ACT I: Final Exam

---

1. All the statements proven in the class, or given in the exercise sheet/assignments can be used without proof.
2. To solve a sub-problem of a particular problem in the question paper, you can assume all its previous sub-problems without proof.
3. Statements mentioned in the appendix section of the question paper can be assumed without proof.
4. Other than that, anything you use needs to be proven.

---

1. A set of $S \subseteq \mathbb{F}_2^k$ vectors is called *ϵ-biased sample space* if the following property holds: Pick a vector $X = (x_1, x_2, \ldots, x_k)$ uniformly at random from S. Then, X has bias at most ϵ, that is, for every nonempty subset $I \subseteq [k]$,

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

where the sum is over $\mathbb{F}_2$. Observe that $S = \mathbb{F}_2^k$ is an ϵ-biased sample space with $\epsilon = 0$. In this problem, we will look at some connections of ϵ-biased sample space to linear codes over $\mathbb{F}_2$.

   (a) **(4 marks)** 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, show that the set of columns of G forms an ϵ-biased sample space of size n.

   (b) **(6 marks)** Let C be an $[n,k]_2$ code such that all nonzero codewords have Hamming weight in the range $\left[\left(\frac{1-\gamma}{2}\right)n, \left(\frac{1+\gamma}{2}\right)n\right]$ where $\gamma \in (0,1)$. Then, show that for every odd positive integer m, there exists an $[n^m, k]_2$ code C' such that all nonzero codewords have Hamming weight in the range

$$\left[\left(\frac{1-\gamma^m}{2}\right)n^m, \left(\frac{1+\gamma^m}{2}\right)n^m\right].$$

   (c) **(3 marks)** Let C be an $[n,k]_2$ code such that all nonzero codewords have Hamming weight in the range $\left[\left(\frac{1-\gamma}{2}\right)n, \left(\frac{1+\gamma}{2}\right)n\right]$ where $\gamma \in (\epsilon, 1)$. Then, show that there exists an ϵ-biased sample space of size

$$n^{O\left(\frac{\log 1/\epsilon}{\log 1/\gamma}\right)}.$$

2. Let $q \ge 2$ be an integer. As we have seen in the class, the *Gilbert-Varshamov bound* (GV bound) says that for every $\delta \in [0, 1-\frac{1}{q})$, there exists a q-ary code with the rate $R \ge 1 - H_q(\delta)$ and relative distance δ, where $H_q(\cdot)$ denotes the q-array entropy function defined in the class. In the class, we also saw a greedy construction-based proof for GV bound. Here, see a graph-theoretic proof for GV

bound. Let $d = \delta n$, and Σ be an alphabet of size q. Let $G_{n,d,q} = (V, E)$ be a graph whose vertex set is Σ^n. Given vertices $\mathbf{u} \neq \mathbf{v} \in \Sigma^n$, we have the edge $\{\mathbf{u}, \mathbf{v}\} \in E$ if and only if $\Delta(\mathbf{u}, \mathbf{v}) < d$. A subset $I \subseteq V$ of vertices is called an *independent set* of $G_{n,d,q}$, if for every $\mathbf{u} \neq \mathbf{v} \in I$, $\{\mathbf{u}, \mathbf{v}\} \notin E$. Then, solve the following sub-problems.

(a) **(2 marks)** Show that any independent set C of $G_{n,d,q}$ is a q-ary code of distance at least d.

(b) **(5 marks)** The *degree* of vertex in a graph $G = (V, E)$ is the number of edges incident on that vertex. Let D be the maximum degree of any vertex in $G = (V, E)$. Then argue that G has an independent set of size at least $\frac{|V|}{D+1}$.

(c) **(3 marks)** Using the parts (a) and (b), prove the GV bound.

3. **(7 marks)** Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be n distinct elements from the finite field $\mathbb{F}_q$. Let $\mathrm{RS}(n,k,q)$ be the Reed-Solomon code of block length n with the evaluation points are $\alpha_1, \alpha_2, \ldots, \alpha_n$, dimension is k and the alphabet is $\mathbb{F}_q$. Now consider the code $\mathrm{RS}(n,k,q)^{\perp}$, that is, the dual of $\mathrm{RS}(n,k,q)$. Design an error-correction algorithm $\mathcal{A}$ for $\mathrm{RS}(n,k,q)^{\perp}$ that runs in $\mathrm{poly}(n)$ $\mathbb{F}_q$-operations and can correct less than $\frac{k+1}{2}$ many errors. More specifically, given a $\mathbf{y} = (y_1, y_2, \ldots, y_n) \in \mathbb{F}_q^n$ as input to $\mathcal{A}$ with the promise that there exists a codeword $\mathbf{c} \in \mathrm{RS}(n,k,q)^{\perp}$ such that $\Delta(\mathbf{y}, \mathbf{c}) < \frac{k+1}{2}$, it outputs $\mathbf{c}$ in $\mathrm{poly}(n)$ $\mathbb{F}_q$-operations. Observe that this gives an error correction algorithm for BCH codes.

