Shadows Under the Word-Subword Relation

R. Ahlswede

Department of Mathematics, University of Bielefeld

V. Lebedev

Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute))

1 Preliminaries

For an alphabet $\mathcal{X} = \{0, 1, \ldots, q-1\}$ we consider the set $\mathcal{X}^n$ of words $x = (x_1, x_2 \cdots, x_n)$ of length n.

$x = (x_1, x_2, \ldots, x_n)$ is n-subword of $y = (y_1, y_2, \ldots, y_k)$ if there is i, $i \in \{0, 1, \ldots, k-n\}$ such that

$$y_{i+1} = x_1, y_{i+1} = x_1, \ldots, y_{i+n} = x_n.$$

x is n-subword of y if there are l and r such that $y = lxr$ where $|l| = i \; i \in \{0, 1, \ldots, k-n\}$.

The shadow of y is the set of all n-subwords:

$$\text{shad}(y) = \{x : \ x \ is \ n - subword \ of \ y\}.$$

Now for any subset $A \subset \mathcal{X}^k$ we define the shadow

$$\text{shad}(A) = \bigcup_{a \in A} \text{shad}(a).$$

For fixed n and k we are interested in

$$\nabla(N) = \min\{|\text{shad}(A)| : A \subset \mathcal{X}^k, |A| = N\}.$$

The shadow-up of x is the following set:

$$up(x) = \{y : \ x \ is \ n - subword \ of \ y\}.$$

Now for any subset $B \subset \mathcal{X}^n$ we define the shadow-up

$$up(B) = \bigcup_{b \in B} up(b).$$

For fixed n and k we are interested in

$$\triangle(M) = \min\{|up(B)| : B \subset \mathcal{X}^n, |B| = M\}.$$

For an alphabet $\mathcal{X} = \{0, 1, \ldots, q-1\}$ we consider the set $\mathcal{X}^k$ of words $x^k = x_1 x_2 \cdots x_k$ of length k. For a word $a^k = a_1 a_2 \cdots a_k \in \mathcal{X}^k$ we define the left shadow

$$\mathrm{shad}^L(a^k) = a_2 \cdots a_k,$$

that is, the subword resulting from the omission of the first letter a_1 from a^k, and the right shadow

$$\mathrm{shad}^R(a^k) = a_1 \cdots a_{k-1},$$

that is, the subword resulting from the omission of the last letter from a^k.

We define the shadow of a^k by

$$\mathrm{shad}(a^k) = \mathrm{shad}^L(a^k) \cup \mathrm{shad}^R(a^k). \tag{1}$$

Now for any subset $A \subset \mathcal{X}^k$ we define the shadow

$$\mathrm{shad}(A) = \bigcup_{a^k \in A} \mathrm{shad}(a^k). \tag{2}$$

We are interested in finding optimal or at least asymptotically optimal lower bounds on the cardinality of N-sets $A \subset \mathcal{X}^k$, that is, the function

$$\nabla_k(q, N) = \min\{|\mathrm{shad}(A)| : A \subset \mathcal{X}^k, |A| = N\}.$$

We write in short $\nabla_k(N)$, if q is fixed, and $\nabla(N)$, if also k is fixed.

Example 1. For any $q_1 \leq q$ we have

$$\nabla_k(q_1^k) \leq q_1^{k-1}.$$

In particular, for $q_1 = 2$, $k = 4$ we have $\nabla_4(16) \leq 8$.

Example 2. In $\mathcal{X}B\mathcal{X}$ q words xby, $y \in \mathcal{X}$, have the same right shadow and, analogously, for left shadow.

$$\mathrm{shad}(\mathcal{X}0^{k-2}\mathcal{X}) = \mathcal{X}0^{k-2} \cup 0^{k-2}\mathcal{X}.$$

So we have $\nabla_k(q^2) \leq 2q - 1,\ k \geq 3$.

In particular, $\nabla_4(16) \leq 7$ that is better then in Example 1.

Clearly that

$$|\text{shad}(A)| \geq \frac{1}{q}|A|. \tag{3}$$

2 The Concept of Basic Sets

We *improve* the structure to building sets

$$\mathcal{X}^l 0^m \mathcal{X}^r \tag{4}$$

and by taking unions of such sets involving a strong symmetry property. We define now our main concept.

