Sets of n–length 0–1–sequences with minimal shadow in $(n-1)$–length subsequences[1]

R. Ahlswede and N. Cai

For $\mathcal{X}^n = \prod_1^n \mathcal{X}$, the sequences of length n over the alphabet $\mathcal{X}$, we consider for sets $A \subset \mathcal{X}^n$ their shadow $\nabla A = \{x^{n-1} \in \mathcal{X}^{n-1} : x^{n-1}$ is subsequence of some $a^n \in A\}$. The goal is to find for given cardinalities sets of minimal cardinality of the shadow. It is *not optimal* to choose the segments of the B–G order (see Preprint 92–036).

Example: Choose $\mathcal{X} = \{0,1,2,\}$, $n = 3$, and notice that the 12^{th} initial segment in B–G order is $A = \{000, 100, 101, 001, 110, 101, 011, 111, 200, 020, 002, 210\}$ with $\nabla A = \{00, 10, 01, 11, 20, 02, 21\}$, $|\nabla A| = 7$. However, for $B = \{000, 100, 010, 001, 110, 101, 011, 111, 200, 210, 201, 211\}$, $|B| = 12$, we have $\nabla B = \{00, 10, 01, 11, 20, 21\}$, $|\nabla B| = 6$.

However, in the binary case the B–G order coincides with the H–order of [1]:

For any integer $u \in [0, 2^n]$ the u–th initial segment consists of all $x^n \in \{0,1\}^n$ with less than $n-k$ ones and all remaining elements with $n-k$ ones, whose complements are in the initial segment of the squashed order (used for instance in Kruskal–Katona).

As in the vertex isoperimetric problem in binary Hamming space ([1]) it is optimal also for our shadows of sets in $\{0,1\}^n$.

We use the unique binomial representation of an integer u

$$u = \binom{n}{n} + \cdots + \binom{n}{k+1} + \binom{\alpha_k}{k} + \cdots + \binom{\alpha_t}{t}; n > \alpha_k > \cdots > \alpha_t \geq 1, \quad (1)$$

and observe that for an initial, H–order segment S with $|S| = u$

$$|\nabla S| = \binom{n-1}{n-1} + \binom{n-1}{n-2} + \cdots + \binom{n-1}{k} + \binom{\alpha_k - 1}{k-1} + \cdots + \binom{\alpha_t - 1}{t-1} = \overset{\nabla}{G}(n,u), \text{ say.} \quad (2)$$

Theorem. *For every* $A \subset \{0,1\}^n$ $|\nabla A| \geq \overset{\nabla}{G}(n, |A|)$ *and the bound is achieved by the* u*–th initial segment in H–order.*

The proof is an immediate consequence of our main discovery, the

∇–Inequality: If $w_1 \leq w_0 < \overset{\nabla}{G}(n, w)$ and $w \leq w_0 + w_1$, then

$$\overset{\nabla}{G}(n, w) \leq \overset{\nabla}{G}(n-1, w_0) + \overset{\nabla}{G}(n-1, w_1). \quad (3)$$

[1] We are grateful to David Daykin for the communication on this subject.

Proof of Theorem by induction on n **:** For $n = 2$ (3) is readily verified. From the IH for $n-1$ we proceed to n. Define for $A \subset \{0,1\}^n$ and $B \subset \{0,1\}^{n-1}$

$A_i = \{x_1 \dots x_{n-1} : x_1 \dots x_{n-1} i \in A\}$ and $B * i = \{y_1 \dots y_{n-1} i : y_1 \dots y_{n-1} \in B\}$.

Next observe that $\bigcup_{i=0}^{1} (\nabla A_i) * i \subset \nabla A, \bigcap_{i=0}^{1} (\nabla A_i) * i = \emptyset$

and that therefore $|\nabla A| \geq \sum_{i=0}^{1} |\nabla A_i| \geq \sum_{i=0}^{1} \overset{\nabla}{G}(n-1, |A_i|)$ (by the IH).

According to the ∇–inequality this can be lower bounded with the desired $\overset{\nabla}{G}(n, |A|)$, if $|A_0|, |A_1| < \overset{\nabla}{G}(n, |A|)$. Otherwise we have for some i $|A_i| = \max(|A_0|, |A_1|) \geq \overset{\nabla}{G}(n, |A|)$ and we are done again, because $\nabla A \supset A_i$.

Proof of the ∇**–inequality :** Instead of presenting our original proof with fairly lengthly calculations with binomial coefficients, we derive the inequality from Lemma 6 of [2], which in turn makes use of an inequality of Eckhoff and Wegner. Define

$$G(n,u) = \binom{n}{n} + \binom{n}{n-1} + \dots + \binom{n}{k} + \binom{\alpha_k}{k-1} + \dots + \binom{\alpha_t}{t-1}; u \text{ as in (1)}. \quad (4)$$

Lemma 6 of [2]: If $0 \leq u_1 \leq u_2$ and $u \leq u_1 + u_2$, then

$$G(n,u) \leq \max(u_2, G(n-1,u_1)) + G(n-1,u_2).$$

Proof of ∇**–inequality :** If (3) does not hold, then

$$w - \overset{\nabla}{G}(n,w) < w_0 - \overset{\nabla}{G}(n-1,w_0) + w_1 - \overset{\nabla}{G}(n-1,w_1)$$

and with the convention $\overline{u}(n-1) = u - \overset{\nabla}{G}(n,u)$

$$\overline{w}(n-1) < \overline{w}_0(n-1) + \overline{w}_1(n-2). \quad (5)$$

Now (1), (2), and (4) imply that

$$G(n-1, \overline{u}(n-1)) = \overset{\nabla}{G}(n,u), \quad (6)$$

$$G(n-1, \overline{w}(n-1)) > G(n-2, \overline{w}_0(n-2)) + G(n-2, \overline{w}_1(n-2)). \quad (7)$$

Lemma 6, (5), and (7) yield

$$G(n-1, \overline{w}(n-1)) \leq \overline{w}_0(n-2) + G(n-2, \overline{w}_1(n-2))$$

or by our convention and (6) $\overset{\nabla}{G}(n,w) \leq w_0$, a contradiction to our hypothesis.

[1] L.H. Harper, "Optimal numberings and isoperimetric problems on graphs", J. Comb. Theory 1, 385–393, 1966.

[2] G.O.H. Katona, "The Hamming–sphere has minimum boundary", Studia Scientiarum Hungarica 10, 131–140, 1975.