A Matrix q-Analogue of the Parikh Map

Ömer Eğecioğlu and Oscar H. Ibarra*

Department of Computer Science
University of California, Santa Barbara
CA 93106, USA
{omer,ibarra}@cs.ucsb.edu

Abstract. We introduce an extension of the Parikh mapping called the *Parikh q-matrix mapping*, which takes its values in matrices with polynomial entries. The morphism constructed represents a word w over a k-letter alphabet Σ as a k-dimensional upper-triangular matrix with entries that are nonnegative integral polynomials in variable q. We show that by appropriately embedding the k-letter alphabet into the $(k+1)$-letter alphabet and putting $q=1$, we obtain the extension of the Parikh mapping to $(k+1)$-dimensional (numerical) matrices introduced by Mateescu, Salomaa, Salomaa, and Yu. The Parikh q-matrix mapping however, produces matrices that carry more information about w than the numerical Parikh matrix. The entries of the q-matrix image of w under this morphism is constructed by *q-counting* the number of occurrences of certain words as scattered subwords of w.

1 Introduction

Parikh's theorem [6] says that every context-free language is "letter-equivalent" to a regular language. More precisely, the commutative image of any context-free language is always a semilinear set, and is therefore also the commutative image of some regular set. Consider the (ordered) alphabet $\Sigma_{\mathrm{k}} = \{\mathrm{a}_1 < \mathrm{a}_2 < \cdots < \mathrm{a}_{\mathrm{k}}\}$ 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

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

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

The Parikh mapping is a very important concept in the theory of formal languages. Various languages accepted (generated) by automata (grammars) more powerful than pushdown automata (context-free grammars) 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 (i.e., each counter can be incremented/decremented by one and tested for zero, but the number of alternations between nondecreasing and nonincreasing modes is bounded by a fixed constant) has a semilinear Parikh map [3]. The fact that the emptiness problem for semilinear sets is decidable implies that the

* Supported in part by NSF Grants IIS-0101134 and CCR02-08595.

emptiness problem for these automata (grammars) is decidable. This decidabilty of emptiness has been used to show the decidability of many decision questions in formal languages (e.g., [2]) and formal verification (e.g., [4]).

The *Parikh matrix mapping* introduced in [5] is a morphism

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

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.

The *Parikh q-matrix mapping* introduced in this paper is a morphism

$$\Psi_q^k \ : \ \Sigma^* \to \mathcal{M}_{\mathrm{k}}(\mathrm{q})$$

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. Moreover if we embed Σ_{k} into $\Sigma_{\mathrm{k+1}}$ in the obvious way, and put $q = 1$, then we obtain the matrices of the Parikh matrix map of [5]. Thus, viewing $w \in \Sigma_{\mathrm{k}}$ as a word in $\Sigma_{\mathrm{k+1}}$ 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. The injectivity of the Parikh matrix mapping is investigated in [1]. From our construction it is easy to see that 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.

The basic idea in the construction of the entries of the Parikh q-matrix image of w is *q-counting* the number of occurrences of certain words as scattered subwords of w.

The paper has five sections in addition to this section. Sections 2 gives some basic notation and definitions. Section 3 recalls the notion of a Parikh matrix mapping introduced in [5] and the fundamental theorem concerning these mappings. Section 4 presents our new Parikh mapping, called q-matrix mapping, that generalizes the Parikh matrix mapping: whereas the latter produces matrices with nonnegative integer entries, the former produces matrices with nonnegative integral polynomials (in variable q) entries. This extended mapping produces matrices that carry more information about the mapped words than the numerical matrices produced by the Parikh matrix mapping. Section 5 presents the main results, including Theorem 2, which gives the main properties of a q-matrix mapping. Section 6 looks at some matrix operations such as injectivity and inverse concerning q-matrix mapping.

2 Definitions

We start with some basic notation and definitions. Most of these are as they appear in references [5] and [1]. 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 $\Sigma = \{\mathrm{a_1}, \mathrm{a_2}, \ldots, \mathrm{a_k}\}$ with a relation of order ("<") on it. If for instance $a_1 < a_2 < \cdots < a_k$, then we use the notation

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

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 $|w|_{a_i}$. Thus $|w| = |w|_{a_1} + |w|_{a_2} + \cdots + |w|_{a_k}$.

