SOLUTIONS TO ASSIGNMENT 5

Exercise 1 (List decodability of linear codes)**.** Show that with high probability, a random (binary) linear code obtained by choosing an $nR \times n$ generator matrix uniformly at random is (p, L)-list decodable as long as

$$R \leq 1 - H(p) - \frac{1}{\lceil \log_2(L+1) \rceil}.$$

Hint: Argue that any set of $L+1$ vectors in $\mathbb{F}_2^k$ contains at least $\lceil \log_2(L+1) \rceil$ linearly independent vectors. If two messages are linearly independent, then what can you say about the corresponding codewords of the random linear code?

Solution. Let $B(\mathbf{y}, np) \triangleq \{\mathbf{x} \in \mathbb{F}_2^n : d(\mathbf{x}, \mathbf{y}) \leq np\}$ denote the Hamming ball of radius np. For a message $\mathbf{m} \in \mathbb{F}_2^{nR}$, let $\mathbf{c}(\mathbf{m})$ denote the corresponding codeword. It suffices to show that

$$\Pr\left[\exists \mathbf{m}_1, \ldots, \mathbf{m}_{L+1} \in \mathbb{F}_2^{nR}, \mathbf{y} \in \mathbb{F}_2^n : \mathbf{c}(\mathbf{m}_1), \ldots, \mathbf{c}(\mathbf{m}_{L+1}) \in B(\mathbf{y}, np)\right] = o(1).$$

A set of l linearly independent vectors in $\mathbb{F}_2^k$ spans a space of size 2^l. Now, consider a set $\mathcal{S}$ of vectors and let $l(\mathcal{S})$ denote the maximal number of independent vectors of $\mathcal{S}$. Since a maximal set spans $\mathcal{S}$, it must be the case that $2^{l(\mathcal{S})} \geq |\mathcal{S}|$, or $l(\mathcal{S}) \geq \log_2 |\mathcal{S}|$. Therefore, any set of $L+1$ messages contains a linearly independent set of size greater than or equal to $\lceil \log_2(L+1) \rceil$. Define $l \triangleq \lceil \log_2(L+1) \rceil$.

Now, let $\mathbf{m}_1, \ldots, \mathbf{m}_l$ denote l linearly independent messages. Let g^k denote a column of the generator matrix (this column is randomly generated). Then, for any fixed $(a_1, \ldots, a_l)$ we have

$$P((g^k)^T[\mathbf{m}_1, \ldots, \mathbf{m}_l] = [a_1, \ldots, a_l]) = 2^{-l}.$$

To see this note that the system of linear equations (with g^k unknown)

$$(g^k)^T[\mathbf{m}_1, \ldots, \mathbf{m}_l] = [a_1, \ldots, a_l]$$

always has 2^{k-l} solutions. Hence, since g^k is uniformly distributed we get

$$P((g^k)^T[\mathbf{m}_1, \ldots, \mathbf{m}_l] = [a_1, \ldots, a_l]) = \frac{2^{k-l}}{2^k} = 2^{-l}.$$

Hence, the codewords corresponding to $\mathbf{m}_1, \ldots, \mathbf{m}_l$ are statistically independent and uniformly distributed over $\mathbb{F}_2^n$. Hence,

$$\begin{aligned} &\Pr\left[\exists \mathbf{m}_1, \ldots, \mathbf{m}_{L+1} \in \mathbb{F}_2^{nR}, \mathbf{y} \in \mathbb{F}_2^n : \mathbf{c}(\mathbf{m}_1), \ldots, \mathbf{c}(\mathbf{m}_{L+1}) \in B(\mathbf{y}, np)\right] \\ &\leq \Pr\left[\exists \text{ linearly independent } \mathbf{m}_1, \ldots, \mathbf{m}_l \in \mathbb{F}_2^{nR}, \mathbf{y} \in \mathbb{F}_2^n : \mathbf{c}(\mathbf{m}_1), \ldots, \mathbf{c}(\mathbf{m}_l) \in B(\mathbf{y}, np)\right] \end{aligned}$$

For any fixed set of linearly independent $\mathbf{m}_1, \ldots, \mathbf{m}_l$, and any $\mathbf{y}$,

$$\Pr[\mathbf{c}(\mathbf{m}_1), \ldots, \mathbf{c}(\mathbf{m}_l) \in B(\mathbf{y}, np)] \leq 2^{nl(1-H(p)-o(1))}.$$

There are at most $\binom{2^{nR}}{l}$ sets of l linearly independent messages. Using this, and taking union bound over the messages and $\mathbf{y}$'s gives us the result.

