my deeply gratitude to Professor K. G. Subramanian. He suggested the topic for my dissertation and guided me in my work by clarifying concepts that I found difficult to notice. I thank him for his tireless effort and I will never forget his support and help.

I would also like to offer my heartfelt thanks to the Dean of School of Mathematical Sciences, Prof. Madya Ahmad Izani Md. Ismail in helping me by listing supervisors who possibly can guide me which led to my choosing the current supervisor.

Furthermore, I also wish to express my sincere thanks from the bottom of my heart to all the lecturers who took courses for me and assisted me during the course of my study here. I would like to specially thank the academic and administrative staff of the School of Mathematical Sciences, Universiti Sains Malaysia who always have a concern for the welfare of the students of the school.

PARIKH MATRIKS

ABSTRAK

Tajuk disertasi ini ialah "Parikh Matriks", iaitu suatu pembelajaran tentang ayat-ayat dan kombinatorik cirinya dalam topic "Kombinatorik atas ayat-ayat" yang luas. Dengan menggunakan matematikal perjelasan ayat, ayat dikatakan ialah satu turutan simbol-simbol yang diambil daripada huruf-huruf ajab yang terdiri daripada suatu set yang mengandungi abjab-abjab symbol. Maka ayat-ayat membentuk berbagai objektif utama dalam pembelajaran berbagai jenis bahasa formal yang mana diketahui sebagai satu cabangan daripada computer sains teoritikal yang telah dikembangkan pada masa empat puluh tahun lalu. Konsep satu ayat punya Parikh vektor ialah mengira bilangan dan jenis kewujudan simbol huruf abjab yang ada dalam satu ayat maka merupakan suatu alat pembelajaran bahasa formal yang penting. Baru-baru ini, Mateescu *et.al*(2001) memperkenalkan suatu teknik yang boleh dijelas dengan mudah, iaitu rekaan segi empat

matrik khas untuk simbol atas huruf abjab tersusun supaya operasi pendaraban matrik dengan matrik manghasilkan matrik ayat. Matrik ini dipanggil Parikh Matrik ayat yang mempunyai Parikh vektor ayat di mana terletak di pepenjuru kedua atas pepenjuru utamanya dan kemasukan element-element yang lain memberi maklumat atas bilangan dan jenis sub-ayat yang ada pada ayat itu. Sesetengah ciri-ciri metematikal Parikh matriks telah dijalankan oleh penyelidik baru-baru lepas ini.

Objektif disertasi ini ialah menghuraikan konsep Parikh Matrik ayat-ayat dan menunjukkan beberapa ciri maklumat yang penting dalam matrik-matrik, di mana telah diperkenalkan dalam Mateescu *et.al* (2001). Bab dua, memperkenalkan fahaman asas dan perkembangan ciri-ciri yang berdasarkan dalam pembelajaran di Mateescu *et.al* (2001). Bab tiga, memperkenal tentang masalah injektif Parikh matriks dan perbincangan keputusan-keputusan berkaitan dalam pembelajaran Atanasiu *et.al* (2002) dan Fosse dan Richomme (2004). Bab empat memperkenalkan jenis generasi Parikh vector yang berlainan dalam Egecioglu (2004) yang berdasarkan fikiran "q-kiraan". Generasi ini boleh diringkaskan kepada Parikh matrik-matrik ketika $q = 1$. Bab akhir memperkenalkan hubungan antara Parikh matriks dengan "ayat-ayat membulat" (Rozenberg dan Salomaa, 1997) , di mana telah mempertimbangkan dalam kontek pembelajaran sifat perbaikian rangkaian sesetengah DNA molekul.

ABSTRACT

This dissertation titled "Parikh Matrices" is a study on words and their combinatorial properties and falls under the broad topic of "Combinatorics on Words". A word, mathematically speaking, is a sequence of symbols taken from an alphabet which is a set of symbols. Words constitute the central objects in the study of Formal languages which has evolved in the past four decades as a branch of theoretical computer science. The concept of a Parikh vector of a word, which counts the number of occurrences of symbols of the alphabet in the word, is an important tool in the study of formal languages. Recently Mateescu *et.al* (2001) introduced an apparently simple but ingenius technique of associating a specific kind of square matrix with every symbol of an ordered alphabet and associated a matrix with every word, using the operation of product of matrices. This matrix is called the Parikh matrix of the word which has the Parikh vector of the word in the second diagonal above the main diagonal and other entries give information on the number of subwords of the given word. Several investigations on the mathematical properties of Parikh matrices have been made by researchers in the recent past.

