REVUE D'ANALYSE NUMÉRIQUE ET DE THÉORIE DE L'APPROXIMATION
Tome 33, N° 2, 2004, pp. 123–139

PROPERTIES OF THE COMPLEXITY FUNCTION FOR FINITE WORDS

MIRA-CRISTIANA ANISIU* and JULIEN CASSAIGNE†

Dedicated to Professor Elena Popoviciu on the occasion of her 80th birthday.

Abstract. The subword complexity function p_w of a finite word w over a finite alphabet A with $\operatorname{card} A = q \geq 1$ is defined by $p_w(n) = \operatorname{card}(F(w) \cap A^n)$ for $n \in \mathbb{N}$, where $F(w)$ represents the set of all the subwords or factors of w. The shape of the complexity function, especially its piecewise monotonicity, is studied in detail.

The function h defined as $h(n) = \min\{q^n, N - n + 1\}$ for $n \in \{0, 1, ..., N\}$ has values greater than or equal to those of the complexity function p_w for any $w \in A^N$, i.e., $p_w(n) \leq h(n)$ for all $n \in \{0, 1, ..., N\}$. As a first result regarding h, it is proved that for each $N \in \mathbb{N}$ there exist words of length N for which the maximum of their complexity function is equal to the maximum of the function h; a way to construct such words is described. This result gives rise to a further question: for a given N, is there a word of length N whose complexity function coincides with h for each $n \in \{0, 1, ..., N\}$? The problem is answered in affirmative, with different constructive proofs for binary alphabets ($q = 2$) and for those with $q > 2$. This means that for each $N \in \mathbb{N}$, there exist words w of length N whose complexity function is equal to the function h. Such words are constructed using the de Bruijn graphs.

MSC 2000. 68R15.

Keywords. Subword complexity function, finite words, de Bruijn graph.

1. DEFINITIONS

Let an alphabet A with $\operatorname{card} A = q \geq 1$ be given. A factor (subword) u of an infinite sequence or finite word w has the *right valence* j if there are j and only j distinct letters x_i such that ux_i, $1 \leq i \leq j$ are also in $F(w)$ (the set of all the subwords, or factors, of w); if a factor has the right valence j it can be extended on the right in exactly j ways. The *left valence* is defined in a similar way. A factor having the right (left) valence ≥ 2 is called *right (left) special*; a factor which is both right and left special is called *bispecial*. The length of a word w will be denoted by $|w|$.

* "T. Popoviciu" Institute of Numerical Analysis, P.O. Box 68, 400110 Cluj-Napoca, Romania, e-mail: mira@math.ubbcluj.ro.

†Institut de Mathématiques de Luminy, CNRS UPR 9016 / FRUMAM, Case 907, 13288 Marseille, France, e-mail: cassaign@iml.univ-mrs.fr.

For an infinite sequence U any factor u can always be extended on the right in a factor of U. For a finite word w there are subwords which cannot be extended on the right. Such words have to be suffixes of w. Let us denote by w_0 the suffix of w of *minimal length* which cannot be extended on the right and by K the length of w_0. Then any other subword λw_0 also cannot be extended on the right. Considering the prefix of w of minimal length which cannot be extended on the left, we shall denote its length by H. The constants K and H were defined by de Luca [15].

Let us denote by $S_0(w)$ the set of all suffixes of w which cannot be extended on the right in $F(w)$, i.e., their right valence is 0. If the length of w is N, then we set for any $0 \leq n \leq N$

$$s_0(n) = \mathrm{card}(S_0(w) \cap A^n) = s(0, n).$$

For all $0 \leq n \leq N$, one has $s_0(n) \leq 1$. Moreover, K being the length of w_0 (the suffix of w of minimal length which cannot be extended on the right), s_0 is given by

$$s_0(n) = \begin{cases} 0, & 0 \leq n \leq K-1, \\ 1, & K \leq n \leq N. \end{cases}$$

It follows that the number of subwords which cannot be extended on the right is

$$\mathrm{card}(S_0(w)) = N - K + 1.$$

For an infinite sequence U, the *(subword) complexity function* $p_U : \mathbb{N} \longrightarrow \mathbb{N}$ (defined in [17] as the *block growth*, then named *subword complexity* in [6]) is given by $p_U(n) = \mathrm{card}(F(U) \cap A^n)$ for $n \in \mathbb{N}$, so it maps each nonnegative number n to the number of factors of length n of U; it verifies the iterative equation

$$p_U(n+1) = p_U(n) + \sum_{j=2}^{q} (j-1)s(j, n), \tag{1}$$

$s(j, n)$ being the cardinal of the set of the factors of U having the length n and the right valence j.

For a finite word w of length N, the *complexity function* $p_w : \mathbb{N} \longrightarrow \mathbb{N}$ given by $p_w(n) = \mathrm{card}(F(w) \cap A^n)$, $n \in \mathbb{N}$, has the property that $p_w(n) = 0$ for $n > N$. The corresponding iterative equation is

$$p_w(n+1) = p_w(n) + \sum_{j=2}^{q} (j-1)s(j, n) - s_0(n). \tag{2}$$

Since $s_0(n) = s(0, n)$ we can write (2) in a condensed form

$$p_w(n+1) = p_w(n) + \sum_{j=0}^{q} (j-1)s(j, n). \tag{3}$$

The above relations have their correspondents in terms of left extensions of the subwords.

For a finite word w of length N over the alphabet A with $\operatorname{card} A = q$, the subword complexity $p_w(n)$ will be less than or equal to the number q^n of all the possible words of length n over the q-letter alphabet and also less than or equal to the number $N - n + 1$ of all occurrences of subwords of length n in w. The map $h : \{0, 1, ..., N\} \longrightarrow \mathbb{N}$ defined in [15]

$$h(n) = \min\{q^n, N - n + 1\} \tag{4}$$

will have values greater than or equal to those of any complexity function p_w for $w \in A^N$, i.e., $p_w(n) \leq h(n)$, $n \in \{0, 1, ..., N\}$. The fact that $p_w(n) \leq N-n+1$ was stated in [11], while $p_w(n) \leq h(n)$ appeared in [20].

We recall that for infinite sequences U one has

$$p_U(n) \leq q^n,\ n \in \mathbb{N},$$

and that there exist sequences, called *complete*, for which the complexity is precisely q^n for all $n \in \mathbb{N}$. An example is the Champernowne sequence

$$0.1.10.11.100.101.110.111.1000.\ldots$$

containing successively all the nonnegative integers written in base 2, and, more generally, in base q (it was used in [4] to construct a normal number in base ten).

