PRISM COMPLEXITY OF MATRICES

Antal Iványi (Budapest, Hungary)
Zoltán Kása (Târgu Mureș, Romania)

Dedicated to Professor K.-H. Indlekofer
on occasion of his 70th birthday

Communicated by Imre Kátai
(Received December 30, 2012; accepted January 17, 2013)

Abstract. Let $d$, $m$, and $q$ be positive integers and $A(q) = \{0, \ldots, q-1\}$ be an alphabet. We investigate a generalization of the well-known subword complexity of $d$-dimensional matrices containing the elements of $A(q)$. Let $L = (L_1, \ldots, L_m)$ be a list of distinct $d$-dimensional vectors, where $L_i = (a_{i1}, \ldots, a_{id})$. The prism complexity of a $d$-dimensional $q$-ary matrix $M$ is denoted by $C(d, q, L, M)$ and is defined as the number of distinct $d$-dimensional $q$-ary submatrices, whose permitted sizes are listed in $L$. We review and extend the earlier results, first of all results concerning maximum complexity of matrices and performance parameters of the construction algorithms.

1. Introduction

Let $d$, $m$, $n$, and $q$ be positive integers, $A(q) = \{0, \ldots, q-1\}$ be an alphabet, $\varepsilon$ be the empty matrix, $A(q, d)^*$ be the set of $d$-dimensional $q$-ary matrices, $A(q, d)^+$ be the set of nonempty $d$-dimensional $q$-ary matrices. Let $L = (L_1, \ldots, L_m)$ be a list of $d$-dimensional vectors, where $L_i$ gives the size of an $a_{i1} \times \cdots \times a_{id}$ sized submatrix of $M$.

The $(q, d, L)$-complexity (or shortly prism complexity) $C(q, d, L, M)$ of a matrix $M$ is defined as follows.

Key words and phrases: Subword complexity, $d$-complexity, scattered complexity, prism complexity, complexity of arrays, De Bruijn graphs.

2010 Mathematics Subject Classification: 68R15.
Definition 1.1. Let $q$, $d$, and $L$ be given. The $(q, d, L)$-complexity of a given $d$-dimensional $q$-ary matrix $M$ is denoted by $C(q, d, L, M)$ and is defined as the number of distinct submatrices (containing neighboring rows and neighboring columns) of $M$ whose permitted sizes are given by $L$, that is

\begin{equation}
C(q, d, L, M) = \sum_{i=1}^{m} f(q, d, L_i, M),
\end{equation}

where $f(q, d, L_i, M) = |S(q, d, L_i, M)|$ and $S(q, d, L_i, M)$ is the set of the distinct $a_{i1} \times a_{i2} \times \cdots \times a_{id}$ sized submatrices of $M$.

For example if $q = 3$, $d = 2$, $L = \{(2, 2), (2, 3)\}$, and $M = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix}$, then $C(q, d, L, M) = 5$, since $M$ contains the following five submatrices having permitted size:

\begin{equation}
(1.3) \begin{pmatrix} 0 & 1 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix}.
\end{equation}

We remark, that sometimes $M$ is considered as a periodic matrix. In this case $C(q, d, L, M) = 12$, since there are further submatrices: four $2 \times 2$ sized matrices

\begin{equation}
(1.4) \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 2 & 0 \end{pmatrix}, \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix},
\end{equation}

and three $3 \times 3$ sized submatrices

\begin{equation}
(1.5) \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 0 \\ 1 & 0 & 0 \end{pmatrix}.
\end{equation}

If a matrix consists of distinct elements of $A(q)$, then it is called a rainbow matrix.

In this paper we deal with some special cases of this new combinatorial complexity measure of $q$-ary matrices—first of all with the characterization of extremal values and construction of extremal matrices.

The structure of the paper is the following. After this introductory Section 1 one-dimensional matrices (words) are considered in Section 2. Then, in Section 3 some construction algorithms of words having maximum subword complexity are presented. In Section 4 a further one-dimensional complexity measure ($d$-complexity) is analyzed. In Section 5 we deal with prism complexity of two-dimensional rectangular matrices.
2. Subword complexity of words

If \( n \geq 1 \), then we consider the \( 1 \times n \) sized matrices as one-dimensional matrices and call words or sequences. In this and the further three sections we deal with the complexity of words.


