Carnegie Mellon University Instructor: Venkatesan Guruswami
15-859Y: Coding Theory Term: Fall 2018

PROBLEM SET 2
Due date: Friday, October 12

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 (and are encouarged to) 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).
- You should not search for solutions on the web. More generally, you should try and solve the problems without consulting any reference material other than the course notes and what we cover in class. If for some reason you feel the need to consult some source, *please acknowledge the source* and try to articulate the difficulty you couldn't overcome before consulting the source and how it helped you overcome that difficulty. Alternatively, before turning to any such material, we encourage you to ask the instructor for hints or clarifications.
- Please start work on the problem set early. The problem set has **six** problems worth 20 points each. There is also a bonus problem (which is open-ended, and you can email or meet with the instructor to discuss any promising ideas you might have).

---

1. For a field $\mathbb{F}$ with $|\mathbb{F}| \geq n$, an n-tuple $\vec{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ of n *distinct* elements of $\mathbb{F}$, and a vector $\mathbf{v} = (v_1, v_2, \ldots, v_n) \in (\mathbb{F}^*)^n$ of n (not necessarily distinct) nonzero elements from $\mathbb{F}$, the *Generalized Reed-Solomon code* $\mathrm{GRS}_{\mathbb{F}}(\vec{\alpha}, \mathbf{v}, k)$ is defined as follows:

$$\mathrm{GRS}_{\mathbb{F}}(\vec{\alpha}, \mathbf{v}, k) = \{(v_1 \cdot p(\alpha_1), v_2 \cdot p(\alpha_2), \ldots, v_n \cdot p(\alpha_n)) \mid p(X) \in \mathbb{F}[X] \text{ has degree} < k\} .$$

   (a) Check that $\mathrm{GRS}_{\mathbb{F}}(\vec{\alpha}, \mathbf{v}, k)$ is an $[n, k, n-k+1]_{\mathbb{F}}$ linear code.

   (b) Prove that the dual code of $\mathrm{GRS}_{\mathbb{F}}(\vec{\alpha}, \mathbf{v}, k)$ is

$$\mathrm{GRS}_{\mathbb{F}}(\vec{\alpha}, \mathbf{v}, k)^{\perp} = \mathrm{GRS}_{\mathbb{F}}(\vec{\alpha}, \mathbf{u}, n-k)$$

   for $\mathbf{u} = (u_1, u_2, \ldots, u_n) \in (\mathbb{F}^*)^n$ where for $i = 1, 2, \ldots, n$,

$$u_i = \frac{1}{v_i \prod_{j \neq i}(\alpha_i - \alpha_j)} .$$

2. Let us recall the notion of k-wise independence from Problem Set 1. For integers $1 \le k \le n$, call a (multi)set $S \subseteq \{0,1\}^n$ to be k-wise independent if for every $1 \le i_1 < i_2 < \cdots < i_k \le n$ and $(a_1, a_2, \ldots, a_k) \in \{0,1\}^k$

$$\mathrm{Prob}_{x \in S}[x_{i_1} = a_1 \wedge x_{i_2} = a_2 \wedge \cdots \wedge x_{i_k} = a_k] = \frac{1}{2^k}$$

where the probability is over an element x chosen uniformly at random from S. Small sample spaces of k-wise independent sets are of fundamental importance in derandomization.

   (a) Using BCH codes and Problem 3 of Problem set 1, show how one can construct a k-wise independent subset of $\{0,1\}^n$ of size at most $2 \cdot (2n)^{\lfloor k/2 \rfloor}$.

   (b) Prove an almost matching lower bound, namely any k-wise independent set $S \subseteq \{0,1\}^n$ satisfies

$$|S| \ge \sum_{i=0}^{\lfloor \frac{k}{2} \rfloor} \binom{n}{i} . \qquad (1)$$

   Suggestion: Find a set of linearly independent vectors in $\mathbb{R}^{|S|}$ of cardinality at least the R.H.S of (1). Specifically, for $T \subseteq \{1, 2, \ldots, n\}$ of size $\le \lfloor k/2 \rfloor$, consider the vector $\langle \chi_T(x) \rangle_{x \in S}$ where $\chi_T(x) = (-1)^{\sum_{i \in T} x_i}$.