The *objective* of this dissertation is to describe the concept of the Parikh matrix of a word and deal with some important properties of these matrices that have been established in Mateescu *et.al* (2001). Chapter 2 introduces the basic notion and

develops properties based on the study done in Mateescu *et.al* (2001). Chapter 3 deals with the problem of injectivity of the Parikh matrices and discusses results pertaining to this notion based on the study in (Atanasiu et al (2002), Fosse and Richomme (2004)). Chapter 4 deals with another kind of generalization of the Parikh vector done in Egecioglu (2004) using a notion of "q-counting". This generalization reduces to the Parikh matrix when $q = 1$. In the last chapter a beginning is made to associate Parikh matrices with "circular words" (Rozenberg and Salomaa, 1997), which have been considered in the context of the study of certain recombinant behaviour of DNA molecules.

CHAPTER 1

INTRODUCTION

1.1 Words

The topic of words and their combinatorial properties is a relatively new but rapidly growing area of discrete mathematics. The basic object in such a study is a word which is a sequence of symbols taken from a finite set, usually called an alphabet. For example if the alphabet is $\Sigma = \{0,1\}$, then 100110101001 is a word over Σ. The mathematical structure involved in the study of words is that of a free monoid and noncommutativity is a fundamental feature of words.

The theory of words is related to many areas of sciences but computer science is perhaps mainly responsible for the interest and research on words. Many aspects of computers and computing stand as real motivation to study words. In fact digital computers operate on sequences of bits which are really words.

Axel Thue's classical work in 1906 on repetition-free words (Lothaire, 1997) is generally considered as a starting point of mathematical research on words. But a systematic study and research on words perhaps started in 1950s. The first book titled

"Combinatorics on Words" (Lothaire, 1997) was published in 1983. Since then a number of problems related to words have been studied by many researchers.

1.2 Languages

Formal language theory (Rozenberg & Salomaa, 1997) was fist developed in the mid 1950's in an attempt to develop theories of natural language acquisition. It was then found that this theory was quite relevant to the artificial languages that had originated in computer science. Since then, the theory of formal languages has been developed extensively. The notion of a formal language was conceived basically for capturing the common features of a natural language such as English or an artificial language such as a programming language. Every language has an alphabet. Based on several rules of formation, sentences or statements are formed that constitute a language. The sentences or statements are also, like words, sequences of symbols. A formal language or simply a language, mathematically speaking, is thus a finite or an infinite set of words over an alphabet. Thus words are also the central objects in the study of formal language theory.

1.3 Parikh Matrix Mapping

Parikh mappings, also called Parikh vectors, introduced in (Parikh, 1966) express properties of words as numerical properties of vectors yielding some fundamental language-theoretic consequences. In fact a Parikh Vector counts in a word w the number of occurrences of the letters of the alphabet. For example if the alphabet is $\Sigma = \{a,b,c\}$ with the ordering $a < b < c$ then the Parikh vector of *acbbcacca* is (3,2,4).

Parikh Matrix Mapping, also called Parikh Matrx, is a newly developed tool for studying numerical properties of words in terms of their subwords or also called scattered-subwords. It was introduced by Mateescu *et.al* (2001) and since then has continuously received attention from the research community working in on words and their application.

The objective of this dissertation is to introduce the concept of a Parikh matrix of a word, discuss various properties of Parikh matrices relying on the research on these matrices carried out in (Atanasiu *et.al*, (2002), Egecioglu, (2004), Fosse and Richomme, (2004), Mateescu *et.al*, (2001)) and finally point out the possibility of associating Parikh matrices with another kind of words, called "circular words" (Rozenberg & Salomaa, 1997) studied in formal language theory.

The dissertation is divided into five chapters with the first chapter constituting the Introduction which includes a section on certain preliminary notions and terms needed

for the remaining chapters. In Chapter 2, the main concept of a Parikh matrix is introduced and several basic properties are discussed. The material presented in this chapter is based on Mateescu *et.al* (2001). In chapter 3 an important notion of "injectivity" of a Parikh matrix is dealt with. Conditions relating to injectivity or non-injectivity of a Parikh matrix mapping that have been established in Atanasiu *et.al* (2002) and Fosse and Richomme (2004) are discussed . Chapter 4 deals with another kind of generalization of the Parikh vector done in Egecioglu (2004) using a notion of "q-counting". Chapter 5 introduces the concept of circular words (Rozenberg & Salomaa, 1997). Some properties of Parikh matrices that could be associated with circular words are then obtained.

1.4 Preliminaries

We denote the set of non-negative integers by N, i.e. $N = \{ 0, 1, 2, 3, \dots \}$

An *alphabet* Σ is a finite set of symbols. For example, if $\Sigma = \{0, 1\}$, $A = \{a, b\}$, then each of Σ and A is an alphabet with two symbols. Σ and A are often called *binary alphabet*.

