

Lecture 11: Shannon vs. Hamming

September 21, 2007

Lecturer: Atri Rudra

Scribe: Kanke Gao & Atri Rudra

In the last lecture, we proved the positive part of Shannon's capacity theorem for the BSC. We showed that by the probabilistic method, there exists an encoding function E and a decoding function D such that

$$\mathbb{E}_{\mathbf{m}} \Pr_{\substack{\text{noise } \mathbf{e} \\ \text{of } BSC_p}} [D(E(\mathbf{m}) + \mathbf{e}) \neq \mathbf{m}] \leq 2^{-\delta'n}. \quad (1)$$

In other words, the *average* decoding error probability is small. However, we need to show that the *maximum* decoding error probability over all messages is small. In the last lecture, we quickly went over how (1) implies the latter. We will start today's lecture by going over this argument again.

1 Shannon's Capacity Theorem (Cont.)

As was mentioned in the last lecture, the trick is to throw away all the messages that have high error probability. In particular, we only keep the messages with probability error at most the median error probability.

Claim 1.1. *Let the messages be ordered by $\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_{2^k}$ and define*

$$P_i = \Pr_{\substack{\text{noise } \mathbf{e} \\ \text{of } BSC_p}} [D(E(\mathbf{m}_i) + \mathbf{e}) \neq \mathbf{m}_i].$$

Assume that $P_1 \leq P_2 \leq \dots \leq P_{2^k}$ and (1) holds, then $P_{2^{k-1}} \leq 2 \cdot 2^{-\delta'n}$

Proof. By the definition of P_i ,

$$\begin{aligned} \frac{1}{2^k} \sum_{i=1}^{2^k} P_i &= \mathbb{E}_{\mathbf{m}} \Pr_{\substack{\text{noise } \mathbf{e} \\ \text{of } BSC_p}} [D(E(\mathbf{m}) + \mathbf{e}) \neq \mathbf{m}] \\ &\leq 2^{-\delta'n}, \end{aligned} \quad (2)$$

where (2) follows from (1). For the sake of contradiction assume that

$$P_{2^{k-1}} > 2 \cdot 2^{-\delta'n}. \quad (3)$$

So,

$$\frac{1}{2^k} \sum_{i=1}^{2^k} P_i \geq \frac{1}{2^k} \sum_{i=2^{k-1}+1}^{2^k} P_i \quad (4)$$

$$> \frac{2 \cdot 2^{-\delta' n} \cdot 2^{k-1}}{2^k} \tag{5}$$

$$> 2^{-\delta' n}, \tag{6}$$

where (4) follows by dropping half the summands from the sum. (5) follows (3) and the assumption on the sortedness of P_i . The proof is now complete by noting that (6) contradicts (2). \square

Thus, our final code will have as its messages $\mathbf{m}_1, \dots, \mathbf{m}_{2^{k-1}}$ and thus, has dimension $k' = k - 1$. Define $\delta = \delta' + \frac{1}{n}$. In the new code, maximum error probability is at most $2^{-\delta n}$. Also if we picked $k \leq \lfloor (1 - H(p + \varepsilon)) n \rfloor + 1$, then $k' \leq \lfloor (1 - H(p + \varepsilon)) n \rfloor$, as required.

Remark 1.2. One can also show that for the q -SC $_p$, the capacity is $1 - H_q(p)$ and for the BEC $_\alpha$, the capacity is $1 - \alpha$.

Remark 1.3. We have shown that a random code can achieve capacity. However, we do not know of even an succinct representation of general codes. A natural question to ask is if random linear codes can achieve the capacity of BSC $_p$. The answer is yes and the proof is left as an exercise.

For linear code, representation and encoding are efficient. But the proof does not give an explicit construction. Further, Shannon's proof uses MLD for which only exponential time implementations are known. Thus, the biggest question left unsolved by Shannon's work is the following.

Question 1.4. Can we come up with an explicit construction of a code of rate $1 - H(p + \varepsilon)$ with efficient decoding and encoding algorithms that achieve reliable communication over BSC $_p$?

As a baby step towards the resolution of the above question, one can ask the following question:

Question 1.5. Can we come up with an explicit construction with $R > 0$ and $p > 0$?

Note that the question above is similar to the $R > 0$ and $\delta > 0$ question in Hamming's world. Elias, answered the above question in the affirmative [1]. His code construction uses a clever combination of Hadamard codes. Unfortunately, we do not have the time to go through the construction.

