A q-Matrix Encoding Extending the Parikh Matrix Mapping

Ömer Eğecioğlu

Abstract: We introduce a generalization of the Parikh mapping called the *Parikh q-matrix encoding*, which takes its values in matrices with polynomial entries. The encoding represents a word w over a k-letter alphabet as a $(k+1)$-dimensional upper-triangular matrix with entries that are nonnegative integral polynomials in variable q. Putting $q = 1$, we obtain the morphism introduced by Mateescu, Salomaa, Salomaa, and Yu [6] which extends the Parikh mapping to $(k+1)$-dimensional (numerical) matrices. The Parikh q-matrix encoding however, produces matrices that carry more information about w than the numerical Parikh matrix. In fact it is injective. The entries of the q-matrix image of w under this encoding is constructed by q-*counting* the number of occurrences of certain words as scattered subwords of w. This construction is distinct from the Parikh q-matrix mapping into k-dimensional upper-triangular matrices with integral polynomial entries introduced by Eğecioğlu and Ibarra [2].

Keywords: Parikh mapping, Parikh matrix mapping, scattered subword, q-analogue.

1 Introduction

By Parikh's theorem [7], the commutative image of any context-free language is a semilinear set, and is therefore also the commutative image of some regular set. Consider the (ordered) alphabet $\Sigma_k = \{a_1 < a_2 < \cdots < a_k\}$ and for $w \in \Sigma^*$, define by $|w|_{a_i}$ the number of occurrences of a_i in w. The *Parikh mapping* is a morphism

$$\Psi \ : \ \Sigma^* \to \mathbb{N}^k$$

where $\mathbb{N}$ is the set of nonnegative integers and $\Psi(w) = (|w|_{a_1}, |w|_{a_2}, \cdots, |w|_{a_k})$ is the Parikh vector.

The Parikh mapping is an important concept in the theory of formal languages. Various languages accepted by automata that are more powerful than pushdown automata have been shown to have effectively computable semilinear sets. For example, it is known that every language accepted by a pushdown automaton augmented with reversal-bounded counters has a semilinear Parikh map [4]. The fact that the emptiness problem for semilinear sets is decidable implies that the emptiness problem for these automata is decidable. This fact has a number of applications in formal languages (e.g., [3]) and formal verification (e.g., [5]).

The *Parikh matrix mapping* introduced in [6] is a morphism $\Psi_{\mathcal{M}_k} : \Sigma^* \to \mathcal{M}_{k+1}$ where $\mathcal{M}_{k+1}$ is a collection of $(k+1)$-dimensional upper-triangular matrices with nonnegative integral entries and unit diagonal. The classical Parikh vector $\Psi(w)$ appears in the image matrix as the second diagonal.

In [2] the *Parikh q-matrix mapping* was introduced. It is a morphism $\Psi_q^k : \Sigma^* \to \mathcal{M}_k(q)$ where $\mathcal{M}_k(q)$ is a collection of k-dimensional upper-triangular matrices with nonnegative integral polynomials in q as entries. The diagonal entries of $\Psi_q^k(w)$ are

$$(q^{|w|_{a_1}}, q^{|w|_{a_2}}, \cdots, q^{|w|_{a_k}})$$

which readily encodes the Parikh vector. In addition, it can be shown [2] that viewing $w \in \Sigma_k$ as a word in Σ_{k+1} with $|w|_{a_{k+1}} = 0$, the Parikh q-matrix $\Psi_q^{k+1}(w)$ evaluated at $q = 1$ is precisely the $(k+1)$-dimensional numerical Parikh matrix $\Psi_{\mathcal{M}_k}(w)$.

It is a basic property of the Parikh matrix mapping that two words with the same Parikh matrix have the same Parikh vector, but two words with the same Parikh vector in many cases have different Parikh matrices [1]. Thus, the Parikh matrix gives more information about a word than a Parikh vector. Similarly, two words with the same Parikh q-matrix have the same Parikh matrix (and therefore the same Parikh vector), but there are cases in which two words with the same Parikh matrix have different q-matrices. Thus the Parikh q-matrix gives more information about a word than the Parikh matrix. We show that the Parikh q-matrix encoding introduced here is injective, and for $q = 1$, it also reduces to the $(k+1)$-dimensional numerical Parikh matrix mapping.

