TUTORIAL XI

1 Error-correcting VS error-detecting codes

1. Show the two following implications for a code $C$ and even integer $d \geq 2$:
   1. If $C$ has minimum distance at least $d$, then $C$ can correct at least $\frac{d}{2} - 1$ errors.
   2. If $C$ can correct at least $\frac{d}{2} - 1$ errors, then $C$ has minimum distance at least $d - 1$.

2. Show that the following statements are equivalent for a code $C$ and an integer $d \geq 2$:
   1. $C$ has minimum distance $d$.
   2. $C$ can detect $d - 1$ errors.
   3. $C$ can correct $d - 1$ erasures (in the erasure model, the receiver knows where the errors have occurred).

2 Generalized Hamming bound

Prove the following bound: for any $(n, k, d)_q$ code $C \subseteq \left( \Sigma \right)^n$ with $|\Sigma| = q$,

$$k \leq n - \log_q \left( \sum_{i=0}^{\lfloor \frac{d-1}{2} \rfloor} \binom{n}{i} (q-1)^i \right)$$

3 Parity check matrix

Let $C$ be a $[n, k, d]_q$-linear code and $G \in \mathbb{F}_q^{k \times n}$ be a generator matrix. That is, $C = \{xG, x \in \mathbb{F}_q^k\}$. We call a parity check matrix of the code $C$ a matrix $H \in \mathbb{F}_q^{(n-k) \times n}$ such that for all $c \in \mathbb{F}_q^n$ we have $cH^T = 0$ if and only if $c \in C$. The objective of this exercise is to show how to construct a parity check matrix from a generator matrix.

1. Show that $H$ is a parity check matrix if and only if $GH^T = 0$ and $\text{rank}(H) = n - k$.

2. Show that, from $G$ we can construct a generator matrix $G'$ of the form $G' = [I_k | P]$ for some $P \in \mathbb{F}_q^{k \times (n-k)}$. (If $n$ is not optimal, we may have to permute the coefficients of the vectors).

3. Construct a parity check matrix from $G'$.

4. Construct a parity check matrix of the code given by the generator matrix $G = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix}$ in $\mathbb{F}_2$. 
4 Almost-universal hash-functions: link between almost-universal hash-functions and codes with a good distance

A hash function is generally a function from a large space to a small one. A desirable property for a hash function is that there are few collisions. A family of functions \( \{f_y\}_{y \in Y} \) from \( f_y : \mathcal{X} \to \mathcal{Z} \) is called \( \epsilon \)-almost universal if for any \( x \neq x' \), we have \( \mathbb{P}_y \{ f_y(x) = f_y(x') \} \leq \epsilon \) for a uniformly chosen \( y \in Y \). In other words, for any \( x \neq x' \),

\[
|\{ y \in Y : f_y(x) = f_y(x') \}| \leq \epsilon |Y| .
\]

The objective of the exercise is to show that almost-universal hash-functions and codes with a good distance are equivalent: from one you can construct the other efficiently.

**Definition 4.1.** Let \( \mathcal{H} = \{f_1, \ldots, f_n\} \) be a family of hash-functions, where for each \( 1 \leq i \leq n \), \( f_i : \mathcal{X} \to \mathcal{Z} \). We define the code \( C_{\mathcal{H}} : \mathcal{X} \to \mathcal{Z}^n \) by

\[
C_{\mathcal{H}}(x) = (f_1(x), \ldots, f_n(x))
\]

for all \( x \in \mathcal{X} \).

On the contrary, given a code \( C : \mathcal{X} \to \mathcal{Z}^n \), we define the family of hash-functions \( \mathcal{H}_C = \{f_1, \ldots, f_n\} \), from \( \mathcal{X} \) to \( \mathcal{Z} \) by

\[
f_i(x) = C(x)_i,
\]

where \( x \in \mathcal{X} \) and \( C(x)_i \) is the \( i \)-th letter of \( C(x) \) in the alphabet \( \mathcal{Z} \).

1. Let \( \mathcal{H} = \{f_1, \ldots, f_n\} \) be a family of \( \epsilon \)-almost universal hash-functions. Prove that \( C_{\mathcal{H}} \) has minimum distance \( (1 - \epsilon)n \).

2. On the other way, let \( C \) be a code from \( \mathcal{X} \) to \( \mathcal{Z}^n \) with minimum distance \( \delta n \), prove that \( \mathcal{H}_C \) is a family of \( (1 - \delta) \)-almost universal hash-functions.