2D Words and Generalized Parikh Matrices

K. Janaki¹, R. Arulprakasam²∗ and V. R. Dare³

¹,² Department of Mathematics, College of Engineering and Technology, Faculty of Engineering and Technology, SRM Institute of Science and Technology, SRM Nagar, Kattankulathur-603 203, Chennai, Tamilnadu, India.
³ Department of Mathematics, Madras Christian College, Chennai-600 059, Tamilnadu, India.
E-mail: r.aruljeeva@gmail.com

Abstract. The extension of Parikh vector is Parikh matrix which is a useful tool in arithmetizing words by numbers. Another interesting problem is the extension of Parikh matrix of a words to 2D words. However, Parikh matrix fails to tell about the number of subword occurrences in 2D word. Therefore, in this paper we introduce the notion of generalized Parikh matrix of 2D words which fulfill the number of subwords occurrences in an 2D words and also we obtain the properties of the generalized Parikh matrices.

Keywords: Parikh matrix, M-ambiguity, Generalized Parikh matrix, 2D words
Mathematics Subject Classification: 68Q45

1. Introduction
In formal language theory [9], arithmetizing words is one of the useful way in the study of word. In this spirit, Parikh vector of a word introduced by R.J.Parikh [8] which counts the number of occurrence of the symbols $\Sigma$ in a word. Since Parikh vector is not an injective, Mateescu et al [5] introduced Parikh matrix which gives more information about a word than a Parikh vector. Parikh vector gives the details about the number of occurrence of the symbols of $\Sigma$ in a word whereas Parikh matrix gives the details about the number of subwords of a word. Various aspects of Parikh matrix are studied in many novel ways in [1, 2, 3, 4, 6, 7, 10, 11, 12, 14, 15, 16, 17]. A. Salomaa [13] recently developed the idea of generalized Parikh matrix, which gives much more information about a word than Parikh matrix. Recently K.G.Subramanian et al [18] extended Parikh matrix of words to 2D words. Since Parikh matrix fails to tell about the number of subword occurrences in 2D words, therefore in this paper, we introduce generalized Parikh matrix of 2D words. This paper comprises four sections. In section 2, few elementary notions which are used throughout the paper are defined. In section 3, horizontal and vertical generalized Parikh matrix are introduced with conclusion in section 4.

2. Preliminaries
In this section, we recollect fundamental definitions and notations of words, scattered subwords, Parikh matrix for words and Parikh matrix for 2D words.

Consider a non-empty finite set of symbols as $\Sigma$. These symbols are termed as letters and the set is termed as an alphabet. Any string over the alphabet is termed as word and $\Sigma^*$ denoted the set of all words.

If $u \in \Sigma^*$ then $|u|$ called as length of $u$. Consider an ordered alphabet $\Sigma_k = \{a_1 < a_2 < ... < a_k\}$ where $k \geq 1$. The monoid morphism is the Parikh matrix mapping defined as $\delta_{P_k} : \Sigma^* \to M_{k+1}$, $\delta_{P_k}(a_q) = (p_{r,s})_{1 \leq r, s \leq k+1}$. For all $1 \leq r \leq k + 1$, $p_{r,r} = 1$ and $p_{q,q+1} = 1$ and remaining elements of...
the matrices are zero. Consider \( u = aabccab \) over \( \Sigma = \{ a < b < c \} \) then the Parikh matrix for \( u \) is

\[
\delta_{P_3}(abbc) = \delta_{P_3}(a)\delta_{P_3}(b)\delta_{P_3}(c)\delta_{P_3}(a)\delta_{P_3}(b) = \\
\begin{bmatrix}
1 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
\end{bmatrix} \begin{bmatrix}
1 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
\end{bmatrix} = \\
\begin{bmatrix}
1 & 3 & 4 & 4 \\
0 & 1 & 2 & 2 \\
0 & 0 & 1 & 2 \\
0 & 0 & 0 & 1 \\
\end{bmatrix}.
\]