2. PROPERTIES OF THE FUNCTION h

Remark 1. We mention at first the trivial case when $q = 1$, for which $h(n) = 1$ for all $n \in \{0, 1, ..., N\}$; there is a word, namely $w_1 = a^N$, whose complexity satisfies $p_{w_1}(n) = 1$ for all $n \in \{0, 1, ..., N\}$.

For $n \geq 1$, $q \geq 2$ and $N \leq q$, we have

$$N - n + 1 \leq q \leq q^n,$$

hence $h(n) = N - n + 1$ for all $n \in \{1, ..., N\}$. For each word w_2 containing $N \leq q$ distinct elements of A the complexity function is $p_{w_2}(n) = N-n+1$ for all $n \in \{1, ..., N\}$, and $p_{w_2}(0) = h(0) = 1$, hence in this case it also coincides with h. □

In what follows we shall consider $q \geq 2$ and $N > q$. The values of the function h are given by the minimum of the values of an increasing exponential and of a descending line, so at the beginning h will follow the exponential, and then the descending line. The following result is presented without proof in [15]:

Proposition 1. *If e_N denotes the first point where $(e_N, h(e_N))$ is on the descending line, the maximum $h_{\max}$ of the function h is attained at e_N, and e_N is given by $\left[\log_q N\right]$ or $\left[\log_q N\right] + 1$ (for a real x, $[x]$ denotes the largest integer which is less than or equal to x).*

We shall determine precisely the point $(e_N, h(e_N))$ where the maximum of the function h is taken.

PROPOSITION 2. *Let* $e_N = \min\{m \in \{1, ..., N\} : h(m) = N - m + 1\}$. *If* $N \in \mathbb{N}$ *is given so that* $q^k + k < N < q^{k+1} + k + 1$, *then* $e_N = k + 1$ *and this is the unique point where* h *attains its maximum* $N - k$; *for* $N = q^k + k$, *we have* $e_N = k + 1$ *and the function* h *attains its maximum* $N - k$ *at both* e_N *and* $e_N - 1$.

In fact, if $q^k + k \leq N < q^{k+1} + k + 1$, *the function* h *is given by*

$$h(n) = \begin{cases} q^n & \text{if } n \leq k \\ N - n + 1 & \text{if } n \geq k + 1 \end{cases} \tag{5}$$

and $h_{\max} = N - k = h(k + 1)$, *for* $N = q^k + k$ *the maximum being attained also at the point* k.

Proof. Let $N \in \mathbb{N}$ be given so that $q^k + k \leq N < q^{k+1} + k + 1$. The function h being increasing on $\{0, 1, ..., e_N - 1\}$ and decreasing on $\{e_N, ..., N\}$, we have only to compare its values on $e_N - 1$ and e_N. From the definition of h and of e_N we have

$$q^{e_N - 1} < N - (e_N - 1) + 1 \tag{6}$$

and

$$q^{e_N} \geq N - e_N + 1, \tag{7}$$

which means that

$$q^{e_N - 1} + e_N - 1 \leq N < q^{e_N} + e_N. \tag{8}$$

But $h(e_N - 1) = q^{e_N - 1}$ and $h(e_N) = N - e_N + 1$, hence from (6) it follows that the maximum of h is taken at e_N. The function $f(x) = q^x + x$ being increasing, from (8) one obtains $e_N = k + 1$ and $h(e_N) = N - k$.

If $q^k + k < N < q^{k+1} + k + 1$ we have $h(k + 1) > h(k)$, so $e_N = k + 1$ is the unique point where h attains its maximum which is equal to $N - k$; if $N = q^k + k$, we have $h(k + 1) = h(k)$, so the maximum $N - k$ of h is taken at two points $e_N = k + 1$ and $e_N - 1$. □

The description of h given in (5) was established in [12]. The value of e_N being related to the integer part of $\log_q N$, we can give a more precise result than that in the above Proposition 1.

REMARK 2. For $q^k + k \leq N < q^{k+1} + k + 1$, we have $k = [\log_q N]$ for $q^k + k \leq N < q^{k+1}$, and $k + 1 = [\log_q N]$ for $q^{k+1} \leq N < q^{k+1} + k + 1$; it follows that in the first case $e_N = [\log_q N] + 1$, and in the second case $e_N = [\log_q N]$. □

Given the number N, the maximum of the function h can be easily determined.

EXAMPLE 1. Let $N = 6$, $q = 2$; we obtain $e_N = 3$ and $h_{\max} = 4 = h(2) = h(3)$. For $N = 7$, $q = 2$, we have $e_N = 3$ and $h_{\max} = 5 = h(3)$. □

3. PROPERTIES OF THE COMPLEXITY FUNCTION p_w

For an infinite sequence U, the complexity function p_U is nondecreasing; if there exists $m \in \mathbb{N}$ such that $p_U(m+1) = p_U(m)$, then p_U is constant for $n \geq m$. The complexity function for a finite word w of length N has obviously a different behaviour, because of $p_w(N) = 1$ (there is a unique factor of length N, namely w). The study of the shape of p_w was considered by Heinz [11] and then by de Luca in [15], the results in [15] being briefly exposed in what follows. Then the piecewise monotonicity of p_w is established in Theorems 5 and 6.

Let us consider for $n \in \{0, ..., N\}$ the number $R_w(n)$ of all right special factors of length n. Any suffix of a right special factor is still a right special factor. Since $R_w(N-1) = R_w(N) = 0$, one can define an integer R by

$$R = \min\{n \in \mathbb{N} : R_w(n) = 0\}.$$

One has $0 \leq R \leq N-1$; thus $R-1$ represents the *maximal length of a right special factor* of w (excepting the case of the word a^N which has no special factor and for which $R = 0$). If $R = 1$, in w there are no right special factors of length $n \geq 1$; such an example is $w = (ab)^k$, $k \geq 1$. Similarly, there exists a number $0 \leq L \leq N-1$ so that $L-1$ represents the *maximal length of a left special factor* of w (except if $w = a^N$). Remember the number $K\,(H)$ representing the minimal length of a suffix (prefix) of w which cannot be extended on the right (left). The numbers K and R (or their duals H and L) play an important role in the description of the shape of p_w.