A *word w* over an alphabet Σ is a finite sequence of symbols from Σ. The length, denoted by $|w|$, of a word w is the number of symbols in w, counting repetitions also. For example, if $\Sigma_1 = \{a,b\}$, then $u = aabbaabab$ is a word over Σ_1 and its length is $|u| = 9$. If $\Sigma_2 = \{a,b,c\}$, then $v = bbaccbaabc$, $w = abbaab$ are words over Σ_2 . The

lengths of v and w are $|v|=10$ and $|w|=6$. An *empty word*, denoted λ, is a word with no symbols and $|\lambda|=0$. The *mirror image or reversal* denoted by *mi(w)* or w^R of a word $w=a_1a_2\ldots a_n$ is $w^R=a_na_{n-1}\ldots a_2a_1$. For example if $u=aabbaabab$ then $u^R=$ *babaabbaa.*

If *a* is a symbol of an alphabet Σ, then $|w|_a$ denotes the number of occurrences of the symbol *a* in a word *w* over Σ. For example, if $\Sigma=\{a,b\}$ and $w=abaabbabbab$, then $|w|=11$, $|w|_a=5$, $|w|_b=6$. Note that $|w|=|w|_a+|w|_b$. If u,v are two words over an alphabet Σ, where $u=a_1a_2a_3\ldots a_n$, $a_i\in\Sigma$ for $i=1,2,\ldots,n$, and $v=b_1b_2\ldots b_m$, $1\le m\le n$, such that each b_i $(1\le i\le m)$ is some $a_j\in\Sigma$ and $b_1,b_2,\ldots,b_m$ is a subsequence of $a_1,a_2,\ldots,a_n$, then v is called a scattered-subword or simply a *subword* of u.

Example 1.4.1

Let $\Sigma=\{a,b\}$. $w=abbaabaaabbbab$ is a word over Σ, then $v=babbab$ is a subword . The positions of the symbols of v in w can be shown as $a\underline{b}b\underline{aa}baaa\underline{b}bb\underline{ab}$, $a\underline{b}b\underline{aa}baaab\underline{b}b\underline{ab}$, $a\underline{b}b\underline{aa}baaabb\underline{bab}$, $a\underline{b}b\underline{aa}baaabb\underline{bab}$, $a\underline{b}b\underline{aa}baaab\underline{b}b\underline{ab}$, and so on, where the underlined positions of w are the positions of the successive symbols of v.

CHAPTER 2

PARIKH MATRIX MAPPING

2.1 Introduction

An extension of the Parikh mapping or Parikh vector was introduced in Mateescu *et.al* (2001) based on certain type of square matrices. The extension is called Parikh matrix mapping or simply Parikh matrix and is intended to provide more information about a word than a Parikh vector which counts the number of occurrences in the given word of the symbols or letters of the alphabet. Various properties of the Parikh matrix have been established in Mateescu *et.al* (2001). In this chapter the concept of a Parikh matrix of a word is considered and the associated properties are discussed giving a number of examples and elaborating on the proof details of the results following Mateescu *et.al* (2001).

2.2 Parikh Matrix

The concept of a Parikh matrix mapping introduced in Mateescu *et.al* (2001) is first recalled in this section. For doing this, the notion of a triangle matrix is considered now.

Definition 2.2.1

A *triangle matrix* is a $k \times k$ $(k \geq 1)$ square matrix $m = (m_{ij})$ such that all the entries in the main diagonal are 1 i.e. $m_{i,i} = 1$ for $i = 1, 2, ..., k$; all the entries below the main diagonal are zero i.e. $m_{i,j} = 0$, for $1 \leq j < i \leq k$ and all the entries above the main diagonal are non-negative integers. The set of all such $k \times k$ matrices is denoted by $\mathcal{M}_\kappa$. In fact matrices in $\mathcal{M}_\kappa$ are square matrices of dimension k.

Example 2.2.1

The matrix $\quad M = \begin{pmatrix} 1 & 5 & 0 & 3 \\ 0 & 1 & 6 & 12 \\ 0 & 0 & 1 & 9 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ is a 4x4 triangle matrix.

Remark

A triangle matrix is a special type of what are called upper triangular matrices in the literature.

Theorem 2.2.1

The set $\mathcal{M}_\kappa$ is a monoid with respect to multiplication of matrices and the unit matrix of dimension k is the identity element.

Proof: Let M_1, M_2 be two matrices in $\mathcal{M}_\kappa$. M_1, M_2 are triangular matrices. The product $M_1 M_2$ of M_1 and M_2 is indeed a triangular matrix. For instance, the i th row

7

of M_1 is of the form $0,0,...,0,1,x_1,x_2,...,x_{k-i}$ where $x_1,x_2,...,x_{k-i}$ are non-negative integers and the entry 1 is the ith entry of the row. There are $(i-1)$ 0s before this entry 1. The ith column of M_2 is of the form $y_1,y_2,...,y_{i-1},1,0,...,0$ where $y_1,y_2,...,y_{i-1}$ are non-negative integers. Again the entry 1 is the ith entry of the column and there are $(k-i)$, 0s after this entry 1. Hence the ith row, ith column entry of M_1M_2 is $0 \cdot y_1 + 0 \cdot y_2 + ... + 0 \cdot y_{i-1} + 1 \cdot 1 + x_1 \cdot 0 + ... + x_{k-i} \cdot 0 = 1$. Thus the main diagonal entries of M_1M_2 are all 1. Similarly, we can show that all the entries below the main diagonal of M_1M_2 are zero and above the main diagonal are non-negative integers.