If \( u \) and \( v \) in \( \Sigma^* \) are M-equivalent iff \( \delta_k(u) = \delta_k(v) \), for all \( k \). The triangle matrix \( R = P \oplus Q \) formed by the sum of all elements of two triangle matrices \( P \) and \( Q \) such that the elements in the leading diagonals of \( P \) and \( Q \) are excluded while adding. Consider a word \( u = b_1b_2...b_t \) of length \( t \) over \( \Sigma \). Consider \( \delta_u \) be generalized Parikh matrix mapping and the morphism \( \delta_u : \Sigma^* \rightarrow M_{t+1} \) and consider the letter \( b \) take place in the word \( u \) in positions \( q_1, q_2, ..., q_p \). The morphism of the letter \( b \) is denoted by \( \delta_u(a) = (p_{r,s})_{1 \leq r \leq s \leq t+1} \). Then for all \( 1 \leq r \leq s \leq t+1, p_{r,r} = 1 \). Furthermore

\[
p_{r,r+1} = \begin{cases} 
1 & \text{if } r \text{ is one of the numbers } q_1, q_2, ..., q_p \\
0 & \text{otherwise} 
\end{cases}
\]

(1)

The equation \( (\lfloor u \rfloor_a).(|u|_b) = (\lfloor u \rfloor_{ab}) + (\lfloor u \rfloor_{ba}) \) have some general information about occurrences of subwords over \( \Sigma = \{ a, b \} \). An arrangement in \( m \) rows and \( n \) columns of letters called 2D words or \( m \times n \) 2D words over a finite alphabet. Let \( \varphi \) be the set of all 2D words over \( \Sigma_k \). For example, 2D word say \( P \) over \( \Sigma_2 = \{ a < b \} \) is

\[
P = \\
\begin{array}{ccc}
a & a & b & a & a \\
b & a & a & a \\
a & b & a & a & b \\
b & b & a & a & a \\
\end{array}
\]

where \( P \in \varphi \). In horizontal Parikh matrix, the number of horizontal subword \( ab \) occurs whereas in vertical Parikh matrix, the number of vertical subword \( a \) occurs. Both horizontal Parikh matrix and vertical Parikh matrix tells the number of occurrences of \( a \) and the number of occurrences of \( b \) in \( P \). The column concatenation of 2D words say \( P \) and \( Q \) is denoted by \( P \odot Q \) where \( P \) and \( Q \) having equal rows. Similarly, the row concatenation is denoted by \( P \odot Q \) where \( P \) and \( Q \) having equal columns.

3. Generalized Parikh Matrices of 2D word

In this section we define horizontal generalized Parikh matrix and vertical generalized Parikh matrix of 2D words and we prove the properties of addition for row concatenation and column concatenation of two generalized Parikh matrices.

Definition 3.1 Consider a word \( u = r_1r_2...r_t \) of length \( t \) over \( \Sigma \). For \( m, n \geq 1 \), let \( P \in \varphi \) be a \( m \times n \) 2D word over \( \Sigma_k = \{ a_1 < a_2 < ... < a_k \} \) and \( x_r \) is the word in the \( r \)th row of \( P \) where \( r \in [1, m] \) and \( y_s \) is the word in the \( s \)th column of \( P \) where \( s \in [1, n] \). Then the generalized Parikh matrices of \( x_r \) and \( y_s \) be \( M_u(x_r) \) and \( M_u(y_s) \) respectively where \( 1 \leq r, s \leq t+1 \). The horizontal generalized Parikh matrix of \( P \) is defined as \( M_u^x(P) = M_u(x_1) \oplus M_u(x_2) \oplus ... \oplus M_u(x_m) \) and the vertical generalized Parikh matrix of \( P \) is defined as \( M_u^y(P) = M_u(y_1^T) \oplus M_u(y_2^T) \oplus ... \oplus M_u(y_n^T) \).