Many papers deal with special combinatorial complexity measures of matrices (especially with \( 1 \times n \) sized matrices). For example arithmetical (Frid [42, 43, 44]), d (Iványi [69], Kása [80, 81]), factor (Ilie [60]), I (Becher and Heiber [13], Kreinovich and Nava [87]), inconstancy (Allouche and Maillard-Teyssier [6]), Lempel–Ziv (Constantinescu and Ilie [21]), joint subword (Jacquet and Szpankowski [75]), linguistic (Popov, Segal, and Trifonov [102], Troyanskaya, Arbell, Koren, Landau, and Bolshoy [114]), palindromic (Allouche, Baake, Cassaigne, and Damanik [5], Anisiu, Anisiu and Kása [7]), pattern (Kamae and Salimov [78], Qu, Rao, Wen, and Xie [103]), repetition (Crochemore and Iliev [23], Ilie, Yu, and Zhang [64]), scattered substring (Okhotin [98]), scattered subword (Fazekas and Nagy [36], Gruber, Holzer and Kutrib [48], Kása [82, 85], Okhotin [98]), square (Ilie [61]), subarray (Iványi [70, 71, 72], Iványi and Tóth [74, 86], Ma [94], MacWilliams and Sloane [95], Siu [109], van Lint, MacWilliams and Sloane [116]), subsequence (Apostolico and Cunial [11], Kuliamin [88]), substring (Elzinga [31], Ilie and Smith [63]), subword (Ehrenfeucht and Rozenberg [28, 29, 30], Ivanko [67, 68], Lempel and Ziv [89], Rote [106]), and word complexity (Ilie, Yu and Zhang [65]).

In this paper we deal first of all with the above defined prism complexity and its special cases.

In the construction of the matrices having maximum complexity the De Bruijn graphs and their generalizations play important role (these graphs are intensively studied also as possible models of networks, therefore there are many connected papers, e.g. [1, 2, 4, 8, 12, 13, 14, 17, 19, 26, 32, 34, 35, 37, 39, 40, 41, 46, 49, 50, 54, 55, 56, 57, 59, 69, 70, 71, 72, 73, 74, 80, 81, 84, 86, 90, 94, 107, 108, 111, 113, 114, 118].

Let \( A(q, 1)^n = A(q)^n \) be the language (set) of all \( n \)-length words (sequences) \( w = w_1 \ldots w_n \) over \( A(q) \), \( A(q, 1)^+ = A(q)^+ \) be the language of nonempty words over \( A(q) \).

A subword of a \( q \)-ary word \( w = w_1 \ldots w_n \in A(q)^n \) is defined as a contiguous part \( w_i \ldots w_j \) (\( 1 \leq i \leq j \leq n \)) of \( w \). We remark that this definition corresponds to \( L = \{(1, 1), \ldots, (1, n)\} \).
A $k$-length subsequence of $w$ is defined as $w_{j_1} \ldots w_{j_k}$, where $1 \leq j_1 < \cdots < j_k \leq n$. According to these definitions the empty matrix (word) $\varepsilon$ is neither subword nor subsequence.

Hunyadvári and Iványi proposed the following join generalization of the subword and subsequence complexity.

**Definition 2.1.** (Hunyadvári, Iványi 1984 [54, 55, 69]) Let $d, r$ and $s$ be positive integers, $u = u_1 \ldots u_r$ and $v = v_1 \ldots v_s$ be elements of $A(q)^+$. $u$ is a $d$-subword of $(u \subseteq_d v)$ if there exists a sequence $j_1, \ldots, j_r$ with $1 \leq j_1 < \cdots < j_r \leq s$ and $j_{i+1} - j_i \leq d$ for $i = 1, \ldots, s - 1$ such that $u_i = v_{j_i}$ for $i = 1, \ldots, r$. If for given $d$, $r$ and $s$ there exist several such sequences then the lexicographically smallest one belongs to the given $d$, $r$ and $s$.

The $d$-complexity of a word $w \in A(q)^n$ is the number of its distinct $d$-subwords which can be computed as the sum of the multiplicities of the distinct 1-, $\ldots$, $n$-length $d$-subwords of $w$. We remark that in this section $d$ means the distance parameter of the complexity, while in the other sections $d$ means the number of dimensions of the considered matrices.