Let $\Sigma = \{\mathrm{a_1} < \mathrm{a_2} < \cdots < \mathrm{a_k}\}$ be an ordered alphabet. The Parikh vector of $w \in \Sigma^*$ is the vector of occurrences $(|w|_{a_1}, |w|_{a_2}, \cdots, |w|_{a_k})$. The Parikh mapping

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

is defined by setting

$$\Psi(w) = (|w|_{a_1}, |w|_{a_2}, \cdots, |w|_{a_k}).$$

Let v, w be words over Σ. As defined in [5], 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 shall also find it useful to denote $|w|_{scatt-v}$ by $S_{w,v}$. Using this notation, we write $S_{acbb,ab} = 2$, $S_{acba,ab} = 1$, and $S_{w,a_i} = |w|_{a_i}$ 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 [5], 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 [7].

3 Parikh matrix mapping

We first describe the extension of the Parikh mapping to matrices as originally defined in [5]. 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 matrix I_k.

The main notion introduced in [5] is as follows:

Definition 1. *Let $\Sigma = \{\mathrm{a}_1 < \mathrm{a}_2 < \cdots < \mathrm{a}_\mathrm{k}\}$ be an ordered alphabet, where $k \geq 1$. The Parikh matrix mapping, denoted by $\Psi_{\mathcal{M}_k}$, is the morphism:*

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

defined as follows:

If $\Psi_{M_k}(a_l) = (m_{i,j})_{1 \leq i,j \leq (k+1)}$, 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 $\Psi_{M_k}(a_l)$ are zero.

Example 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. We compute some special cases: $\Psi_{\mathcal{M}_3}(ab^2) = \Psi_{\mathcal{M}_3}(a)\Psi_{\mathcal{M}_3}(b)\Psi_{\mathcal{M}_3}(b)$ and $\Psi_{\mathcal{M}_3}(abca) = \Psi_{\mathcal{M}_3}(a)\Psi_{\mathcal{M}_3}(b)\Psi_{\mathcal{M}_3}(c)\Psi_{\mathcal{M}_3}(a)$. Thus

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

and

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

and consequently

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

Remark 1. The Parikh matrix mapping is not an injective mapping. For instance over the ordered alphabet $\{a < b < c\}$ one has

$$\Psi_{\mathcal{M}_3}(acb) = \Psi_{\mathcal{M}_3}(cab) = \begin{bmatrix} 1&1&1&0\\ 0&1&1&0\\ 0&0&1&1\\ 0&0&0&1 \end{bmatrix}$$

Conditions for two words α and β to possess the same Parikh matrix was studied for the binary alphabet in [1]. We will discuss some of these conditions later in the paper.

The main property of the Parikh matrix mapping proved in [5] is the following theorem:

Theorem 1. *([5], Theorem 3.1) Let $\Sigma = \{\mathrm{a_1} < \mathrm{a_2} < \cdots < \mathrm{a_k}\}$ 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. $m_{i,j+1} = S_{w,a_{i,j}}$, *for all* $1 \leq i \leq j \leq k$.

As a corollary

Corollary 1. *([5], Corollary 3.1) Let $\Sigma = \{\mathrm{a_1} < \mathrm{a_2} < \cdots < \mathrm{a_k}\}$ The matrix $\Psi_{M_k}(w)$ has the second diagonal (i.e., the vector $(m_{1,2}, m_{2,3}, \ldots, m_{k,k+1})$) 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 q-counting scattered subwords

Next we introduce a collection of polynomials $S_{w,a_{i,j}}(q)$ indexed by pairs of words $a_{i,j}, w \in \Sigma^*$, with $j \leq k-1$. These polynomials will "q-count" the quantities $S_{w,a_{i,j}}$ defined above for general v and w as will be explained shortly in the case $a_{i,j}$ a scattered subword of w. To construct $S_{w,a_{i,j}}(q)$, we consider each factorization