Example 3.2 Let $u = abba$ over $\Sigma = \{a,b\}$ and $P = b a a b a$ in $\varphi$ over $\Sigma_2 = \{a < b\}$. Then the generalized Parikh matrices $M_u(x_r)$, $1 \leq r \leq 3$ of the horizontal words $x_1 = ababa$, $x_2 = baaba$ and $x_3 = aaabb$ are

$$M_u(x_1) = \begin{bmatrix} 1 & 3 & 1 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$M_u(x_2) = \begin{bmatrix} 1 & 3 & 2 & 0 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and

$$M_u(x_3) = \begin{bmatrix} 1 & 3 & 6 & 3 & 0 \\ 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 6 & 0 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The horizontal generalized Parikh matrix $M_u^r(P)$ of $P$ is

$$M_u^r(P) = \begin{bmatrix} 1 & 9 & 1 & 1 \\ 0 & 1 & 6 & 3 \\ 0 & 0 & 1 & 9 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Similarly the vertical generalized Parikh matrix $M_u^c(P)$ of $P$ is

$$M_u^c(P) = \begin{bmatrix} 1 & 9 & 3 & 0 \\ 0 & 1 & 6 & 2 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Theorem 3.1 For all $P, Q \in \varphi$ over $\Sigma_k$, the column and row concatenation of generalized Parikh matrices of $P$ and $Q$ will be of the form $M_u^r(P \circ Q)$ and $M_u^c(P \circ Q)$ where $u$ be the word over $\Sigma$ then

(i). $M_u^r(P \circ Q) = M_u^r(P) + M_u^r(Q)$

(ii). $M_u^c(P \circ Q) = M_u^c(P) + M_u^c(Q)$.
Proof. Let the words $x_r$ and $y_r$ be in the $r^{th}$ rows of $P$ and $Q$ respectively. From Definition 3.1
\[
M_u^r(P) = M_u(x_1) \oplus M_u(x_2) \oplus \ldots \oplus M_u(x_m)
\]
\[
M_u^r(Q) = M_u(y_1) \oplus M_u(y_2) \oplus \ldots \oplus M_u(y_m).
\]
Then the column concatenation of generalized Parikh matrix of $P$ and $Q$ be
\[
M^r_u(P \circ Q) = M_u(x_1) \oplus M_u(x_2) \oplus \ldots \oplus M_u(x_m) \oplus M_u(y_1) \oplus M_u(y_2) \oplus \ldots \oplus M_u(y_m)
\]
\[
= M^r_u(P) + M^r_u(Q).
\]
The proof of row concatenation of generalized Parikh matrix of $P$ and $Q$ be $M^c_u(P \circ Q)$ being $M^c_u(P)$ with $M^c_u(Q)$ is similar.

Example 3.3 Let $u = aba$ over $\Sigma = \{a, b\}$ and $P = \begin{array}{ccc} a & b & a \\ b & a & b \end{array}$, $Q = \begin{array}{ccc} b & b & a \\ a & b & a \end{array}$ where $P, Q \in \wp$ over $\Sigma_2 = \{a < b\}$. The horizontal generalized Parikh matrices for $P$ and $Q$ are
\[
M^r_u(P) = \begin{bmatrix} 1 & 6 & 3 & 1 \\ 0 & 1 & 3 & 3 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 1 \end{bmatrix}
\]
\[
M^r_u(Q) = \begin{bmatrix} 1 & 4 & 4 & 0 \\ 0 & 1 & 5 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 1 \end{bmatrix}.
\]
The vertical generalized Parikh matrices for $P$ and $Q$ are
\[
M^c_u(P) = \begin{bmatrix} 1 & 6 & 3 & 1 \\ 0 & 1 & 3 & 3 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 1 \end{bmatrix}
\]
\[
M^c_u(Q) = \begin{bmatrix} 1 & 4 & 2 & 0 \\ 0 & 1 & 5 & 4 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 1 \end{bmatrix}.
\]
Then the column concatenation of generalized Parikh matrix of $P$ and $Q$ be
\[
M^r_u(P \circ Q) = M^r_u(P) + M^r_u(Q)
\]
\[
= \begin{bmatrix} 1 & 10 & 7 & 1 \\ 0 & 1 & 8 & 5 \\ 0 & 0 & 1 & 10 \\ 0 & 0 & 0 & 1 \end{bmatrix}
\]
and the row concatenation of generalized Parikh matrix of $P$ and $Q$ be
\[
M^c_u(P \circ Q) = M^c_u(P) + M^c_u(Q)
\]
\[
= \begin{bmatrix} 1 & 10 & 5 & 1 \\ 0 & 1 & 8 & 7 \\ 0 & 0 & 1 & 10 \\ 0 & 0 & 0 & 1 \end{bmatrix}.
\]
Proposition 3.1 If the column and row concatenation of generalized Parikh matrices of \( P \) and \( Q \) where \( u \) be the word over \( \Sigma \) and \( P, Q \in \wp \) over \( \Sigma_k \) will be of the form \( M_c^u (P \circ Q) \) and \( M^r_u (P \diamond Q) \) then