**Definition 2.2.** (Hunyadvári, Iványi 1984 [54, 55, 69]) For given $w \in A(q)^n$ the $d$-complexity $C(q, d, w)$ of $w$ is

$$C(q, d, w) = \sum_{i=1}^{n} f(q, d, w, i),$$

where $f(q, d, w, i) = |S(q, d, w, i)|$, $S(q, d, w, i) = S(u \subseteq_d v) \cap A(q)^i$ for $i = 1, \ldots, n$.

If in Definition 2.1 and Definition 2.2 $d = 1$, then we get the usual definitions of subword, resp. subword complexity introduced by Morse and Hendlund in 1938 [97] and redefined later by others [27, 49]. It is worth to remark that some authors (as Shallit [108], and Flaxman, Harrow, Sorkin [38]) count the empty word too as a subword.

In some cases we will suppose that the investigated matrices are periodic.

Let $m(q, n)$ denote the maximum number of 1-subwords of $w \in A(q)^n$. $w$ is called $d$-complex if $C(q, 1, w) = m(q, n)$. An infinite word $w = w_1 w_2 \ldots$ is called 1-supercomplex if $C(q, 1, w^{(k)}) = m(q, n)$ for all prefixes $w^{(k)} = w_1, \ldots, w_k$ ($k = 1, 2, \ldots$) of $w$.

In 1984 Nóra Vörös [117] gave the following bounds.
Theorem 2.1. (Vörös [117]) If \( n \) and \( q \) are positive integers and \( w \in A(q)^n \), then
\[
(2.2) \quad n \leq C(q, 1, w) \leq \sum_{i=1}^{n} \min(q^i, n - i + 1)
\]
and the bounds are sharp.

Proof. See [9, 13, 37, 117].

(2.2) was reproved by Ferenczi and Kása [37] in 1999, by Anisiu and Cassaigne in 2004 [9], by Becher and Heiber in 2012 [13]. Theorem 2.1 was reproved in the case \( q = 2 \) by Shallit in 1993 [108].

In 1984 Nóra Vörös proved the following sufficient and necessary condition for the existence of supercomplex words.

Theorem 2.2. (Vörös [117]) If \( n \geq 1 \) then there exists a \( 1 \)-supercomplex word if and only if \( q = 1 \) or \( q \geq 3 \).

Proof. See [12, 24, 38, 69, 117].

This assertion in 1986 was reproved by Cummings and Wiedemann [24], in 1987 by Iványi [69], in 2004 by Flaxman, Harrow and Sorkin [38], and in 2011 by Becher and Heiber [12].

In 1993 Shallit proved the following closed form for the maximum \( 1 \)-subword complexity of binary sequences.

Theorem 2.3. (Shallit [108]) If \( n \) is a positive integer then
\[
(2.3) \quad m(2, n) = \left( \frac{n - k + 1}{2} \right)^2 + 2^{k+1} - 2,
\]
where \( k \) is the unique integer such that \( 2^k + k - 1 \leq n < 2^{k+1} + k \).

Proof. See [37, 108].

In 1999 Ferenczi and Kása gave the following closed upper bound for \( m(q, n) \).

Theorem 2.4. (Ferenczi, Kása [37]) If \( q \geq 3 \) then
\[
(2.4) \quad m(q, n) \leq \left( \frac{n - k + 1}{2} \right)^2 + q^{k+1} - 1,
\]
where \( k \) is the unique integer such that \( q^k + k - 1 \leq n < q^{k+1} + k \), further
\[
(2.5) \quad m(q, n) = \frac{n^2}{2} - \Omega(n \log n).
\]

Proof. See [37].
In 2004 Anisiu and Cassaigne analyzed the 1-complexity function

\[ h(n, q, i) = \min(q^i, n - i - 1) \quad \text{for } i = 1, \ldots, n \]

and proved the following theorem.

