How great codes can you find?

Ádám Kunos and Fanni Nedényi

Winter School, University of Würzburg, Germany
Bolyai Institute, University of Szeged, Hungary

Basic setup

We consider the following fundamental problem.
Sender: has a message, encodes it to a codeword, and sends it to a recipient through a noisy channel. Some parts of the codeword might get corrupted.
Recipient: decodes the received word. He wants to get the original message.
Goal: construct a coding and decoding method that returns the original message with high probability.
It is reasonable to suppose that our encoded messages, called **codewords**, are just sequences of bits of the same length, so the set C of codewords is a subset of $\mathbb{F}_2^n$ ($\mathbb{F}_2$ denotes the field of two elements) for some positive integer n.

Measuring the goodness of codes

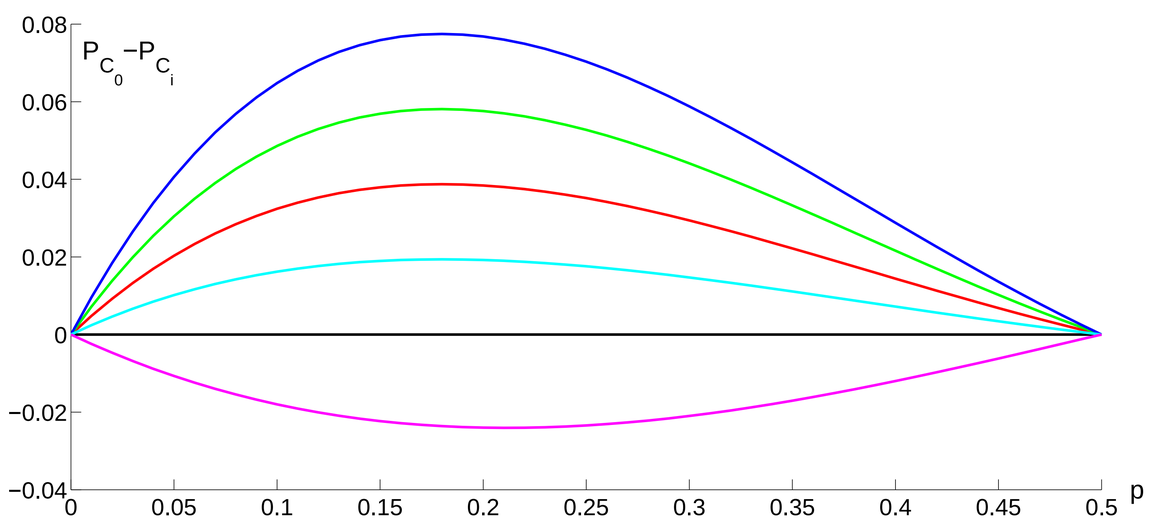
Binary Symmetric Channel (BSC): the bits are corrupted with a probability $0 \leq p \leq 1$, independently of each other. This p is called the **bit error rate**. If we have a BSC with $p < 1/2$ then the best possible decoding is just taking the codeword that differs in the least number of bits from the received word.
A way to measure goodness: Let $P_{w,C}$ denote the probability of wrongly decoding the codeword $w \in C$. The probability that a randomly chosen codeword gets wrongly decoded is the **word error rate**

$$P_C = \frac{1}{|C|} \sum_{w \in C} P_{w,C}.$$

The smaller the probability P_C, the better the code C.
Note that there exist other ways to measure the goodness of codes.

Example

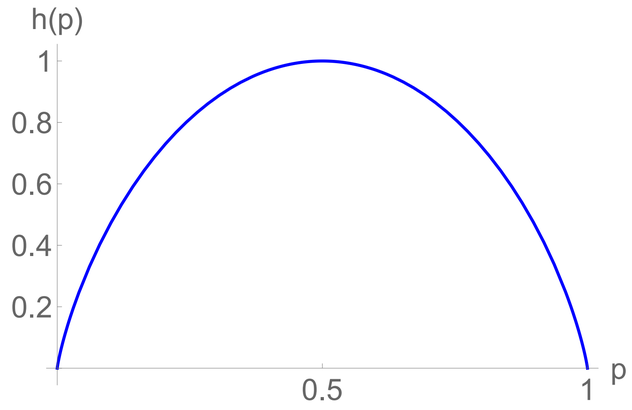
Suppose that we have a BSC with bit error rate p. Let us have four messages, and codewords of length 4. We wish to compare some such codes in terms of the goodness defined above.
Fix $C_0 = \{0000, 0011, 1100, 1111\}$. One can easily calculate $P_{C_0} = 2p - p^2$. We take some other codes $C_i \subseteq \mathbb{F}_2^4$, $i = 1, \ldots, 5$. The following figure shows the functions $P_{C_0} - P_{C_i}$, $i = 0, \ldots, 5$.



Shannon's theorems

Linear code: A code $C \subseteq \mathbb{F}_2^n$ is said to be linear if C is a linear subspace of $\mathbb{F}_2^n$. The dimension of the subspace C is called the dimension of the linear code.
Balanced family of linear codes: Consider a family of linear codes $\mathcal{F}$ with dimension k and length n. $\mathcal{F}$ is called balanced if every $0 \neq v \in \mathbb{F}_2^n$ is contained in the same number of codes in $\mathcal{F}$. Note that the set of k-dimensional subspaces of $\mathbb{F}_2^n$ is a balanced family of linear codes.
Entropy function: Consider the entropy function

$$h(p) = -p \log_2 p - (1-p) \log_2 (1-p), \qquad 0 \leq p \leq 1.$$



The function captures the uncertainty of the outcome depending on p.

- $h(0) = 0 = \min_{0 \leq p \leq 1} h(p)$;
- $h(1/2) = 1 = \max_{0 \leq p \leq 1} h(p)$;
- $h(1) = 0 = \min_{0 \leq p \leq 1} h(p)$;

Shannon's first theorem

Let $0 < R < 1 - h(p)$ and $\mathcal{F}_n$ be a balanced family of linear codes with codewords of length n and dimension $\lfloor Rn \rfloor$. Then

$$\min_{C \in \mathcal{F}_n} P_C \to 0, \qquad n \to \infty.$$

Information rate: $R_C = \dim(C)/n$.
The theorem means that if the information rate is small enough, then by increasing the length of the codewords one can achieve arbitrarily small word error rate.

Shannon's second theorem

If $C_n \subseteq \mathbb{F}_2^n$ is a sequence of codes such that for some fixed K $1 - h(p) < K \leq R_{C_n} \leq 1$ holds, then $\lim_{n \to \infty} P_{C_n} = 1$.

Conclusion

Together the two theorems of Shannon mean that for a fixed bit error rate p, the constant $1 - h(p)$ serves as a bound for which the following holds. For a rate $R < 1 - h(p)$ there exist linear codes such that their word error rate gets arbitrarily close to 0. On the other hand, for a rate bigger than $1 - h(p)$ one can not find good long codes.
Roughly speaking, this means that it suffices to consider only linear codes in terms of the goodness defined above. This is good news, since linear codes have much more structure than arbitrary codes, making them easier to deal with.

University of Szeged
H-6720 Szeged, Dugonics tér 13
www.u-szeged.hu
www.palyazat.gov.hu

TÁMOP-4.1.1.C-12/1/KONV-2012-0005 project
"Preparation of the concerned sectors for educational and R&D activities related to the Hungarian ELI project"

SZÉCHENYI 2020
HUNGARIAN GOVERNMENT
European Union
European Social Fund
INVESTING IN YOUR FUTURE