The basic idea in the construction of the entries of the Parikh q-matrix encoding of w is q-*counting* the number of occurrences of certain words as scattered subwords of w. This is done by keeping track of the indices (positions) of the letters of the scattered subword in w as exponents of monomials in q.

The structure of this paper is as follows. Sections 2 gives some basic notation and definitions. Section 3 recalls the notion of a Parikh matrix mapping introduced in [6] and the fundamental theorem concerning these mappings. Section 4 presents the polynomials which generalize the count of scattered subwords. Section 5 presents our Parikh q-matrix encoding and the main results, including Theorem 2, which gives the main properties of the encoding. Section 6 looks at properties such as injectivity and gives an interpretation of the inverse matrix in terms of mirror images.

2   Definitions

We start with some basic notation and definitions. Most of these are as they appear in references [6, 1, 2]. The set of all nonnegative integers is denoted by $\mathbb{N}$. We denote by $\mathbb{N}[q]$ the collection of polynomials in the variable q with coefficients from $\mathbb{N}$. $\mathbb{Z}$ denotes integers, and $\mathbb{Z}[q]$ denotes the ring of polynomials in the variable q with integral coefficients. For an alphabet Σ, we denote the set of all words over Σ by Σ^* and the empty word by λ. We use "ordered" alphabets. An ordered alphabet is an alphabet $\Sigma = \{a_1, a_2, \ldots, a_k\}$ with a relation of order ("$<$") on it. If for instance $a_1 < a_2 < \cdots < a_k$, then we use the notation

$$\Sigma = \{a_1 < a_2 < \cdots < a_k\}.$$

If $w \in \Sigma^*$ then $|w|$ denotes the length of w. For $a_i \in \Sigma$ and $w \in \Sigma^*$ the number of occurrences of the letter a_i in w is denoted by $|w|_{a_i}$.

Let v, w be words over Σ. As defined in [6], the word v is called a *scattered subword* of w if there exists a word u such that $w \in u \sqcup\!\sqcup v$, where $\sqcup\!\sqcup$ denotes the shuffle operation. If $v, w \in \Sigma^*$, then the number of occurrences of v in w as a scattered subword is denoted by $|w|_{scatt-v}$. Partially overlapping occurrences of a word as a scattered subword of a word are counted as distinct occurrences. For example, $|acbb|_{scatt-ab} = 2$, $|acba|_{scatt-ab} = 1$.

Notation We also use the symbol $P_{w,v}$ to denote $|w|_{scatt-v}$. Using this notation, $P_{acbb,ab} = 2$, $P_{acba,ab} = 1$, and $P_{w,a_i} = |w|_{a_i}$ for any letter $a_i \in \Sigma$.

Notation Consider the ordered alphabet $\{a_1 < a_2 < \cdots < a_k\}$ where $k \geq 1$. As in [6], we denote by the symbol $a_{i,j}$ the word $a_i a_{i+1} \cdots a_j$ where $1 \leq i \leq j \leq k$.

For motivation and further issues about the Parikh mapping as well as language-theoretic notions not considered here, we refer the reader to [8].

3   Parikh matrix mapping

We first describe the extension of the Parikh mapping to matrices as originally defined in [6]. The extension involves special types of triangular matrices. These are square matrices $m = (m_{i,j})_{1 \leq i,j \leq k}$ such that $m_{i,j} \in \mathbb{N}$, for all $1 \leq i, j \leq k$, $m_{i,j} = 0$, for all $1 \leq j < i \leq k$, and moreover, $m_{i,i} = 1$, for all $1 \leq i \leq k$. The set of all these matrices of dimension k is denoted by $\mathcal{M}_k$. Thus $\mathcal{M}_k$ is the collection $k \times k$ upper-triangular matrices with entries from $\mathbb{N}$ and unit diagonal. The set $\mathcal{M}_k$ is a monoid with respect to multiplication of matrices and has a unit which is the identity matrix I_k. The main notion introduced in [6] is as follows:

Definition 3. Let $\Sigma = \{a_1 < a_2 < \cdots < a_k\}$ be an ordered alphabet, where $k \geq 1$. The Parikh matrix mapping, denoted by $\Psi_{\mathcal{M}_k}$, is the morphism:

$$\Psi_{\mathcal{M}_k} \: : \: \Sigma^* \to \mathcal{M}_{k+1},$$

defined as follows: If $\Psi_{\mathcal{M}_k}(a_l) = (m_{i,j})_{1 \leq i,j \leq (k+1)}$, then for each $1 \leq i \leq k + 1$, $m_{i,i} = 1$, $m_{l,l+1} = 1$ and all other elements of the matrix $\Psi_{\mathcal{M}_k}(a_l)$ are zero.

