Problem Set 3

- Due Date: **26 March, 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 1 is from the book Essential Coding Theory and problem 2 is from a course by Guruswami.

---

1. **[Improved construction of linear codes on GV bound]** (3+5+5+3)

   In the class, we saw a greedy construction of a q-ary code with block length n and message length k on the GV bound in deterministic time $q^{O(n)}$. In this problem, we will try to get a similar construction for linear codes on the GV bound. We already saw the existence of such codes in the class via a probabilistic argument.

   (a) **(3 points)**. Argue that the probabilistic argument from the class can be derandomized to give a deterministic construction of a linear code of dimension k and block length n over $\mathbb{F}_q$ on the GV bound in time $q^{O(kn)}$.

   (b) **(5 points)**. A $k \times n$ matrix A is a Toeplitz Matrix if it satisfies the property that for all $i \in \{2, 3, \ldots, k\}$ and $j \in \{2, 3, \ldots, n\}$, $A_{i,j} = A_{i-1,j-1}$, or in other words, every diagonal has the same value. In particular, a random Toeplitz matrix can be picked by just picking the entries in the first row and column uniformly and independently at random, and then using the relation $A_{i,j} = A_{i-1,j-1}$ to deduce the other entries.

   Prove that for any non-zero vector $x \in \mathbb{F}_q^k$, the vector $A^T \cdot x$ is distributed uniformly over $\mathbb{F}_q^n$, where A is a random Toeplitz matrix picked as described above.

   (c) **(5 points)**. Argue that the above question implies that a random linear code picked by picking a random Toeplitz matrix as its generator matrix lies on the GV bound with non-zero probability (in fact, with high probability).

   (d) **(3 points)**. Conclude from the above discussion that there is a deterministic construction of a linear code on the GV bound with block length n and dimension k over $\mathbb{F}_q$ in time $q^{O(k+n)}$.

2. **[Tensor product decoding]** (4+4+4+4+4)

   In this problem, we will adapt Forney's GMD decoding for concatenated codes to unique-decode tensor-product codes all the way up to their unique-decoding radius.

Given a $(n_1, k_1, d_1)_q$-code $\mathcal{C}_1$ and $(n_2, k_2, d_2)_q$-code $\mathcal{C}_2$, the tensor-product code $\mathcal{C}_1 \otimes \mathcal{C}_2$ is the set of codewords $c \in [q]^{n_1 \times n_2}$, such that for all $i \in [n_1]$, $c(i, \cdot) \in \mathcal{C}_2$ and for all $j \in [n_2]$, $c(\cdot, j) \in \mathcal{C}_1$. Recall from problem set 2 that the distance of the tensor-product code is $d_1 \cdot d_2$. In this problem, we will design a decoder for $\mathcal{C}_1 \otimes \mathcal{C}_2$ that corrects any pattern of less than $d_1 \cdot d_2/2$ errors. We assume that we have access to two decoders Dec_1 and Dec_2 with the following properties.

- $\textsc{Dec}_1$ is a decoder for $\mathcal{C}_1$ that can correct any pattern of errors provided the number of errors is less than $d_1/2$.
- $\textsc{Dec}_2$ is a decoder for $\mathcal{C}_2$ that can correct any pattern of errors *and* erasures provided the number of erasures plus twice the number of errors is less than d_2.
- If the corresponding promises are not met, then $\textsc{Dec}_1$ and $\textsc{Dec}_2$ return some arbitrary codewords in $\mathcal{C}_1$ and $\mathcal{C}_2$ respectively.

Consider Algorithm 1 based on Forney's GMD Decoding. This is a randomized algorithm and can be derandomized as in Forney's GMD Decoding. Prove either via the following parts or otherwise that this is a unique-decoder for $\mathcal{C}_1 \otimes \mathcal{C}_2$ all the way upto half the minumum distance $d_1 d_2$.

Algorithm 1: GMD Decoder for Tensor-product code $\mathcal{C}_1 \otimes \mathcal{C}_2$

Input: A received word $r \in [q]^{n_1 \times n_2}$

Output: The *unique* codeword $c \in \mathcal{C}_1 \otimes \mathcal{C}_2$ such that

$$\Delta(r, c) < d_1 d_2/2$$

if one exists and $\perp$ otherwise.