(i) \( M^c_u (P \circ Q) \neq M^c_u (P) + M^c_u (Q) \)

(ii) \( M^r_u (P \circ Q) \neq M^r_u (P) + M^r_u (Q) \).

Theorem 3.2 Let \( P \) and \( P' \) be 2D words. For all horizontal generalized Parikh matrix \( M^r_u (P) \), consider the word \( u \) of length \( t \) over an ordered alphabet \( \Sigma_t = \{ a_1 < a_2 < \ldots < a_t \} \) where \( P' \in \Sigma_t \) such that \( M^r_u (P) = M^s_r (P') \) can be effectively constructed.

Proof. Consider the word \( u \) of length \( t \) over \( \Sigma_t \). Let \( 1, \ldots, t \) represents the occurrences of the letters in \( u \). A morphism \( \delta : \Sigma \rightarrow \Sigma_t \) is indicated by \( \delta (a) = q_p q_{p-1} \ldots q_1 \) where \( a \in \Sigma \) appear in \( u \) from the positions \( q_1, q_2, \ldots, q_p \), \( 1 \leq p \leq t, q_i < q_{i+1}, 1 \leq r \leq p - 1 \). Choose \( P' = h (P) \) where \( P, P' \in \wp \) over \( \Sigma \). By Example 3.2, \( h (a) = 41 \) and \( h (b) = 32 \).

For 2D word

\[
P = \begin{array}{c}
a b a b a \\
b b a b a \\
a a a b b 
\end{array}
\]

we obtain

\[
P' = \begin{array}{c}
41 32 41 32 41 \\
32 41 32 41 \\
41 41 32 32
\end{array}
\]

For ordered alphabet \( \Sigma_4 = \{ 1, 2, 3, 4 \} \), we have

\[
M^s_r (P') = \begin{bmatrix}
1 & 9 & 11 & 4 & 1 \\
0 & 1 & 6 & 3 & 2 \\
0 & 0 & 1 & 6 & 7 \\
0 & 0 & 0 & 1 & 9 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
\]

Using induction hypothesis on the length of each horizontal words of \( P \) we will prove the remaining part of the theorem. If \( P \) is \( 1 \times 1 \) array then the theorem satisfied. Assume that the theorem satisfies for each horizontal words of \( P \) be length \( n \). Consider \( \delta (w) \in \Sigma^+_t \) and \( w \in \Sigma \). Now we have \( M^r_u (P \diamond w) = M^s_r (P) + M_u (w) \). By induction hypothesis, \( M^s_r (P) = M^s_r (h (P)) \). Therefore

\[
M^r_u (P \diamond w) = M^r_s (h (P)) + M_u (w) = M^s_r (h (P)) + M^s_r (h (w)) = M^r_s (h (P \diamond w)) .
\]