Example 3.1. Let Σ be the ordered alphabet $\{a < b < c\}$. Then the Parikh matrix mapping Ψ_{M_3} represents each word over Σ^* as a 4×4 upper triangular matrix with unit diagonal with nonnegative integral entries. As an example $\Psi_{\mathcal{M}_3}(ab^2) = \Psi_{\mathcal{M}_3}(a)\Psi_{\mathcal{M}_3}(b)\Psi_{\mathcal{M}_3}(b)$. Thus

$$\Psi_{\mathcal{M}_3}(ab^2) = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 2 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Furthermore

$$
\Psi_{\mathcal{M}_3}(ab^2ac^2a) = \begin{bmatrix} 1 & 1 & 2 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \Psi_{\mathcal{M}_3}(a)\Psi_{\mathcal{M}_3}(c)\Psi_{\mathcal{M}_3}(c)\Psi_{\mathcal{M}_3}(a) = \begin{bmatrix} 1 & 3 & 2 & 4 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{1}
$$

The main property of the Parikh matrix mapping is the following theorem:

Theorem 1. ([6], Theorem 3.1) Let $\Sigma = \{a_1 < a_2 < \cdots < a_k\}$ be an ordered alphabet, where $k \geq 1$ and assume that $w \in \Sigma^*$. The matrix $\Psi_{M_k}(w) = (m_{i,j})_{1 \leq i,j \leq (k+1)}$, has the following properties

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

2. $m_{i,i} = 1$, for all $1 \leq i \leq k+1$,

3. $m_{i,j+1} = P_{w,a_{i,j}}$, for all $1 \leq i \leq j \leq k$.

As a corollary

Corollary 1. *([6], Corollary 3.1) Let* $\Sigma = \{a_1 < a_2 < \cdots < a_k\}$ *The matrix* $\Psi_{M_k}(w)$ *has as its the second diagonal the Parikh vector of* w, *i.e.,* $(m_{1,2}, m_{2,3}, \ldots, m_{k,k+1}) = \Psi(w) = (|w|_{a_1}, |w|_{a_2}, \cdots, |w|_{a_k})$.

4 Index enumerator of scattered subwords

Next we introduce a collection of polynomials $P_{w,a_{i,j}}(q)$ indexed by $w \in \Sigma^*$, and $a_{i,j} = a_i a_{i+1} \cdots a_j$, with $1 \leq i \leq j \leq k$. These polynomials will "q-count" the numbers $P_{w,a_{i,j}}$ defined above for general v and w. To construct $P_{w,a_{i,j}}(q)$, we consider a factorization

$$
w = u_i a_i u_{i+1} a_{i+1} \cdots u_j a_j u_{j+1} \tag{2}
$$

with $u_s \in \Sigma^*$ for $i \leq s \leq j+1$. The numbers

$$
I_i = |u_i| + 1, \ I_{i+1} = |u_i| + |u_{i+1}| + 2, \ldots, \ I_j = |u_i| + |u_{i+1}| + \cdots + |u_j| + j - i + 1 \tag{3}
$$

are simply the indices (positions) in w where the letters $a_i, a_{i+1}, \ldots a_j$ appear respectively, in the factorization (2). To construct the polynomial $P_{w,a_{i,j}}(q)$, we form the monomial

$$
q^{I_i + I_{i+1} \cdots + I_j} \tag{4}
$$

in $\mathbb{N}[q]$ corresponding to each such factorization, and add up the resulting monomials. Thus

$$
P_{w,a_{i,j}}(q) = \sum_{w = u_i a_i u_{i+1} \cdots u_j a_j u_{j+1}} q^{I_i + I_{i+1} + \cdots + I_j} \tag{5}
$$

where the exponents are the indices of $a_i, a_{i+1} \ldots, a_j$ in w as given in (3).

