CS6845 - Pseudorandomness
Assignment 4

Due date: April 3, in class

Instructions

- For each day's delay in submission, you will lose 25% marks.
- Discussing the questions on the assignment with your classmates is allowed, but **no** collaboration is allowed while writing up the solution. If you discuss an assignment problem with someone, please acknowledge that while writing the solution. This will **not** affect your grade.
- Checking online sources for answers is **strongly discouraged**. If you need any clarification, you are welcome to discuss with me or Dinesh. Any academic dishonesty will result in zero marks for the assignment.
- Include all the steps/arguments in your answer/proof.

---

1. (2 points) Let $\mathcal{C}_1$ be an $[n_1, k_1, d_1]_2$ code, and $\mathcal{C}_2$ an $[n_2, k_2, d_2]_2$ code. Let $\mathcal{C} \subseteq \mathbb{F}_2^{n_1 \times n_2}$ be the subset of $n_1 \times n_2$ matrices whose columns belong to $\mathcal{C}_1$ and rows belong to $\mathcal{C}_2$. Show that $\mathcal{C}$ is an $[n_1 n_2, k_1 k_2, d_1 d_2]_2$ code.

2. Let $\mathcal{C}$ be an $[n, k, d]_q$ linear code. Suppose that we are in the setting that the channel erases certain bits, but does not change any bit. In other words the received word is a string $y \in (\mathbb{F} \cup \{?\})^n$.
   (a) (1 point) Show that if the number of erasures in y is less than d, then there exists a unique $x \in \mathcal{C}$ such that if $y_i \neq \{?\}$, then $x_i = y_i$.
   (b) (3 points) Give an $O(n^3)$ algorithm to compute the unique x from the received message y.

3. Let $\mathcal{C}_1$ be an $[n, k_1, d_1]_q$ code and $\mathcal{C}_2$ be an $[n, k_2, d_2]_q$ code. Define a new code $\mathcal{C}_1 \circ \mathcal{C}_2 = \{(c_1, c_1 + c_2) \mid c_1 \in \mathcal{C}_1, c_2 \in \mathcal{C}_2\}$. We will now prove some properties about this code.
   (a) (2 points) Show that $\mathcal{C}_1 \circ \mathcal{C}_2$ is an $[2n, k_1 + k_2, \min(2d_1, d_2)]_q$ code.
   (b) (2 points) If G_i is the generator matrix for $\mathcal{C}_i$ for $i \in \{1, 2\}$, write down the generator matrix for $\mathcal{C}_1 \circ \mathcal{C}_2$.
   (c) (3 points) Assume that there exists algorithms $\mathcal{A}_i$ for code $\mathcal{C}_i$ such that: (i) $\mathcal{A}_1$ can decode from e errors and s erasures if $2e + s < d_1$, and (ii) $\mathcal{A}_2$ can decode from from $\lfloor \frac{d_2 - 1}{2} \rfloor$ errors. Design an algorithm that can correct $\lfloor \frac{d-1}{2} \rfloor$ errors for $\mathcal{C}_1 \circ \mathcal{C}_2$, where $d = \min\{2d_1, d_2\}$.
   (Hint: On receiving a word (y_1, y_2), first apply $\mathcal{A}_2$ on $y_2 - y_1$. Then create a word for $\mathcal{A}_1$.)

4. In this question we will prove some properties of the Reed Solomon code.
   (a) (2 points) For any $[n, k, d]_q$ Reed Solomon code, exhibit two codewords that are at distance exactly $n - k + 1$.
   (b) (3 points) Let $RS_{n,k,q}$ denote the $[n, k, n - k + 1]_q$ Reed Solomon code. Show that the dual code of $RS_{n,k,q}$ is the Reed Solomon code $RS_{n,n-k,q}$.

5. A *t-burst error pattern* is a string $e \in \{0,1\}^n$ such that all the 1s in e occur between the indices i and $i+t-1$ for some $1 \le i \le n$. In this exercise, we will see how to correct burst error patters.
   (a) (1 point) Show that if there exists an $[n,k,d]_{2^m}$ code, then there exists an $[nm, km, d' \ge d]_2$ code.
   (b) (3 points) Show that for every $R > 0$, there is a large enough n such that there is a binary code with rate R and block length n that can correct any t-burst errors patterns if $t \le \left(\frac{1-R}{2}\right) \cdot n$.
   (Hint: Use Reed Solomon codes.)

6. We will now look at connections between linear codes and some other pseudorandom objects.
   (a) (2 points) A set $S \subseteq \mathbb{F}_q^n$ is k-wise independent if for every set of positions I with $|I| = k$, the set S projected to I has each of the vectors in $\mathbb{F}_q^k$ appear the same number of time.
   Let $\mathcal{C}$ be a linear code such that $\mathcal{C}^\perp$ has distance $d^\perp$. Show that the set $\mathcal{C}$ is $(d^\perp - 1)$-independent.
   (b) (2 points) A set of vectors $S \subseteq \mathbb{F}_2^k$ is called an ϵ-biased space if for every $I \subseteq [k]$ the following holds:
   $$\left| \Pr_{x \in S}\left[\sum_{i \in I} x_i = 0\right] - \Pr_{x \in S}\left[\sum_{i \in I} x_i = 1\right] \right| \le \epsilon.$$
   Let $\mathcal{C}$ be an $[n,k,d]_2$ code such that every non-zero codeword has hamming weight in the range $\left[\frac{1-\epsilon}{2}n, \frac{1+\epsilon}{2}n\right]$. If $G \in \mathbb{F}_2^{k \times n}$ is a generator matrix of $\mathcal{C}$, show that the set of columns of G is an ϵ-biased space of size n.