Coding for channels with localized errors

Coding for channels with localized errors

J. H. VAN LINT

I. Introduction.

The following problem was discussed in a lecture by S. I. Gelfand in Oberwolfach (Information Theory, May 1989), based on joint work with L. A. Bassalygo and M. S. Pinsker. We consider a binary channel and we are interested in codes of length n. Let t be given, $0 < t < n$. Before a message is transmitted the *sender* is given a subset E of cardinality at most t of the positions $\{1, 2, \ldots, n\}$ in which errors *may* occur (i.e. outside E all bits are received correctly). The receiver does not know E but sender and receiver have a prearranged codebook that is used for transmission of M possible messages. The question is to determine $F_t(n) :=$ the maximal value of M for which communication over this channel with a code of length n is possible.

More formally we have a message set $\mathcal{M} := \{1, 2, \ldots, M\}$ and a coding function $\phi : \mathcal{M} \times \mathcal{P}_t(n) \to \{0,1\}^n$, where $\mathcal{P}_t(n)$ denotes the collection of subsets of $\{1, 2, \ldots, n\}$ of size $\leq t$. The decoding function ψ is such that for all $m \in \mathcal{M}$ and all $E \in \mathcal{P}_t(n)$ and each error vector $\mathbf{e}$ that is 0 outside E we have

$$\psi(\phi(m, E) + \mathbf{e}) = m. \tag{1.1}$$

The maximal value of M for which such functions ϕ and ψ exist is $F_t(n)$.

Let $t/n = \tau$. We define the asymptotic rate $R(\tau)$ by

$$R(\tau) := \limsup_{n \to \infty} n^{-1} \log F_t(n). \tag{1.2}$$

The following theorem was announced.

THEOREM 1. *We have*

$$\frac{1}{2n} \cdot \frac{2^n}{\sum_{i=0}^{t} \binom{n}{i}} \leq F_t(n) \leq \frac{2^n}{\sum_{i=0}^{t} \binom{n}{i}}. \tag{1.3}$$

COROLLARY 1. $R(\tau) = 1 - h(\tau)$.

II. The case $t = 1$.

The situation for $t = 1$ is quite surprising. Say that just before transmission, the sender is told that the fifth bit *may* be received in error. If the sender was *certain* that this would happen, he could simply change

that bit before sending it, ensuring correct reception. So, he misses only *one* bit of information, namely whether the error will occur or not. An optimist would think that a rate of $1 - \frac{1}{n}$ might be achievable. On the other hand, the sender can be lazy and agree with the receiver to use a 1-error- correcting code. In that case he does not even have to know in which position the possible error can occur. This sounds like a stupid scheme, i. e. one expects to be able to do much better. As the right-hand side of (1.3) shows, this is *not* the case! Hamming codes are optimal for this problem. We present our own proof of this fact.

LEMMA 1. *Let $1 \leq k \leq n$. Let a subset S of size k of $\{1, 2, \ldots, n\}$ be given. Let $\mathbf{e}_i$ denote the error vector of weight 1 with a one in position i and define*

$$R_i(m) := \{\phi(m, \{i\}), \phi(m, \{i\}) + \mathbf{e}_i\},$$

$$R(m) := \bigcup_{i \in S} R_i(m). \tag{2.1}$$

Then for all m we have $|R(m)| \geq k + 1$.

PROOF: Note that $R(m)$ is the set of all possible received messages if m is sent and the error position is restricted to S. Let $m \in \mathcal{M}$. For $k = 1$ the assertion is trivial. Let the assertion be true for all values of k less than l. We prove (2.1) for $k = l$. Let $|S| = l$. Form a bipartite graph on $S \cup R(m)$ with an edge from i to the two elements of $R_i(m)$. By the induction hypothesis we know that $|R(m)| \geq l$. Assume that $|R(m)| = l$. Now, for any subset A of S there are at least $|A|$ vertices in $R(m)$ joined to some vertex $a \in A$. (In fact one more if $A \neq S$.) So, Hall's condition is satisfied, showing that there is a *matching* from S to $R(m)$. Without loss of generality this is a matching from $i \in S$ to $\phi(m, \{i\})$. Since every vertex of S has degree 2, we can form a circuit by alternating between edges $\{i, \phi(m, \{i\})\}$ and edges $\{i, \phi(m, \{i\}) + \mathbf{e}_i\}$. This implies that for some subset A of S, we have $\sum_{i \in A} \mathbf{e}_i = \mathbf{0}$, which is absurd. □.

Clearly, the lemma shows that $F_1(n) \leq 2^n/(n+1)$, since the sets $R(m)$, $m \in \mathcal{M}$, must be disjoint.

III. The upper bound.

We did not succeed in generalizing the idea of using Hall's theorem to prove the upper bound of Theorem 1. However, we shall give a simple proof of a generalization of Lemma 1.

LEMMA 2. *Let P be any collection of error patterns and suppose that the sender knows that some error pattern from P can occur and in fact*

is told which one just before transmission. As in the previous section, we denote by $R(m)$ the set of all possible received messages (under these conditions) if m is sent (for some code that works). Then $|R(m)| \geq |P|$.

PROOF: Let $\phi^*(m, E)$ denote the word in $\{0, 1, *\}^n$ that we obtain from the codeword $\phi(m, E)$ for message m and error pattern E by replacing the entries in positions of E by the symbol $*$. The possible received messages form the set $R_E(m)$ obtained by "*filling*" the $*$'s in all possible ways by 0's and 1's. The assertion of the lemma states that a set of p elements of $\{0, 1, *\}^n$ such that no two of these have the same "$*$-pattern" yields a set of at least p different fillings. This is obviously true if the p words are from $\{0, *\}^n$, since replacing the $*$'s by 1's yields at least p different words. Now suppose that we have an arbitrary collection of p words from $\{0, 1, *\}^n$. We order these as follows: first those ending in 0, then those ending in $*$, and finally those ending in 1. We make a second list by replacing the final 1 in the bottom of the list by a 0. For each of these lists, we make the corresponding set of fillings, starting at the top and working our way down; any filling that has appeared earlier is of course not listed a second time. These two procedures are clearly identical until the part is reached where final 1's were replaced by 0's. If at this stage a filling of an element of the original list is rejected because it appeared earlier, then the corresponding filling of the (corresponding) element of the new list will also be rejected. (The earlier appearances are caused by elements ending in a $*$.) By repeating this procedure, we see that the trivial situation, where the symbol 1 is not used, is actually the worst case. □

Obviously, the right-hand side of Theorem 1 is an immediate consequence of Lemma 2. This same lemma was used by Gelfand in his proof but the proof of the lemma was different (I think). The idea of replacing 1's by 0's was suggested to me by L. Tolhuizen.

IV. The lower bound.

PROPOSITION. *I cannot prove the lower bound (yet).*