Nonstandard coding method for nonbinary codes correcting localized errors

R. Ahlswede, L. Bassalygo, M. Pinsker

I. Abstract

In [1] we claimed that the transmission rate of a nonbinary (q-ary) code of length n which corrects τn localized errors asymptotically equals the Hamming bound on an interval $\tau \in [0, \tau_0]$, $\tau_0 \leq 1/2$, and conjectured that $\tau_0 = 1/2$ (transmission rate is zero if the number of errors is greater than or equal to $n/2$). Though we have not succeeded in proving our conjecture we decided to promulgate the derivation of the incomplete result on the Hamming bound, the more so, as the method of the proof itself is of independent interest.

The situation with localized errors (when the encoder knows the positions where errors can occur) is characterized by the following fact: to attain the Hamming bound asymptotically, it suffices to provide the decoder with "small" information. Indeed, let us divide the transmitted segment into a growing number of segments of equal length (the length is also growing) and arrange the segments in an ascending order with respect to the number of possible errors on them. The number of the first segment and the number of a set that covers the positions of possible errors on this segment are precisely the "small" information that should be known to the decoder. Now, let us explain why this information is sufficient to attain the Hamming bound. On the first segment, outside the positions of the covering set, we transmit to the decoder the number of the second segment and the number of a set that covers the error positions on this segment. The remaining positions of the first segment are used for message transmission. On the second segment we do the same procedure, and so on. Surely one can ask why we transmit covering sets but not the actual positions of possible errors, which, in fact, would leave more positions for message transmission. The answer is simple: with the optimal choice of covering sets, the gain due to their rather small number compensates the loss of positions for message transmission. Simple computations show that thus we attain the Hamming bound.

Thus, the main problem is to transmit the mentioned "small" information to the decoder. We overcome this problem with the help of special nonstandard encoding and decoding procedures. In encoding, we first construct three auxiliary code words from which the transmitted word is constructed. In decoding, we construct three auxiliary "de-code" words from the word received and then reconstruct the transmitted message using them.

Transmission of "small" information is performed with the help of binary constant-weight

codes that correct localized errors and defects (see [2]). To employ these codes, we proceed as follows: start the above-mentioned encoding procedure not from the first segment but from the $(k+1)$st, leaving the first k segments free (the local length of the first k segments is small). Then this encoding procedure generates the first q-ary codeword outside the first k segments and in positions outside the covering sets on the other segments. The unity positions of this q-ary codeword (i.e. the positions with the symbol 1 in them) are considered as the defect positions for the second binary codeword that we are going to construct (we can fix the number of unity positions of the first codeword beforehand). As the set of positions of localized errors for the second word we consider, together with actual such positions, all positions of the first k segments. According to [2], in this case we can choose the second binary word to be of small weight w (since, in fact, we have to transmit "small" information, namely, information on the $(k+1)$st segments and also the numbers of the first k segments), where these w unity symbols lie outside the positions of defects and localized errors. On the first k segments, we construct the third codeword of the binary localized-error-correcting code in order to transmit information on the location of w unity symbol of the second codeword and which q-ary symbols of the first codeword are in these positions. Now, from three codewords constructed, we can construct a q-ary codeword to be transmitted:

(a) on the first k segments, we transmit the third codeword;

(b) from the second codeword, we transmit only w unity symbols;

(c) in the remaining positions, we transmit symbols of the first codeword.

In decoding, from the received word we first construct the second de-code binary word which corresponds to the second codeword transmitted over the above-described channel with defects and localized errors. To do this, we replace all symbols from 2 to $q-1$ of the received word by 0. From this word, the decoder reconstructs the "small" information to transmit which was the main difficulty. In particular, this "small" information includes the nubmers of the first k segments which become known to the decoder, and thus he can construct the third de-code word which coincides on the first k segments with the received word. From this word, the decoder reconstructs the symbols of the first codeword in those w positions where the unity symbols of the second codeword were transmitted. It is clear now how to construct the first de-code word: in these w positions it coincides with the first codeword and in the remaining positions outside the first k segments it coincides with the received word. Now, from this first de-code word, the decoder successively reconstructs the message starting from the $(k+1)$st segment since the "small" information previously reconstructed includes both the number of the $(k+1)$st segment and the number of the set that covers the possible error positions on this segment.

Now we proceed to the strict statement. Let $Q = \{0, 1, \ldots, q-1\}$ be the alphabet, B be the set of q-ary sequences of length n, $M = \{m\}$ be the set of messages, $\mathcal{E}_t\{E \mid E \subseteq [1, 2, \ldots, n], |E| = t\}$ be the set of all possible collections of error positions of multiplicity $t(|\mathcal{E}_t| = \binom{n}{t})$, and let $V(E)$ be the set of q-ary words of length n that are equal to zero in positions outside W $(|V(E)| = q^t)$. Since, while encoding, we know those t positions where errors can occur, a codeword $x(m,E)$ depends on $m \in M$ and $E \in \mathcal{E}_t$. A code $X = \{x(m,E), m \in M, E \in \mathcal{E}_t\}$ corrects t localized errors is the condition

$$x(m,E) + e \neq x(m',E') + e'$$

holds for all $E, E' \in \mathcal{E}_t$, $e \in V(E)$, $e' \in V(E')$, $m, m' \in M$, $m \neq m'$ (the addition is made modulo q). It is known [1] that the maximum transmission rate R of such a code does not exceed the Hamming bound

$$R \leq 1 - h_q(\tau),$$

where

$$h_q(\tau) = -\tau \log_q \tau - (1-\tau)\log_q(1-\tau_+\tau \log_q(q-1), t = \tau n (0 \leq \tau \leq 1/2).$$

Theorem. *Let $0 < \tau < 1/2 - \frac{q-2}{2q(2q-3)}$. Then, for any $\varepsilon > 0$, there is a number $n(\varepsilon)$ such that for $n > n(\varepsilon)$ a code of length n with transmission rate $1 - h_q(\tau) - \varepsilon$ exists which corrects τn localized errors.*

References

[1] R. Ahlswede, L.A. Bassalygo, M.S. Pinsker, Nonbinary codes correcting localized errors, IEEE Trans. Inform. Theory, Vol. 39, No. 4, 1413-1416, 1993.

[2] R. Ahlswede, L.A. Bassalygo, M.S. Pinsker, Binary constant weight codes correcting localized errors and defects, Probl. Inf. Trans., Vol. 30, No. 2, 10-13, 1994.