Let us denote

$$r\,(n) = \sum_{j=2}^{q} (j-1)\, s\,(j, n)\,, n \in \{1, ..., N\}\,.$$

The function r has the property that $r(n) > 0$ for $n \in [0, R-1]$, and $r\,(n) = 0$ for $n \in [R, N]$. The recurrence relation (2) can be written as

$$p_w(n+1) = p_w(n) + r(n) - s_0(n). \tag{9}$$

If $R < K$, for $n \in [0, R-1]$ one has $s_0\,(n) = 0$ and $r\,(n) > 0$. From (9) we obtain that p_w is strictly increasing on $[0,\ R]$. For $n \in [R, K-1]$, $s_0\,(n) = 0$ and $r\,(n) = 0$, so p_w is constant on the interval $[R, K]$. For $n \in [K, N-1]$, $s_0\,(n) = 1$ and $r\,(n) = 0$, so p_w is strictly decreasing on $[K, N]$, and, moreover, for $n \in [K, N-1]$, $p_w\,(n+1) = p_w\,(n) - 1$.

If $R \geq K$, for $n \in [0,\ K-1]$ we obtain $s_0\,(n) = 0$, $r\,(n) > 0$, so from (9) p_w will be strictly increasing on $[0, K]$. For $n \in [K, R-1]$, $s_0\,(n) = 1$ and $r\,(n) > 0$, hence p_w is non-decreasing on $[K, R]$. For $n \in [R, N-1]$ one has $s_0\,(n) = 1$ and $r\,(n) = 0$, which implies that p_w is strictly decreasing on $[R, N]$ and, for $n \in [R, N-1]$, $p_w\,(n+1) = p_w\,(n) - 1$.

It follows

PROPOSITION 3. [15]. *The subword complexity p_w takes its maximum at R and, moreover,*

$$p_w(R) = N - \max\{R, K\} + 1.$$

Proof. In both cases analyzed above, p_w has its maximum at R. If $R \geq K$, $p_w(n+1) = p_w(n) - 1$ for all $n \in [R, N-1]$, so $1 = p_w(N) = p_w(R) - (N-R)$ and then $p_w(R) = N-R+1$. If $R < K$, $p_w(n+1) = p_w(n) - 1$ for all n in $[K, N-1]$ and $1 = p_w(N) = p_w(K) - (N-K)$. Since $p_w(K) = p_w(R)$ the result follows. □

In a similar way, one can prove

PROPOSITION 4. [15]. *The subword complexity p_w takes its maximum at L and, moreover,*

$$p_w(L) = N - \max\{L, H\} + 1,$$

hence $\max\{R, K\} = \max\{L, H\}$.

From the analysis before Proposition 3, we have the following information on the shape of the function p_w [15]:

For $R < K$, it is strictly increasing (starting from $p_w(0) = 1$ and $p_w(1) = q = \operatorname{card} A$), then constant, and then strictly decreasing (with $p_w(n+1) = p_w(n) - 1$ on the last interval).

For $R \geq K$, p_w is at first strictly increasing, then non-decreasing, and at last strictly decreasing also with $p_w(n+1) = p_w(n) - 1$.

So in both cases, there is an interval on which p_w is increasing and one on which p_w is strictly decreasing. The only problem is that in the second case it could be that after becoming constant, p_w would increase again. We show that this is not the case.

Let us consider $n \in [K, R-1]$, so $s_0(n) = 1$, $r(n) > 0$ and $p_w(n+1) \geq p_w(n)$. Suppose that there exists n so that

$$K \leq n < n+1 < n+2 \leq R,$$

$$p_w(n+1) = p_w(n), \tag{10}$$

$$p_w(n+2) > p_w(n+1). \tag{11}$$

From (10) one obtains that $s(2, n) = 1$ and $s(j, n) = 0$ for $j \geq 3$, i.e., there exists a unique right special factor having length n, and its valence is 2: let it be denoted v_n. From (11) it follows that

$$r(n+1) = \sum_{j=2}^{q} (j-1)\, s(j, n+1) > 1,$$

which is possible for two situations:

I. $s(2, n+1) = 2$;

II. $\exists\, j \geq 3$, $s(j, n+1) \neq 0$.

If II. is true, there will be a right special factor of length $n+1$ having valence at least 3, and then the factor obtained by excluding its first letter will have length n and valence at least 3, contradicting the uniqueness of v_n.

If I. is true, there will exist two different right special factors of length $n+1$ and valence at least 2. They can differ only by their first letter, otherwise there would exist two different factors of length n and valence 2. So they will have the form

$$av_n,\ bv_n,\ a \neq b,$$

i.e., v_n will be bispecial, and in the word w there will be also

$$av_nc,\ av_nd,\ bv_nc,\ bv_nd,\ a \neq b,\ c \neq d. \tag{12}$$

The subword v_n cannot be a suffix of w since v_n is extendable to the right and there is no extendable suffix of length greater than or equal to K. Let us consider the last occurrence of v_n, suppose it is followed by c. Then

$$w = z_1 v_n c z_2, \tag{13}$$

and, v_nc being left special, v_nc will have another occurrence in w, so

$$w = z_1' v_n c z_2',\ \left|z_2'\right| > |z_2|\,.$$

Let u be the longest common prefix of v_ncz_2 and v_ncz_2', which will satisfy

$$n+1 \leq |u| \leq |v_n c z_2|\,.$$

Since the subword u is a proper prefix of v_ncz_2', u is right extendable; then it cannot be a suffix of w, hence it is also a proper prefix of v_ncz_2, and thus right special. The suffix of length n of u is then right special, in contradiction with the fact that the last occurrence of v_n, the unique right special factor of length n, was chosen so that $w = z_1v_ncz_2$. It follows that in the case $K < R$, if $p_w(n) = p_w(n+1)$ for a value $n \geq K$, then p_w will remain constant until it will begin (at R) to decrease to 1 (it cannot start increasing again). We mention that Heinz [11] proved that from $p_w(n) = p_w(n+1)$ and $N > p_w(n) + n$ it follows $p_w(n) = p_w(n+2)$.