It is known that multiplication of matrices is associative.

The unit matrix of dimension k is $I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \end{pmatrix}$ and for any matrix M

in $\mathcal{M}_\kappa$, we have $MI = IM = M$. Thus I is the identity element. Hence $\mathcal{M}_\kappa$ is a monoid.

Definition 2.2.2

Let $\Sigma = \{a_1,a_2,a_3,...,a_k\}$ be a set of symbols. Σ is called an alphabet. An alphabet ordered alphabet is an alphabet Σ with an order relation $<$ on Σ. If $a_1 < a_2 < a_3 < ... < a_k$, we write Σ as $\Sigma = \{a_1 < a_2 < a_3 < ... < a_k\}$.

For example $\Sigma_1 = \{a,b\}$ is an alphabet, and if the order relation $<$ on Σ_1 is given by

$a < b$, we write $\Sigma_1 = \{a < b\}$. Likewise $\Sigma_2 = \{a < b < c\}$ is another ordered alphabet.

Definition 2.2.3 (*Parikh Matrix mapping*)

Let $\Sigma = \{a_1 < a_2 < a_3 < ... < a_k\}$ be an ordered alphabet with $k \geq 1$. The Parikh matrix

mapping Ψ_{M_k} is the mapping $\Psi_{M_k} : \Sigma^* \rightarrow \mathcal{M}_{k+1}$ defined as follows:

Let $\Psi_{M_k}(a_q) = (m_{ij})_{(k+1) \times (k+1)}$, $a_q \in \Sigma$,

where   i) $m_{ii} = 1$, for $i=1,2,...,k+1$

   ii) $m_{q,q+1} = 1$

   iii) all other entries of the matrix $\Psi_{M_k}(a_q)$ are zero.

In other words, the main diagonal entries of the matrix $\Psi_{M_k}(a_q)$ are all 1; the entry in

the qth row, $(q+1)$st column is 1 and all other entries are zero.

For a word $w = a_{q_1} a_{q_2} ... a_{q_m}$ where each $a_{q_r} \in \Sigma$, the Parikh matrix mapping Ψ_{M_k} is

given by $\Psi_{M_k}(w) = \Psi_{M_k}(a_{q_1}) \Psi_{M_k}(a_{q_2}) ... \Psi_{M_k}(a_{q_m})$ where the operation on the right

side of this equation is matrix multiplication. The matrix $\Psi_{M_k}(w)$ is called the *Parikh*

matrix of w.

Example 2.2.2

Let Σ be the ordered alphabet $\{a < b < c\}$ and assume that $w = babca$. Note that

$\Psi_{M_3}(w)$ is a 4×4 triangle matrix that can be computed as follows:

$$\Psi_{M_3}(babca) = \Psi_{M_3}(b)\Psi_{M_3}(a)\Psi_{M_3}(b)\Psi_{M_3}(c)\Psi_{M_3}(a)$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 2 & 1 & 1 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

2.3 Properties of Parikh Matrices

A basic property of the Parikh matrix mapping is first dealt with.

Theorem 2.3.1

Let $\Sigma = \{a_1 < a_2 < a_3 < ... < a_k\}$ be an ordered alphabet, where $k \geq 1$. Let $w \in \Sigma^*$. The

matrix $\Psi_{M_k}(w) = (m_{ij})_{(k+1) \times (k+1)}$ has the following properties:

      i) $m_{i,j} = 0$, for all $1 \leq j < i \leq k+1$,

      ii) $m_{i,i} = 1$, for all $1 \leq i \leq k+1$

      iii) $m_{i,j+1} = |w|_{a_{i,j}}$, for all $1 \leq i \leq j \leq k$

where $a_{i,j}$ denotes the word $a_i a_{i+1} \ldots a_j$ for $1 \leq i \leq j \leq k$ and $|w|_{a_{i,j}}$ is the number of occurrences of the subword $a_{i,j}$ in w, counting partially overlapping occurrences as distinct occurrences.

Proof: By definition, the Parikh matrix $\Psi_{M_k}(a_q)$, for $a_q \in \Sigma$, has properties (*i*) and (*ii*). Also, $\Psi_{M_k}(w)$ for a word w over Σ, is computed by matrix multiplication.

Hence $w = w' a_i$, where $|w'| = n$ and $a_i \in \Sigma$ with $1 \leq i \leq k$.