**Theorem 2.5.** (Anisiu, Cassaigne [9]) If \( n \) and \( q \) are positive integers then there exists a word \( w = w_1 \ldots w_n \in A(q)^n \) such that

\[ C(q, n, w) = m(q, n) = \sum_{i=1}^{n} h(n, q, i). \]

**Proof.** See [9]. \( \blacksquare \)

In 2004 Flaxman, Harrow and Sorkin proved the following upper bound of \( m(q, n) \).

**Theorem 2.6.** (Flaxman, Harrow, Sorkin [38]) If \( q \geq 3 \) and \( n \geq 1 \) are positive integers, then

\[ m(q, n) \leq \left( \frac{n - k + 1}{2} \right) + \frac{q^{k+1} - 1}{q - 1} - 1, \]

where \( k = \lfloor \log_q n \rfloor \).

**Proof.** See [38]. \( \blacksquare \)

Higgins in 2012 published the following upper bound for \( m(q, n) \).

**Theorem 2.7.** (Higgins, [50]) If \( q \) and \( n \) are positive integers, then

\[ m(q, n) = \frac{n^2}{2} - O(n \log n). \]

**Proof.** See [50]. \( \blacksquare \)

The previous seven theorems (Theorem 2.1, 2.2, \ldots, 2.7) are consequences of the following new theorem.

**Theorem 2.8.** If \( q \) and \( n \) are positive numbers, then

\[ m(q, n) \leq \sum_{i=1}^{n} \min(q^i, n - i + 1) \]

and

\[ \sum_{i=1}^{n} \min(q^i, n - i + 1) = \left( \frac{n - k + 1}{2} \right) + \frac{q^{k+1} - 1}{q - 1} - 1, \]
where \( k \) is the unique integer such that \( q^k + k - 1 \leq n < q^{k+1} + k \) and

\[
m(q, n) = \frac{n^2}{2} - n \log_q n + n + \Theta(\log^2 n),
\]

further if \( q \neq 2 \) then there exists an infinite word \( w = w_1w_2\ldots \) whose prefixes \( w^{(k)} = w_1\ldots w_k \) have the 1-subword complexity \( C(q, 1, w^{(k)}) = m(q, k) \).

**Proof.** The proof of (2.10) can be found e.g. in [54, 117].

\[
q^k \leq n - k + 1 \quad \text{if and only if} \quad k \text{ is the unique integer such that } q^k + k - 1 \leq n < q^{k+1} + k.
\]

Therefore (2.10) implies

\[
m(q, n) = q + \cdots + q^k + (n - k) + \cdots + 1.
\]

Since

\[
q + \cdots + q^k = \frac{q^k - 1}{q - 1},
\]

and

\[
1 + \cdots + n - k = \binom{n-k}{2},
\]

(2.13), (2.14) and (2.15) imply (2.11).

In 1999 in the paper [37] appeared the following assertion: if \( n \) and \( q \) are positive integers, then

\[
m(q, n) \leq \frac{(n-k)(n-k+1)}{2} + q^{k+1} - 1
\]

where \( k \) is the unique integer such that \( 2^k + k - 1 \leq n < 2^{k+1} + k \).

In 2004 in [38] the following assertion was published: if \( n \) and \( q \) are positive integers, then

\[
m(q, n) = \frac{(n-k)(n-k+1)}{2} + \frac{q^{k+1} - 1}{q - 1} - 1
\]

where \( k = \lfloor \log n \rfloor \).

The following example shows that (2.16) and (2.17) are not exact. Let \( q = 3 \) and \( n = 2 \). The word \( w = (01) \) has maximal complexity \( |\{(0, 1, 01)\}| = 3 \). In this case (2.16) results 9 and (2.17) gives 4 while the word \( w = (0, 1) \) has maximal complexity and so \( |\{(0, 1, 01)\}| = 3 \) is the correct value.

In the following we describe further interesting results.
In 1988 Sridhar [111] proved that De Bruijn graphs $B(q, n)$ are $q$-connected. In 1991 Li and Zhang [91] counted the number of spanning trees and Eulerian tours. Blażewicz, Formanowicz Kasprzak and Kobler [15] in 2002 proposed a polynomial algorithm to decide whether a directed graph is a De Bruijn graph or the subgraph of a De Bruijn graph with given $a$ (length of vertex names).

