Coding Theory: Assignment II

16th August 2025

---

1. **Due date:** 20th November 2025
2. Write the solutions in LaTeX.
3. All statements proven in the class, given on the exercise sheet, or in the appendix of the assignment can be used without proof. Other than that, anything you use needs to be proven.
4. Although strongly discouraged, if you do use an external source (for example, other than the main text book and problems in the exercise sheet, lecture notes, or any material available online), **acknowledge all your sources** in your writeup. This will not affect your grades; failure to acknowledge sources will be treated as a serious case of academic dishonesty.
5. Allowed to discuss with others but write the solutions independently, and **acknowledge your collaborators**. Failure to do so will be treated as a serious case of academic dishonesty.
6. **Total marks: 70**

---

1. For any positive integer p, we use $\mathbb{Z}_p$ to denote the set of integers $\{0, 1, 2, \ldots, p-1\}$. For two positive integers m and p, let $[m]_p$ denote the unique positive integer in $\mathbb{Z}_p$ we get as a remainder after dividing m by p.

   Let $1 \leq k \leq n$ be positive integers and $p_1 < p_2 < p_3 < \cdots < p_n$ be n distinct primes. Let $K = \prod_{i=1}^{k} p_i$ and $N = \prod_{i=1}^{n} p_i$. Let $C \subseteq \mathbb{Z}_{p_1} \times \mathbb{Z}_{p_2} \times \cdots \times \mathbb{Z}_{p_n}$ be a code defined as follows:

   Message space: $\mathbb{Z}_K$, that is, every message word is treated as an integer in $\mathbb{Z}_K$.

   Encoding: The encoding function $\mathrm{E} : \mathbb{Z}_K \to \mathbb{Z}_{p_1} \times \mathbb{Z}_{p_2} \times \cdots \times \mathbb{Z}_{p_n}$ is defined in the following way:
   For any $m \in \mathbb{Z}_K$,
   $$\mathrm{E}(m) = \left([m]_{p_1}, [m]_{p_2}, [m]_{p_3}, \ldots, [m]_{p_n}\right).$$

   This code can be seen as the number-theoretic counterpart of Reed-Solomon codes. It is known as the Chinese Remainder code and is based on the Chinese Remainder Theorem (CRT) in number theory (see Appendix).

   For any two distinct messages $m_1 \neq m_2 \in \mathbb{Z}_K$, let
   $$\Delta(\mathrm{E}(m_1), \mathrm{E}(m_2)) = \#\left\{i \in [n] \;\middle|\; [m_1]_{p_i} \neq [m_2]_{p_i}\right\}.$$

   Then show the following:

   (i) (**5 marks**) $\min_{m_1 \neq m_2 \in \mathbb{Z}_K} \Delta(\mathrm{E}(m_1), \mathrm{E}(m_2)) = n - k + 1$.

   (ii) (**3 marks**) For any $m_1 \neq m_2 \in \mathbb{Z}_K$, $\prod_{i \in [n] : [m_1]_{p_i} \neq [m_2]_{p_i}} p_i \geq \frac{N}{K-1}$.

   In the next part of the problem, we prove that there exists an efficient error correction algorithm for E. The setup of the error-correction algorithm is the following:

Input: As input, we are given $\mathbf{y} = (y_1, y_2, y_3, \ldots, y_n) \in \mathbb{Z}_{p_1} \times \mathbb{Z}_{p_2} \times \cdots \times \mathbb{Z}_{p_n}$ with the promise that there exists a message $m \in \mathbb{Z}_K$ such that

$$\prod_{i \in [n] : [m]_{p_i} \neq y_i} p_i < \sqrt{\frac{N}{K-1}}. \tag{1}$$

Output: An $m \in \mathbb{Z}_K$ satisfying Equation 1.

Then, show the following:

(a) (**4 marks**) Given a $\mathbf{y} = (y_1, y_2, y_3, \ldots, y_n) \in \mathbb{Z}_{p_1} \times \mathbb{Z}_{p_2} \times \cdots \times \mathbb{Z}_{p_n}$, there exists a unique $m \in \mathbb{Z}_K$ satisfying Equation 1.

(b) (**4 marks**) Design an $\mathrm{poly}(\log p_n, n)$-time error detection algorithm for E. That is, given any $\mathbf{y} = (y_1, y_2, y_3, \ldots, y_n) \in \mathbb{Z}_{p_1} \times \mathbb{Z}_{p_2} \times \cdots \times \mathbb{Z}_{p_n}$, in time $\mathrm{poly}(\log p_n, n)$, decide whether $\mathbf{y} \in C$.

(c) (**2 marks**) There exists a positive integer $1 \leq r < \sqrt{\frac{N}{K-1}}$ such that for every $i \in [n]$, $[r]_{p_i} = 0$ if and only if $[m]_{p_i} \neq y_i$. It is analogous to the "error-locator" polynomial in Reed-Solomon decoding.