It follows that:

$$\Psi_{M_k}(w) = \Psi_{M_k}(w' a_i) = \Psi_{M_k}(w') \Psi_{M_k}(a_i)$$

Assume that:

$$\Psi_{M_k}(w') = \begin{pmatrix} 1 & m'_{1,2} & \cdots & \cdots & m'_{1,k+1} \\ 0 & \ddots & \ddots & \vdots & m'_{2,k+1} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & m'_{k,k+1} \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix}$$

Note that by the inductive hypothesis the matrix $\Psi_{M_k}(w')$ has the property (*iii*).

From Definition, we deduce that

$$\Psi_{M_k}(a_i) = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & 0 & \cdots & \cdots & \vdots \\ \vdots & \ddots & 1 & 1 & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 & \vdots \\ \vdots & \vdots & \vdots & \ddots & 1 & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & 1 \end{pmatrix}$$

Note that all elements in the matrix $\Psi_{M_k}(a_i)$ are zero except that the elements on the main diagonal are 1 and also the element on the position $(i, i+1)$ is 1.

Therefore,

$$\Psi_{M_k}(w')\Psi_{M_k}(a_i) = \begin{pmatrix} 1 & m'_{1,2} & \cdots & \cdots & m'_{1,k+1} \\ 0 & \ddots & \ddots & & m'_{2,k+1} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & m'_{k,k+1} \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & 0 & \cdots & \vdots \\ \vdots & \ddots & 1 & 1 & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix} = M$$

Note that the resulting matrix, $M = (m_{p,q})_{1 \le p,q \le k+1}$ has the property that:

$m_{j,i+1} = m'_{j,i} + m'_{j,i+1}$, for all $1 \le j \le i$ and for all other indices, $m_{p,q} = m'_{p,q}$.

This proves Theorem 2.3.1

Theorem 2.3.2

The Parikh mapping is *not injective.*

Proof: Consider the ordered alphabet $\Sigma = \{a < b\}$ and the words $w_1 = aabab$, $w_2 = bababbaaaa$. Then the Parikh matrices of $w_1 w_2 = aababbababbaaaa$ and $w_2 w_1 = bababbaaaaaabab$ are the same, namely

$$\Psi_{M_3}(w_1 w_2) = \Psi_{M_3}(w_2 w_1) = \begin{pmatrix} 1 & 9 & 22 \\ 0 & 1 & 6 \\ 0 & 0 & 1 \end{pmatrix}.$$

This shows that Parikh mapping is not injective.

Theorem 2.3.3

The set of all triangle matrices of order $k \geq 2$ with integer entries is a noncommutative group with respect to the operation of multiplication of matrices and with unit element, the unit matrix of order k.

Corollary:

For each Parikh matrix A, the inverse matrix A^{-1} exists.

Definition 2.3.1

Let $\Sigma = \{a_1 < a_2 < ... < a_k\}$ be an ordered alphabet. Let w be a word and let the Parikh matrix of w be $\Psi_{M_k}(w) = (m_{ij})_{(k+1)\times(k+1)}$. The *alternative Parikh matrix of w is* $\overline{\Psi}_{M_k}(w)$ is the matrix $(m'_{ij})_{(k+1)\times(k+1)}$, where $m'_{ij} = (-1)^{i+j} m_{i,j}$, for all $1 \leq i, j \leq k+1$.

Example 2.3.1

Let $\Sigma = \{a < b\}$ and $w = abaab$.

$$\text{Then } \Psi_{M_3}(w) = \begin{pmatrix} 1 & 3 & 4 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} \text{ and } \overline{\Psi}_{M_3}(w) = \begin{pmatrix} 1 & -3 & 4 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix}$$

We now state a relation between the inverse of the Parikh matrix of a word w and the alternate Parikh matrix of the mirror image or reversal of w.

Theorem 2.3.4

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

$$[\Psi_{M_k}(w)]^{-1} = \overline{\Psi}_{M_k}(mi(w)).$$

Proof. The proof is by induction on $n = |w|$.

If $n = 1$, then $w = a_q$ for some $1 \le q \le k$. Note that by Definition, $\Psi_{M_k}(a_q) = (m_{i,j})_{1 \le i, j \le (k+1)}$, $m_{i,i} = 1$, $m_{q,q+1} = 1$ and all other elements of the matrix $\Psi_{M_k}(a_q)$ are zero.

It is easy to verify that $[\Psi_{M_k}(a_q)]^{-1} = (m'_{i,j})_{1 \le i, j \le (k+1)}$, such that for each $1 \le i \le (k+1)$, $m'_{i,i} = 1$, $m'_{q,q+1} = -1$ and all other elements of the matrix $[\Psi_{M_k}(a_q)]^{-1}$ are zero.

Hence, $[\Psi_{M_k}(a_q)]^{-1} = \overline{\Psi}_{M_k}(mi(w))$.