Example 4.1. Suppose $\Sigma = \{a < b < c\}$ and $w = ab^2ac^2a$. If $i = 1, j = 3$, then $a_{i,j} = abc$ and writing the indices and underlining the positions of a, b and c for each appearance of abc as a scattered subword of w, we obtain

	1	2	3	4	5	6	7	Positions	Contribution
Factorization :	$\underline{a}$	$\underline{b}$	b	a	$\underline{c}$	c	a	$1, 2, 5$	q^8
Factorization :	$\underline{a}$	$\underline{b}$	b	a	c	$\underline{c}$	a	$1, 2, 6$	q^9
Factorization :	$\underline{a}$	b	$\underline{b}$	a	$\underline{c}$	c	a	$1, 3, 5$	q^9
Factorization :	$\underline{a}$	b	$\underline{b}$	a	c	$\underline{c}$	a	$1, 3, 6$	q^{10}

Consequently

$$
P_{w,abc}(q) = q^8 + 2q^9 + q^{10}.
$$

Similarly, we find that

$$
\begin{array}{ll}
P_{w,a}(q) = & q + q^4 + q^7 \qquad\qquad P_{w,ab}(q) = & q^3 + q^4 \\
P_{w,b}(q) = & q^2 + q^3 \qquad\qquad\quad P_{w,ac}(q) = & q^7 + q^8 \\
P_{w,c}(q) = & q^5 + q^6 \qquad\qquad\quad P_{w,bc}(q) = & q^7 + 2q^8 + q^9
\end{array}
$$

Since the summation in the definition (5) of $P_{w,a_{i,j}}(q)$ is over all occurrences of $a_{i,j}$ in w as a scattered subword, the following proposition is immediate:

Proposition 1. Let $\Sigma = \{a_1 < a_2 < \cdots < a_k\}$ and $1 \le i \le j \le k$. Then $P_{w,a_{i,j}}(1) = P_{w,a_{i,j}}\ (= |w|_{scatt-a_{i,j}})$.

Thus the polynomials $P_{w,a_{i,j}}(q)$ "q-count" the number of occurrences of $a_i a_{i+1} \cdots a_j$ as a scattered subword of w.

5   The Parikh q-matrix encoding

We denote by $\mathcal{M}_k(q)$ the collection of k-dimensional upper-triangular matrices with entries in $\mathbb{N}[q]$. Let I_k denote the identity matrix of dimension k. We define a mapping

$$\overline{\Psi} : \Sigma \times \mathbb{N} \ \to \ \mathcal{M}_{k+1}(q)$$

as follows. The matrix $\overline{\Psi}(a_l, j)$ corresponding to a pair $a_l \in \Sigma = \{a_1 < a_2 < \cdots < a_k\}$ and $j \in \mathbb{N}$, is defined as the matrix obtained from I_{k+1} by changing the $(l, l+1)$-st entry in I_{k+1} to q^j. Thus if $\overline{\Psi}(a_l, j) = (m_{i,j})_{1 \le i,j \le k+1}$, then $m_{i,i} = 1$ for $1 \le i \le k + 1$, $m_{l,l+1} = q^j$, and all other entries of the matrix $\overline{\Psi}(a_l, j)$ are zero.

Example 5.1. When the alphabet is $\Sigma = \{a < b < c\}$,

$$\overline{\Psi}(a,j) = \begin{bmatrix} 1 & q^j & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \overline{\Psi}(b,j) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & q^j & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \overline{\Psi}(c,j) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & q^j \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Remark 2. The matrix $\overline{\Psi}(a_l, j)$ differs from the matrix $\Psi(a_l)$ introduced in [6] in the definition of the Parikh matrix mapping only in the $(l, l + 1)$-st entry: In $\Psi(a_l)$ this entry is 1, whereas in $\overline{\Psi}(a_l, j)$ it is q^j. We note that the matrices $\overline{\Psi}(a_l, j)$ specialize to $\Psi(a_l)$ for $q = 1$, for any value of j.

We now define a mapping (which we again denote by $\overline{\Psi}$) from Σ^* to $\mathcal{M}_{k+1}(q)$ by setting $\overline{\Psi}(\lambda) = I_{k+1}$ and

$$\overline{\Psi}(w_1 w_2 \cdots w_n) = \overline{\Psi}(w_1, 1)\overline{\Psi}(w_2, 2) \cdots \overline{\Psi}(w_n, n), \quad w_i \in \Sigma, \ 1 \le i \le n.$$

We refer to $\overline{\Psi}$ as the *Parikh q-matrix encoding*.

