# Problem Set 2

- Due Date: **March 05, 2024**
- Turn in your problem sets electronically (LaTeX, pdf or text file) by email. If you submit handwritten solutions, start each problem on a fresh page.
- Collaboration is encouraged, but all writeups must be done individually and must include names of all collaborators.
- Refering sources other than the text book and class notes is strongly discouraged. But if you do use an external source (eg., other text books, lecture notes, or any material available online), ACKNOWLEDGE all your sources (including collaborators) in your writeup. This will not affect your grades. However, not acknowledging will be treated as a serious case of academic dishonesty.
- The points for each problem are indicated on the side. The total for this set is 80.
- Be clear in your writing.
- Problem 2 is due to Shangguan and Tamo while problems 3-5 are adaptations of similar problems from the book "Essential Coding Theory" (Guruswami, Rudra and Sudan) and Guruswami's course.

1. **[Dual of Reed-Solomon codes for arbitrary evaluation sets]** (6+4+2)

   Let $S \subseteq \mathbb{F}$. Define $a : S \to \mathbb{F}^*$ as follows:

   $$a(\alpha) = \prod_{\substack{\alpha' \in S \\ \alpha' \neq \alpha}} \frac{1}{\alpha - \alpha'} \ .$$

   (a) Show that for any polynomial $p$ of degree $< |S| - 1$, we have $\sum_{\alpha \in S} a(\alpha) p(\alpha) = 0$.

   (b) Define the bilinear form

   $$\begin{aligned} \langle \cdot, \cdot \rangle_S : \mathbb{F}^S \times \mathbb{F}^S &\to \mathbb{F} \\ (f, g) &\mapsto \sum_{\alpha \in S} a(\alpha) \cdot f(\alpha) \cdot g(\alpha). \end{aligned}$$

   Show that for any two polynomials $p, q$ such that $p \in RS_{\mathbb{F}}[S, k]$ and $q \in RS_{\mathbb{F}}[S, |S| - k]$, we have $\langle p, q \rangle_S = 0$.

   Observe that for the special case when $S = \mathbb{F}$, this bilinear form is identical to the standard bilinear form $\langle f, g \rangle = \sum_{\alpha \in \mathbb{F}^*} f(\alpha) \cdot g(\alpha)$ (upto scaling by a constant).

   (c) Observe that the bilinear form is full-rank. In particular, if $V$ is a $k$-dimensional subspace of $\mathbb{F}^S$, note that the dimension of $V^{\perp_S}$, the dual of $V$ with respect to this bilinear form is exactly $|S| - k$. Here,

   $$V^{\perp_S} := \{u \in \mathbb{F}^S | \langle u, v \rangle_S = 0, \forall v \in V\}.$$

   Under this bilinear form $\langle \cdot, \cdot \rangle_S$, what is the dual of $RS_{\mathbb{F}}[S, k]$?

   **Note:** The dual of $RS_{\mathbb{F}}[\mathbb{F}^*, k]$ obtained this way is different from that obtained in class using the standard bilinear form $\langle f, g \rangle = \sum_{\alpha \in \mathbb{F}^*} f(\alpha) \cdot g(\alpha)$.

2. **[Generalization of Singleton bound]** (15)

   The Singleton bound states that $R \leq 1-\delta$, where $R$ is the rate and $\delta$ is the fractional minimum distance of a code $\mathcal{C}$. Equivalently, we may state the following: for any positive integer $L$ and any code $\mathcal{C}$, let $\rho_L$ be the largest $\rho \in [0,1]$ such that any ball of fractional radius $\rho$ has at most $L$ codewords. Note that $\rho_1 = \delta/2$. Hence, the Singleton bound in terms of $\rho_1$ is $R \leq 1 - 2\rho_1$, or equivalently,

   $$|\mathcal{C}| \leq q^{n-2\rho_1 \cdot n}.$$

   Prove the following generalization of the Singleton bound: For any code $\mathcal{C} \subseteq [q]^n$ and any positive integer $L$,

   $$|\mathcal{C}| \leq Lq^{n-\lfloor \frac{(L+1)\cdot \rho_L \cdot n}{L} \rfloor}.$$

3. **[Tensor codes]** (6+8+8)

   Given a $(n_1, k_1, d_1)_q$ code $C_1$ and a $(n_2, k_2, d_2)_q$ code $C_2$, the direct product of $C_1$ and $C_2$, denoted $C_1 \otimes C_2$, is an $(n_1n_2, k_1k_2, d)_q$ code constructed as follows. View a message of $C_1 \otimes C_2$ as a $k_2$-by-$k_1$ matrix $M$. Encode each row of $M$ by the code $C_1$ to obtain an $k_2$-by-$n_1$ intermediary matrix. Encode each column of this intermediary matrix with the $C_2$ code to get an $n_2$-by-$n_1$ matrix representing the codeword encoding $M$.

   In this problem, we first show that the resulting code has distance at least $d_1d_2$ in either case. Then we show that if $C_1$ and $C_2$ are linear, then the resulting code is also linear, and furthermore is the same as the code that would be obtained by encoding the columns with $C_2$ first and then encoding the rows with $C_1$.

   (a) Prove that the distance of the code $C_1 \otimes C_2$ is at least $d_1d_2$.

