

ASYMPTOTICALLY DENSE NONBINARY CODES CORRECTING A CONSTANT NUMBER OF LOCALIZED ERRORS

R. AHLWEDE, L.A. BASSALYGO, M.S. PINSKER

ABSTRACT

The binary case was studied in [1], but the method used there doesn't give the tight answer for nonbinary cases and we presented in [2] another method for the corresponding result. Here we formulate the main theorem and prove the auxiliary statements used in [2].

During the transmission of q -ary words of length n over the channel at most t errors occur, and the encoder knows the set E of t positions, where these errors are possible. The decoder doesn't know anything about these positions. Let $\mathcal{E}_t = \{E \mid E \subseteq \{1, 2, \dots, n\}, |E| = t\}$ be the set of all subsets from $\{1, 2, \dots, n\}$ of size t and let \mathcal{M} be a set of messages ($|\mathcal{M}| = M$). A code word $x(m, E)$ depends not only on the message $m \in \mathcal{M}$ but also on the configuration of possible errors E . So there exists the natural correspondence between the message $m \in \mathcal{M}$ and the list of code words $\bigcup_{E \in \mathcal{E}_t} \{x(m, E)\}$, which we use for the transmission of this message. Thus the code X for the set of messages \mathcal{M} represents a collection of M lists $\{\bigcup_{E \in \mathcal{E}_t} \{x(m, E)\}, m \in \mathcal{M}\}$. Since we can use the same word for different configurations, the size of a list can be essentially smaller than the size of the set \mathcal{E}_t ($|\mathcal{E}_t| = \binom{n}{t}$). Let us define the cylinder $C(a, A)$ with the base $a = (a_1, \dots, a_n)$ and the support A ($A \subseteq \{1, 2, \dots, n\}$) as the set of words (y_1, \dots, y_n) with $y_i = a_i$, if $i \in A$. It is clear that the size of the cylinder $C(a, A)$ is equal to $q^{|A^c|}$ and the number of different cylinders with the same support A is equal to $q^{n-|A|}$.

As a result of the transmission of the codeword $x(m, E)$ every word of $C(x(m, E), E)$ can appear as output of the channel. The code X corrects t localized errors, if the decoder can correctly recover every message $m \in \mathcal{M}$. The following condition is necessary and sufficient for it:

$$C(x(m, E), E) \cap C(x(m', E'), E') = \emptyset \text{ for all } E, E' \in \mathcal{E}_t, m, m' \in \mathcal{M}, m \neq m'. \quad (1)$$

The maximal number of messages, which we can transmit by a code correcting t localized errors, is denoted by $L_q(n, t)$.

Proposition 1:

$$L_q(n, t) \leq \frac{q^n}{S_t}$$

where $S_t = \sum_{i=0}^t (q-1)^i C_n^i$ is the size of a sphere of radius t in the Hamming n -space.

A proof of this bound in the q -ary case can be given as for the binary case in [3] or [4]. The key inequality there has the following generalization.

Lemma 1. *Let $C(a_i, A_i), \dots, C(a_T, A_T)$ be cylinders with pairwise different supports $A_i \neq A_j, i \neq j$. Then for the size of the union of the cylinders*

$$\left| \bigcup_{i=1}^T C(a_i, A_i) \right| \geq \sum_{i=1}^T (q-1)^{|A_i|}.$$

Proof: We proceed by an induction on n . For $n = 1$ the statement is obvious. Let $C(a_1, A_1), \dots, C(a_T, A_T)$ satisfy the condition of the Lemma. We consider now the new family \mathbb{C} of cylinders:

- a) if $n \notin A_i$, then $C(a_i, A_i) \in \mathbb{C}$
- b) if $n \in A_i$, then $C(a_i^{(k)}, A_i \setminus n) \in \mathbb{C}$ for all $k (k = 0, 1, \dots, q-1)$, where $a_i^{(k)} = (a_{i1}, \dots, a_{in-1}, k)$.

We have

$$\mathbb{C} = \bigcup_{k=0}^{q-1} \mathbb{C}^{(k)}$$

where $\mathbb{C}^{(k)}$ – all cylinders from \mathbb{C} whose last coordinate is equal to $k (k = 0, 1, \dots, q-1)$. It is easy to show that

$$\left| \bigcup_{C \in \mathbb{C}} C \right| = \sum_{k=0}^{q-1} \left| \bigcup_{C \in \mathbb{C}^{(k)}} C \right|.$$

