University of California, Berkeley Instructor: Venkatesan Guruswami
CS 294-226: Advances in Error-Correcting Codes Term: Fall 2022

PROBLEM SET 1
Due by 11.59pm PDT Friday, September 23

INSTRUCTIONS

- You are allowed to collaborate with up to two students taking the class in solving problem sets. But here are some rules concerning such collaboration:

    1. You should think about each problem by yourself for at least 30 minutes before commencing any collaboration.
    2. Collaboration is defined as discussion of the lecture material and solution approaches to the problems. Please note that *you are not allowed to share any written material and you must write up solutions on your own*. You must clearly acknowledge your collaborator(s) in the write-up of your solutions.
    3. Of course, if you prefer, you can also work alone.

- Solutions typeset in LaTeX are encouraged, but not required. If you are submitting handwritten solutions, please write clearly and legibly (you might want to first write the solution sketch in rough, before transferring it to the version you turn in which should be a ”finished product”).
- Solutions should be turned in via gradescope.
- You may consult any of the class notes or videos that were assigned for initial preparation, and the “Essential Coding Theory” text linked from the course page, but please refrain from consulting other sources. In particular, you should not search for solutions on the web. Do not hesitate to contact the GSI or instructor should you have any questions.
- Please start work on the problem set early. The problems vary in difficulty and length and it will be a good idea to give yourself enough time for the trickier problems.

---

1. (20 points) Let $G = (V, E)$ be any undirected graph (assume no loops or multiple edges). A *cut* in the graph is the subset of all edges that connect a vertex in S to vertex in $V \setminus S$, for some subset $S \subseteq V$. Let Cuts$(G) \subseteq \{0, 1\}^E$ consist of the characteristic vectors of all cuts of G.

    (a) Prove that Cuts(G) is an $[|E|, |V| - 1]_2$ binary linear code. What parameter of G equals the distance of Cuts(G)?

    (b) Describe the dual code Cuts$(G)^\perp$ of Cuts(G). What is its dimension? What parameter of G equals the distance of Cuts$(G)^\perp$?

    (c) Given a graph $G = (V, E)$ and a parameter c, it is NP-hard to tell if G has a cut of size at least c (this is the Max Cut problem). Via a reduction from Max Cut, show that the nearest codeword problem (NCP) is NP-hard. In NCP, the input is a generator matrix $A \in \mathbb{F}_2^{n \times k}$ of a binary linear code C, a vector $y \in \mathbb{F}_2^n$, and a nonnegative integer e, and one must decide if there is a codeword in C within Hamming distance e from y.

2. (15 points) We proved in lecture that if a code $C \subset \{0,1\}^n$ has relative distance $\delta \in (0, 1/2)$, then any Hamming ball of radius $J(\delta)n$ has at most $O(n)$ codewords of C, where $J(\delta) := \frac{1}{2}(1 - \sqrt{1 - 2\delta})$. Prove that this is tight, in the sense that for every $\delta \in (0, 1/2)$ and $\epsilon > 0$ and large enough n, there is a code $C' \subset \{0,1\}^n$ (not necessarily linear!) of relative distance δ that has exponentially many codewords fall within a Hamming ball of radius $(J(\delta) + \epsilon)n$.

   Hint: If we pick $2^{\alpha \cdot \epsilon^2 n}$ random samples from Binomial$(n, \rho + \epsilon)$ (the random distribution on $\{0,1\}^n$ where each coordinate is equal to 1 with probability $\rho + \epsilon$ and 0 otherwise) for a sufficiently small constant $\alpha > 0$, then w.h.p. all points have Hamming weight at most $(\rho + 2\epsilon)n$, and every pair of points will have Hamming distance at least $2\rho(1 - \rho)n$.

3. (20 points) In this problem, we will explore binary versions of the Reed-Solomon codes.

   (a) Let $q = 2^m$ and consider the "full" Reed-Solomon code consisting of evaluations of degree $q/2$ polynomials at all of $\mathbb{F}_q$ (so it is a $[q, q/2 + 1, q/2]$ linear code over $\mathbb{F}_q$). Prove that no matter which $\mathbb{F}_2$-basis we express the field elements in, the resulting binary code has distance only $q/2$.
   Hint: There is a polynomial $T \in \mathbb{F}_{2^m}[X]$ of degree 2^{m-1} that takes only values in $\mathbb{F}_2$. Also for $a \in \mathbb{F}_{2^m}$, $a \in \mathbb{F}_2$ iff $a^2 = a$.

   (b) For any $[n, k]_q$ Reed-Solomon code, and $w \geq n - k + 1$, prove that the number of codewords of weight w is at most $\binom{n}{w} q^{w-n+k}$.[1]

   (c) Consider a Reed-Solomon code of block length n and rate R over $\mathbb{F}_{2^m}$. Convert this Reed-Solomon code to a *binary* code of block length m as follows. Pick matrices $G_1, G_2, \ldots, G_n \in \mathbb{F}_2^{m \times m}$ uniformly and indepenently at random. Each codeword $(c_1, c_2, \ldots, c_n) \in \mathbb{F}_{2^m}^n$ is transformed into $(G_1 c_1, G_2 c_2, \cdots, G_n c_n) \in \mathbb{F}_2^{mn}$ where we interpret c_i as a vector in $\mathbb{F}_2^m$ by expressing elements of $\mathbb{F}_{2^m}$ in some fixed basis over $\mathbb{F}_2$.

      i. Fix $\epsilon > 0$. Prove that, with probability at least $1 - 2^{-\Omega(nm)}$ over the choice of the G_i's, the above binary code has distance at least $(h^{-1}(1 - R) - \epsilon)nm$.
      Hint: For a fixed Reed-Solomon codeword of weight w, what is the probability that after encoding by the G_i's, it is mapped to a binary string of low weight? Then union bound over all weight w codewords.

      ii. Conclude that the above sampling procedure yields codes approaching the Gilbert-Varshamov bound with high probability.

