Assignment 4

Exercise 1 (Zyablov bound)**.** We will show a low complexity procedure based on code concatenation to design an explicit code which achieves $R > 0, \delta > 0$. By low complexity we mean subexponential in the block length.

From Exercise 7 Assignment 2 there exists linear codes over $[q]$ whose asymptotic rate $r = \lim_{n\to\infty} \frac{k(n)}{n}$ and relative minimum distance $\delta = \lim_{n\to\infty} \frac{d(n)}{n}$ satisfy the GV bound

$$r \geq 1 - H_q(\delta).$$

1. Argue that to find a length n code whose rate and relative minimum distance satisfy the

   $$r \geq 1 - H_q(\delta) - \varepsilon$$

   it takes $q^{O(kn)}$ time, as opposed to $q^{O(q^k n)}$ time if the code has no structure. Hint: how many generator matrices are there with paramters k, n?

2. Consider concatenating a linear code approaching the GV bound (inner code) and a Reed Solomon code (outer code). Show that such a construction yields an asymptotic rate

   $$\mathcal{R} \geq \sup_{r\geq 0} r \left(1 - \frac{\delta}{H_q^{-1}(1 - r - \varepsilon)}\right)$$

   for any $\varepsilon > 0$, where δ represents the relative minimum distance of the concatenated code and where r denotes the rate of the inner code. This bound is called the Zyablov bound.

3. Plot the Zyablov bound and the GV bound (rate as a function of relative minimum distance).

4. Argue that it is possibe to construct an explicit code achieving the Zyablov bound with time complexity $\mathcal{N}^{\mathcal{O}(\log \mathcal{N})}$ where $\mathcal{N}$ denotes the length of the concatenated code.

   Hence, although the Zyablov bound is lower than the GV bound, it is easier to construct a code that achieves the Zyablov bound (by concatenation) than to construct a linear code achieving the GV bound (which takes $O(q^{\mathcal{N}})$ time).

Exercise 2 (Binary symmetric channel)**.** Let us examine the performance of linear codes against random errors. The binary symmetric channel with crossover probability $p < 1/2$ is defined by the following process: Given a codeword $\mathbf{c} \in \mathbb{F}_2^n$, we generate a random vector $\mathbf{y}$ where y_i is obtained by flipping c_i with probability p, independently of everything else. Equivalently,

$$\mathbf{y} = \mathbf{c} + \mathbf{z},$$

where $\mathbf{z}$ is a random vector whose components are independent and follow a Bernoulli(p) distribution. Here $\mathbf{y}$ is called the received vector, and $\mathbf{z}$ the noise vector.

We will measure the performance of a code $\mathcal{C} \subset \mathbb{F}_2^n$ of size 2^{nR} using the *average probability of error* under a minimum distance decoder $\mathrm{DEC}(\mathbf{y}) = \arg\min_{\mathbf{c}\in\mathcal{C}} d(\mathbf{y}, \mathbf{c})$:

$$\begin{aligned} P_e(\mathcal{C}) &= \frac{1}{2^{nR}} \sum_{\mathbf{c}\in\mathcal{C}} \Pr_{\mathbf{z}}[\exists \mathbf{c}' \in \mathcal{C}\backslash\{\mathbf{c}\} : \mathrm{DEC}(\mathbf{y}) = \mathbf{c}'] \\ &= \frac{1}{2^{nR}} \sum_{\mathbf{c}\in\mathcal{C}} \Pr_{\mathbf{z}}[\exists \mathbf{c}' \in \mathcal{C}\backslash\{\mathbf{c}\} : d(\mathbf{y}, \mathbf{c}') \le d(\mathbf{y}, \mathbf{c})], \end{aligned}$$

where $d(\cdot, \cdot)$ denotes Hamming distance. This is the average probability that there exists a codeword different from $\mathbf{c}$, that is closer to the received vector.

The goal of this and the next exercise is to show that for every $\epsilon > 0$ there exist linear codes of rate $R = 1 - H(p) - \epsilon$ whose probability of error is $2^{-\Omega(n)}$.

1. First, show that the Hamming distance between $\mathbf{y}$ and $\mathbf{c}$ is approximately np:

$$\Pr[d(\mathbf{c}, \mathbf{y}) > np(1 + \epsilon/2)] \le 2^{-\Omega(n)}$$

   Hint: Find the probability that $\mathbf{z}$ has Hamming weight greater than $np(1 + \epsilon/2)$. You can use Chernoff bound, or directly compute the probability and then use Stirling's approximation.

2. Next, show that the probability of error can be bounded from above as $P_e(\mathcal{C}) \le P_e^{(1)} + P_e^{(2)}$, where

$$P_e^{(1)} = \frac{1}{2^{nR}} \sum_{\mathbf{c}\in\mathcal{C}} \Pr_{\mathbf{z}}[\exists \mathbf{c}' \in \mathcal{C}\backslash\{\mathbf{c}\} : d(\mathbf{y}, \mathbf{c}') \le np(1 + \epsilon/2)]$$

   and

$$P_e^{(2)} = \Pr[d(\mathbf{c}, \mathbf{y}) > np(1 + \epsilon/2)] \le 2^{-\Omega(n)}$$

3. Let us now find the probability of error for a random linear code obtained by choosing a generator matrix G uniformly. Show that for any two nonzero message vectors $\mathbf{u}_1 \neq \mathbf{u}_2$, the corresponding codeword $\mathbf{u}_1 G$ and $\mathbf{u}_2 G$ are statistically independent.

4. For fixed messages $\mathbf{u}_1 \neq \mathbf{u}_2$, show that

$$\Pr_{G,\mathbf{z}}\left[d(\mathbf{u}_1 G, \mathbf{u}_2 G + \mathbf{z}) < np(1 + \epsilon/2)\right] \le 2^{-n(1-H(p(1+\epsilon/2))+o(1))}$$

   Hint: First compute $\Pr_G\left[d(\mathbf{u}_1 G, \mathbf{x}) < np(1 + \epsilon/2)\right]$ for a fixed $\mathbf{x} \in \mathbb{F}_2^n$. Then average over $\mathbf{z}$.

5. Use part 4 to show that if $R < 1 - H(p) - \epsilon$, then $P_e^{(2)} = 2^{-\Omega(n)}$.

6. Combine everything to prove that there exists a linear code with rate $R \ge 1 - H(p) - \epsilon$ and $P_e = o(1)$.