Definition 1. *For non-negative integers l (left), m, (middle), and r (right) satisfying $l \geq r$ and $k = l + m + r$, we define the basic set $\mathcal{B}(k, l, r)$ in $\mathcal{X}^k$ as follows:*

$$\mathcal{B}(k, l, r) = \bigcup_{s=0}^{l-r} \mathcal{X}^{l-s} 0^m \mathcal{X}^{r+s}. \tag{5}$$

$$\mathcal{B}(k,l,r) = \bigcup_{s=0}^{l-r} \mathcal{X}^{l-s} 0^m \mathcal{X}^{r+s}.$$

For example $\mathcal{B}(7,3,1)$ is the union of the rows in the matrix

$$\begin{matrix} \mathcal{X} & \mathcal{X} & \mathcal{X} & 0 & 0 & 0 & \mathcal{X} \\ \mathcal{X} & \mathcal{X} & 0 & 0 & 0 & \mathcal{X} & \mathcal{X} \\ \mathcal{X} & 0 & 0 & 0 & \mathcal{X} & \mathcal{X} & \mathcal{X} \end{matrix}$$

for $q = 2$

$|\mathcal{B}(7,3,1)| = 16 + 8 + 8 = 32$

$\operatorname{shad} \mathcal{B}(7,3,1) = \mathcal{B}(6,3,0)$

$$\begin{matrix} \mathcal{X} & \mathcal{X} & \mathcal{X} & 0 & 0 & 0 \\ \mathcal{X} & \mathcal{X} & 0 & 0 & 0 & \mathcal{X} \\ \mathcal{X} & 0 & 0 & 0 & \mathcal{X} & \mathcal{X} \\ 0 & 0 & 0 & \mathcal{X} & \mathcal{X} & \mathcal{X} \end{matrix}$$

$|\operatorname{shad} \mathcal{B}(7,3,1)| = 8 + 4 + 4 + 4 = 20$

Lemma *For every* $l \geq r \geq 1$, $m + r > l$, *that is,* $k = l + m + r > 2l$ *we have*

(i) $|\mathcal{B}(k,l,r)| = 2^{l+r} + 2^{l+r-1}(l-r) = 2^{l+r-1}(l-r+2)$ *for* $q = 2$

(i') $|\mathcal{B}(k,l,r)| = q^{l+r} + q^{l+r-1}(l-r)(q-1)$

(ii) $shad\,\mathcal{B}(k,l,r) = \mathcal{B}(k-1,l,r-1)$

(iii) $|shad\,\mathcal{B}(k,l,r)| = 2^{l+r-2}(l-r+3)$ *for* $q = 2$

(iii') $|shad\,\mathcal{B}(k,l,r)| = \frac{N}{q} + q^{l+r-2}(q-1)$

So we have the

Theorem 1. *For* $N = q^{l+r} + q^{l+r-1}(l-r)(q-1)$ *and* $k = l + m + r > 2l \geq 2r \geq 2$

$$\frac{1}{q}N \leq \nabla_k(q,N) \leq \frac{1}{q}\left(1 + \frac{1}{l-r+1}\right)N.$$

Actually, the lower bound holds for all N.

Definition 2. *Consider a set C of sequences of length n with cardinality M ($|C| = M$). Then $s_n(C, M)$ is the number of pairs (a, x), $a \in \mathcal{X}$, $x = (x_1, x_2, \ldots, x_n) \in C$ such that we have $(a, x_1, x_2, \ldots, x_{n-1}) \in C$ also. Denote by*

$$s_n(M) = max_C s_n(C, M). \tag{6}$$

Theorem 2. *For any q and k*

$$\nabla_k(s_{k-1}(M)) = M.$$

We have also

Theorem 3. *For any q and k*

$$\triangle(M) = 2qM - s(M).$$

Theorem 4. *For any q and k*

$$s_k(q^k - M) = q^{k+1} - 2qM + s_k(M).$$

References

[1] R. Ahlswede and N. Cai, Shadows and isoperimetry under the sequence-subsequence relation, Combinatorica 17 (1), 11-29, 1997.

[2] G. Katona, A theorem on finite sets, in: Theory of Graphs, Proc. Colloq. Tihany 1966, Akadémiai Kiadó, 187-207, 1968.

[3] J.B. Kruskal, The number of simplices in a complex, in: Mathematical optimization Techniques, Berkeley and Los Angeles, 251-278, 1963.

[4] B.A. Lindström and H.O. Zetterström, A combinatorial problem in the k-adic number systems, Proc. Amer. Math. Soc., Vol. 18, 166-170, 1967.

[5] B. Patkos and A. Chowdhury, Shadows and intersections in vector spaces, JCT A, to appear.