4. 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 can be treated as an integer in $\mathbb{Z}_K$.

Encoding: The encoding function $\mathrm{E}_{\mathrm{CRT}} : \mathbb{Z}_K \to \mathbb{Z}_{p_1} \times \mathbb{Z}_{p_2} \times \cdots \times \mathbb{Z}_{p_n}$ is defined in the following way: For any $m \in \mathbb{Z}_K$,

$$\mathrm{E}_{\mathrm{CRT}}(m) = \left([m]_{p_1}, [m]_{p_2}, [m]_{p_3}, \ldots, [m]_{p_n}\right).$$

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 the point 3 in the Appendix).

For any two distinct messages $m_1 \neq m_2 \in \mathbb{Z}_k$, let

$$\Delta\left(\mathrm{E}_{\mathrm{CRT}}(m_1), \mathrm{E}_{\mathrm{CRT}}(m_2)\right) = \#\left\{i \in [n] \mid [m_1]_{p_i} \neq [m_2]_{p_i}\right\}.$$

Then show the following:

$$\textbf{(5 marks)} \qquad \min_{m_1 \neq m_2 \in \mathbb{Z}_k} \Delta\left(\mathrm{E}_{\mathrm{CRT}}(m_1), \mathrm{E}_{\mathrm{CRT}}(m_2)\right) = n - k + 1.$$

In the next part of the problem, we prove that there exists an efficient error correction algorithm for $\mathrm{E}_{\mathrm{CRT}}$. 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

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

$i \in [n] : [m]_{p_i} \neq y_i$

2

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

Then, show the following:

(a) **(3 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) **(3 marks)** Design an $\text{poly}(\log p_n, n)$-time error detection algorithm for E_{CRT}. 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 $\text{poly}(\log p_n, n)$, decide whether $\mathbf{y} \in C$.

(c) **(4 marks)** There exists a positive integer $1 \le r \le E$ such that for every ~~some~~ $i \in [n]$, $[r]_{p_i} = 0$ if and only if $[m]_i \neq y_i$. It is analogous to the "error-locator" polynomial in Reed-Solomon decoding.

(d) **(4 marks)** There exists $1 \le R \le E$ and $0 \le M < N/E$ integers such that

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

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

Note: Using the above problem, you can show that E_{CRT} 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 at home.

1 Appendix

1. **Reducing Bias:** Let $S = (s_1, s_2, \ldots, s_n) \in \{0,1\}^n$ be a binary string with pn many 1's for some $p \in (0,1)$. For some positive integer m, let $S' \in \{0,1\}^{n^m}$ be a binary string such that

$$S' = \left(\bigoplus_{j=1}^{m} s_{i_j} \right)_{i_1, i_2, \ldots, i_m \in [n]}.$$

Then, the number of 1's in the string S' is

$$\Delta_p = \frac{1}{2} \cdot (1 - (1-2p)^m) \cdot n^m.$$

Furthermore, if m is odd, then Δ_p is a non-decreasing function of p.

2. **Upper bound for the volume of Hamming ball:** Let $q \ge 2$ be an integer, and $p \in [0, 1 - \frac{1}{q}]$. Then,

$$\cancel{\sum_{i=0}^{pn} \binom{n}{n} \le q^{H_q(p)n}}, \quad \sum_{i=0}^{pn} \binom{n}{i} (q-1)^i \le q^{H_q(p)n}$$

where $H_q(\cdot)$ is the q-ary entropy function defined in the class.

3. **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 $\Phi : Z_L \to \mathbb{Z}_{p_1} \times \mathbb{Z}_{p_2} \times \cdots \times \mathbb{Z}_{p_\ell}$ 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 $m \in \mathbb{Z}_L$, $\Phi(m)$ can be computed in time $\text{poly}(\ell, \log p_\ell)$. Similarly, given a point $\mathbf{v} \in \mathbb{Z}_{p_1} \times \mathbb{Z}_{p_2} \times \cdots \times \mathbb{Z}_{p_\ell}$, $\Phi^{-1}(\mathbf{v})$ can be computed in time $\text{poly}(\ell, \log p_\ell)$.

4. For any two positive integers M and N, we can compute $M + N$, $M \cdot N$, $\lfloor M/N \rfloor$ and $M \bmod N$ in time $\mathrm{poly}(\log M + \log N)$.