3. (a) Recall the definition of "tensor product" of codes from Problem Set 1. If C_1 is an $[n_1, k_1, d_1]_2$ binary linear code, and C_2 an $[n_2, k_2, d_2]$ binary linear code, then $C = C_1 \otimes C_2 \subseteq \mathbb{F}_2^{n_2 \times n_1}$ is defined to subspace of $n_2 \times n_1$ matrices whose rows belong to C_1 and whose columns belong to C_2.

   Suppose C_2 has an efficient algorithm to correct $< d_2/2$ errors and C_1 has an efficient errors-and-erasures decoding algorithm to correct any combination of e errors and s erasures provided $2e + s < d_1$. Show how one can efficiently decode C up to $< d_1 d_2/2$ errors using these algorithms as subroutines.

   (b) Consider the bivariate version of the Reed-Solomon code, which encodes a polynomial $f \in \mathbb{F}_q[X, Y]$ with degree less than k in *both* X and Y by its evaluations at all q^2 points $(\alpha, \beta) \in \mathbb{F}_q \times \mathbb{F}_q$.

      i. What are the block length, dimension, and minimum distance of this code?

      ii. Describe how one can efficiently decode this code up to (almost) half its minimum distance.

   ((The natural) hint: Relate to Part (a) of this question.)

4. In this problem, we will look at some binary "BCH-like" subfield subcodes of Reed-Solomon codes that meet the Gilbert-Varshamov bound.

   Let $\mathbb{F} = \mathbb{F}_{2^m}$. Fix positive integers k, n with $(n-k)m < n < 2^m$, and a tuple $\vec{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ of n *distinct* elements of $\mathbb{F}$. For a vector $\mathbf{v} = (v_1, \ldots, v_n) \in (\mathbb{F}^*)^n$ of n not necessarily distinct nonzero elements from $\mathbb{F}$, recall the *Generalized Reed-Solomon code* $\mathrm{GRS}_{\mathbb{F}}(\vec{\alpha}, \mathbf{v}, k)$ defined as follows:

$$\mathrm{GRS}_{\mathbb{F}}(\vec{\alpha}, \mathbf{v}, k) = \{(v_1 \cdot p(\alpha_1), v_2 \cdot p(\alpha_2), \ldots, v_n \cdot p(\alpha_n)) \mid p(X) \in \mathbb{F}[X] \text{ has degree } < k\} .$$

   (a) Argue that $\mathrm{GRS}_{\mathbb{F}}(\vec{\alpha}, \mathbf{v}, k) \cap \mathbb{F}_2^n$ is a binary linear code of rate at least $1 - \frac{(n-k)m}{n}$.

   (b) Let $\mathbf{c} \in \mathbb{F}_2^n$ be a nonzero binary vector. Prove that (for every choice of $\vec{\alpha}, k$) there are at most $(2^m - 1)^k$ choices of the vector $\mathbf{v}$ for which $\mathbf{c} \in \mathrm{GRS}_{\mathbb{F}}(\vec{\alpha}, \mathbf{v}, k)$.

(c) Prove that if the integer D satisfies

$$(2^m - 1)^{n-k} > \sum_{i=0}^{D-1} \binom{n}{i} ,$$

then there exists a vector $\mathbf{v} \in (\mathbb{F}^*)^n$ such that the minimum distance of the binary linear code $\mathrm{GRS}_{\mathbb{F}}(\vec{\alpha}, \mathbf{v}, k) \cap \mathbb{F}_2^n$ is at least D.

(d) Using Parts (4a) and (4c), show how to conclude that the family of codes $\mathrm{GRS}_{\mathbb{F}}(\vec{\alpha}, \mathbf{v}, k) \cap \mathbb{F}_2^n$ contains binary linear codes that meet the Gilbert-Varshamov bound.

5. For this problem, assume the NP-hardness of the following problem (this can be shown via a reduction from Subset Sum):

Instance: A set $S = \{\alpha_1, \ldots, \alpha_n\} \subseteq \mathbb{F}_{2^m}$, an element $\beta \in \mathbb{F}_{2^m}$, and an integer $1 \leq k < n$.