Let us denote by J the smallest number greater than or equal to K for which w has precisely one right special factor of that length, with valence 2 (if this is not the case, take $J = R$). We have established the following

THEOREM 5. *For a finite word of length N, the complexity function is at first strictly increasing, then constant and at last decreasing having the slope -1. If $R < K$, the successive intervals are $[0, R]$, $[R, K]$ and $[K, N]$, while for $R \geq K$ they are $[0, J]$, $[J, R]$ and $[R, N]$.*

One can easily avoid to analyze two cases by simply considering instead of a word $w \in A^N$, a word $W \in (A \cup \{*, \#\})^{N+2}$ obtained by adding two different symbols which are not in A at the beginning and at the end of w, i.e., $W = *w\#$. The complexity functions for w and W are related by $p_W(n) = p_w(n) + 2$ for $n \in \{1, ..., N+1\}$ (and obviously $p_W(N+2) = 1$). So the graph of p_W is the graph of p_w shifted by two units parallel to the y-axis,

and the two functions have the same monotonicity. For W we have $K_W = 1$, $R_W \geq R_w$ and, similarly, $H_W = 1$, $L_W \geq L_w$, hence in this case we have always $R_W \geq K_W$; from Proposition 4 it follows also $R_W = L_W$. The advantage of considering the word W is that instead of the four parameters K, H, R, L we are left with only one, namely the common value M of $R_W = L_W$. Denoting by J the smallest positive number for which W has precisely one right special factor of that length, with valence 2 (if there is not such a factor, $J = M$), we obtain

THEOREM 6. *For a finite word w of length N, the intervals of monotonicity of p_w are $[0, J]$, $[J, M]$ and $[M, N]$, the function increasing at first, being constant and then decreasing with the slope -1; the maximum of p_w is $p_w(M) = N - M + 1$. The numbers J and M are those defined above for the word $W = *w\#$.*

EXAMPLE 2. Let $w = babbabbbaa$, so $N = 10$, $K = 2$, $R = 5$, $H = 5$, $L = 4$, $J = 4$, $M = 5$. In this case $p_w(0) = 1$, $p_w(1) = 2$, $p_w(2) = 4$, $p_w(3) = 5$, $p_w(4) = 6$, and $p_w(n) = 11 - n$ for $n \in \{5, ..., 10\}$.

For $w = aababbabab$, we have $N = 10$, $K = 5$, $R = 4$, $H = 2$, $L = 5$, $J = 4$, $M = 5$ and $p_w(0) = 1$, $p_w(1) = 2$, $p_w(2) = 4$, $p_w(3) = 5$, $p_w(4) = 6$, and $p_w(n) = 11 - n$ for $n \in \{5, ..., 10\}$. □

REMARK 3. For the words in example 2, the function p_w is concave on $[1, N]$, i.e.,

$$p_w(k + 2) - p_w(k + 1) \leq p_w(k + 1) - p_w(k), k \in \{1, ..., N - 2\}.$$

However this is not the rule, as the following examples show for both $K < R$ and $K > R$. □

EXAMPLE 3. Let $w = abbabbbaaba$, so $N = 11, K = 3, R = 4, H = 4$, $L = 4$, $J = 4$, $M = 4$. In this case $p_w(0) = 1$, $p_w(1) = 2$, $p_w(2) = 4$, $p_w(3) = 7$, and $p_w(n) = 12 - n$ for $n \in \{4, ..., 11\}$.

Let $w = abbabbbaababbba$, so $N = 15$, $K = 7$, $R = 4$, $H = 4$, $L = 7$, $J = 4$, $M = 7$. In this case $p_w(0) = 1$, $p_w(1) = 2$, $p_w(2) = 4$, $p_w(3) = 7$, $p_w(4) = p_w(5) = p_w(6) = p_w(7) = 9$, and $p_w(n) = 16 - n$ for $n \in \{8, ..., 15\}$. □

We mention that the refinement of de Luca's result has been proved independently by Levé and Séébold [14] while studying k-reachable integers.

4. THE FUNCTION h AND RELATED WORDS

In section 2 we found the point where the function h takes its maximum. A problem to be considered is the following: are there any words w of length N such that $h(e_N) = \max\{p_w(n) : n \in \{0, ..., N\}\}$? If such words do exist, they have the property that the maximum of their complexity function cannot be exceeded by the maximum of the complexity function of any other word of length N.

The answer to this problem is in affirmative and it relies on the following result which was stated by Good in [10] for $q = 2$. The enumeration of the words whose existence is proved was given by de Bruijn [2], who later [3] acknowledged the priority of C. Flye Sainte-Marie [7].

LEMMA 7. *Given an alphabet A with* card $A = q$, *for each $k \in \mathbb{N}$ the shortest word containing all the q^k words of length k has $q^k + k - 1$ letters.*

Proof. The existence of such a word (which is usually named de Bruijn word of order k) is proved by considering the de Bruijn graph B_{k-1} (which is strongly connected) with q^{k-1} vertices labelled with the elements of A^{k-1}, and q^k arcs (an arc from u to v exists if and only if there exist two letters $x, y \in A$ such that $ux = yv \in A^k$). Each vertex has the same number q of inward and outward arcs; therefore, there exists an Eulerian cycle, and each path, starting from any vertex and following the cycle until coming back to that vertex, will provide a word (obviously the shortest) of length $q^k + k - 1$ which contains exactly one occurrence of all the q^k words of length k. The word of length $q^k + k - 1$ is often identified with the cycle formed by its first q^k letters. □

REMARK 4. For the de Bruijn word of order k, whose existence was proved in Lemma 7, we have $R = K = J = k$, and the maximum of its complexity function is attained at k and equals q^k. Such a word can be represented in the form $x_1...x_{q^k}...x_{q^k+k-1}$ (with $x_{q^k+1}...x_{q^k+k-1} = x_1...x_{k-1}$), or as a cycle $(x_1...x_{q^k})$ or as an infinite periodic sequence with the period q^k. □

The first algorithm which constructs such a word was given by Martin [16]. Considering the alphabet $A = \{i_1, ..., i_q\}$, the algorithm in question is built up out the following three rules.

I. *Each of the first $k - 1$ symbols is chosen equal to i_1.*

II. *The symbol a_m to be added to the sequence $a_1a_2...a_k...a_{m-k+1}...a_{m-1}$, where $a_1 = ... = a_{k-1} = i_1$, $m \geq k$ and the a's stand for the i's in a certain order, is the i_j with the greatest subscript consistent with the requirement that the section $a_{m-k+1}...a_{m-1}a_m$ duplicate no previously occurring section of k symbols in the above sequence.*

III. *Rule* II. *is first applied for $m = k$ (in which case $a_m = a_k = i_q$) and is then applied repeatedly until a further application is impossible.*

This algorithm needs a very large memory (for all the subwords of length k which have already been obtained), but there exist also some memoryless algorithms exposed, for example, in [8], [9] and [18].