Initialize r', c' to be empty words in $[q]^{n_1 \times n_2}$
Comment: r' will be a partially decoded word while c' will eventually be the final decoded codeword.
for $j \in [n_2]$ **do**
  Run Decoder $\textsc{Dec}_1$ on word $r(\cdot, j) \in [q]^{n_1}$ to obtain (row) codeword $c_j \in \mathcal{C}_1$.
  Set $F_j \leftarrow \min\{\Delta(r(\cdot, j), c_j), d_1/2\}$.
  With probability $2F_j/d_1$, set $r'(\cdot, j) \leftarrow \underbrace{???\ldots?}_{n_1 \text{times}}$, otherwise set $r'(\cdot, j) \leftarrow c_j$.
for $i \in [n_1]$ **do**
  Run Decoder $\textsc{Dec}_2$ on (possibly partially erased) word $r'(i, \cdot) \in ([q] \cup \{?\})^{n_2}$ to obtain (column) codeword $c^{(i)} \in \mathcal{C}_2$.
  Set $c'(i, \cdot) \leftarrow c^{(i)}$.
if $\Delta(r, c') < d_1 d_2/2$ **then**
  return c'
else
  return $\perp$

Let $r \in [q]^{n_1 \times n_2}$ be the received word with the promise that there exists a codeword $c \in \mathcal{C}_1 \otimes \mathcal{C}_2$ such that $\Delta(r, c) < d_1 d_2/2$. In this case, we will prove below that the word c' output by the algorithm is in fact c.

For each $j \in [n_2]$, define

$$E_j := \Delta(r(\cdot, j), c(\cdot, j)).$$

Observe that $\sum_{j\in[n_2]} E_j$ is the total number of errors which is promised to be less than $d_1d_2/2$. Furthermore, if $E_j < d_1/2$, then the codeword c_j obtained in Line 3 is exactly the (row) codeword $c(\cdot, j) \in \mathcal{C}_1$.

(a) Prove that if $E_j \geq d_1/2$, then $E_j + F_j \geq d_1$. Also show that the number of such j's for which $E_j \geq d_1/2$ is less than d_2.

(b) Define indicator random variables U_j and V_j for each $j \in [n_2]$ as follows:

$$U_j = \begin{cases} 1 & \text{if } j^{th} \text{ row is erased in Line 5,} \\ 0 & \text{otherwise.} \end{cases}$$

$$V_j = \begin{cases} 1 & \text{if } j^{th} \text{ row is } \textit{not} \text{ erased in Line 5 and } r'(\cdot, j) \neq c(\cdot, j), \\ 0 & \text{otherwise.} \end{cases}$$

Prove that for each $j \in [n_2]$, $\mathbb{E}[U_j + 2V_j] \leq 2E_j/d_1$.

(c) Conclude that there exists a particular choice of random coins such that

$$\sum_{j\in[n_2]} (U_j + 2V_j) < d_2.$$

(d) Define indicator random variables $U_j^{(i)}$ and $V_j^{(i)}$ for each $i \in [n_1]$ and $j \in [n_2]$ as follows:

$$U_j^{(i)} = \begin{cases} 1 & \text{if } r'(i, j) =?, \\ 0 & \text{otherwise.} \end{cases}$$

$$V_j^{(i)} = \begin{cases} 1 & \text{if } r'(i, j) \neq? \text{ and } r'(i, j) \neq c(i, j), \\ 0 & \text{otherwise.} \end{cases}$$

Clearly, $U_j^{(i)} = U_j$ for each i and j. Prove that furthermore, for each i and j we have $V_j^{(i)} \leq V_j$.

(e) Use the above to argue that there exists a choice of random coins such that simultaneously for each $i \in [n_1]$, the decoded (column) codeword $c^{(i)} = c'(i, \cdot)$ is in fact the (column)codeword $c(i, \cdot) \in \mathcal{C}_2$ in the tensor-product codeword c. Hence $c' = c$.

3. **[List-decodability of the Hadamard code via Fourier analysis]** (4+3+3+4)

Recall that the Hadamard code is the $[2^k, k, 2^{k-1}]_2$-code which consists of the evaluations of all linear functions. More precisely, the Hadamard codewords are precisely the linear functions

$$\mathbf{x} = (x_1, x_2, \ldots, x_n) \longmapsto \sum a_i x_i \pmod 2$$

for each $\mathbf{a} \in \{0, 1\}^k$. In this problem, we will prove that the Hadamard code is $(1/2 - \delta, 1/\delta^2)$-list-decodable using Fourier analysis for every $\delta \in (0, 1)$.