Flaxman, Harrow and Sorkin in 2004 [38] proved bounds for average 1-complexity and subsequence complexity. Szpankowski gave a more detailed analysis of average 1-complexity in 2011 [113]. Ivanko also investigated the average 1-complexity [67, 68].

3. Construction of De Bruijn words

In this section several construction algorithms of De Bruijn words are presented.

**Definition 3.1.** If $q \geq 1$ and $n \geq 1$ then the $(q, a)$-type De Bruijn word is defined so that it contains all possible $q$-ary words $w \in A(q)^n$ exactly once as a 1-subword.

This definition implies the length of the $(q, a)$-type nonperiodic and periodic De Bruijn words.

**Corollary 3.1.** The length of a $(q, n)$-type nonperiodic De Bruijn word is $q^n + a - 1$ and the length of a $(q, n)$-type periodic De Bruijn word is $q^n$.

The first known proof of the existence of $(2, a)$-type De Bruijn words appeared in 1894 and was published by T. Flye-Sante Marie [39]. This assertion was proved again many times afterwards, e.g. in 1946 independently by Good [47] and by De Bruijn [25]. For example Fredricksen, Kessler and Maiorana, Etzion proposed construction algorithms for binary and later $q$-ary De Bruijn words [34, 40, 41, 104].

3.1. Algorithm Martin

Generating De Bruijn words is a common task with respectable number of algorithms. Let $q \geq 2$ and $A(q) = \{0, \ldots, q - 1\}$ be an alphabet. Our goals are to generate from one side a $(q, k)$-type De Bruijn word, and also to generate all $(q, k)$-type De Bruijn words for given $q$ and $k$.

We present here a natural version Natural-Martin [84] of the classical Martin algorithm [96].

We begin with the word $0^n$, and add at its right end the greatest possible letter, such that the suffix of length $a$ of the obtained word does not duplicate
a previously occurring subword of length \( a \). The algorithm repeats this until such a prolongation is impossible.

When we cannot continue, a nonperiodic De Bruijn word is obtained, with the length \( q^a + a - 1 \). In the following algorithm the input is \( q \), the size of the alphabet and \( a \), the pattern size. The output is \( w = w_1, \ldots, w_{q^a + a - 1} \), a \((q,a)\)-type nonperiodic De Bruijn word. The working variables are \( l \), which is a logical variable, signalizing whether the last suffix is a new subword or not, and the cycle variables \( i \) and \( k \).

The pseudocodes are written according to the conventions described in [22].

Natural-Martin\((q,a)\)

01 for \( i \leftarrow 1 \) to \( a \)  
02 \( w_i \leftarrow 0 \)  
03 \( i \leftarrow a \)  
04 repeat  
05 \( l \leftarrow TRUE \)  
06 \( j \leftarrow q \)  
07 while \( j \geq 1 \)  
08 \( \text{if } b_{i-j+2}w_{i-j+3} \ldots w_i(j-1) \not\subset w_1w_2 \ldots w_i \)  
09 \( i \leftarrow i + 1 \)  
10 \( w_i \leftarrow j - 1 \)  
11 \( l \leftarrow FALSE \)  
12 else \( j \leftarrow j - 1 \)  
13 until \( l = TRUE \)  
14 return \( w \)

Because this algorithm generates all elements of a De Bruijn sequence of length \( q^a + a - 1 \), further \( q \) and \( a \) are independent, the time complexity Natural-Martin is \( \Omega(q^a) \). The more precise characterization of the running time depends on the implementation of line 08. The repeat statement is executed \( q^a - 1 \) times. The while statement is executed at most \( q \) times for each step of the repeat. The test \( w_{i-a+2}w_{i-a+3} \ldots w_iw_k \not\subset w_1w_2 \ldots w_i \) can be made in the worst case in \( aq^a \) steps. So, the total number of steps is not greater than \( aq^{2a+1} \), resulting the worst case bound \( O(q^{a+1}) \). If we use the Knuth-Morris-Pratt string matching algorithm [22], then the worst case running time is \( O(q^{2a}) \).

3.2. Algorithm Quick-Martin

Algorithm Quick-Martin also generates one-dimensional perfect arrays (De Bruijn words). Its inputs are the alphabet size \( q \) and the window size \( a \). Its
output is a $q$-ary perfect sequence of length $q^a$. The output begins with $a$ zeros and always continues with the maximal permitted element of the alphabet.