   (b) Suppose $C_1$ and $C_2$ are linear codes. Let $G_1 \in \mathbb{F}_q^{n_1 \times k_1}$ be a generator matrix for the code $C_1$ and $G_2 \in \mathbb{F}_q^{n_2 \times k_2}$ be a generator matrix for the code $C_2$. Show that the direct product code $C_1 \otimes C_2$ is a linear code that has as its codewords

   $$\{G_2 M G_1^T \;\mid\; M \in \mathbb{F}_q^{k_2 \times k_1}\}.$$

   Conclude that the code $C_1 \otimes C_2$ is linear if $C_1$ and $C_2$ are. Also, that the same code is obtained by encoding the columns with $C_2$ first and then encoding the rows in the intermediate matrix with $C_1$.

   (c) Suppose $C_1$ and $C_2$ are linear codes. Show that the code $C_1 \otimes C_2$ is equivalent to the following code whose codewords are all $n_2 \times n_1$ matrices whose rows are codewords of $C_1$ and columns are codewords of $C_2$. What is the dual of the tensor-code?

4. **[NP-hardness of RS decoding]** (15)

   Consider the following problem:

   **Input 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 $\sum_{i \in T} \alpha_i = \beta$.

   [Note: It can be shown that this problem is NP-hard via a reduction from subset sum.]

   Consider the $[n, k, n-k+1]_{2^m}$ Reed-Solomon code $RS_{n,k,S}$ 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 $RS_{n,k,S}$ 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 $\sum_{i \in T} \alpha_i = \beta$.

This implies 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.)

5. **[Polynomial-based MDS codes]** (8+2+4+4)

In this problem we will see that Reed-Solomon codes, univariate multiplicity codes and folded Reed-Solomon codes are all essentially special cases of a large family of codes that are based on polynomials. We begin with a definition of these codes.

Let $m \geq 1$ be an integer parameter and define $m < k \leq n$. Further, let $E_1(X), E_2(X), \ldots, E_n(X)$ be $n$ polynomials over $\mathbb{F}_q$, each of degree $m$. Further, these polynomials pair-wise do not have any non-trivial factors (that is, $\gcd(E_i(X), E_j(X))$ has degree 0 for every $i \neq j \in [n]$.) Consider any message $\mathbf{m} = (m_0, m_1, \ldots, m_{k-1}) \in \mathbb{F}_q^k$ and let $f_{\mathbf{m}}(X)$ be the message polynomial as defined for the Reed-Solomon code (In other words, $f_{\mathbf{m}}(X) = \sum_{i=0}^{k-1} m_i X^i$). Then the codeword for $\mathbf{m}$ is given by

$$(f_{\mathbf{m}}(X) \pmod{E_1(X)}, f_{\mathbf{m}}(X) \pmod{E_2(X)}, \ldots, f_{\mathbf{m}}(X) \pmod{E_n(X)})$$

In the above we think of $f_{\mathbf{m}}(X) \pmod{E_1(X)}$ as an element of $\mathbb{F}_{q^m}$. In particular, given a polynomial of degree at most $m-1$, we will consider any bijection between the $q^m$ such polynomials and $\mathbb{F}_{q^m}$. We will first see that this code is MDS and then we will see why it contains Reed-Solomon and related codes as special cases.

(a) Prove that the above code is an $[n, k/m, n - \lfloor (k-1)/m \rfloor]_{q^m}$-code (and is thus MDS).

(b) Let $\alpha_0, \alpha_1, \ldots, \alpha_{n-1} \in \mathbb{F}_q$ be distinct elements. Define $E_i(X) = X - \alpha_i$. Argue that for this special case, the above code (with $m = 1$) is the Reed-Solomon code.

(c) Let $\alpha_0, \alpha_1, \ldots, \alpha_{n-1} \in \mathbb{F}_q$ be distinct elements. Define $E_i(X) = (X - \alpha_i)^m$. Argue that for this special case, the above code is equivalent to the following generalization of the Reed-Solomon coded called the *univariate multiplicity code*. The encoding of the message $\mathbf{m}$ at location $\alpha \in S$ is

$$(f_{\mathbf{m}}^{(0)}(\alpha), f_{\mathbf{m}}^{(1)}(\alpha), f_{\mathbf{m}}^{(2)}(\alpha), \ldots, f_{\mathbf{m}}^{(m-1)}(\alpha)),$$

where $f^{(i)}(\alpha)$ refers to the $i^{th}$ derivative of $f_{\mathbf{m}}$. In other words, in addition to giving the evaluation of $f_{\mathbf{m}}(X)$ at the location $\alpha$ (as in the Reed-Solomon code), we also give the evaluation of the low-order derivatives.

(d) Let $\alpha_0, \alpha_2, \ldots, \alpha_{n-1} \in \mathbb{F}_q$ be elements such that the $mn$ elements $\{\alpha_i \gamma^j : i \in [n], j \in [m]\}$ are all distinct. Define $E_i(X) = \prod_{j=0}^{m-1}(X - \alpha_i \gamma^j)$. Argue that for this special case, the above code is is equivalent to the following generalization of the Reed-Solomon coded called the *folded Reed-Solomon code*. The encoding of the message $\mathbf{m}$ at location $\alpha \in S$ is

$$(f_{\mathbf{m}}(\alpha), f_{\mathbf{m}}(\alpha\gamma), f_{\mathbf{m}}(\alpha\gamma^2), \ldots, f_{\mathbf{m}}(\alpha\gamma^{m-1})).$$

In other words, in addition to giving the evaluation of $f_{\mathbf{m}}(X)$ at the location $\alpha$ (as in the Reed-Solomon code), we also give the evaluation of the related polynomials $f_{\mathbf{m}}(\gamma X)$, $f_{\mathbf{m}}(\gamma^2 X), \ldots, f_{\mathbf{m}}(\gamma^{m-1} X)$.