Assume \( A = \) \[ a b a b a \] and \( A \subseteq P \) and \( w = ababb \). We have

\[
A = \begin{array}{c}
a b a b a \\
\end{array}, \quad \text{and} \quad A \subseteq P \text{ and } w = ababb .
\]

\[
M^r_u (A) = \begin{bmatrix}
1 & 8 & 10 & 6 & 4 \\
0 & 1 & 7 & 8 & 6 \\
0 & 0 & 1 & 8 & 10 \\
0 & 0 & 0 & 1 & 7 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
\]
and

\[
M_r^c(w) = \begin{bmatrix}
1 & 2 & 5 & 1 & 2 \\
0 & 1 & 3 & 2 & 2 \\
0 & 0 & 1 & 2 & 5 \\
0 & 0 & 0 & 1 & 3 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
\]

Then

\[
M_r^c(A \diamond w) = \begin{bmatrix}
1 & 10 & 15 & 7 & 6 \\
0 & 1 & 10 & 10 & 8 \\
0 & 0 & 1 & 10 & 15 \\
0 & 0 & 0 & 1 & 10 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}.
\]

Finally, we have

\[
M_r^c(A \diamond w) = M_r^c(h(A \diamond w)).
\]

**Theorem 3.3** Let \(P\) and \(P'\) be 2D words. For every vertical generalized Parikh matrix \(M_r^c(P)\), consider the word \(u\) of length \(t\) over an ordered alphabet \(\Sigma_t = \{a_1 < a_2 < \ldots < a_t\}\) where \(P' \in \Sigma_t\) such that \(M_r^c(P) = M_s^c(P')\) can be effectively constructed.

**Proof.** Consider the word \(u\) of length \(t\) over \(\Sigma_t\). Let \(1, \ldots, t\) represents the occurrences of the letters in \(u\). A morphism \(\delta : \Sigma \to \Sigma_t\) is indicated by \(\delta(a) = q_pq_{p-1} \ldots q_1\) where \(a \in \Sigma\) appear in \(u\) from the positions \(q_1, q_2, \ldots, q_p, 1 \leq p \leq s, q_r < q_{r+1}, 1 \leq r \leq p - 1\). Choose \(P' = h(P)\) where \(P, P' \in \wp\) over \(\Sigma\). By Example 3.2, \(h(a) = 41\) and \(h(b) = 32\).

For 2D word

\[
P = \begin{bmatrix}
\begin{array}{cccc}
abab & abab & abab & abab & abab
\end{array}
\end{bmatrix}
\]

we obtain

\[
P' = \begin{bmatrix}
4 & 3 & 4 & 3 & 4 \\
1 & 2 & 1 & 2 & 1 \\
3 & 4 & 4 & 3 & 4 \\
2 & 1 & 1 & 2 & 1 \\
4 & 4 & 4 & 3 & 3
\end{bmatrix}
\]

For ordered alphabet \(\Sigma_4 = \{1, 2, 3, 4\}\), we have

\[
M_s^c(P') = \begin{bmatrix}
1 & 9 & 3 & 0 & 0 \\
0 & 1 & 6 & 2 & 0 \\
0 & 0 & 1 & 6 & 3 \\
0 & 0 & 0 & 1 & 9 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}.
\]

Using induction hypothesis on the length of each vertical words of \(P\) we will prove the remaining part of the theorem. If \(P\) is \(1 \times 1\) array then the theorem satisfied. Assume that the theorem satisfies
for each vertical words of $P$ be length $n$. Consider $\delta (w) \in \Sigma^+_k$ and $w \in \Sigma$. Now we have $M_u^c (P \circ w) = M_u^c (P) + M_u (w)$. By induction hypothesis, $M_u^c (P) = M_u^c (h (P))$. Therefore

\[
M_u^c (P \circ w) = M_u^c (h (P)) + M_u (w) = M_u^c (h (P)) + M_s (h (w)) = M_u^c (h (P \circ w)).
\]

