Aalto University
School of Science
and Technology

Coding Theory on Non-Standard Alphabets

A Brief Introduction into Traditional Coding Theory

Marcus Greferath

Department of Mathematics and Systems Analysis
Aalto University School of Sciences
marcus.greferath@aalto.fi

May 2023

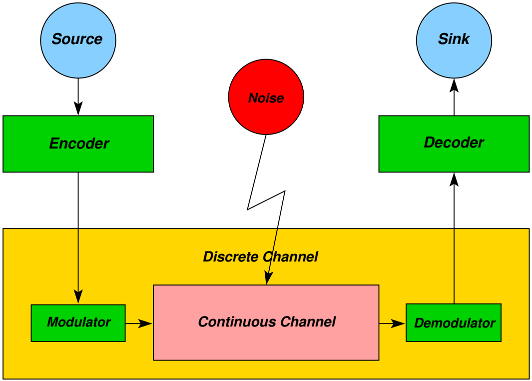
Block Codes

- Let F be a q-element set, referred to as the alphabet.
- A block code of length n is a non-empty subset $C \subseteq F^n$.
- The size of C is $M := |C|$, and

  $$R := \frac{1}{n} \log_q(M)$$

  is called the rate of C.
- The elements of C are referred to as codewords.
- F^n is called the ambient space of C.

- The following illustrates the general structure of a communication scheme.



- Let $C \subseteq F^n$ be a block code. A maximum likelihood decoder is a partial mapping $f : F^n \longrightarrow C$ such that for all $z \in C$ there holds

$$\mathrm{Prob}(f(z) \text{ trans} \mid z \text{ rec}) = \max_{c \in C} \mathrm{Prob}(c \text{ trans} \mid z \text{ rec}).$$

- If $f : F^n \longrightarrow C$ is a decoder, the probability of a decision error whenever $c \in C$ was transmitted as

$$P_{\mathrm{err}}(f, c) := \sum_{\substack{z \in F^n \\ f(z) \neq c}} \mathrm{Prob}(z \text{ rec} \mid c \text{ trans}),$$

and by averaging, we set

$$P_{\mathrm{err}}(f, C) := \frac{1}{|C|} \sum_{c \in C} P_{\mathrm{err}}(f, c).$$

- **Observation:** The maximum likelihood decoder will minimize the error probability among all possible decoders.

- **Shannon's Theorem:** Given a discrete channel on the alphabet F with capacity $\mathcal{C}$. For $0 < R < \mathcal{C}$, there exists a family $(C_n)_{n\in\mathbb{N}}$ of block codes over F, together with maximum likelihood decoders $f_n : F^n \longrightarrow C_n$, such that:

  - C_n is a code of length n of rate at least R.

  - $\lim_{n\to\infty} P_{\mathrm{err}}(f_n, C_n) = 0$.

- **Summary:** Keeping the rate at given level below the capacity, an investment in the length of the code used will yield arbitrary good reliability of communication.

- Although not obvious to most beginners, both the encoding and the decoding function are generally hard to evaluate in terms of complexity theory.
- Hence, efficient ways to evaluate these functions are desired.
- Restricting to block codes with structure will easily provide highly efficient schemes to perform the encoding process.
- This will often be beneficial for the complexity of the desired decoding schemes.
- Shannon's theorem has a converse, that says that exceeding the capacity will be punished on the spot.

- Shannon's theorem in the above form deals with capacities, rates, and error probabilities.

- It does not say anything yet about one other important parameter of a code: the minimum distance.