For the inductive step assume that the Theorem is true for all words $u \in \Sigma^*$ be a word with $|w| = n+1$.

Clearly, $w = x a_p$ such that $|x| = n$ and $a_p \in \Sigma$ for some $1 \le p \le k$.

Now notice that:

$$[\Psi_{M_k}(w)]^{-1} = [\Psi_{M_k}(x a_p)]^{-1} = [\Psi_{M_k}(x)\Psi_{M_k}(a_p)]^{-1} = [\Psi_{M_k}(a_p)]^{-1}[\Psi_{M_k}(x)]^{-1}.$$

Assume that:

$$\overline{\Psi}_{M_k}(mi(x)) = \begin{pmatrix} 1 & m_{1,2} & \cdots & \cdots & m_{1,k+1} \\ 0 & 1 & \ddots & & m_{2,k+1} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & m_{k,k+1} \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

By the inductive hypothesis $[\Psi_{M_k}(x)]^{-1} = \overline{\Psi}_{M_k}(mi(x)) = (m_{i,j})_{1 \leq i,j \leq k+1}$, where

$m_{i,i} = 1$, $1 \leq i \leq (k+1)$, $m_{i,j} = 0$, $1 \leq j < i \leq k+1$ and $m_{i,j+1} = (-1)^{i+j+1} |mi(x)|_{a_{i,j}}$, for all

$1 \leq i \leq j \leq k$.

From Definition, we deduce that

$$[\Psi_{M_k}(a_p)]^{-1} = \begin{pmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & 0 & \ddots & & \vdots \\ \vdots & \ddots & 1 & -1 & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & 0 & 0 \\ \vdots & & & & \ddots & 1 & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & 1 \end{pmatrix}$$

Note that all elements in the matrix $[\Psi_{M_k}(a_p)]^{-1}$ are zero except that the

elements on the main diagonal are 1 and also the element on the position $(p, p+1)$ is -1.

Therefore,

$$[\Psi_{M_k}(w)]^{-1} = \begin{pmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & 0 & \ddots & & \vdots \\ \vdots & \ddots & 1 & -1 & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & 0 & 0 \\ \vdots & & & & \ddots & 1 & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & m_{1,2} & \cdots & \cdots & \cdots & m_{1,k+1} \\ 0 & \ddots & \ddots & & & m_{2,k+1} \\ \vdots & \ddots & 1 & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & & \ddots & 1 & m_{k,k+1} \\ 0 & \cdots & \cdots & \cdots & 0 & 1 \end{pmatrix} = M'$$

Note that the resulting matrix, $M' = (m'_{i,j})_{1 \leq i,j \leq k+1}$ has the property that:

$$m'_{p,j} = m_{p,j} - m_{p+1,j} = (-1)^{p+j} |mi(x)|_{a_{p,j-1}} -(-1)^{p+j+1} |mi(x)|_{a_{p+1,j-1}} =$$

$$= (-1)^{p+j} \, | \, a_p mi(x) |_{a_{p,j-1}} = (-1)^{p+j} \, | \, mi(w) |_{a_{p,j-1}} \, , \quad \text{for all} \quad p < j \le k+1 \text{ and for all}$$

other indices, $m'_{i,q} = m_{i,q}$.

This proves the Theorem.

We now illustrate Theorem 2.3.4 in the following example.

Example 2.3.2

Let Σ be the ordered alphabet $\{a<b<c\}$ and assume that $w=babca$. Note that:

$$\Psi_{M_3}(babca) = \begin{pmatrix} 1 & 2 & 1 & 1 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Note that $mi(babca) = acbab$ and by Theorem the inverse of the above matrix

is:

$$[\Psi_{M_3}(babca)]^{-1} = \overline{\Psi}_{M_3}(acbab) = \begin{pmatrix} 1 & -2 & 3 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

There is another method of computing the inverse of a Parikh matrix. We introduce some definitions and notations before we describe this method.

Let $(A,<)$ be an ordered set. The *dual* order of the order $<$, denoted $<^\circ$, is defined as follows:

$$a <^\circ b \text{ if and only if } b < a$$

Let $\Sigma = \{a_1 < a_2 < ... < a_k\}$ be an ordered alphabet. The *dual* ordered alphabet, denoted

$\Sigma_\circ$, is: $\Sigma_\circ = \{a_k < a_{k-1} < ... < a_1\}$.

Consider the ordered alphabet $\Sigma = \{a_1 < a_2 < ... < a_k\}$ Let $w \in \Sigma^*$ be a word.

The Parikh matrix associated to w with respect to the *dual* order on Σ is denoted by

$\Psi_{M_{k,\circ}}(w)$.

Let $v = (v_1, v_2, ... v_n)$ be a vector. The *reverse* of v, denoted $v^{(rev)}$, is the vector:

$v^{(rev)} = (v_n, v_{n-1}, ... v_1)$.