It follows from the condition of the Lemma that the support $A_i \setminus n$ of q cylinders $C(a_i^{(k)}, A_i \setminus n), k = 0, 1, \dots, q-1$ differs from the support of other cylinders at least for $q-1$ subfamilies $\mathbb{C}^k, k = 0, 1, \dots, q-1$. Thus one proves the Lemma using the induction step to estimate $|\bigcup_{C \in \mathbb{C}^{(k)}} C|$. It is easy to obtain Proposition 1 from Lemma 1. In fact for every union of cylinders $\bigcup_{E \in \mathcal{E}_t} C(\cdot, E)$ there exists some union of cylinders $\bigcup_{E \in \bigcup_{i=0}^t \mathcal{E}_i} C(\cdot, E)$ such that

$$\bigcup_{E \in \mathcal{E}_t} C(\cdot, E) = \bigcup_{E \in \bigcup_{i=1}^t \mathcal{E}_i} C(\cdot, E)$$

and therefore by Lemma 1

$$\left| \bigcup_{E \in \mathcal{E}_t} C(\cdot, E) \right| = \left| \bigcup_{E \in \bigcup_{i=1}^t \mathcal{E}_i} C(\cdot, E) \right| \geq S_t. \quad (2)$$

Now we have from the condition (1) that

$$\left(\bigcup_{E \in \mathcal{E}_t} C(x(m, E), E) \right) \cap \left(\bigcup_{E' \in \mathcal{E}_t} C(x(m', E'), E') \right) = \emptyset \text{ for } m \neq m'.$$

From here and (2) the Proposition follows.

The following lower bound can be easily deduced by the standard greedy algorithm (maximal coding).

Proposition 2:

$$L_q(n, t) \geq \frac{q^n}{q^{2t} \binom{n}{t}}.$$

Proof: Let X be the code $X = \{\bigcup_{E \in \mathcal{E}_t} \{x(m, E)\}, m \in \mathcal{M}\}$ for M messages, correcting t localized errors. As

$$\left| \bigcup_{m \in \mathcal{M}} \bigcup_{E \in \mathcal{E}_t} C(x(m, E), E) \right| \leq M \cdot \binom{n}{t} \cdot q^t$$

and the number of different cylinders with the same support $E' (|E'| = t)$ is equal to q^{n-t} , for any support $E' \in \mathcal{E}_t$ there exists a cylinder $C(a, E')$ with

$$C(a, E') \cap C(x(m, E), E) = \emptyset \text{ for all } m \in \mathcal{M} \text{ and } E \in \mathcal{E}_t,$$

if

$$M \binom{n}{t} q^t < q^{n-t}. \quad (3)$$

Therefore, if the inequality (3) takes place, it is possible, according to the condition (1), to construct the code for $M + 1$ messages, correcting t localized errors. Hence Proposition 2 follows.

Already Proposition 1 and 2 imply the asymptotic equivalence within a constant

$$L_q(n, t) \asymp \frac{q^n}{n^t}, \text{ when } t \text{ is fixed and } n \rightarrow \infty.$$

We draw attention to the fact that this equivalence is known for nonbinary error-correcting codes, except $t = 1$, only for $t = 2$ and $q = 3, 4$ [5].

The following theorem gives the precise constant in the equivalence.

Theorem. *For every constant t*

$$L_q(n, t) = \frac{q^n}{S_t} (1 + o(1)) = \frac{t!}{(q-1)^t} \cdot \frac{q^n}{n^t} (1 + o(1)),$$

where $o(1) \rightarrow 0$ as $n \rightarrow \infty$ ($o(1)$ depends certainly on t and q).

When proving this theorem [2] we referred to the following recurrence relation, having an independent interest.

Lemma 2. *If $N \leq (q - 1)T + 1$, then*

$$L_q(N + T, 1) \geq q^{N-1} L_q(T, 1).$$

Proof: On the first N positions we always transmit a parity check ($\bmod q$) code of the size q^{N-1} . The last T positions we reserve for the code, which allows us to transmit $L_q(T, 1)$ messages and to correct a single localized error. The method of transmission depends on the position of the localized error in the following way:

- a) The error is on the last T positions. We use our code as a code, correcting a single localized error.
- b) The error is in the first N positions. We use our code for the transmission of both, the message and the number of the position, where the error can occur. According to Proposition 1 and Lemma 2 at least $(q - 1)T + 1$ different words in the output of the channel of length T correspond to everyone of $L_q(T, 1)$ messages and therefore we can make the successful transmission, if

$$N \leq (q - 1)T + 1.$$

The proof is complete.

REFERENCES

1. L.A. Bassalygo, S.I. Gelfand, M.S. Pinsker, Simple methods of deduction of lower bounds in coding theory, *Probl. Peredaci Inform.*, vol. 27, 1991, no. 4, 3–8.
2. R. Ahlswede, L.A. Bassalygo, M.S. Pinsker, Nonbinary codes correcting localized errors, *IEEE Inf. Theory*, in print.
3. L.A. Bassalygo, S.I. Gelfand, M.S. Pinsker, Coding for channels with localized errors, *Proc. 4-th Soviet-Swedish Workshop in Information Theory*, Gotland, Sweden, 1989, 95–99.
4. L.A. Bassalygo, S.I. Gelfand, M.S. Pinsker, Coding for partially localized errors, *IEEE Inf. Theory*, vo. 37, no. 3, May 1991, 880–884.
5. I.I. Dumer, V.A. Zinoviev, New maximum codes over $GF(4)$, *Probl. Peredaci Inform.*, vol. 14, 1978, no. 3, 24–34.