- **Definition:** For an alphabet F, define the Hamming distance

$$d_H : F \times F \longrightarrow \{0,1\}, \;\; (x,y) \mapsto \begin{cases} 1 & : \; x \neq y \\ 0 & : \; \text{otherwise.} \end{cases}$$

For positive length n, extend this function additively to

$$d_H : F^n \times F^n \longrightarrow \mathbb{N}, \;\; (x,y) \mapsto \#\{i \mid x_i \neq y_i\}.$$

- **Observation:** d_H is a metric on F^n, as it is symmetric, strictly positive, and satisfies the triangle inequality

$$d_H(x,z) \leq d_H(x,y) + d_H(y,z) \text{ for all } x, y, z \in F^n.$$

- **Definition:** The minimal distance of $d = d_H(C)$ of a code C is defined as

$$d_H(C) \ := \ \min\{d_H(x,y) \mid x, y \in C, \ x \neq y\}$$

- For $i \in \{0, \ldots, n\}$ consider

$$B_i \ := \ \frac{1}{|C|} \#\{(u,v) \in C \times C \mid d_H(u,v) = i\},$$

and $B(x,y) := \sum_{i=0}^{n} B_i x^i y^{n-i}$, the distance enumerator of C.

- With all what we previously defined, we describe a block code C by a triple (n, M, d) where n is the length and M is the number of elements in C, and where $d = d_H(C)$.
- Such a code can detect up to $d-1$ errors, while it can correct up to $\lfloor \frac{d-1}{2} \rfloor$ errors.
- **Example:** Consider the binary code

  $$C = \{00000000, 11011001, 00111110, 11100111\}.$$

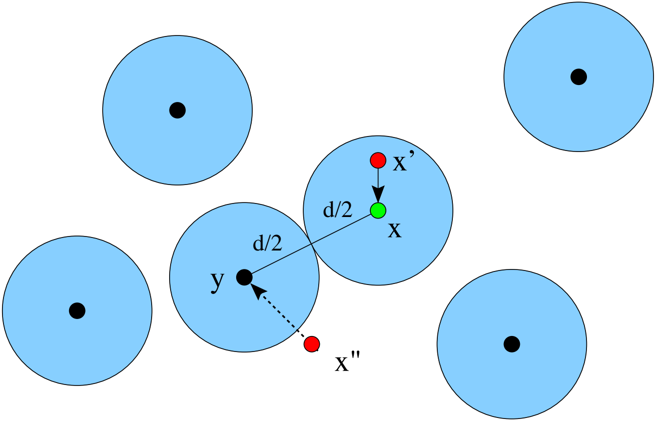
  Here, $n = 8$, $M = 4$ and the minimum distance is 5.
- We could use these words to communicate four different messages, where the most efficient way to represent these is to think of the words $00, 01, 10$ and 11.

- An encoder will therefore assign:

$$\begin{array}{lclclcl} 00 & \mapsto & 00000000 & \quad & 01 & \mapsto & 00111110 \\ 10 & \mapsto & 11011001 & \quad & 11 & \mapsto & 11100111 \end{array}$$

- At the receiving end of our channel we are told that the word 11000101 has been received.
- This is not a word in C! There must have been an error. Which word has most likely been sent?
- This of course depends on the structure of the channel, and the probabilities involved.
- By hand, we find out that among all words in C the word 11100111 is closest to the received word 11000101.
- Using a minimum distance decoder, we decide that the word 11100111 was the one originally sent.

- Illustration of the packing, and the decoding process.



- When the underlying noisy channel π has the form

$$\pi = \begin{bmatrix} 1-p & \frac{p}{q-1} & \frac{p}{q-1} & \cdots & \frac{p}{q-1} \\ \frac{p}{q-1} & 1-p & \frac{p}{q-1} & \cdots & \frac{p}{q-1} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \frac{p}{q-1} & \frac{p}{q-1} & \cdots & \frac{p}{q-1} & 1-p \end{bmatrix}$$

  then this is the right procedure for decoding.

- More-over, the distance enumerator can particularly easily be used to compute the error probability $P_{\mathrm{u}}(C)$.

- Here $P_{\mathrm{u}}(C)$ the probability, that, if the code is used for error detection only, that an undetected error occurs.

- This probability will then be $P_{\mathrm{u}}(C) = B(p, 1-p)$.