Now we introduce the notion of a reverse of a triangle matrix.

Let $M = (m_{i,j})_{1 \le i, j \le n}$ be a triangle matrix.

The *reverse* of M, denoted $M^{(rev)}$, is the matrix $M^{(rev)} = (m'_{i,j})_{1 \le i, j \le n}$, where

$m'_{i,j} = m_{n+1-j, n+1-i}$, for all $1 \le i < j \le n$.

Note that $M^{(rev)}$ is also a triangle matrix.

One easy way to obtain $M^{(rev)}$ is to reverse in M all diagonals that are parallel to the main diagonal.

Example 2.3.3

$$\text{If } M = \begin{pmatrix} 1 & 5 & 4 & 6 \\ 0 & 1 & 9 & 8 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix} \text{ then } M^{(rev)} = \begin{pmatrix} 1 & 2 & 8 & 6 \\ 0 & 1 & 9 & 4 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

A further method to compute $M^{(rev)}$ is to consider the transpose of $M = (m_{i,j})_{1 \le i, j \le n}$, i.e. $M' = (m'_{i,j})_{1 \le i, j \le n}$, where $m'_{i,j} = m_{j,i}$, for all $1 \le i, j \le n$.

The reverse $M^{(rev)}$ of M has the following properties.

Theorem 2.3.5

Let A, B be two triangle matrices of the same dimension.

Then:

(i) $[A^{(rev)}]^{(rev)} = A$

(ii) $(AB)^{(rev)} = B^{(rev)} A^{(rev)}$

We illustrate statements (i) and (ii).

i):

$$\text{Let} \quad A = \begin{pmatrix} 1 & 5 & 4 & 6 \\ 0 & 1 & 9 & 8 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad A^{(rev)} = \begin{pmatrix} 1 & 2 & 8 & 6 \\ 0 & 1 & 9 & 4 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\text{Then,} \quad [A^{(rev)}]^{(rev)} = \begin{pmatrix} 1 & 5 & 4 & 6 \\ 0 & 1 & 9 & 8 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

ii):

Now, let $A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 5 & 6 \\ 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 3 & 2 & 3 \\ 0 & 1 & 4 & 9 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 1 \end{pmatrix}$,

and so $AB = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 5 & 6 \\ 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 3 & 2 & 3 \\ 0 & 1 & 4 & 9 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 5 & 13 & 43 \\ 0 & 1 & 9 & 45 \\ 0 & 0 & 1 & 13 \\ 0 & 0 & 0 & 1 \end{pmatrix}$,

Then $AB^{(rev)} = \begin{pmatrix} 1 & 13 & 45 & 43 \\ 0 & 1 & 9 & 13 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

$A^{(rev)} = \begin{pmatrix} 1 & 7 & 6 & 4 \\ 0 & 1 & 5 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ and $B^{(rev)} = \begin{pmatrix} 1 & 6 & 2 & 3 \\ 0 & 1 & 4 & 9 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix}$.

so that,

$B^{(rev)}A^{(rev)} = \begin{pmatrix} 1 & 6 & 9 & 3 \\ 0 & 1 & 4 & 2 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 7 & 6 & 4 \\ 0 & 1 & 5 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

$= \begin{pmatrix} 1 & 13 & 45 & 43 \\ 0 & 1 & 9 & 13 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

Now we deal with another property of the inverse of a Parikh matrix.

Theorem 2.3.6

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

$$[\Psi_{M_k}(w)]^{-1} = [\overline{\Psi}_{M_{k,\circ}}(w)]^{(rev)}.$$

Proof: The proof is by induction on $n = |w|$. If $n = 1$, then $w = a_q$ for some $1 \le q \le k$.

It follows that $\Psi_{M_k}(a_q)$ is the triangle matrix having only 1 on the main diagonal, the

value on the position $(q, q+1)$ is also 1 and all other elements are 0.

Note that the dual order the letter a_q appears on the position $k+1-q$. Hence,

the matrix $\Psi_{M_{k,\circ}}(a_q)$ is the triangle matrix having only 1 on the main diagonal, the

value on the position $(k+1-q, k+2-q)$ is also 1 and all other elements are 0.

The alternate Parikh matrix $\overline{\Psi}_{M_{k,\circ}}(a_q)$ is the triangle matrix having only 1 on

the main diagonal, the value on the position $(k+1-q, k+2-q)$ is -1 and all other

elements are 0. Thus $\overline{\Psi}_{M_{k,\circ}}(a_q)$ is the triangle matrix having only 1 on the main

diagonal, the value on the position $(k+2-(k+2-q), k+2-(k+1-q)) = (q, q+1)$ is

-1 and all other elements are 0.

It can be verified that this is exactly the inverse matrix of the matrix $\Psi_{M_k}(a_q)$.

For the inductive step assume that $w = w'a_i$, where $|w'| = n$ and $a_i \in \Sigma$.