The following effective implementation of Martin algorithm is due to Horváth and Iványi [53].

**Quick-Martin($q, a$)**

```plaintext
01 for $i = 0$ to $q^a - 1$ // line 01–02: initialization of $C$
02 $C[i] = q - 1$
03 for $i = 1$ to $a$ // line 03–04: initialization of $w$ and $k$
04 $w[i] = 0$
05 for $i = a + 1$ to $q^a$ // line 05–11: generation of $w$
06 $k = w[i - a + 1]$
07 for $j = 1$ to $a - 1$
08 $k = kq + w[i - a + j]$
09 $w[i] = C[k]$
10 $C[k] = C[k] - 1$
11 return $w$
```

This algorithm runs in $\Theta(aq^a)$ time.

### 3.3. Algorithm Optimal-Martin

The following implementation [72] of Martin algorithm requires even smaller running time than Quick-Martin.

The **input** of Opt-Martin is $q$: the size of the alphabet; and $a$: the length of the pattern. The **output** is $w = w_1 \ldots w_{q^a}$. The **working variables** are $C = C[0], \ldots, C[q^a - 1]$: the counters belonging to the vertices of the De Bruijn graph; $k$: the decimal value of the label of the current vertex of the De Bruijn graph. $i$ cycle variable.

**Optimal-Martin($q, a$)**

```plaintext
01 for $i = 0$ to $q^a - 1$
02 $C[i] = q - 1$
03 for $i = 1$ to $a$
04 $w[i] = 0$
05 $k = 0$
06 for $i = a + 1$ to $q^a$
07 $k = q(k - w[i - a]q^{a-2}) + w[i - 1]$
08 $w[i] = C[k]$
09 $C[k] = C[k] - 1$
10 return $w$
```
The running time of any algorithm which constructs a one-dimensional cyclical perfect array is $\Omega(q^a)$, since the sequence contains $q^a$ elements. The running time of $\textsc{Optimal-Martin}$ is $\Theta(q^a)$.

4. Scattered complexity of words

Scattered subwords of a word were defined by Kása in [82] as follows. This definition is not a special case of the general definition of prism complexity, since here the list $J$ contains the permitted differences of the indices of the choosed letters, that is $J \subseteq \{1, \ldots, n-1\}$.

**Definition 4.1.** Let $n$ and $q$ be positive integers, $J \subseteq \{1, \ldots, n-1\}$ and $u = x_1 \ldots x_n \in A(q)^n$. A $J$-subword of length $s$ of $u$ is defined as $v = x_{i_1} \ldots x_{i_s}$, where

- $i_1 \geq 1$,
- $i_{j+1} - i_j \in J$ for $j = 1, \ldots, s-1$,
- $i_s \leq n$.

Using Definition 4.1 we can formalize the concept of the scattered subword complexity.

**Definition 4.2.** For given $J$ the scattered subword complexity (shortly $J$-complexity) of a word $u \in A(q)^n$ is the number of $J$-subwords of $u$.

In the case $1 \leq d \leq n-1$ and of $J = \{1, \ldots, d\}$ the $J$-subword is the $d$-subword defined in [69], while in the case $J = \{d, \ldots, n-1\}$ is the super-$d$-subword defined in [83]. The corresponding $d$-complexity and super-$d$-complexity are similarly defined.

The scattered subword complexity for rainbow words can be easily computed by a graph method [82]. The letters $x_i$ of the rainbow word are the vertices of the graph, and two vertices, $x_i$ and $x_j$, are joined by an arc from $x_i$ to $a_j$ if these letters can be neighbors in this order in a scattered subword. The scattered subword complexity is equal to the number of directed paths in this attached graph (here each vertex is considered as a null length path). The number of directed paths in this graph with $n$ vertices can be computed by a Floyd-Warshall type algorithm with worst case complexity $\Theta(n^3)$ [82]. If $1 \leq d_1 \leq d_2 \leq n-1$, and the list $J$ is $\{d_1, d_1 + 1, \ldots, d_2\}$, then the scattered subword complexity (the so called $(d_1, d_2)$-complexity [84]) can be computed by a linear algorithm [85].

