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

PROBLEM SET 0

Due by midnight PDT Friday, September 2

INSTRUCTIONS

- Since this problem set is intended to help you self-assess your preparedness for the course, you should solve the problems and write-up the solutions by yourself.
- 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; details will be forthcoming.
- You may consult any of the notes/videos that were assigned for initial preparation, but please refrain from consulting other sources. Do not hesitate to contact the GSI or instructor should you have any questions.

---

1. (a) How can you convert a code $C \subseteq \{0,1\}^n$ of minimum distance $d \geq 2$ into a code $C' \subseteq \{0,1\}^{n-1}$ of the same size as C and minimum distance at least $d-1$?

   (b) How can you convert a code $C \subseteq \{0,1\}^n$ of minimum distance d that is an odd number into a code $C' \subseteq \{0,1\}^{n+1}$ of the same size as C and minimum distance equal to $d+1$?

   (c) What is the code that you get when you apply the transformation (b) above to the $[2^r - 1, 2^r - 1 - r, 3]_2$ Hamming code to get a code of length 2^r? What is the dual of the resulting code?

   (d) Argue that in any binary linear code, either all codewords begin with a 0 or exactly half of the codewords begin with a 0.

   (e) Given a code $C \subseteq \{0,1\}^n$ a *puncturing* of C is another code C' obtained by dropping some codeword positions of C. More precisely, if $C \subseteq \{0,1\}^n$ and the set of punctured positions is $P \subseteq \{1, 2, \ldots .n\}$, then the punctured code is $\{(c_i)_{i \in P} \mid (c_1, \ldots, c_n) \in C\}$.
   Prove that a binary linear code with no repetitions (i.e., there are no two positions $i \neq j$ such that for every codeword $c \in C$, $c_i = c_j$) and without a symbol that is always 0 (i.e., there are no positions i for which every $c \in C$, $c_i = 0$), is a puncturing of the Simplex code. Hence, the simplex code is the "longest" linear code that does not repeat codeword symbols or have an always 0 symbol.

2. Suppose G is an $n \times k$ generator matrix of a binary linear code $C \subseteq \mathbb{F}_2^n$ (so the columns of G form a basis of C).

   (a) Suppose B is an $k \times k$ matrix of 0-1 entries which is invertible over the field $\mathbb{F}_2$. Prove that GB is also a generator matrix for C.

   (b) Argue that all generator matrices for C can be obtained this way from G, i.e., as GB for some choice of $k \times k$ invertible matrix B.

   (c) Argue that C admits an encoding that is systematic, i.e., there are k codeword positions $1 \leq i_1 < i_2 < \cdots < i_k \leq n$ such that the codeword encoding any message places the j'th bit of the message, $1 \leq j \leq k$, in position i_j.

(d) Suppose the distance of C equals d. Argue why G must have at least kd 1's in it.

3. Let $\epsilon > 0$, k be a large enough integer, and $n = \lceil 10k/\epsilon^2 \rceil$. Let G be a random $n \times k$ matrix whose entries are i.i.d uniform bits. Prove that with high probability (tending to 1 as $k \to \infty$) the binary linear code C generated by the columns of G has (i) dimension k, and (ii) every nonzero codeword of C has Hamming weight in the range $[(1/2 - \epsilon)n, (1/2 + \epsilon)n]$. (Such codes are known as *ϵ-balanced codes*.)

4. A set of vectors $S \subseteq \{0,1\}^n$ is called t-wise independent if for every set of positions I with $|I| = t$, the set S projected to I has each of the vectors in $\{0,1\}^t$ appear the same number of times. (In other words, if one picks a vector $(s_1, \ldots, s_n)$ from S at random, then for every $1 \leq i_1 < i_2 < \cdots < i_t \leq n$, the vector $(s_{i_1}, s_{i_2}, \ldots, s_{i_t})$ is uniformly distributed in $\{0,1\}^t$.)

   Prove that a linear code C is t-wise independent if and only if its dual $C^\perp$ has minimum distance at least $t + 1$.

5. Let $\mathbb{F}_q$ denote the finite field of size q, where $q = p^k$ for some prime number p and positive integer k. Let $\mathbb{F}_q^*$ denote the nonzero elements of $\mathbb{F}_q$.

   (a) Show that $a^{q-1} = 1$ for any a in $\mathbb{F}_q^*$. (Hint: consider the map $x \mapsto ax$ over $\mathbb{F}_q^*$).

   (b) Let d be the smallest natural number such that $a^d = 1$. Show that d divides $q - 1$. The value d is known as the *order of* a and is denoted $\mathrm{ord}(a)$.

   (c) Show that the number of elements in $\mathbb{F}_q^*$ of order exactly d is at most $\varphi(d)$, where $\varphi(d)$ denotes the number of integers m in $\{1, 2, \ldots, d\}$ such that $\gcd(m, d) = 1$. (Hint: how many roots can the polynomial $x^d - 1$ have?)

   (d) (Not to turn in) Prove that $\sum_{d \text{ divides } q-1} \varphi(d) = q - 1$ by considering the number of natural numbers $n \leq q - 1$ such that $\gcd(n, q - 1) = \frac{q-1}{d}$.

   (e) Conclude that $\mathbb{F}_q^*$ must be a multiplicative cyclic group. i.e. there exists an $\alpha \in \mathbb{F}_q^*$ such that $\mathbb{F}_q^* = \{1, \alpha, \alpha^2, \ldots, \alpha^{q-2}\}$.