We can prove now

THEOREM 8. *For each $N \in \mathbb{N}$, there exists a word of length N over an alphabet A with* card $A = q$ *for which the maximum of the complexity function is equal to the maximum $h_{\max}$ of h; the maximum is taken at the same points for both functions. Such words can be easily constructed using the de Bruijn words.*

Proof. Keeping in mind the considerations in Remark 1, which mean precisely that the theorem is true for $q = 1$ and for $q \geq 2$, $N \leq q$, we shall consider $q \geq 2$ and $N > q$. Let k be the unique natural number so that

$$q^k + k \leq N \leq q^{k+1} + k.$$

If $N = q^k + k$, we apply Lemma 7 for k, obtaining a word of length $N - 1$ containing as factors all the q^k words of k letters, and $q^k - 1$ distinct words of length $k + 1$. The word v obtained by adding a letter from A at its end will contain q^k words of k letters and q^k distinct words of length $k + 1$, hence $p_v(k) = p_v(k+1) = N-k$. This is the maximum of the function p_v, it is equal to $h_{\max}$ and it is attained at the same points as the maximum of h given in Proposition 2. Actually, in this case we have $p_v = h$.

Let us now consider the case $N = q^{k+1} + k - m$, $m \in \{0, 1, ..., q^{k+1} - q^k - 1\}$. Applying Lemma 7 for the number $k + 1$, we obtain a shortest word w containing all the q^{k+1} words of length $k + 1$, having $q^{k+1} + k$ letters. So for each $m \in \{0, 1, ..., q^{k+1} - q^k - 1\}$, the prefix w_m of w obtained by deleting m final letters will satisfy $p_{w_m}(k + 1) = q^{k+1} - m > q^k \geq p_{w_m}(k)$, this being the maximum of the complexity function for the considered word. The maximum is attained only for $k + 1$.

Applying Proposition 2 for $N = q^{k+1} + k - m$ we obtain $h_{\max} = h(k+1) = q^{k+1}+k-m-k = q^{k+1}-m$, which means that the maximum of the complexity function for w_m is equal to the maximum of the complexities of all possible words of length $q^{k+1} + k - m$. □

Example 4. Let us consider for the 2-letter alphabet $A = \{a, b\}$ the values $N = 6$ and $N = 7$. For $N = 6 = 2^2 + 2$ we have $k = 2$ and, by adding a letter (for example a) to the Martin word of order 2, $abbaa$, we obtain $v = abbaaa$; the maximum of p_v is $p_v(2) = p_v(3) = 4$. For $N = 7 = 2^2 + 3$ we can consider the Martin word of order 3, $aabbbabaaa$, and delete three symbols from the end. The word $w_3 = aabbbab$ has the maximum of its complexity function given by $p_{w_3}(3) = 5$.

The maximum of the function h for $N = 6$ and $N = 7$ was calculated in Example 1 and it coincides with that of p_v, respectively p_{w_3} and is taken at the same points. □

5. THE REPRESENTATION OF h AS A COMPLEXITY FUNCTION

An interesting problem is: Let $q \geq 1$ and $N \in \mathbb{N}$ be given and the function $h : \{0, 1, ..., N\} \longrightarrow \mathbb{N}$ defined as in (4). Is there a word w of length N over the q-letter alphabet A such that

$$h(n) = p_w(n) \text{ for all } n \in \{0, 1, ..., N\}, \tag{14}$$

i.e., h is the complexity function for that word? If such a word does exist, how can it be constructed?

The question has an affirmative answer for the trivial cases $q = 1$ and $q \geq 2$, $N \leq q$, mentioned in Remark 1, so it has to be studied for $q \geq 2$, $N > q$. In the proof of Theorem 8 it was shown that, given the number $N = q^k + k$, $k \geq 1$, there exists a word v of length N containing q^k distinct words of length $k+1$, and also q^k words of length k. This means that h and p_v coincide on k and $k+1$. One the one hand, $p_v(k) = h(k) = q^k$ means that v contains all possible words of length k as factors, and this implies that it also contains all possible words of shorter lengths, hence $h(n) = p_v(n) = q^n$ for $n \in \{0, 1, ..., k\}$. On the other hand, $p_v(k+1) = h(k+1) = N-k$ means that each of the $N-k$ factors of length $k+1$ of v occurs exactly once, as there are precisely $N-k$ available positions for a factor of this length, and this implies that longer factors occur only once too, hence $h(n) = p_v(n) = N+1-n$ for $n \in \{k+1, k+2, ..., N\}$. We have shown that $h(n) = p_v(n)$ for all $n \in \{0, 1, ..., N\}$, and the question is positively answered for $N = q^k + k$, $k \geq 1$.

If we consider now $N = q^{k+1} + k$, $k \geq 1$, case which corresponds to the choice $m = 0$ in the proof of Theorem 8, we obtain the existence of a word $w = w_0$ of length N containing all q^{k+1} words of length $k+1$. The point $(k+1, p_w(k+1))$ being on both the curves (n, q^n) and $(n, N+1-n)$, it follows that $h(n) = p_w(n)$ for all $n \in \{0, 1, ..., N\}$.

We mention at first a sufficient condition for the existence, for $q \geq 2$ and $N > q$, of a word w of length N whose complexity function is equal to h.

LEMMA 9. *Given an alphabet with* $\operatorname{card} A = q \geq 2$, *if for each* $k \geq 1$ *there exists a de Bruijn word* v *of order* $k+1$ *from which it is possible to obtain successively words shorter with one symbol so that the number of subwords of length* $k+1$ *decreases by one, but the number of words of length* k *remains* q^k, *until we are left with a word of length* $q^k + k$, *then for each* $N \in \left\{q^k + k, ..., q^{k+1} + k\right\}$ *there exists a word* v_N *with* $p_{v_N} = h$.