4. (25 points) In class, we have seen the Gilbert-Varshamov (GV) bound, which proved the existence of (linear) codes $C \subseteq \{0,1\}^n$ of distance d of size

$$|C| \geq \frac{2^n}{\sum_{i=0}^{d-1} \binom{n}{i}}$$

   In this problem, we are going to show a minor improvement on the GV bound. In particular, we are going to show the existence of linear $[n, n/2]_2$ codes C with distance at least d, where d is the largest integer such that $\sum_{i=0}^{d-1} \binom{n}{i} < \frac{1}{1000} \cdot n 2^{n/2}$.

   We say a matrix $A \in \mathbb{F}_2^{k \times k}$ is *circulant* if there exists a vector $c = (c_0, \ldots, c_{k-1}) \in \mathbb{F}_2^k$ such that $A_{i,j} = c_{i-j \pmod{k}}$. For example, when $k = 5$, a circulant matrix is going to be of the following

[1] It is possible, using inclusion-exclusion, to determine the exact count of weight w codewords in any MDS code.

form:

$$\begin{bmatrix} c_0 & c_4 & c_3 & c_2 & c_1 \\ c_1 & c_0 & c_4 & c_3 & c_2 \\ c_2 & c_1 & c_0 & c_4 & c_3 \\ c_3 & c_2 & c_1 & c_0 & c_4 \\ c_4 & c_3 & c_2 & c_1 & c_0 \end{bmatrix}$$

For any vector $c \in \mathbb{F}_2^k$, let $A_c \in \mathbb{F}_2^{k \times k}$ be the circulant matrix defined by $(A_c)_{i,j} = c_{i-j \pmod{k}}$. We will show that for a uniformly random $c \in \mathbb{F}_2^k$, the linear code C whose generator matrix is

$$G = \left[\frac{I_{k\times k}}{A_c}\right]$$

is an $[n, n/2]_2$ code with distance at least d with high probability, where d is the largest integer such that $\sum_{i=0}^{d-1} \binom{n}{i} < \frac{1}{1000} \cdot n2^{n/2}$.

(a) Given two circulant matrices $A, B \in \mathbb{F}_2^{k\times k}$, prove that AB is also a circulant matrix and $AB = BA$. Conclude that the set of $k \times k$ circulant matrices forms a commutative ring R with a unit.

(b) For any vector $c = (c_0, \ldots, c_{k-1}) \in \mathbb{F}_2^k$, let $p_c(x) \in \mathbb{F}_2[x]/(x^k - 1)$ be the polynomial defined by $p_c(x) := c_0 + c_1 x + \ldots + c_{k-1}x^{k-1}$. Prove that the map $A_c \mapsto p_c$ forms a ring isomorphism between R and $\mathbb{F}_2[x]/(x^k - 1)$.

(c) From here on after, let k be a prime number such that 2 generates the multiplicative group $\mathbb{F}_k^*$. Prove that the polynomial $p(x) := 1 + x + \ldots + x^{k-1} \in \mathbb{F}_2[x]$ is irreducible. (Hint: Let α be a root of $p(x)$ over a field extension $\mathbb{F}_{2^t}$. What can you say about the order of α? If $p(x)$ was reducible, what would be the size of the smallest field containing α?)

(d) Prove that for any vector $m \in \mathbb{F}_2^k$ that is not the all-zeroes or all-ones vector, we have that $\text{rank}(A_m) \geq k - 1$. (Hint: How are $\text{rank}(A_m)$ and $\gcd(p_m(x), x^k - 1)$ related?)

(e) For any two vectors $u, v \in \mathbb{F}_2^k$, prove that $A_v u = A_u v$.[2]

(f) Prove that for any vector $m \in \mathbb{F}_2^k$ that is not the all-zeroes or all-ones vector, we have for a uniformly random $c \in \mathbb{F}_2^k$ that

$$\Pr_{c \xleftarrow{R} \mathbb{F}_2^k} [\text{wt}(A_c m) < d] \leq \frac{3\sum_{i=0}^{d-1}\binom{k}{i}}{2^k}.$$

(g) Given any two vectors $x, y \in \mathbb{F}_2^k$ such that $y_i = x_{i+\ell \pmod{k}}$ for all $0 \leq i \leq k-1$ and some natural number ℓ, prove that $\text{wt}(Gx) = \text{wt}(Gy)$. (Recall that G is the generator matrix of the code C.)

(h) Assuming that there are infinitely many primes k such that 2 generates $\mathbb{F}_k^*$[3], conclude that are infinitely many linear $[n = 2k, k, \geq d]_2$ codes provided $\sum_{i=0}^{d-1}\binom{n}{i} < \frac{1}{1000} \cdot n2^{n/2}$.
(Hint: There should be only $\approx 2^k/k$ events you have to worry about rather than 2^k events. That reduction in number of events is exactly what gets us to beat the GV bound.)

[2] Regarding notation, the vector u is a column vector with its entries ordered from top to bottom as $u_0, u_1, \ldots, u_{k-1}$.

[3] This assumption is known as Artin's conjecture and is widely believed to be true. The actual proof circumvents this assumption, but we will assume it for simplicity of the proof.