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

with $u_s \in \Sigma^*$ for $i \leq s \leq j+1$, and construct the corresponding monomial

$$q^{|u_i|_{a_i} + |u_{i+1}|_{a_{i+1}} + \cdots + |u_j|_{a_j} + |u_{j+1}|_{a_{j+1}}} \tag{2}$$

in $\mathbb{N}[q]$, and add up these monomials. Note that $a_{j+1} \in \Sigma$ since $j < k$, so that the last term in the exponent in (2) is defined. Thus

$$S_{w,a_{i,j}}(q) = \sum_{w = u_i a_i \cdots u_j a_j u_{j+1}} q^{|u_i|_{a_i} + |u_{i+1}|_{a_{i+1}} + \cdots + |u_j|_{a_j} + |u_{j+1}|_{a_{j+1}}} \tag{3}$$

Example 2. Suppose $\Sigma = \{\mathrm{a} < \mathrm{b} < \mathrm{c} < \mathrm{d}\}$ and $i = 2, j = 2$. Then $a_{i,j} = b$ and for $w \in \Sigma^*$,

$$S_{w,b}(q) = \sum_{w = xby} q^{|x|_b + |y|_c}$$

For example for $w = baccbcdab$, the relevant factorizations of w are

$$(\lambda)b(accbcdab), \quad (bacc)b(cdab), \quad (baccbcda)b(\lambda),$$

and therefore

$$S_{w,b}(q) = q^{0+3} + q^{1+1} + q^{2+0} = 2q^2 + q^3.$$

Example 3. Suppose $\Sigma = \{\mathrm{a} < \mathrm{b} < \mathrm{c} < \mathrm{d}\}$ and $i = 2, j = 3$. Then $a_{i,j} = bc$ and for $w \in \Sigma^*$,

$$S_{w,bc}(q) = \sum_{w=xbycz} q^{|x|_b+|y|_c+|z|_d}$$

For example for $w = baccbcdab$, the relevant factorizations of w are

$$(\lambda)b(a)c(cbcdab), \quad (\lambda)b(ac)c(bcdab), \quad (\lambda)b(accb)c(dab), \quad (bacc)b(\lambda)c(dab),$$

and therefore

$$S_{w,bc}(q) = q^{0+0+1} + q^{0+1+1} + q^{0+2+1} + q^{1+0+1} = q + 2q^2 + q^3.$$

Example 4. Suppose $\Sigma = \{\mathrm{a} < \mathrm{b} < \mathrm{c} < \mathrm{d}\}$ and $i = 1, j = 3$. Then $a_{i,j} = abc$ and for $w \in \Sigma^*$,

$$S_{w,abc}(q) = \sum_{w=taxbycz} q^{|t|_a+|x|_b+|y|_c+|z|_d}$$

For example for $w = baccbcdab$, the only relevant factorization of w is

$$(b)a(cc)b(\lambda)c(dab),$$

and therefore

$$S_{w,abc}(q) = q^{0+0+1} = q.$$

Since the summation in the definition (3) is over all occurrences of $a_{i,j}$ in w as a scattered subword, the following proposition is immediate:

Proposition 1. *Let* $\Sigma = \{\mathrm{a}_1 < \mathrm{a}_2 < \cdots < \mathrm{a}_k\}$ *and* $1 \leq i \leq j < k$. *Then*

$$S_{w,a_{i,j}}(1) = S_{w,a_{i,j}} (= |w|_{scatt-a_{i,j}}).$$

This is the sense in which the polynomials $S_{w,a_{i,j}}(q)$ "q-count" the number of occurrences of $a_i a_{i+1} \cdots a_j$ as a scattered subword of w. These polynomials are the "q-analogues" of the numbers $S_{w,a_{i,j}}$.

5 Parikh q-matrix mapping

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. The matrix $\Psi_q(a_l)$ corresponding to a $a_l \in \Sigma$ is defined as the matrix obtained from I_k first by changing the l-th diagonal element from 1 to q. Then if $l < k$, we also change the entry immediately to the right of the q from 0 to a 1. Thus if $\Psi_q(a_l) = (m_{i,j})_{1 \leq i,j \leq k}$, then

1. $m_{l,l} = q$,
2. $m_{i,i} = 1$ for $1 \leq i \leq k, i \neq l$,
3. $m_{i,i+1} = 1$ if $l < k$,
4. all other entries of the matrix $\Psi_q(a_l)$ are zero.