Now using the inductive hypothesis and Theorem we obtain the following:

$$[\Psi(w)]^{-1} = [\Psi(w'a_i)]^{-1} = [\Psi(a_i)]^{-1}[\Psi(w')]^{-1} = [\overline{\Psi}_{M_{k,\circ}}(a_i)]^{(rev)}[\overline{\Psi}_{M_{k,\circ}}(w')]^{(rev)} =$$

$$= [\overline{\Psi}_{M_{k,\circ}}(w')\overline{\Psi}_{M_{k,\circ}}(a_i)]^{(rev)} = [\overline{\Psi}_{M_{k,\circ}}(w'a_i)]^{(rev)} = [\overline{\Psi}_{M_{k,\circ}}(w)]^{(rev)}.$$

This concludes the proof.

Example 2.3.4

Let Σ be the ordered alphabet $\{a < b < c\}$ and assume that $w = babca$.

Note that:

$$\Psi_{M_3}(babca) = \begin{pmatrix} 1 & 2 & 1 & 1 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Note that the dual order alphabet is $\Sigma_\circ = \{c < b < a\}$. Thus:

$$\Psi_{M_{3,\circ}}(babca) = \Psi_{M_{3,\circ}}(b)\Psi_{M_{3,\circ}}(a)\Psi_{M_{3,\circ}}(b)\Psi_{M_{3,\circ}}(c)\Psi_{M_{3,\circ}}(a)$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The alternate Parikh matrix of the matrix $\Psi_{M_{3,\circ}}(babca)$ is:

$$\overline{\Psi}_{M_{3,\circ}}(babca) = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -2 & 3 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Finally, the reverse matrix of the above matrix is:

$$[\overline{\Psi}_{M_{3,\circ}}(babca)]^{(rev)} = \begin{pmatrix} 1 & -2 & 3 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

This last matrix is indeed the inverse of the matrix $\Psi_{M_3}(babca)$.

We recall that for the order $a<b$, the *dual* order is $b<a$. Also, the Parikh matrix mapping over $\Sigma = \{a < b\}$ is Ψ_{M_3} which we will simply write as Ψ_M and then we will denote the Parikh matrix mapping over the *dual* order $b<a$ by $\Psi_{M,\circ}$.

Theorem 2.3.7

Let $\Sigma = \{a < b\}$ and $w \in \Sigma^*$.

Then (*i*) $[\Psi_M(w)]^{(rev)} = \Psi_M(w)$ if and only if $|w|_a = |w|_b$.

(*ii*) $[\Psi_{M,\circ}(w)]^{(rev)} = \Psi_M(mi(w))$.

This follows from the definition of the operation *rev* and the *dual* order.

Example 2.3.5

Let $w=aabba$. Then, here $|w|_a = 3$ and $|w|_b = 2$; $|w|_a \neq |w|_b$.

$$\Psi_M(w) = \begin{pmatrix} 1 & 3 & 4 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} \neq [\Psi_M(w)]^{(rev)} = \begin{pmatrix} 1 & 2 & 4 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix}$$

Also, $mi(w)=abbaa$. Then, here $|mi(w)|_{ab} = 2$ and note that this is with respect to $a<b$.

So,
$$\Psi_M(mi(w)) = \begin{pmatrix} 1 & 3 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}.$$

However, for $b<a$,
$$\Psi_{M,\circ}(w) = \Psi_{M,\circ}(aabba) = \begin{pmatrix} 1 & 2 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix}.$$

so that
$$[\Psi_{M,\circ}(w)]^{(rev)} = \Psi_M(mi(w)).$$

CHAPTER 3

INJECTIVITY OF PARIKH MATRIX MAPPING

3.1 Introduction

The Parikh mapping or Parikh vector (Parikh, 1966) is not an injective mapping as many words may have the same Parikh vector. For instance if the alphabet is V = {a,b}, then all the ten words *abaab, ababa, abbaa, aabab, aabba, aaabb, baaab, baaba, babaa, bbaaa* of length five have the same Parikh vector (3,2) indicating that there are three *a*'s and two *b*'s in all these words. The Parikh matrix is intended to provide more information than the Parikh vector in the sense that the Parikh matrix has the Parikh vector in the second main diagonal but also gives a count of the number of subwords in a given word with respect to an ordering of the alphabet. But the Parikh matrix is also not injective as was noticed in the earlier chapter. In fact the two words *abba, baab* have the same Parikh matrix

$$\begin{pmatrix} 1 & 2 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}$$

So the problem of injectivity of the Parikh matrix mapping or the Parikh matrix is of importance and has been studied in detail in the literature. In this chapter, we discuss properties of word relating to injectivity and non-injectivity of Parikh matrix mapping based on the study undertaken in Atanasiu *et.al*(2002) and Fosse *et.al* (2004).