2 Hamming vs. Shannon

As a brief interlude, let us compare the salient features of the works of Hamming and Shannon,

HAMMING	SHANNON
Focus on codewords itself	Directly deals with encoding and decoding functions
Looked at explicit codes	Not explicit at all
Fundamental trade off: rate vs. distance (easier to get a handle on this)	Fundamental trade off: rate vs. error
Worst case errors	Stochastic errors

We note the connection between finding codes of high distance and achieving reliable communication over BSC $_p$ in the following proposition

Proposition 2.1. *Let $0 \leq p < \frac{1}{2}$ and $0 < \varepsilon \leq \frac{1}{2} - p$. If an algorithm A can handle $p + \varepsilon$ fraction of worst case errors, then it can be used for reliable communication over BSC_p*

Proof. By the Chernoff bound, with probability $\leq 1 - 2^{-\frac{\varepsilon^2 n}{3}}$, fraction of errors in BSC_p is $\leq p + \varepsilon$. Then by assumption on A , it can be used to recover the transmitted message. \square

Note that the above result implies that one can have reliable transmission over BSC_p with any code of relative distance $2p + \varepsilon$ (for any $\varepsilon > 0$).

Remark 2.2. *The converse of Proposition 2.1 is also true. More precisely, if the decoding error probability is exponentially small for the BSC, then the corresponding code must have constant relative distance. The proof of this claim is left as an exercise.*

3 Singleton Bound

Recall that in the Hamming world we are interested in the trade-off between the rate and the distance of a code. We will approach this trade-off in the following way: If we fix the relative distance of the code to be δ , what is the best rate R ? We will first look at some upper bounds.

We begin by considering the trade-off between R and δ given by the Hamming bound. Recall that we proved the following:

$$\frac{k}{n} \leq 1 - \frac{\log_q \text{Vol}_q(\mathbf{0}, \lfloor \frac{d-1}{2} \rfloor)}{n}$$

By the following lower bound on the volume of a Hamming ball (which we have seen in the proof of the converse of the Shannon theorem):

$$\text{Vol}_q(\mathbf{0}, \lfloor \frac{d-1}{2} \rfloor) \geq q^{H_q(\frac{\delta}{2})n - o(n)},$$

we obtain the following asymptotic version of the Hamming bound:

$$R \leq 1 - H_q\left(\frac{\delta}{2}\right) + o(1).$$

We will now prove another upper bound called the Singleton bound.

Theorem 3.1 (Singleton Bound). *For every $(n, k, d)_q$ code, $k \leq n - d + 1$.*

Proof. Let c_1, c_2, \dots, c_M be the codewords of an $(n, k, d)_q$ code C . Note that we need to show $M \leq q^{n-d+1}$. To this end, we define c'_i to be the prefix of the codeword c_i of length $n - d + 1$. See Figure 1 for a pictorial description.

We now claim that for every $i \neq j$, $c'_i \neq c'_j$. For the sake of contradiction, assume that there exists an $i \neq j$ such that $c'_i = c'_j$. Note that this implies that $\Delta(c_i, c_j) \leq d - 1$, which contradicts the fact that C has distance d . Thus, M is the number of prefixes of codewords in C of length $n - d + 1$, which implies that $M \leq q^{n-d+1}$ as desired. \square

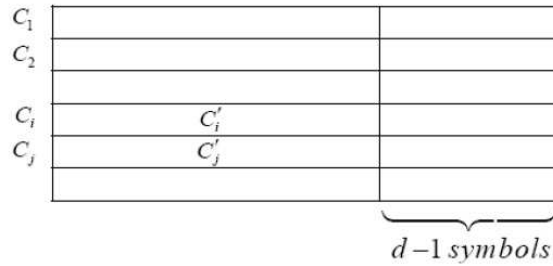


Figure 1: Construction of a new code in the proof of the Singleton bound.

Remark 3.2. *The asymptotic version of the Singleton bound states that $R \leq 1 - \delta + o(1)$. It is worth noting that the bound is independent of the alphabet size. It turns out that the upper bound is worse than the Hamming bound for binary codes. However, this bound is better for larger alphabet sizes. In fact, we will look at a family of codes called Reed-Solomon codes (which are defined over large alphabets) that meet the Singleton bound.*

References

- [1] Peter Elias. Error-free coding. *IEEE Transactions on Information Theory*, 4(4):29–37, 1954.