Proof. Let v_N be the word of length $N \in \left\{q^k + k, ..., q^{k+1} + k\right\}$ obtained from v after having removed $q^{k+1} + k - N$ letters, at each step the number of subwords of length $k+1$ being diminished by 1, while the number of subwords of length k remains constant. Then $p_{v_N}(k+1) = N - k = h(k+1)$ and $p_{v_N}(k) = q^k = h(k)$, hence $p_{v_N}(n) = h(n)$ for each $n \in \{0, ..., N\}$. □

REMARK 5. The condition in Lemma 9 is fulfilled if there exists a de Bruijn word of order $k+1$ whose prefix is a de Bruijn word of order k. In this case we can simply delete in turn one letter from the end of the word of order $k+1$. □

The existence of words which satisfy the conditions in Lemma 9 (in fact those in Remark 5) was proved for $q \geq 3$ by Vörös [20]. It follows also as a consequence of a stronger result obtained by Cummings and Wiedemann in Proposition 2 from [5]. In fact the overlap of the two de Bruijn sequences in [5] is even longer than it is needed in Remark 5. We remind that the de Bruijn graph B_k has as vertices the elements in A^k and an arc from any vertex $x_1...x_k$ to $x_2...x_kx_{k+1}$, where $x_i \in A$ for $i \in \{1, ..., k+1\}$. The graph

B_{k+1} has as vertices the arcs of B_k, and the arcs in this graph are obtained by joining two consecutive arcs in B_k. An Eulerian circuit in B_k corresponds to a Hamiltonian one in B_{k+1} and conversely. The result of Cummings and Wiedemann follows from the fact that if one removes from the Eulerian circuit in B_k, which corresponds to a de Bruijn sequence of order $k+1$, the circuit corresponding to a de Bruijn sequence of order k, the remaining graph is still Eulerian and connected (it is essential that $q \geq 3$).

LEMMA 10. [5]. *If $q \geq 3$ and $k \geq 1$ each de Bruijn sequence of order k can be strongly embedded in a de Bruijn sequence of order $k+t$ with $t \geq 1$, i.e., the two sequences have the same symbols on the first $q^k + k + t - 1$ positions.*

It follows that, for $q \geq 3$, there exist infinite sequences whose prefixes of length N satisfy (14) for each $N \in \mathbb{N}$. Such sequences were called in [13] and [20] *supercomplex*; similarly, a word of length M was called supercomplex if all its prefixes of length $N \leq M$ satisfied (14). In [13] and [20] it was shown that supercomplex sequences do not exist for binary alphabets, more precisely it was verified that a binary supercomplex word has the length at most 9. This means that no de Bruijn sequence of order 2 can be embedded in a de Bruijn sequence of order 3. In [5] a general negative result is given for binary alphabets: in this case no de Bruijn sequence of order $k \geq 2$ ever embeds in a de Bruijn sequence of order $k+1$ (even if we ask the coincidence to take place only for the first $2^k + k - 1$ positions, hence a weak embedding). It follows that for a binary alphabet we cannot obtain a word as that in the sufficient condition in Remark 5, unless $k = 1$. Nevertheless we can construct in this case a de Bruijn word of order $k+1$ from which the sequences in Lemma 9 can be obtained, even if this word has not as a prefix a de Bruijn word of order k.

LEMMA 11. *A finite number of cycles can be appended to any de Bruijn cycle of order k over a binary alphabet in order to make it a de Bruijn cycle of order $k+1$.*

Proof. Let $w = (x_1 ... x_{2^k})$ be a de Bruijn cycle of order k. It will be also a Hamiltonian circuit in the de Bruijn graph B_k. The graph G, formed by all the vertices in A^k and the arcs in B_k which are not in the Hamiltonian circuit determined by w, has each vertex of degree 2 (one outward and one inward arc). It follows that G will be a union of vertex disjoint cycles, and w and G are arc disjoint. Each of these cycles will have common vertices with the Hamiltonian circuit determined by w, hence they can be appended one by one to it to form finally an Eulerian circuit in B_k, that is a de Bruijn cycle of order $k+1$. □

Now we can state

THEOREM 12. *For each alphabet A with* $\operatorname{card} A = q \geq 1$ *and for each $N \in \mathbb{N}$ there exists a word of length N whose complexity function coincides with the function h.*

Proof. For $q = 1$, or $q \geq 2$ and $N \leq q$, the result was already proved in Remark 1.

Let $q = 2$ and $k \geq 1$ so that $N \in \left\{2^k + k, ..., 2^{k+1} + k\right\}$. Consider a de Bruijn cycle of order k (constructed for example using Martin's algorithm) and extend it as in Lemma 11, by adding vertex disjoint cycles, to a de Bruijn cycle of order $k+1$. Write it as a de Bruijn word such that it ends with the letters of one of the appended cycles. When we remove one by one all the symbols in that cycle, the number of subwords of length $k+1$ will decrease at each step by one, but the number of subwords of length k will remain the same (all these subwords are included in the initial de Bruijn cycle). Write again the obtained cycle as a word which ends with another appended cycle and delete in turn the last symbol until the cycle disappears. Finally we are left with a word of length $2^k + k$ obtained from the initial de Bruijn cycle of order k, which contains 2^k words of length $k+1$ and 2^k words of length k.

If we are not interested to obtain all the words of length $N \in \{q^k + k, ..., q^{k+1} + k\}$, but only a specific one, we can apply a result of Shallit [19]: For each $i \in \left\{1, ..., 2^k\right\}$ the graph B_k contains a cycle of length i that can be used to construct a closed chain of length $2^k + i$ which visits every vertex at least once.

Finally, let $q \geq 3$ and $k \geq 1$ so that $N \in \left\{q^k + k, ..., \ q^{k+1} + k\right\}$. Applying Lemma 10 for $t = 1$, we obtain the existence of a de Bruijn word of order $k+1$ which has as prefix a de Bruijn word of order k, hence it satisfies the conditions in Lemma 9. It follows that for each $N \in \{q^k + k, ..., \ q^{k+1} + k\}$ there exists a word of length N (obtained by successively deleting a symbol from the end of the de Bruijn word of order $k+1$) whose complexity function is the function h corresponding to that N. □

Example 5. Let us first consider the case of a binary alphabet $A = \{a, b\}$. We shall construct, as in the proof of Theorem 12, words u_N with $N \in \{1, 2, ..., 10\} \cup \{37, ..., 69\}$ for which $p_{u_N} = h$. We can obviously consider $u_1 = a$ and $u_2 = ab$. We have a weak embedding, marked by a gap, of the de Bruijn word of order 1, ab, in the de Bruijn word of order 2, $ab\ baa$ (situation which is no longer possible for words of order k, respectively $k+1$ for $k \geq 2$). We obtain in turn $u_5 = abbaa$, $u_4 = abba$ and $u_3 = abb$. Let us now consider for $k = 2$ the Martin cycle $w = (abba)$, which corresponds to the word $u_5 = abbaa$. The graph G obtained from B_2 by removing all the arcs of w is the union of the cycles (a), (b) and (ab), i.e., $aa \to aa$, $bb \to bb$, respectively $ab \to ba \to ab$. Appending these to the cycle w we obtain for instance the de Bruijn cycle of order 3 $u = (a\underline{a}b\underline{a}\underline{b}b\underline{b}a)$, where we underlined the appended cycles; we write it as

$$u_{10} = abb\underline{b}aa\underline{a}b\underline{ab};$$

deleting the symbols of (ab) from the end, and then those in the loops (a) and (b) (which can be deleted without shifting the cycle) we obtain in turn

$$u_9 = abb\underline{b}aa\underline{a}ba, \ u_8 = abb\underline{b}aa\underline{a}b, \ u_7 = abb\underline{b}aab, \ u_6 = abbaab.$$

For all these words obtained from u we have $p_{u_N}(2) = 4$ and $p_{u_N}(3) = N - 2$, $N \in \{6, ..., 10\}$.