(d) (**4 marks**) There exists $1 \leq R < \sqrt{\frac{N}{K-1}}$ and $0 \leq M < \sqrt{N(K-1)}$ integers such that

$$y_i \cdot R = M \bmod p_i \text{ for all } i \in [n]. \tag{2}$$

This is analogous to setting the system of linear equations in the Reed-Solomon decoding algorithm.

(e) (**4 marks**) For any (R_1, M_1) and (R_2, M_2) satisfying Equation 2, show that $\frac{M_1}{R_1} = \frac{M_2}{R_2}$.

(f) (**4 marks**) Given an (R, M) satisfying Equation 2, we can compute the message m in time $\mathrm{poly}(n, \log p_n)$.

Note: Using the above problem, you can show that E can correct up to $\frac{\log p_1}{\log p_1 + \log p_n} \cdot (n-k)$ many errors. You can try it as an exercise.

As an exercise, you can also describe the RS codes and its decoding algorithm in language of Chinese Remainder Theorem (for univariate polynomials).

2. In this problem, we design a polynomial time decoding algorithm for Reed-Muller code. More specifically, we show a reduction of the decoding of Reed-Muller codes to the decoding of Reed-Solomon codes. As you remember, for positive integers r, m, q with q is a prime power, the Reed-Muller code $\mathsf{RM}(m, r, q)$ is defined as follows:

Message space: The set of all m-variate polynomials $f(x_1, x_2, \ldots, x_m) \in \mathbb{F}_q[x_1, x_2, \ldots, x_m]$ of degree at most r and individual degree $< q$, that is, $\deg_{x_i}(f) < q$ for all $i \in [m]$.

Encoding: For every polynomial $f(x_1, x_2, \ldots, x_m)$ in the message space, the encoding of $f(x_1, x_2, \ldots, x_m)$ is the evaluations of f at all the points in $\mathbb{F}_q^m$, that is,

$$f \mapsto (f(\boldsymbol{\alpha}))_{\boldsymbol{\alpha} \in \mathbb{F}_q^m}.$$

We also assume the *low degree* set up of the Reed-Muller code, that is, $r < q$. As discussed in the class, the *Polynomial Zero Lemma (low-degree case)*, that is **Lemma 9.2.2** in Essential Coding Theory, implies the *distance* of $\mathsf{RM}(m, r, q)$ is $(q-r) \cdot q^{m-1}$. In the following problems, we design a polynomial time decoding algorithm for $\mathsf{RM}(m, r, q)$ that can correct less than $\frac{(q-r)}{2} \cdot q^{m-1}$ many errors. Let $\mathbb{F}_{q^m}$ be an degree m extension of the finite field $\mathbb{F}_q$.

(a) Observe that both $\mathbb{F}_{q^m}$ and $\mathbb{F}_q^m$ form vector spaces over $\mathbb{F}_q$ of dimension m. Let $\Phi : \mathbb{F}_{q^m} \to \mathbb{F}_q^m$ be a $\mathbb{F}_q$-linear bijection, that is, Φ is a bijection map and for all $\alpha, \beta \in \mathbb{F}_q$ and $u, v \in \mathbb{F}_{q^m}$, $\Phi(\alpha u + \beta v) = \alpha\Phi(u) + \beta\Phi(v)$. Then, show the following:

i. (**3 marks**) For any $\mathbb{F}_q$-linear bijection $\Phi : \mathbb{F}_{q^m} \to \mathbb{F}_q^m$ can be viewed as tuple of m linear functions $(\Phi_1, \Phi_2, \ldots, \Phi_m)$ with $\Phi_i : \mathbb{F}_{q^m} \to \mathbb{F}_q$ so that $\Phi(u) = (\Phi_1(u), \ldots, \Phi_m(u))$ for all $u \in \mathbb{F}_{q^m}$.

ii. (**7 marks**) If $\Phi = (\Phi_1, \ldots, \Phi_m) : \mathbb{F}_{q^m} \to \mathbb{F}_q^m$ is an $\mathbb{F}_q$-linear bijection, then each $\Phi_i(y)$ can be represented as a polynomial of degree at most q^{m-1}.

(b) (**10 marks**) Show that $\mathsf{RM}(m, r, q) \subseteq \mathsf{RS}(q^m, k, q^m)$, where $\mathsf{RS}(q^m, k, q^m)$ is the Reed-Solomon codes of block length q^m, dimension $k \leq r \cdot q^{m-1} + 1$, the alphabet set $\mathbb{F}_{q^m}$ and $\mathbb{F}_{q^m}$ is the set of evaluation points.

Observe that we now can apply the decoding algorithm for the Reed-Solomon codes and can correct less than $\frac{q-r}{2} \cdot q^{m-1}$ many errors of $\mathsf{RM}(m, r, q)$.