When the alphabet is $\Sigma = \{\mathrm{a} < \mathrm{b} < \mathrm{c}\}$, then

$$\Psi_q(a) = \begin{bmatrix} q & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad \Psi_q(b) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & q & 1 \\ 0 & 0 & 1 \end{bmatrix}, \qquad \Psi_q(c) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & q \end{bmatrix}$$

We extend the mapping from Σ to Σ^* by setting

1. $\Psi_q(\lambda) = I_k$,
2. $\Psi_q(w_1 w_2 \cdots w_n) = \Psi_q(w_1)\Psi_q(w_2)\cdots\Psi_q(w_n)$, $w_i \in \Sigma, 1 \leq \mathrm{i} \leq \mathrm{n}$

We will refer to $\Psi_q = \Psi_q^k$ as the Parikh q-matrix mapping. Note that the parameter k, which is $|\Sigma|$ is implicit in our notation.

Remark 2. Just as the Parikh mapping is a morphism from the monoid $(\Sigma^*, \cdot, \lambda)$ to the monoid $(\mathbb{N}^k, +, (0, 0, \ldots, 0))$, the set of matrices $\mathcal{M}_k(q)$ is a monoid with respect to matrix multiplication and I_k as its unit.

Thus the Parikh q-matrix mapping is a morphism

$$\Psi_q \ : \ \Sigma^* \rightarrow \mathcal{M}_{\mathrm{k}}(\mathrm{q}).$$

As examples, we have $\Psi_q(ab^2) = \Psi_q(a)\Psi_q(b)\Psi_q(b)$ and $\Psi_q(abca) = \Psi_q(a)\Psi_q(b)\Psi_q(c)\Psi_q(a)$. Thus

$$\Psi_q(ab^2) = \begin{bmatrix} q & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & q & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & q & 1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} q & q^2 & 1+q \\ 0 & q^2 & 1+q \\ 0 & 0 & 1 \end{bmatrix}$$

$$\Psi_q(abca) = \begin{bmatrix} q & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & q & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & q \end{bmatrix} \begin{bmatrix} q & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} q^2 & 2q & q \\ 0 & q & q \\ 0 & 0 & q \end{bmatrix}$$

Consequently, for $w = ab^2abca$, we compute that

$$\Psi_q(w) = \begin{bmatrix} q & q^2 & 1+q \\ 0 & q^2 & 1+q \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} q^2 & 2q & q \\ 0 & q & q \\ 0 & 0 & q \end{bmatrix} = \begin{bmatrix} q^3 & 2q^2+q^3 & q+2q^2+q^3 \\ 0 & q^3 & q+q^2+q^3 \\ 0 & 0 & q \end{bmatrix}$$

Remark 3. For the Parikh q-matrix mapping it is not true 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 ([5], Remark 3.2).

Proposition 2. *Let $\Sigma = \{\mathrm{a}_1 < \mathrm{a}_2 < \cdots < \mathrm{a}_\mathrm{k}\}$ and $w \in \Sigma^*$. Then the vector of diagonal entries of the matrix $\Psi_q(w)$ is*

$$(q^{|w|_{a_1}}, q^{|w|_{a_2}}, \cdots, q^{|w|_{a_k}}) \in \mathbb{N}[q]^k.$$

Proof. The matrices $\Psi_q(a_l)$ are all upper-triangular. It is easy to see that the diagonal entries of a product of two upper-triangular matrices depend only on the diagonal elements of each of the matrices. Since diagonal matrices commute, and each occurrence of the letter a_l in w has the effect of multiplying the l-th diagonal entry of the k-dimensional identity matrix I_k by q, the result follows immediately.

Remark 4. We note that the Parikh vector of w is given by the formal derivative of

$$(q^{|w|_{a_1}}, q^{|w|_{a_2}}, \cdots, q^{|w|_{a_k}}) \in \mathbb{N}[q]^k$$

with respect to q at $q = 1$.