Example 5.2. Let Σ be the ordered alphabet $\{a < b < c\}$. Then $\overline{\Psi}(ab^2) = \overline{\Psi}(a, 1)\overline{\Psi}(b, 2)\overline{\Psi}(b, 3)$. Thus

$$\overline{\Psi}(ab^2) = \begin{bmatrix} 1 & q & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & q^2 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & q^3 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & q & q^3 + q^4 & 0 \\ 0 & 1 & q^2 + q^3 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Furthermore

$$\overline{\Psi}(ab^2 ac^2 a) = \begin{bmatrix} 1 & q & q^3 + q^4 & 0 \\ 0 & 1 & q^2 + q^3 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \overline{\Psi}(a, 4)\overline{\Psi}(c, 5)\overline{\Psi}(c, 6)\overline{\Psi}(a, 7)$$

$$= \begin{bmatrix} 1 & q + q^4 + q^7 & q^3 + q^4 & q^8 + 2q^9 + q^{10} \\ 0 & 1 & q^2 + q^3 & q^7 + 2q^8 + q^9 \\ 0 & 0 & 1 & q^5 + q^6 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Note that for $q = 1$, this is the matrix $\Psi_{\mathcal{M}_3}(ab^2 ac^2 a)$ displayed as (1).

Remark 3. The Parikh q-matrix encoding is not a morphism for arbitrary q. For example over $\Sigma = \{a < b\}$ $\overline{\Psi}(ab) = \overline{\Psi}(a, 1)\overline{\Psi}(b, 2) \ne \overline{\Psi}(a, 1)\overline{\Psi}(b, 1) = \overline{\Psi}(a)\overline{\Psi}(b)$

Remark 4. Neither the Parikh matrix mapping of [6] nor the Parikh q-matrix mapping of [2] is injective. We will later show that the Parikh q-matrix encoding $\overline{\Psi}$ is injective. On the other hand, both Parikh matrix mapping and the Parikh q-matrix mapping are morphisms, whereas Parikh q-matrix encoding is not.

Remark 5. Just as in the case of the Parikh matrix mapping and the Parikh q-matrix mapping, it is not true for the Parikh q-matrix encoding that if $\mathcal{L}$ is a context-free language, then its image is some suitable extension of the notion of semilinearity to matrices over $\mathbb{N}[q]$. This is a direct consequence of Theorem 3 and the similar negative result concerning the Parikh matrix mapping ([6], Remark 3.2).

Now we give a characterization of the entries of the matrix $\overline{\Psi}(w)$ in terms of the polynomials $P_{w,a_{i,j}}(q)$ introduced in section 4.

Theorem 2. Let $\Sigma = \{a_1 < a_2 < \cdots < a_k\}$ be an ordered alphabet, where $k \geq 1$ and assume that $w \in \Sigma^*$. The matrix $\overline{\Psi}(w) = (m_{i,j}(q))_{1 \leq i,j \leq k+1}$, has the following properties

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

2. $m_{i,i} = 1$, for all $1 \leq i \leq k+1$,

3. $m_{i,j+1} = P_{w,a_{i,j}}(q)$, for all $1 \leq i \leq j < k+1$.

Proof. The proof of the parts 1. and 2. are immediate. We now prove property 3. Assume that $|w| = n$. The proof is by induction on n. If $n \leq 1$, the assertion holds. Assume now that part 3. holds for all words of length n and let w be of length $n+1$. Write $w = w'a_j$ where $|w'| = n$ and $a_j \in \Sigma$. Then

$$\overline{\Psi}(w) = \overline{\Psi}(w')\overline{\Psi}(a_j, n+1)$$

Assume that

$$\overline{\Psi}(w') = \begin{bmatrix} 1 & m'_{1,2} & \cdots & \cdots & m'_{1,k+1} \\ 0 & 1 & \cdots & \cdots & m'_{2,k+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & 1 & m'_{k,k+1} \\ 0 & 0 & \cdots & \cdots & 1 \end{bmatrix} = M'$$

By the inductive hypothesis the matrix $M' = \overline{\Psi}(w')$ has property 3. We have

$$\overline{\Psi}(a_j, n+1) = \begin{bmatrix} 1 & 0 & & \cdots & & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & q^{n+1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & & \cdots & & 1 \end{bmatrix}$$

where the matrix differs from I_{k+1} only in the entry in position $(j, j+1)$ where it is q^{n+1} instead of 0. Let $M = \overline{\Psi}(w)$. Then

$$M = \begin{bmatrix} 1 & m'_{1,2} & \cdots & \cdots & m'_{1,k+1} \\ 0 & 1 & \cdots & \cdots & m'_{2,k+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & 1 & m'_{k,k+1} \\ 0 & 0 & \cdots & \cdots & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & & \cdots & & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & q^{n+1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & & \cdots & & 1 \end{bmatrix}$$