In general, besides the loops (a) and (b), for greater values of k we can have in G more than one cycle. We shall consider now $k = 5$ and we shall construct the words of length $N \in \{37, ..., 69\}$. To avoid too lengthy words we shall write x^i for the concatenation of i letters x. For $k = 5$, the Martin cycle is $w = (a^4b^5ab^3a^2b^2abab^2a^3baba^2ba)$. The graph G obtained from B_5 by removing the arcs of w is the union of the cycles

$$(a), (b), (ab), (ab^2), (a^4ba^2b^3aba^3b^2a^2bab^4), \tag{15}$$

where, for example, (a) represents $a^5 \to a^5$, and (ab^2) is $ab^2ab \to b^2ab^2 \to bab^2a \to ab^2ab$. A de Bruijn cycle of order 6 obtained by appending the five cycles (which will be underlined) is

$$u = (a^4\underline{a}b^4\underline{a^4ba^2b^3aba^3b^2a^2bab^4}\underline{b}\underline{b}ab^2\underline{ab^2}ba^2b^2abab\underline{ab}ba^3baba^2ba)$$

and we can write

$$\begin{aligned}
u_{69} &= ab^5\underline{b}ab^2\underline{ab^2}ba^2b^2abab\underline{ab}ba^3baba^2ba^5\underline{a}b^4\underline{a^4ba^2b^3aba^3b^2a^2bab^4}\\
u_{68} &= ab^5\underline{b}ab^2\underline{ab^2}ba^2b^2abab\underline{ab}ba^3baba^2ba^5\underline{a}b^4a^4ba^2b^3aba^3b^2a^2bab^3\\
&...\\
u_{44} &= ab^5\underline{b}ab^2\underline{ab^2}ba^2b^2abab\underline{ab}ba^3baba^2ba^5\underline{a}b^4,
\end{aligned}$$

the last one corresponding to the cycle $(\underline{a}b^5\underline{b}ab^2\underline{ab^2}ba^2b^2abab\underline{ab}ba^3baba^2ba^5)$. We write it as a word ending with the next cycle, (ab^2), in the union (15)

$$u'_{44} = b^2ab^3a^2b^2abab\underline{ab}ba^3baba^2ba^5\underline{a}b^5\underline{b}ab^2\underline{ab^2},$$

and we obtain

$$\begin{aligned}
u_{43} &= b^2ab^3a^2b^2abab\underline{ab}ba^3baba^2ba^5\underline{a}b^5\underline{b}ab^2ab\\
u_{42} &= b^2ab^3a^2b^2abab\underline{ab}ba^3baba^2ba^5\underline{a}b^5\underline{b}ab^2a\\
u_{41} &= b^2ab^3a^2b^2abab\underline{ab}ba^3baba^2ba^5\underline{a}b^5\underline{b}ab^2.
\end{aligned}$$

We write now the corresponding word ending with the cycle (ab)

$$u'_{41} = babab^2a^3baba^2ba^5\underline{a}b^5\underline{b}ab^3a^2b^2abab\underline{ab}$$

and from this we get

$$\begin{aligned}
u_{40} &= babab^2a^3baba^2ba^5\underline{a}b^5\underline{b}ab^3a^2b^2ababa\\
u_{39} &= babab^2a^3baba^2ba^5\underline{a}b^5\underline{b}ab^3a^2b^2abab.
\end{aligned}$$

Deleting the loop (b) and then the loop (a) we obtain at last

$$\begin{aligned}
u_{38} &= babab^2a^3baba^2ba^5\underline{a}b^5ab^3a^2b^2abab\\
u_{37} &= babab^2a^3baba^2ba^5b^5ab^3a^2b^2abab.
\end{aligned}$$

We have $p_{u_N}(5) = 32$ and $p_{u_N}(6) = N - 5$ for $N \in \{37, ..., 69\}$.

Let now a 3-letter alphabet $A = \{a, b, c\}$ be given (the situation is similar for any $q > 3$). We have obviously $w_1 = a$, $w_2 = ab$, $w_3 = abc$. From a de

Bruijn word of order 2 which contains a strongly embedded de Bruijn word of order 1

$$abca\ accbba,$$

the gap marking the end of the overlapping, we can obtain the words

$$\begin{aligned} &w_{10} = abcaaccbba \\ &... \\ &w_4 = abca. \end{aligned}$$

Similarly, from the de Bruijn word of order 3 which contains a strongly embedded de Bruijn word of order 2

$$aabbacbccaa\ cababcacccbbbcbaaa,$$

we can obtain successively the words

$$\begin{aligned} &w_{29} = aabbacbccaacababcacccbbbcbaaa \\ &... \\ &w_{11} = aabbacbccaa, \end{aligned}$$

which satisfy $p_{w_N}(2) = 9$ and $p_{w_N}(3) = N - 2$ for $N \in \{11, ..., 29\}$. □

REMARK 6. It is clear now that Theorem 8 is a consequence of the stronger result from Theorem 12. However, if one is interested only in obtaining words w with $\max\{p_w(n) : 1, ..., N\} = h_{\max}$, the constructive methods in the proof of Theorem 8 are simpler and faster. □

6. OTHER COMPLEXITY MEASURES FOR FINITE WORDS

The complexity function p_w which was used throughout the paper has the advantage that it can be defined in the same way both for infinite sequences and for finite words, as it was stated in the introduction. As far as finite words are concerned, the first measure of subword complexity seems to have been introduced by Heinz [11] as the total number of factors of w,

$$K(w) = \sum_{n=0}^{N} p_w(n).$$

The problem of studying the maximum of $K(w)$ over all the words of length N over a finite alphabet with q elements was a central one. It is easy to see that the maximum of $K(w)$ over the words in A^N is attained at w_0 if for each $n \in \{0, ..., N\}$ the maximum of $p_w(n)$ is attained at $p_{w_0}(n)$. One has then the delimitation (obtained in [20])

$$K(w) \leq \sum_{n=0}^{N} h(n),$$

with h defined in (4), and, having in mind the explicit form of h in (5),

$$K(w) \leq \frac{q^{k+1}-1}{q-1} + \frac{(N-k)(N-k+1)}{2}, \tag{16}$$

where k is the unique natural number for which $q^k + k \leq N \leq q^{k+1} + k$. The bound in (16) appeared in [12] and [19]. In view of Theorem 12, there are words of length N whose total complexity $K(w)$ equals the value in the right hand side of (16). The existence of such words, as it was already mentioned in the proof of Theorem 12, was established for binary alphabets in [19].