Theorem 2. *Let $\Sigma = \{\mathrm{a}_1 < \mathrm{a}_2 < \cdots < \mathrm{a}_\mathrm{k}\}$ be an ordered alphabet, where $k \geq 1$ and assume that $w \in \Sigma^*$. The matrix $\Psi_q(w) = (m_{i,j}(q))_{1\leq i,j\leq k}$, has the following properties*

1. $m_{i,j} = 0$, *for all* $1 \leq j < i \leq k$,
2. $m_{i,i} = q^{|w|_{a_i}}$, *for all* $1 \leq i \leq k$,
3. $m_{i,j+1} = S_{w,a_{i,j}}(q)$, *for all* $1 \leq i \leq j < k$.

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

$$\Psi_q(w) = \Psi_q(w')\Psi_q(a_j)$$

Assume that

$$\Psi_q(w') = \begin{bmatrix} q^{|w'|_{a_1}} & m'_{1,2} & \cdots & \cdots & m'_{1,k} \\ 0 & q^{|w'|_{a_2}} & \cdots & \cdots & m'_{2,k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & m'_{k-1,k} \\ 0 & 0 & \cdots & \cdots & q^{|w'|_{a_k}} \end{bmatrix} = M'$$

By the inductive hypothesis the matrix $\Psi_q(w')$ has property 3. The proof has two cases depending on whether $j = k$, or $j < k$. For $j < k$, we have

$$\Psi_q(a_j) = \begin{bmatrix} 1 & 0 & & \cdots & & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & q & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & & \cdots & & 1 \end{bmatrix}$$

where the matrix differs from I_k only in two entries: The entry in position (j,j) is q, and the entry in position $(j,j+1)$ is 1. Let $M = \Psi_q(w)$. Then

$$M = \begin{bmatrix} q^{|w'|_{a_1}} & m'_{1,2} & \cdots & \cdots & m'_{1,k} \\ 0 & q^{|w'|_{a_2}} & \cdots & \cdots & m'_{2,k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & m'_{k-1,k} \\ 0 & 0 & \cdots & \cdots & q^{|w'|_{a_k}} \end{bmatrix} \begin{bmatrix} 1 & 0 & & \cdots & & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 \ldots & & q & 1 & & \ldots 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & & \cdots & & 1 \end{bmatrix}$$

If $M = (m_{p,q})_{1 \le p,q \le k}$, then

$$m_{i,j} = q m'_{i,j} \ for\ 1 \le i \le j,$$
$$m_{i,j+1} = m'_{i,j} + m'_{i,j+1} \ for\ 1 \le i \le j$$

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

$$S_{w'a_j,a_{i,j-1}}(q) = qS_{w,a_{i,j-1}}(q) \ for\ 1 \le i \le j,$$
$$S_{w'a_j,a_{i,j}}(q) = S_{w',a_{i,j-1}}(q) + S_{w',a_{i,j}}(q) \ for\ 1 \le i < j$$

and are unchanged otherwise. In the case $j = k$ the only change that appears in going from M' to M is that the last column M is obtained from M' by multiplying the elements of the last column of M' by q. This corresponds to the fact that the number of occurrences of a_k in u_{k+1} in any factorization of the form (1) is increased by 1: i.e.,

$$S_{w'a_k,a_{1,k-1}}(q) = qS_{w',a_{1,k-1}}(q)$$

and the proof follows by induction.

Remark 5. The structure of how the polynomials in the matrix are indexed can be mnemonically recorded as shown below for the alphabet $\Sigma = \{\mathrm{a}_1 < \mathrm{a}_1 < \mathrm{a}_3 < \mathrm{a}_4\}$:

$$\begin{array}{c} a_1 \\ a_2 \\ a_3 \\ a_4 \end{array} \begin{bmatrix} q^{|w|_{a_1}} & a_1 & a_1a_2 & a_1a_2a_3 \\ 0 & q^{|w|_{a_2}} & a_2 & a_2a_3 \\ 0 & 0 & q^{|w|_{a_3}} & a_3 \\ 0 & 0 & 0 & q^{|w|_{a_4}} \end{bmatrix}$$

As an example, the entry in second row and the fourth column is a shorthand for the polynomial $S_{w,a_2a_3}(q)$, the q-count of the number of occurrences of a_2a_3 as a scattered subword of w as developed in section 4.