This method combined with the classical Latin square method yields an algorithm by which even the scattered subwords can be obtained [82].
(\(d_1, d_2\))-subwords are related to compositions of integers. Compositions are partitions in which the order of the components does matter. A \((d_1, d_2)\)-composition is a restricted composition in which the components are natural numbers from the interval \([d_1, d_2]\).

For example, for the word \(abcdefg\) the \((2,4)\)-subwords, which begin in \(a\) and end in \(g\) are: \(aeg\), \(aceg\), \(adg\), \(acg\), which correspond to the following compositions in which the components are the distances between the letters in the original word:

\[ 6 = 4 + 2 = 2 + 2 + 2 = 3 + 3 = 2 + 4. \]

In general, if \(a_1a_2\ldots a_{n+1}\) is a \((d_1, d_2)\)-subword of the rainbow word \(a_1a_2\ldots a_{n+1}\), then this subword corresponds to a composition:

\[ n = (i_1 - 1) + (i_2 - i_1) + \cdots + (i_s - i_{s-1}) + (n + 1 - i_s). \]

Definition of the \((d_1, d_2)\)-subword can be generalized for rainbow words as we choose letters not only going ahead in the word, but back too, at every step [85].

**Definition 4.3.** Let \(n, d_1 \leq d_2, q, \) and \(s\) be positive integer numbers, and let \(u = x_1\ldots x_n \in A(q)^n\) be a rainbow word over the alphabet \(A(q)\). A rainbow word \(v = x_{i_1}\ldots x_{i_s}\), where

- \(i_1 \geq 1,\)
- \(d_1 \leq |i_{j+1} - i_j| \leq d_2, \) for \(j = 1, \ldots, s - 1,\)
- \(i_s \leq n,\) is an \(s\)-length duplex \((d_1, d_2)\)-subword of \(u.\)

For example \(acfbc\) and \(beadfc\) both are duplex \((2,4)\)-subwords, and the duplex subwords are rainbow words too, of the word \(abcdef.\)

The number of all duplex \((d_1, d_2)\)-subwords of a word is the *duplex \((d_1, d_2)\)-complexity* of that rainbow word.

We remark that the term *scattered subword complexity* was used earlier for example by Fazekas and Nagy in 2008, further by Gruber, Holzer and Kutrib in 2009 [48], but they defined the scattered complexity of languages, while Kása [83] defined the scattered complexity of words.


Flaxman, Harrow and Sorkin in 2004 [38], Szpankowski in 2001 [113] characterized the average subsequence complexity.
5. Prism complexity of two-dimensional rectangular matrices

In this section we consider the complexity of the usual $q$-ary matrices, that is the $d=2$ case of the prism complexity.

**Definition 5.1.** Let $q$, $a$, $b$, $A$, $B$ be positive integers, $M$ an $A \times B$ sized, periodic $q$-ary matrix, $A \geq a$, $B \geq b$, $a \leq b$, $\mathcal{L} = \{(a, b)\}$ and $q^{ab} = AB$. $M$ is called $(q, a, b, M, N)$-type De Bruijn matrix, if it contains every possible $a \times b$ sized $q$-ary submatrix exactly once.

We remark, that De Bruijn matrices are called also perfect maps [105, 99, 100] or De Bruijn tori [57]. The first result belongs to Reed and Stewart [105] proving the existence of a $4 \times 4$ sized periodic binary matrix containing all possible $2 \times 2$ sized binary submatrix exactly once. A connected empirical result is due to Péter Selmezi, whose program gives that there are 800 $5 \times 5$ sized non periodic binary array containing the $2 \times 2$ sized submatrices exactly once, 256 such matrices having identical first and fifth columns and 32 such matrices having additionally identical first and fifth rows too, so the problem solved by Reed and Stewart has 32 solutions.

Ma [94] in 1984, Fan Fan, Ma and Siu in 1985 proposed an algorithm which constructs a binary matrix containing every $a_1 \times a_2$ sized binary matrix as submatrix exactly once.

Iványi in 1989 proved the following theorem.

**Theorem 5.1.** (Iványi, 1989 [70]) If $q \geq 1$, $a \geq 1$, $b \geq 1$, then there exist matrices $A$ and $B$ such that there exists a $(q, a, b, A, B)$-type de Bruijn matrix.