Assume $A = a b a b a b$ and $A \subseteq P$ and $w = a b a b a b$. We have

\[
M_u^c (A) = \begin{bmatrix}
1 & 8 & 4 & 2 & 0 \\
0 & 1 & 7 & 6 & 2 \\
0 & 0 & 1 & 8 & 4 \\
0 & 0 & 0 & 1 & 7 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
\]

and

\[
M_u^c (w) = \begin{bmatrix}
1 & 2 & 5 & 1 & 2 \\
0 & 1 & 3 & 2 & 2 \\
0 & 0 & 1 & 2 & 5 \\
0 & 0 & 0 & 1 & 3 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
\]

Then

\[
M_u^c (A \circ w) = \begin{bmatrix}
1 & 10 & 9 & 3 & 2 \\
0 & 1 & 10 & 8 & 4 \\
0 & 0 & 1 & 10 & 9 \\
0 & 0 & 0 & 1 & 10 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}.
\]

Finally, we have $M_u^c (A \circ w) = M_u^c (h (A \circ w))$.

4. Conclusion

In this paper, horizontal generalized Parikh matrix and vertical generalized Parikh matrix for 2D words are defined with illustrations. Also we obtained the properties of addition for row concatenation and column concatenation of two generalized Parikh matrices. This work can be further extended to horizontal Parikh-friendly permutation and vertical Parikh-friendly permutation for 2D words.

5. References

[1] A. Atanasiu, C. Martin-Vide and A. Mateescu, On the injectivity of the Parikh matrix mapping, Fund. Inform. 46, 2001: 1-11.
[2] A. Atanasiu, Binary amiable words, Int. J. Fund. Comput. Sci. 18, 2007: 387-400.
[3] A. Atanasiu, R. Atanasiu and I. Petre, Parikh matrices and amiable words, Theor. Comput. Sci. 390, 2008: 102-109.
[4] S. Fosse and G. Richomme, Some characterizations of Parikh matrix equivalent binary words, Inf. Process. Lett. 92, 2004: 077-082.
[5] A. Mateescu, A. Salomaa, K. Salomaa and S. Yu, A Sharpening of the Parikh mapping, Theor. Inform. Appl. 35, 2001: 551-564.
[6] A. Mateescu and A. Salomaa, Matrix indicators for subword occurrences and ambiguity, Int. J. Fund. Comput. Sci. 15, 2004: 1-21.
[7] A. Mateescu, Algebraic aspects of Parikh matrices, Theory is Forever. 10, 2004: 170-180.
[8] R.J.Parikh, On context-free languages, J.Assoc.Comput.Mach.13, 1966: 570-581.
[9] A.Salomaa, Formal Languages, Academic Press, New York, 1973.
[10] A.Salomaa, On the injectivity of Parikh matrix mappings, Fund.Inform.64, 2005: 391-404.
[11] A.Salomaa, Subword histories and associated matrices, Theor.Comput.Sci.407, 2008: 233-247.
[12] A.Salomaa, Criteria for the matrix equivalence of words, Theor.Comput.Sci.411, 2010: 1818-1827.
[13] A.Salomaa, Parikh Matrices: Subword indicators and degrees of ambiguity, LNCS.1101, 2018: 100-112.
[14] V.N.Serbanuta and T.F.Serbanuta, Injectivity of the Parikh matrix mappings revisited, Fund.Inf.73, 2006: 265-283.
[15] V.N.Serbanuta, On Parikh matrices, ambiguity and prints, Int.J.Fund.Comput.Sci.20, 2009: 151-165.
[16] T.F.Serbanuta, Extending Parikh matrices, Theor.Comput.Sci.310, 2004: 233-246.
[17] K.G.Subramanian, A.Miin Huey and A.K.Nagar, On Parikh matrices, Int.J.Fund.Comput.Sci.20, 2009: 211-219.
[18] K.G.Subramanian, K.Mahalingam, R.Abdullah and A.K.Nagar, Two-dimensional digitized picture arrays and Parikh matrices, Int.J.Fund.Comput.Sci.24, 2013: 393-408.