Proposition 3. *Let $\Sigma = \{\mathrm{a}_1 < \mathrm{a}_2 < \cdots < \mathrm{a}_k\}$ and $w \in \Sigma^*$. Suppose the vector of super diagonal entries of the matrix $\Psi_q(w)$ is*

$$(m_{1,2}(q), m_{2,3}(q), \cdots, m_{k-1,k}(q)) \in \mathbb{N}[q]^{k-1}.$$

Then at $q = 1$, this vector evaluates to

$$(q^{|w|_{a_1}}, q^{|w|_{a_2}}, \cdots, q^{|w|_{a_{k-1}}}).$$

Proof. This proposition is a special case of a stronger result that characterizes the whole matrix $\Psi_q(w)$ at $q = 1$ that we give as Theorem 3.

Theorem 3. *Suppose* $\Sigma = \{\mathrm{a}_1 < \mathrm{a}_2 < \cdots < \mathrm{a}_\mathrm{k}\}$ *and* $w \in \Sigma^*$. *Consider* w *as a word over* $\Gamma = \{a_1 < a_2 < \cdots < a_k < a_{k+1}\}$ *and let* $\Psi_q(w)$ *be the resulting Parikh* q*-matrix in* $\mathbb{N}[q]^{k+1}$. *Then* $\Psi_q(w)$ *evaluated at* $q = 1$ *is the Parikh matrix* $\Psi_{\mathcal{M}_k}(w)$.

Proof. Combine Theorem 2, Theorem 1, and Proposition 1.

6 Injectivity, inverse, and further remarks

Just as the Parikh matrix mapping, the Parikh q-matrix mapping is not an injective mapping either. For instance over the ordered alphabet $\{a < b < c\}$ one has

$$\Psi_q(acb) = \Psi_q(cab) = \begin{bmatrix} q & q & 1 \\ 0 & q & 1 \\ 0 & 0 & q \end{bmatrix}$$

However, there are instances in which two words can have the same Parikh matrix, but distinct Parikh q-matrices.

The injectivity of the Parikh matrix mapping was studied in [1]. In particular it was proved that over a binary alphabet Σ, a pair of *palindromic amiable* words α, β have the same Parikh matrix image. The definition of palindromic amiable pair is as follows:

1. Both α and β are palindromes,
2. α and β have the same Parikh vector, i.e., $\Psi(\alpha) = \Psi(\beta)$.

For example the words $\alpha = aba^2ba$ and $\beta = ba^4b$ over the alphabet $\Sigma = \{\mathrm{a} < \mathrm{b}\}$ are palindromic amiables. Therefore as proved in [1], they have the same 3×3 Parikh matrix image. We calculate directly that indeed

$$\Psi_{\mathcal{M}_2}(\alpha) = \begin{bmatrix} 1 & 4 & 4 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} = \Psi_{\mathcal{M}_2}(\beta) \tag{4}$$

The corresponding matrices given by the Parikh q-matrix mapping Ψ_q are calculated over the alphabet $\{a < b < c\}$ in accordance with Theorem 3. These are also 3×3 upper-triangular matrices, but with entries from $\mathbb{N}[q]$. They are given by

$$\Psi_q(\alpha) = \begin{bmatrix} q^4 & 2q^2 + 2q^3 & 1 + 2q + q^2 \\ 0 & q^2 & 1 + q \\ 0 & 0 & 1 \end{bmatrix} \tag{5}$$

$$\Psi_q(\beta) = \begin{bmatrix} q^4 & q + q^2 + q^3 + q^4 & 1 + q + q^2 + q^3 \\ 0 & q^2 & 1 + q \\ 0 & 0 & 1 \end{bmatrix} \tag{6}$$

Clearly, these two distinct matrices reduce to $\Psi_{\mathcal{M}_2}(\alpha) = \Psi_{\mathcal{M}_2}(\beta)$ given in (4) as guaranteed by Theorem 3. Thus the matrices obtained by the Parikh q-matrix mapping contains finer information that is able to distinguish words that are equal under the ordinary Parikh matrix map.