If $M = (m_{r,s})_{1 \leq r,s \leq k+1}$, then

$$m_{i,j+1} = m'_{i,j+1} + m'_{i,j} \cdot q^{n+1} \quad \text{for all } 1 \leq i \leq j, \tag{6}$$

and for all other indices, $m_{r,s} = m'_{r,s}$. But these are immediate from the definition of the polynomials $P_{w,a_{i,j}}(q)$ which satisfy

$$P_{w'a_j,a_{i,j}}(q) = P_{w',a_{i,j}}(q) + P_{w',a_{i,j-1}}(q) \cdot q^{n+1} \quad \text{for } 1 \leq i \leq j \leq k, \tag{7}$$

and the proof follows by induction. $\qquad\square$

Theorem 3. Suppose $\Sigma = \{a_1 < a_2 < \cdots < a_k\}$ and $w \in \Sigma^*$. Let $\overline{\Psi}(w)$ be the Parikh q-matrix encoding by $(k+1)$-dimensional matrices over $\mathbb{N}[q]$. Then $\overline{\Psi}(w)$ evaluated at $q = 1$ is the Parikh matrix $\Psi_{\mathcal{M}_k}(w)$.

Proof. Combine Theorem 2, Theorem 1, and Proposition 1. $\qquad\square$

Corollary 6. *Let* $\Sigma = \{a_1 < a_2 < \cdots < a_k\}$ *and* $w \in \Sigma^*$. *Suppose the vector of second diagonal entries of the matrix* $\overline{\Psi}(w)$ *is* $(m_{1,2}(q), m_{2,3}(q), \cdots, m_{k,k+1}(q)) \in \mathbb{N}[q]^k$. *Then*

$$(m_{1,2}(q), m_{2,3}(q), \cdots, m_{k,k+1}(q)) = (P_{w,a_1}(q), P_{w,a_2}(q), \cdots, P_{w,a_k}(q))$$

and at $q = 1$, *this vector evaluates to the Parikh vector* $(|w|_{a_1}, |w|_{a_2}, \cdots, |w|_{a_k}) \in \mathbb{N}^k$.

6   Injectivity and matrix inverses

The Parikh matrix mapping and the Parikh q-matrix mapping are not injective mappings. For instance over the ordered alphabet $\{a < b < c\}$ one has $\Psi(acb) = \Psi(cab)$ [6] as well as $\Psi_q(acb) = \Psi_q(cab)$ [2]. Injectivity properties of the Parikh matrix mapping was studied for the binary alphabet in [1].

Note that by Theorem 2, the polynomials which form the second diagonal of the matrix $\overline{\Psi}(w)$ completely determine w. This is because the $(i, i+1)$-st entry in $\overline{\Psi}(w)$ is $P_{w,a_i}(q)$, which encodes the positions of the letter a_i in w. Clearly, knowing the positions of each letter a_i in w for $1 \leq i \leq k$ is equivalent to knowing w completely. Thus we have

Theorem 4. The Parikh q-matrix encoding is injective: i.e. $\overline{\Psi}(u) = \overline{\Psi}(v)$ implies $u = v$.

The notion of the alternate (signed) Parikh matrix was developed in [6]. It has the nice property that the inverse of the matrix $\Psi_{\mathcal{M}_k}(w)$ is the alternate Parikh matrix of the mirror image $mi(w)$ of w. Suppose that $\Sigma = \{a_1 < a_2 < \cdots < a_k\}$ and $w \in \Sigma^*$ with Parikh matrix $M = (m_{i,j})_{1 \leq i,j \leq k+1}$. Then the alternate Parikh matrix of w is the matrix $M' = (m'_{i,j})_{1 \leq i,j \leq k+1}$ where $m'_{i,j} = (-1)^{i+j} m_{i,j}$ for all $1 \leq i \leq j \leq k+1$. Then the following interesting result holds

Theorem 5. (Restatement of [6], Theorem 3.2) Suppose $\Sigma = \{a_1 < a_2 < \cdots < a_k\}$ and $w = w_1 w_2 \cdots w_n \in \Sigma^*$. Then the matrix inverse of the Parikh matrix of $mi(w) = w_n \cdots w_2 w_1$ is the alternate Parikh matrix of w.