Exercise 2 (List decoding from erasures)**.** We say that a code is (p, L)-erasure list-decodable if for any vector $\mathbf{y} \in \{0, 1, *\}^n$ (where $*$ denotes the erasure symbol) with at most pn erasures, there are at most L codewords that agree with $\mathbf{y}$ in the unerased positions. For any vector $\mathbf{c}$ and $T \subset [n]$, let $\mathbf{c}_T$ denote the restriction of $\mathbf{c}$ to T, i.e., it is the $|T|$-length vector $(c_i : i \in T)$. Formally, a code $\mathcal{C} \subset \mathbb{F}_2^n$ is (p, L)-erasure list-decodable if for every $T \subset [n]$ with $|T| \geq (1-p)n$, and $\mathbf{y}' \in \{0,1\}^{|T|}$, we have

$$|\{\mathbf{c} \in \mathcal{C} : \mathbf{c}_T = \mathbf{y}'\}| \leq L.$$

Prove the following:

1. If $\mathcal{C}$ has minimum distance d, then it is $\left(\frac{d-1}{n}, 1\right)$-list decodable.
2. For every $\epsilon > 0$, there exists a (p, L)-erasure list decodable code of rate

   $$R \geq \frac{L}{L+1}(1-p) - \frac{H(p)}{L+1} - \epsilon$$

   Hint: Use random codes. For a fixed $T, \mathbf{y}'$, compute the probability that the codeword for a fixed message is equal to $\mathbf{y}$ when restricted to T. Do this for $L + 1$ messages. Then take a union bound over messages, $\mathbf{y}$', and T.
3. Show that if a code of rate $1 - p + \epsilon$ is (p, L)-erasure list-decodable, then $L = 2^{\Omega(n)}$.

Solution.

1. If the minimum distance of a code is d, then it can correct every pattern of at most $d-1$ erasures. Hence, it is $(\frac{d-1}{n}, 1)$-list decodable.
2. Define $A(\mathbf{y}', T) \triangleq \{\mathbf{x} \in \mathbb{F}_2^n : \mathbf{x}_T = \mathbf{y}'\}$. We need to show that

   $$\Pr_{\mathcal{C}}[\exists T, \mathbf{y}' : |\mathcal{C} \cap A(\mathbf{y}', T)| \geq L + 1] = o(1).$$

   Fix $T, \mathbf{y}'$. Let $|T| = t$. For any message $\mathbf{m}$,

   $$\Pr[\mathbf{c}(\mathbf{m}) \in A(\mathbf{y}', T)] = \frac{1}{2^t},$$

   and for any fixed set of $L + 1$ messages

   $$\Pr[\mathbf{c}(\mathbf{m}_1), \ldots, \mathbf{c}(\mathbf{m}_{L+1}) \in A(\mathbf{y}', T)] = \frac{1}{2^{t(L+1)}}.$$

   Taking union bound over message sets, T and $\mathbf{y}'$,

   $$\begin{aligned}\Pr_{\mathcal{C}}[\exists T, \mathbf{y}' : |\mathcal{C} \cap A(\mathbf{y}', T)| \geq L + 1] &= \Pr_{\mathcal{C}}[\exists T, |T| = n(1-p), \mathbf{y}' : |\mathcal{C} \cap A(\mathbf{y}', T)| \geq L + 1] \\ &\leq \binom{n}{np} 2^{n(1-p)} \binom{2^{nR}}{L+1} \frac{1}{2^{n(1-p)(L+1)}} \\ &\leq 2^{n(H(p)+1-p)}\, 2^{nR(L+1)}\, 2^{-n(1-p)(L+1)},\end{aligned}$$

   which is $o(1)$ if $R \geq \frac{L}{L+1}(1-p) - \frac{H(p)}{L+1} - \epsilon$.

Reference: V. Guruswami: List Decoding of Error-Correcting Codes, LNCS 3282, pp. 251-277, 2004.
`https://link.springer.com/content/pdf/10.1007%2F978-3-540-30180-6_10.pdf`

3. Suppose $\mathcal{C}$ is any code of rate $1-p+\epsilon$ and minimum list size L. Choose T to be a fixed subset of $[n]$ having size $n(1-p)$, and $\mathbf{y}'$ a random vector of length $n(1-p)$ with i.i.d. Bernoulli(1/2) components. Fix any codeword $\mathbf{c} \in \mathcal{C}$. This is in $A(\mathbf{y}', T)$ if $\mathbf{c}_T = \mathbf{y}'$. Hence,

$$\Pr_{\mathbf{y}}[\mathbf{c} \in A(\mathbf{y}', T)] = \frac{1}{2^{n(1-p)}}.$$

Let $\xi = \sum_{\mathbf{c}\in\mathcal{C}} 1_{\{\mathbf{c}\in A(\mathbf{y}',T)\}}$ be the number of codewords in $A(\mathbf{y}', T) \cap \mathcal{C}$. Then,

$$\mathbb{E}_{\mathbf{y}'}[\xi] = \sum_{\mathbf{c}\in\mathcal{C}} \Pr[\mathbf{c} \in A(\mathbf{y}', T)] = 2^{nR}/2^{n(1-p)} = 2^{n\epsilon}.$$

Therefore, there exists at least one $\mathbf{y}'$ such that $|A(\mathbf{y}', T) \cap \mathcal{C}| \geq 2^{n\epsilon}$.