Question: Is there a nonempty subset $T \subseteq \{1, 2, \ldots, n\}$ with $|T| = k + 1$ such that $\sum_{i \in T} \alpha_i = \beta$?

Consider the $[n, k]$ Reed-Solomon code C_{RS} over $\mathbb{F}_{2^m}$ obtained by evaluating polynomials of degree at most $k - 1$ at points in S. Define $y \in (\mathbb{F}_{2^m})^n$ as follows: $y_i = \alpha_i^{k+1} - \beta \alpha_i^k$ for $i = 1, 2, \ldots, n$.

Prove that there is a codeword of C_{RS} at Hamming distance at most $n - k - 1$ from y if and only if there is a set T as above of size $k + 1$ satisfying $\sum_{i \in T} \alpha_i = \beta$.

Conclude that finding the nearest codeword in a Reed-Solomon code over exponentially large fields is NP-hard. (Proving this for polynomial sized fields remains an embarrassing open question.)

6. In this problem, we develop a more abstract view of the Reed-Solomon decoding algorithm that we saw in class. This enables extending the apprpach to other Reed-Solomon-like codes, such as algebraic-geometric codes. First we give some definitions. Let $\mathbb{F}$ be a field. For $u, v \in \mathbb{F}^n$, define $u * v = (u_1 v_1, u_2 v_2, \ldots, u_n v_n) \in \mathbb{F}^n$ be the component-wise product. For $U, V \subseteq \mathbb{F}^n$, define $U * V = \{u * v \mid u \in U,\ v \in V\}$.

The idea of the abstract decoding procedure is that given a code C capable of correcting e errors (i.e., its distance exceeds $2e$) that we want to decode, we construct an *error-locator* code E, such that $E * C$ is contained in another linear code N that has large distance. Specifically, we want codes E and N to have the following properties:

- $\dim(E) > e$.
- $E * C \subseteq N$.
- $\mathrm{dist}(N) > e$.
- $\mathrm{dist}(C) > n - \mathrm{dist}(E)$

Consider the following decoding algorithm for C. Given as input $r \in \mathbb{F}^n$ with Hamming distance at most e from some codeword $c \in C$, the goal of the algorithm is to find c.

Step 1: Find $a \in E$ and $b \in N$, $a \neq 0$, such that $a * r = b$.

Step 2: For each i, if $a_i = 0$, set $s_i = ?$, and oherwise set $s_i = r_i$. Perform erasure decoding (for the code C) on the resulting vector s, to find a $c \in C$ such that $c_i = s_i$ whenever $s_i \neq ?$.
Output c.

The exercises below justify the algorithm, proving its efficiency and correctness. Again, we assume that the input $r \in \mathbb{F}^n$ satisfies the property that there is a $c \in C$ with $\Delta(r, c) \leq e$ (such a c is then unique, due to the assumed e-error correction property of C).

(a) Prove that a, b as in Step 1 exist.

(b) Prove that the algorithm can be implemented in polynomial time, given generator matrices of C, N, E.

(c) Prove that for every (a, b) satisiying the condition of Step 1, $a * c = b$.

(d) Prove that if $a * c' = b$ for some $c' \in C$, then $c' = c$.

(e) Conclude the correctness of the algorithm.

(f) If C is an $[n, n-2e]$ Reed-Solomon code, what are E and N in the above abstraction that correspond to the Welch-Berlekamp algorithm covered in lecture?

7. (Open-ended bonus problem) For the Wozencraft ensemble discussed in class, find an explicit $\alpha \in \mathbb{F}_{2^m}$ (i.e., computable in deterministic $\mathrm{poly}(m)$ time, or even $2^{o(m)}$ time) for which the $[2m, m]_2$ binary linear code C_α mapping $x \in \mathbb{F}_{2^m}$ to $(x, \alpha x) \in \mathbb{F}_{2^m} \times \mathbb{F}_{2^m}$ has distance at least d, for as large a value of d as you are able to establish.