**Proof.** (Sketch) (a) If $q = 1$ then the assertion is straightforward.

(b) If $a = 1$ then algorithm BRUIJN produces the required word.

(c) If $a = b = 2$ then see [74] or the monograph of Knuth [86].

(d) If $q \geq a \geq 3$ and $b \geq 2$ then we construct $\mathcal{M}$ as follows.

(d1) We use OPTIMAL-MARTIN with input data $q$ and $a$ and the $w$ output will the first column of $\mathcal{M}$.

(d2) The $i$th ($i = 2, \ldots, q^{a(b-1)}$) column of $\mathcal{M}$ is generated shifting cyclically downwards its $(i-1)$th column by $w_{i-1}$, where $w_1 \ldots w_s$ ($s = q^{a(b-1)} - 1$) is the output of OPTIMAL-MARTIN for alphabet size $q^b$.

(e) The case $b = 2$ and $a \geq 3$ is similar to (c).

(f) Since in the cases (d) and (e) the height $q^a$ of the constructed matrix is a divisor of the sum of the shift sizes and any two $a \times b$ submatrices are distinct
It is worth to remark that the size of the constructed De Bruijn matrix is $q^a \times q^b$. The papers [99, 100] contain detailed analysis of the possible sizes of De Bruijn matrices as the function of the sizes of the window. sizes $a$ and $b$.

The analog of (2.2), containing the bounds for a one-dimensional word $w$ is the following assertion.

**Lemma 5.1.** Let $q$, $a$, $b$, $A$, $B$ be positive integers. If $\mathcal{M}$ is an $A \times B$ sized $q$-ary matrix and $\mathcal{L} = \{(a,b)\}$ then we get the following bounds:

\[(5.1) \quad 1 \leq C(q, 2, \mathcal{L}, \mathcal{M}) \leq \min(q^{ab}, AB).\]

The lower bound is sharp. In some cases the upper bound is also sharp.

**Proof.** If $\mathcal{M}$ is a homogeneous word (that is $q = 1$ or $\mathcal{M}$ contains only one element of a larger alphabet) then the $a \times b$ sized submatrices are identical so the complexity equals to 1. Since $A \geq a$ and $B \geq b$ therefore the complexity is always positive, so the lower bound is sharp.

The complexity of $\mathcal{M}$ is not larger than the number of possible distinct $a \times b$ sized submatrices, and also is not larger than the number of elements of $\mathcal{M}$ (since each element can be the left upper element of one $a \times b$-sized submatrix), therefore the upper bound is correct. According to Theorem 5.1 the constructed $q^a \times q^b$ sized $\mathcal{M}$ contains distinct submatrices, therefore in e.g. in this case the upper bound also is sharp. ■

For $q$ large enough we can form a crossbow matrix [69] in which the elements are distinct. For such matrices we get the following simple consequence of Lemma 5.1 and Theorem 5.1.

**Corollary 5.1.** If $q \geq 1$, $A \geq a$, $B \geq b$, $\mathcal{M}$ is an $A \times B$ sized $q$-ary matrix, $\mathcal{L} = \{(a,b)\}$ and $q^{ab} \geq AB$ then

\[(5.2) \quad C(q, 2, \mathcal{L}, \mathcal{M}) = AB.\]

**Proof.** At the given conditions the minimum in (5.1) equals to $AB$. ■

**Acknowledgement.** The authors thank the referee for the proposed useful corrections.
References


[88] Kuliamin, V.V., Test sequence construction using minimum information on the tested system, *Programming Comp. Software*, 31(6) (2005), 301-309.


Prism complexity of matrices


http://mathworld.wolfram.com/deBruijnSequence.html

http://en.wikipedia.org/wiki/De_Bruijn_sequence

**Antal Iványi**  
Faculty of Informatics  
Eötvös Loránd University  
H-1117 Budapest, Pázmány P. sétány 1/C  
Hungary  
tony@inf.elte.hu

**Zoltán Kása**  
Department of Mathematics and Informatics  
Sapientia Hungarian University of Transylvania  
RO 540485 Târgu Mureș, Șoseaua Sighișoarei, 1C.  
Romania  
kasa@ms.sapientia.ro