Let $\mathcal{F}$ denote the set of all functions from $\{0, 1\}^k$ to $\mathbb{R}$. Note $\mathcal{F}$ is a 2^k-dimensional vector space over $\mathbb{R}$. Define an inner product on this space as follows:

$$\langle f, g\rangle := \mathop{\mathbb{E}}_{\mathbf{x}}[f(x)g(x)].$$

For any $\mathbf{a} \in \{0, 1\}^k$, define $\chi_{\mathbf{a}} \in \mathcal{F}$ as follows: $\chi_{\mathbf{a}}(\mathbf{x}) := (-1)^{\sum_{i\in[k]} a_i x_i \pmod 2}$.

(a) Show that for all $\mathbf{a} \neq \mathbf{b}$, we have $\langle \chi_{\mathbf{a}}, \chi_{\mathbf{b}} \rangle = 0$. Conclude that the 2^k functions $\{\chi_{\mathbf{a}}\}_{\mathbf{a} \in \{0,1\}^k}$ form an orthonormal basis of functions for the vector space $\mathcal{F}$.
Hence, conclude that any function $f \in \mathcal{F}$ can be expressed uniquely as follows:

$$f(\mathbf{x}) = \sum_{\mathbf{a}} \widehat{f}(\mathbf{a}) \cdot \chi_{\mathbf{a}}(\mathbf{x}).$$

where $\widehat{f}(\mathbf{a}) = \langle f, \chi_{\mathbf{a}} \rangle$. These real numbers $\widehat{f}(\mathbf{a})$ are called the Fourier coefficients of f.

(b) (Parsevals' equation). Show that for $f \in \mathcal{F}$, we have

$$\|f\|_2^2 = \langle f, f \rangle = \sum_{\mathbf{a}} |\widehat{f}(\mathbf{a})|^2.$$

Hence, for any Boolean function $f : \{0,1\}^k \to \{1,-1\}$, we have $\sum_{\mathbf{a}} |\widehat{f}(\mathbf{a})|^2 = 1$.
It will be convenient to express the range of a Boolean function as $\{1,-1\}$ instead of $\{0,1\}$. We move from $\{0,1\}$ to $\{1,-1\}$ using the transformation $b \mapsto (-1)^b$. Observe that with this notation in place, the $\chi_{\mathbf{a}}$'s exactly correspond to all the linear functions (and thus all the Hadamard codewords).

(c) Let $f : \{0,1\}^k \to \{1,-1\}$ be any Boolean function and $\mathbf{a} \in \{0,1\}^k$ such that $\Pr_{\mathbf{x}}[f(\mathbf{x}) = \chi_{\mathbf{a}}(\mathbf{x})] \geq \frac{(1+\delta)}{2}$. Conclude that $\widehat{f}(\mathbf{a}) \geq \delta$.

(d) Let $f : \{0,1\}^k \to \{1,-1\}$ be any Boolean function. Conclude that there at most $1/\delta^2$ linear functions which have agreement at least $\frac{(1+\delta)}{2}$ with f.
We have thus proved that for any Boolean function f, there are at most $1/\delta^2$ linear functions which are within $\frac{1-\delta}{2}$ fractional distance from f. We will give an alternate proof of this fact via the Goldreich-Levin list-decoding algorithm later in the course.

The remaining two problems are based on the list-recovery problem, a generalization of the list-decoding problem, defined as follows:

List-recovery problem: Let $\mathcal{C}$ be a $(n,k)_q$-code. Given n subsets $S_i \subseteq [q], 1 \leq i \leq n$ where $|S_i| \leq \ell$, output all codewords $c = (c_1, \ldots, c_n)$ such that $c_i \notin S_i$ for at most e values of i. If for every valid input the number of such codewords is at most L, then the corresponding code is called $(e/n, \ell, L)$-list recoverable.

Clearly, (ρ, L)-list-decodable codes are $(\rho, 1, L)$-list-recoverable codes.

---

3. **[Polynomial reconstruction and list-recoverability of RS codes]** (5+5+5)

In this problem, we will show that the Guruswami-Sudan list-decoding algorithm discussed in lecture can be modified to prove list-recoverability of the Reed-Solomon code.

Recall the setting of the Guruswami-Sudan algorithm.