The notion of the alternate (signed) Parikh matrix developed in [5] 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. This property also carries over to the case of the Parikh q-matrix mapping with some modifications. Let $\Sigma = \{\mathrm{a}_1 < \mathrm{a}_2 < \cdots < \mathrm{a}_\mathrm{k}\}$. We define a morphism (called the *alternate Parikh q-matrix mapping*) $\overline{\Psi}_q = \overline{\Psi}_q^k$ from Σ^* to a collection of k-dimensional upper-triangular matrices over $\mathbb{Z}[q]$. $\overline{\Psi}_q$ is defined on Σ as follows: If $\overline{\Psi}_q(a_l) = (m_{i,j})_{1\le i,j\le k}$, then

1. $m_{l,l} = 1$,
2. $m_{i,i} = q$ for $1 \le i \le k, i \ne l$,
3. $m_{i,i+1} = -1$ if $l < k$,
4. all other entries of the matrix $\Psi_q(a_l)$ are zero.

Example 5. When the alphabet is $\Sigma = \{\mathrm{a} < \mathrm{b} < \mathrm{c}\}$, then

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

Note that $\Psi_q(a)\overline{\Psi}_q(a) = \Psi_q(b)\overline{\Psi}_q(b) = \Psi_q(c)\overline{\Psi}_q(c) = qI_3$. As an example, for $w = ab^2abca$, we compute that

$$\overline{\Psi}_q(mi(w)) = \begin{bmatrix} q^4 & -2q^3 - q^4 & 2q^3 + 2q^4 + q^5 \\ 0 & q^4 & -q^4 - q^5 - q^6 \\ 0 & 0 & q^6 \end{bmatrix}$$

Then the following result holds (stated without proof in this extended abstract):

Theorem 4. *Suppose $\Sigma = \{\mathrm{a}_1 < \mathrm{a}_2 < \cdots < \mathrm{a}_\mathrm{k}\}$ and $w \in \Sigma^*$. If Then Ψ_q and $\overline{\Psi}_q$ are the Parikh q-matrix, and the alternate Parikh q-matrix mappings from Σ^* to upper-triangular integral matrices over $\mathbb{Z}[q]$, then the k-dimensional matrix identity*

$$\Psi_q(w)\overline{\Psi}_q(mi(w)) = \begin{bmatrix} q^{|w|} & 0 & \cdots & & 0 \\ 0 & q^{|w|} & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & & 0 & q^{|w|} \end{bmatrix} \tag{7}$$

holds.

It can also be shown that the identity (7) reduces to the matrix inverse identity of the Parikh matrix mapping of [5] when we extend the alphabet as in Theorem 3 and put $q = 1$.

References

1. Atanasiu, A., Martin-Vide, C., Mateescu, A.: On the injectivity of the Parikh matrix mapping. Fundamenta Informaticae, 49 (2002) 289-299.
2. Harju, T., Ibarra, O., Karhumaki, J., Salomaa, A: Some decision problems concerning semilinearity and commutation. J. Computer and System Sciences, 65 (2002), 278-294.
3. Ibarra, O.: Reversal-bounded multicounter machines and their decision problems. J. Assoc. Comput. Mach., 24 (1978) 123-137.
4. Ibarra, O., Su, J., Dang, Z., Bultan, T., Kemmerer, R.: Counter machines and verification problems. Theoretical Computer Science, 289 (2002) 165-189.
5. Mateescu, A., Salomaa, A., Salomaa K., Yu, S.: A sharpening of the Parikh mapping. Theoretical Informatics and Applications, 35 (2001) 551-564.
6. Parikh, R.J.: On context-free languages. J. Assoc. Comput. Mach., **13** (1966) 570-581.
7. Rozenberg, G., Salomaa, A. (eds): Handbook of Formal Languages. Springer, Berlin, 1997.