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 encouraged 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 $v = (v_1, v_2, \ldots, v_n) \in (\mathbb{F}^*)^n$ of $n$ (not necessarily distinct) nonzero elements from $\mathbb{F}$, the Generalized Reed-Solomon code $\text{GRS}_{\mathbb{F}}(\vec{\alpha}, v, k)$ is defined as follows:

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

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

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

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

for $u = (u_1, u_2, \ldots, u_n) \in (\mathbb{F}^*)^n$ 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 $1 \leq k \leq n$, call a (multi)set $S \subseteq \{0, 1\}^n$ to be $k$-wise independent if for every $1 \leq i_1 < i_2 < \cdots < i_k \leq n$ and $(a_1, a_2, \ldots, a_k) \in \{0, 1\}^k$

\[
\Pr_{x \in S}[x_{i_1} = a_1 \land x_{i_2} = a_2 \land \cdots \land x_{i_k} = a_k] = \frac{1}{2^k}
\]

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 $2 \cdot (2n)^{[k/2]}$.

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

\[
|S| \geq \sum_{i=0}^{\lfloor \frac{k}{2} \rfloor} \binom{n}{i}.
\]

**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 $\leq \lfloor k/2 \rfloor$, 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 $[n_1, k_1, d_1]$ binary linear code, and $C_2$ an $[n_2, k_2, d_2]$ binary linear code, then $C = C_1 \otimes C_2 \subseteq \mathbb{F}_2^{n_1 \times n_2}$ is defined to subspace of $n_1 \times n_2$ 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 $\text{GRS}_\mathbb{F}(\vec{\alpha}, \mathbf{v}, k)$ defined as follows:

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

(a) Argue that $\text{GRS}_\mathbb{F}(\vec{\alpha}, \mathbf{v}, k)$ 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 $\mathbf{c} \in \text{GRS}_\mathbb{F}(\vec{\alpha}, \mathbf{v}, k)$.
5. For this problem, assume the NP-hardness of the following problem (this can be shown via a reduction.

6. In this problem, we develop a more abstract view of the Reed-Solomon decoding algorithm that

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 \((a, b)\) satisfying the condition of Step 1, \( a \ast c = b \).

(d) Prove that if \( a \ast c' = b \) 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 \( \text{poly}(m) \) time, or even \( 2^{o(m)} \) time) for which the \([2m, m]_2\) binary linear code \( C_\alpha \) 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.