Input:
- A finite field $\mathbb{F}$, integers $n \geq k \geq 1$ and an agreement parameter $t \leq n$.
- n points $(\alpha_1, \beta_1), (\alpha_2, \beta_2), \ldots, (\alpha_n, \beta_n) \in \mathbb{F} \times \mathbb{F}$ such that all the α_i's are distinct.

Output: List of all polynomials $P(X) \in \mathbb{F}_{<k}[X]$ of degree strictly less than k such that

$$|\{i \in [n] \colon P(\alpha_i) = \beta_i\}| \geq t.$$

The Guruswami-Sudan algorithm used the method of multiplicities to reduce the agreement parameter t to $\lceil\sqrt{(k-1)n}\rceil$.

(a) The Guruswami-Sudan used the same multiplicity parameter r for all points (α_i, β_i). Suppose the algorithm were instead given different multiplicity parameters, say r_i for the i^{th} point (α_i, β_i). Show that that this generalization outputs all polnomials $P(X)$ of degree strictly less than k such

$$\sum_{i:P(\alpha_i)=\beta_i} r_i > \sqrt{(k-1)\sum_{i=0}^{n}\binom{r_i+1}{2}}.$$

(b) In this part, we observe that the α_i's need not be distinct for the Guruswami-Sudan algorithm to work. In fact, it suffices if the n points (α_i, β_i) are distinct. Combining with the above, show that if we are given n distinct points $\alpha_1, \ldots, \alpha_n$ and multiplicity weights $r_{i,\beta}$ for each $i \in [n]$ and $\beta \in \mathbb{F}$, the Guruswami-Sudan can be modified to find all polynomails $P(X)$ of degree strictly less than k such

$$\sum_{i} r_{i,P(\alpha_i)} > \sqrt{(k-1)\sum_{i=0}^{n}\sum_{\beta\in\mathbb{F}}\binom{r_{i,\beta}+1}{2}}.$$

(c) Use the above part to show that the $[n,k]_q$-Reed-Solomon code is $(1-t/n, \ell, \mathrm{poly}(n))$-list-recoverable provided $t > \sqrt{(k-1)\ell n}$.

4. **[Obtaining efficient binary list-decodable codes via code concatenation]** (5+5+5)

In Problem 3, we showed that the $[2^k, k, 2^{k-1}]_2$-Hadamard code is $(1/2-\delta, 1/\delta^2)$-list-decodable. This code has excellent distance and list-decodability but has inverse exponential rate. In this problem, we will construct, for every $\delta \in (0,1)$, explicit codes over the binary alphabet with rate $\mathrm{poly}(1/\delta)$ and $(1/2-\delta, \mathrm{poly}(n))$-list-decodability (accompanied with an efficient list-decoder).

(a) Show that if

i. $\mathcal{C}_{\mathrm{outer}}$ is a $(N, K, D)_{q^k}$-code and is (μ, ℓ, L)-list-recoverable and

ii. $\mathcal{C}_{\mathrm{inner}}$ is a $(n, k, d)_q$-code and is (τ, ℓ)-list-decodable,

then

- the concatenated code $\mathcal{C}_{\mathrm{outer}} \circ \mathcal{C}_{\mathrm{inner}}$ is (ρ, L)-list-decodable provided $\rho \le \mu \cdot \tau$.

(b) Instantiate the outer and inner codes $\mathcal{C}_{\mathrm{outer}}$ and $\mathcal{C}_{\mathrm{inner}}$ as follows:

- $\mathcal{C}_{\mathrm{outer}}$ is a Reed-Solomon code with distance $(1-\varepsilon)$ and appropriate list-recoverability given by Problem 3c [you may use the results of 3c as blackbox (even if you have not proved it)].
- In lecture, we showed for every $\eta \in (0,1)$ a random binary code with rate $R = 1 - h_2(1/2-\eta) - 1/L$ is $(1/2-\eta, L)$-list-decodable. Use this as the inner code $\mathcal{C}_{\mathrm{inner}}$. (Note that we do not if this code is *efficiently* list-decodable.)

(c) Given $\delta \in (0,1)$, choose the parameters ε, η, L appropriately to show that the concatenated code is an explicit code over the binary alphabet with rate $\mathrm{poly}(\delta)$ and $(1/2-\delta, \mathrm{poly}(n))$-list-decodability (accompanied with an efficient list-decoder).