3. Recall, in the exam, we have seen that the construction of certain kind of good linear codes over $\mathbb{F}_2$ implies the construction of good *ϵ-biased sample spaces*. Recalling the definition, a set $S \subseteq \mathbb{F}_2^m$ of vectors is called ϵ-biased sample space if the following property holds: Pick a vector $X = (x_1, x_2, \ldots, x_m)$ uniformly at random from S. Then, X has bias at most ϵ, that is, for every *nonempty* subset $I \subseteq [m]$,

$$\left| \Pr\left(\sum_{i \in I} x_i = 0 \right) - \Pr\left(\sum_{i \in I} x_i = 1 \right) \right| \leq \epsilon,$$

where the sum is over $\mathbb{F}_2$.

Let C be an $[n, k]_2$ code such that all non-zero codewords have Hamming weight in the range $\left[\left(\frac{1-\epsilon}{2}\right)n, \left(\frac{1+\epsilon}{2}\right)n\right]$. Let $G \in \mathbb{F}_2^{k \times n}$ be a generator matrix of C. Then, we have seen that the set of columns of G form an ϵ-biased sample space of size n. In this problem, using code concatenation, we see a construction of explicit ϵ-biased sample space.

(a) (**6 marks**) Show that there exists an ϵ-biased space of size $O(\frac{m}{\epsilon^2})$.

(b) (**7 marks**) Let t be a positive integer. Let $\mathsf{RS}(2^t, k, 2^t)$ be the Reed-Solomon code of block length 2^t, dimension $k = \epsilon 2^t$, the alphabet set $\mathbb{F}_{2^t}$, and the set of evaluation points is $\mathbb{F}_{2^t}$. Let Had_t be the *Hadamard code* of dimension t, that is, its generator matrix G_t is a $t \times 2^t$ matrix over $\mathbb{F}_2$ whose columns are the set of all t length binary strings. Let $C_{\text{out}} = \mathsf{RS}(2^t, k, 2^t)$ with $k = \epsilon 2^t$ and $C_{\text{in}} = \mathsf{Had}_t$. Then, show that the concatenation code $C_{\text{out}} \circ C_{\text{in}}$ is a $[4^t, tk, d]_2$ code where $d \in \left[\left(\frac{1-\epsilon}{2}\right)4^t, \left(\frac{1+\epsilon}{2}\right)4^t\right]$.

(c) (**7 marks**) Show that for all sufficiently large m, we can construct an ϵ-biased sample space over $\mathbb{F}_2^m$ of size

$$O\left(\frac{m^2}{\epsilon^2 \cdot \log^2(\frac{m}{\epsilon})}\right).$$

Observe that the size bound of ϵ-biased sample space we get at Problem 3c is much larger the size bound promised at Problem 3a. After a series of works, the paper Explicit, almost optimal, epsilon-balanced codes by *Amnon Ta-Shma* gives a construction of ϵ-biased sample space which is asymptotically almost equal to the bound promised by Problem 3a.

Appendix

Trace function in finite field. Let $\mathbb{F}_q$ be a finite field and $\mathbb{F}_{q^\ell}$ be an extension of $\mathbb{F}_q$. For all $\alpha \in \mathbb{F}_{q^\ell}$, let

$$\mathrm{Tr}(\alpha) := \alpha + \alpha^q + \alpha^{q^2} + \cdots + \alpha^{q^{\ell-1}}.$$

Then, the following hold:

1. Tr is a surjective linear function from $\mathbb{F}_{q^\ell}$ to $\mathbb{F}_q$.
2. For any $\beta \in \mathbb{F}_{p^\ell}$, let $L_\beta : \mathbb{F}_{q^\ell} \to \mathbb{F}_q$ be the linear function defined as $L_\beta(\alpha) = \mathrm{Tr}(\alpha\beta)$ for all $\alpha \in \mathbb{F}_{q^\ell}$. Then, for any two distinct $\beta_1, \beta_2 \in \mathbb{F}_{q^\ell}$, $L_{\beta_1} \neq L_{\beta_2}$. Furthermore, $\{L_\beta \mid \beta \in \mathbb{F}_{q^\ell}\}$ is the set of all linear transformations from $\mathbb{F}_{q^\ell}$ to $\mathbb{F}_q$.

Chinese Remainder Theorem (CRT): Let $p_1, p_2, \ldots, p_\ell$ be ℓ distinct primes. Let $L = \prod_{i=1}^{\ell} p_i$. Then, the mapping $\Phi : \mathbb{Z}_L \to \mathbb{Z}_{p_1} \times \mathbb{Z}_{p_2} \times \cdots \times \mathbb{Z}_{p_\ell}$ defined as

$$\Phi(m) = \left([m]_{p_1}, [m]_{p_2}, \ldots, [m]_{p_\ell}\right) \text{ for all } m \in \mathbb{Z}_L$$

is a bijection. Furthermore, for any $m \in \mathbb{Z}_L$, the image $\Phi(m)$ can be computed in time $\mathrm{poly}(\ell, \log p_\ell)$. Similarly, given a point $\mathbf{v} \in \mathbb{Z}_{p_1} \times \mathbb{Z}_{p_2} \times \cdots \times \mathbb{Z}_{p_\ell}$, the preimage $\Phi^{-1}(\mathbf{v})$ can be computed in time $\mathrm{poly}(\ell, \log p_\ell)$.