It can also be shown that there is a similar identity involving the k-dimensional matrices for w and $mi(w)$ for the Parikh q-matrix mapping (See [2], Theorem 4).

We define the *alternate Parikh q-matrix encoding* $\overline{\Psi}'$ on Σ^* as follows: Suppose $\overline{\Psi}(w) = (m_{i,j}(q))_{1 \leq i,j \leq k+1}$. Then $\overline{\Psi}'(w) = (m'_{i,j}(q))_{1 \leq i,j \leq k+1}$ where $m'_{i,j}(q) = (-1)^{i+j} m_{i,j}(q)$ for all $1 \leq i \leq j \leq k+1$. Then

Theorem 6. Given $w = w_1 w_2 \cdots w_n \in \Sigma^*$, the matrix inverse of

$$\overline{\Psi}(w_n, n)\overline{\Psi}(w_{n-1}, n-1)\cdots\overline{\Psi}(w_1, 1) \tag{8}$$

is the alternate Parikh matrix encoding of w.

Proof. Omitted. $\qquad\square$

Thus not only we need to reverse the order of the letters as in $mi(w)$, but also their indices. For $q = 1$, however, the indices do not matter, and from Theorem 6, we obtain Theorem 3.2 of [6] for the Parikh matrix as a special case.

Example 6.1. For $w = ab^2ac^2a$, the matrix product displayed in (8) evaluates to

$$\begin{bmatrix} 1 & q + q^4 + q^7 & q^6 + q^7 + q^9 + q^{10} & 0 \\ 0 & 1 & q^2 + q^3 & 0 \\ 0 & 0 & 1 & q^5 + q^6 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

with inverse

$$\begin{bmatrix} 1 & -q - q^4 - q^7 & q^3 + q^4 & -q^8 - 2q^9 - q^{10} \\ 0 & 1 & -q^2 - q^3 & q^7 + 2q^8 + q^9 \\ 0 & 0 & 1 & -q^5 - q^6 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The latter matrix is obtained from the matrix $\overline{\Psi}(w)$ in (6) by changing the signs of the entries in a checkerboard fashion as required by the definition of the alternate Parikh q-matrix encoding of w.

Remark 7. The Parikh matrix mapping and the Parikh q-matrix mapping are non-injective morphisms, and the latter is a refinement of the former [2]. The Parikh q-matrix encoding introduced in the present paper is injective, and reduces to the (numerical) Parikh matrix mapping of [6]. It is possible to construct weaker and non-injective versions of the Parikh q-matrix encoding in a number of ways. One is to set $q^{p+1} = q$ for some $p \in \mathbb{N}$ and keep track of the positions of the letters in w modulo p only. The encoding is then with upper-triangular matrices with entries over the quotient ring $\mathbb{Z}[q]/<q^{p+1}-q>$.

References

[1] Atanasiu, A., Martin-Vide, C., Mateescu, A.: On the injectivity of the Parikh matrix mapping. Fundamenta Informaticae, 49 (2002) 289-299.

[2] Eğecioğlu, Ö., and Ibarra, O.: A Matrix q-Analogue of the Parikh Map. UCSB Technical Report, TR 2004-06, February 2004.

[3] Harju, T., Ibarra, O., Karhumaki, J., Salomaa, A: Some decision problems concerning semilinearity and commutation. J. Computer and System Sciences, 65 (2002), 278-294.

[4] Ibarra, O.: Reversal-bounded multicounter machines and their decision problems. J. Assoc. Comput. Mach., 24 (1978) 123-137.

[5] Ibarra, O., Su, J., Dang, Z., Bultan, T., Kemmerer, R.: Counter machines and verification problems. Theoretical Computer Science, 289 (2002) 165-189.

[6] Mateescu, A., Salomaa, A., Salomaa K., Yu, S.: A sharpening of the Parikh mapping. Theoretical Informatics and Applications, 35 (2001) 551-564.

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

[8] Rozenberg, G., Salomaa, A. (eds): Handbook of Formal Languages. Springer, Berlin, 1997.

Sabanci University
Faculty of Engineering and Natural Sciences
Tuzla, 34956 Istanbul, Turkey
E-mail: omer@cs.ucsb.edu

On sabbatical leave from
Department of Computer Science, University of California,
Santa Barbara, CA 93106, USA