There are also other notions of complexity for finite words. The *maximal complexity* of a word $w \in A^N$, defined by Rauzy, is

$$\mathcal{C}(w) = \max\{p_w(n) : n \in \{0, 1, ..., N\}\}.$$

A notion of global complexity for finite words is given in [1], namely the *global maximal complexity* in A^N

$$\mathcal{K}(N) = \max\{\mathcal{C}(w) : w \in A^N\}.$$

By $\mathcal{R}(N)$ it is denoted the set of the values i for which there exists a word $w \in A^N$ such that $p_w(i) = \mathcal{K}(N)$:

$$\mathcal{R}(n) = \left\{i \in \{0, 1, ..., N\} : \text{there exists } w \in A^N,\, p_w(i) = \mathcal{K}(n)\right\}.$$

With these notations we obtain from Theorem 12

Corollary 13. *For each $N \in \mathbb{N}$, the global maximal complexity is given by $\mathcal{K}(N) = \max\{h(n) : n \in \{0, 1, ..., N\}\}$, the function h being defined by (4).*

Proof. We have

$$\begin{aligned}\mathcal{K}(N) &= \max\left\{\max\{p_w(n) : n \in \{0, 1, ..., N\}\} : w \in A^N\right\}\\ &= \max\left\{\max\left\{p_w(n) : w \in A^N\right\} : n \in \{0, 1, ..., N\}\right\}\\ &= \max\left\{h(n) : n \in \{0, 1, ..., N\}\right\},\end{aligned}$$

the last equality following from the fact that h coincides with the complexity function p_w for at least one word of length N. □

Applying the result in Proposition 2, the values for $\mathcal{K}(N)$ and $\mathcal{R}(N)$ given in [1] can be easily obtained.

Corollary 14. *For $q^k + k \leq N < q^{k+1} + k + 1$, we have $\mathcal{K}(N) = N - k$. If $N = q^k + k$, then $\mathcal{R}(N) = \{k, k + 1\}$; if $q^k + k < N < q^{k+1} + k + 1$, then $\mathcal{R}(N) = \{k + 1\}$.*

Acknowledgment. This research was partially supported by the program of academic exchange CNRS-Romanian Academy.

REFERENCES

[1] Anisiu, M.-C., Blázsik, Z. and Kása, Z., *Maximal complexity of finite words*, Pure Math. Appl., **13**, pp. 39–48, 2002.

[2] de Bruijn, N. G., *A combinatorial problem*, Nederl. Akad. Wetensch. Proc., **49**, pp. 758–764, 1946 = Indag. Math., **8**, pp. 461–467, 1946.

[3] de Bruijn, N. G., *Acknowledgement of priority to C. Flye Sainte-Marie on the counting of circular arrangements of 2^n zeros and ones that show each n-letter word exactly once*, T. H. -Report 75-WSK-06, Technological University Eindhoven, the Netherlands, pp. 1–14, 1975.

[4] Champernowne, D. G., *The construction of decimals normal in the scale of ten*, J. London Math. Soc., **8**, pp. 254–260, 1933.
[5] Cummings, L. J. and Wiedemann, D., *Embedded de Bruijn sequences*, Proceedings of the 7th Southeastern international conference on combinatorics, graph theory, and computing (Boca Raton, Florida, 1986), Congr. Numer., **53**, pp. 155–160, 1986.
[6] Ehrenfeucht, A., Lee, K. P. and Rozenberg, G., *Subword complexities of various classes of deterministic developmental languages without interactions*, Theoret. Comput. Sci., **1**, pp. 59–75, 1975.
[7] Flye Sainte-Marie, C., *Solution to question nr. 48*, l'Intermédiaire des Mathématiciens, **1**, pp. 107–110, 1894.
[8] Fredericksen, H., *A survey of full length nonlinear shift register cycle algorithms*, SIAM Review, **24**, pp. 195–221, 1982.
[9] Games, R. A., *A generalized recursive construction for de Bruijn sequences*, IEEE Trans. Inform. Theory, **29**, pp. 843–850, 1983.
[10] Good, I. J., *Normal recurring decimals*, J. London Math. Soc., **21**, pp. 167–169, 1946.
[11] Heinz, M., *Zur Teilwortkomplexität für Wörter und Folgen über einem endlichen Alphabet*, EIK, **13**, pp. 27–38, 1977.
[12] Hunyadváry, L. and Iványi, A., *On some complexity measures of words*, Dep. Math., Karl Marx Univ. Econ., Budapest 1984-2, pp. 67–82, 1984.
[13] Hunyadváry, L. and Iványi, A., *Construction of complex chains sequences and ring sequences*, in Abstracts of Colloquium Theory of Algorithms, Pécs, p. 20, 1984.
[14] Levé, F. and Séébold, P., *Proof of a conjecture on word complexity*, Journées Montoises d'Informatique Théorique (Marne-la-Vallée, 2000), Bull. Belg. Math. Soc. Simon Stevin, **8**, pp. 277–291, 2001.
[15] de Luca, A., *On the combinatorics of finite words*, Theoret. Comput. Sci., **218**, pp. 13–39, 1999.
[16] Martin, M. H., *A problem in arrangements*, Bull. American Math. Soc., **40**, pp. 859–864, 1934.
[17] Morse, M. and Hedlund, G. A., *Symbolic dynamics*, Amer. J. Math., **60**, pp. 815–866, 1938.
[18] Ralston, A., *A new memoryless algorithm for de Bruijn sequences*, J. Algorithms, **2**, pp. 50–62, 1981.
[19] Shallit, J., *On the maximum number of distinct factors in a binary string*, Graph Comb., **9**, pp. 197–200, 1993.
[20] Vörös, N., *On the complexity measures of symbol-sequences*, in Proceedings of the Conference of Young Programmers and Mathematicians (ed. A. Iványi), Eötvös Loránd University, Budapest, pp. 43–50, 1984.

Received by the editors: March 10, 2004.