University of California, Berkeley Instructor: Venkatesan Guruswami
CS 294-226: Advances in Error-Correcting Codes Term: Fall 2022

PROBLEM SET 3

Due by 11.59pm PDT Friday, November 4

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 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 which should be a "finished product").
- Solutions should be turned in via gradescope.
- You may consult any of the class notes or videos that were assigned for initial preparation, and the "Essential Coding Theory" text linked from the course page, but please refrain from consulting other sources. In particular, you should not search for solutions on the web. Do not hesitate to contact the GSI or instructor should you have any questions.
- Please start work on the problem set early. The problems vary in difficulty and length and it will be a good idea to give yourself enough time for the trickier problems.

A. **(This problem need not be turned in)** Let X, Y be random variables with finite supports $\mathcal{X}$ and $\mathcal{Y}$ respectively.

   (a) Show that $\hat{x} = \text{argmax}_{x \in \mathcal{X}} \Pr[X = x]$ satisfies $\Pr[X \neq \hat{x}] \leq H(X)$.

   (b) Define $\hat{X} : \mathcal{Y} \to \mathcal{X}$ as $\hat{X}(y) := \text{argmax}_{x \in \mathcal{X}} \Pr[X = x \mid Y = y]$. Prove that

$$\Pr_{X,Y}[X \neq \hat{X}(Y)] \leq H(X \mid Y) \, .$$

B. **(This problem need not be turned in)** Recall the two invertible linear transforms P_n, R_n on $\mathbb{F}_2^n$, n a power of two, that we defined recursively:

   - $P_0(x) = x$, and $P_n(U, V) = (P_{n/2}(U + V), P_{n/2}(V))$ where U (resp. V) is the first $n/2$ (resp. last $n/2$) bits of the n-bit input to P_n.
   - $R_0(x) = x$, and $R_n(U, V)_{2i} = R_{n/2}(U)_i + R_{n/2}(V)_i$ and $R_n(U, V)_{2i+1} = R_{n/2}(V)_i$, for $i = 0, 1, \ldots, n/2 - 1$, where the X_i denotes the i'th bit of vector X, and we index the bits starting at 0.

Recall that we used the transform P_n to conveniently present and analyze the successive cancellation decoder, and the transform R_n to conveniently analyze the polynomially strong polarizing property of the transform. In this exercise, we will relate these transforms to conclude that the polarizing property of R_n implies that of P_n, as alluded to in class.

(a) Show that for $n = 2^m$, the rows (resp. columns) of P_n can be indexed by subsets I (resp. J) of $\{1, 2, \ldots, m\}$ in some order so that

$$P_n(I, J) = \begin{cases} 1 & \text{if } I \subseteq J \\ 0 & \text{otherwise} \end{cases}$$

(b) Show that $R_n = B_n P_n$ where $B_n : \mathbb{F}_2^n \to \mathbb{F}_2^n$ permutes the coordinates via "bit-reversal," that is, it maps location i with binary representation $b_m b_{m-1} \cdots b_2 b_1$ (where $m = \lg n$) to location j with binary representation $b_1 b_2 \cdots b_{m-1} b_m$.

(c) Prove that P_n and B_n commute, i.e., $B_n P_n = P_n B_n$.

(d) Recap why the above implies that R_n and P_n have identical polarization properties.

---

1. Recall the Wozencraft ensemble of rate $1/2$ codes C_α, $\alpha \in \mathbb{F}_{2^m}^*$, with codewords $(x, \alpha \cdot x)$ for $x \in \mathbb{F}_{2^m}$, viewed as a vector in $\mathbb{F}_2^{2m}$ under some fixed basis of $\mathbb{F}_{2^m}$ over $\mathbb{F}_2$. Let $\epsilon > 0$ and let m be large enough. Prove that with high probablity a random code drawn from the Wozencraft ensemble (i.e., C_α for random $\alpha \in \mathbb{F}_{2^m}^*$), admits a decoder that achieves vanishing (in m) decoding error probability on BSC_p for $p = h^{-1}(1/2) - \epsilon$. Informally put, a random codes from the Wozencraft ensemble achieves the capacity of the binary symmetric channel.

2. In this problem, you will show that codes over large alphabets are uniquely decodable from most error patterns of Hamming weight approaching their distance (as opposed to half the distance). Let $\epsilon > 0$ and $q \geq \exp(a/\epsilon)$ for some large enough absolute constant a. Let $0 < \delta < 1 - 1/q$ and let $C \subseteq \{0, 1, \ldots, q-1\}^n$ be a code with relative distance δ.

   Let $E \subseteq \{1, 2, \ldots, n\}$ be an arbitrary subset of size $(\delta - \epsilon)n$, and $c \in C$ be an arbitrary codeword of C.

   Prove that for all but $q^{-\Omega(\epsilon n)}$ of the error patterns e supported on E (namely, $e_i = 0$ for $i \notin E$ and $e_i \neq 0$ for $i \in E$), the codeword c is the unique codeword of C within Hamming distance at most $(\delta - \epsilon)n$ from $c + e$.

3. Let $Z \in \mathbb{F}_2^n$ be sampled as n i.i.d copies of $\text{Ber}(p)$ and let $W = P_n(Z)$.

   (a) Show how, given $i \in \{1, 2, \ldots, n\}$, one can efficiently sample the random variable $W_{<i}$, i.e., W restricted to the first $i - 1$ coordinates.

   (b) Given an arbitrary string $w_{<i} \in \mathbb{F}_2^{i-1}$, show how one can compute the distribution of W_i conditioned on $W_{<i} = w_{<i}$. (Hint: Use the approach behind the successive cancellation decoder.)

   (c) Using the above parts, show that there is a randomized algorithm running in $\text{poly}(n, 1/\gamma)$ time which with probability at least $1 - 1/n^2$ outputs an estimate of the conditional entropy $H(W_i | W_{<i})$ within an additive error of γ.

(d) Suppose that P_n is $(\epsilon, 1/n^2)$-polarizing, i.e., the set $\{i \mid H(W_i \mid W_{<i}) \geq 1/n^2\}$ has size at most $(h(p)+\epsilon)n$ (we proved this property in class provided n is at least a sufficiently big polynomial function of $1/\epsilon$).

Combine the steps above to give a poly(n) time randomized algorithm that with probability at least $1-1/n$ outputs a set S of size at most $n(h(p)+\epsilon)$ such that for all $i \notin S$, $H(W_i \mid W_{<i}) \leq 3/n^2$.

4. Consider the polarizing matrix P_n for $n = 2^m$ from lecture, with rows indexed as mentioned in Problem B by subsets I of $\{1, 2, \ldots, m\}$.

   (a) For $0 \leq r \leq m$, let $C^{(m,r)}$ be the linear code which is the row-span of the rows of P_n indexed by subsets I of size at most r. What is the dimension of $C^{(m,r)}$? Prove that the minimum distance of $C^{(m,r)}$ equals 2^{m-r}.

   (b) For a family $\mathcal{F}$ of subsets of $\{1, 2, \ldots, m\}$, let $C_{\mathcal{F}}$ denote the code which is the row span of the rows of P_n indexed by all $I \in \mathcal{F}$.

   Prove that the minimum distance of $C_{\mathcal{F}}$ is equal to $\min_{I \in \mathcal{F}} 2^{m-|I|}$.

   (c) Argue that any such code $C_{\mathcal{F}}$ with rate R (i.e., with $|\mathcal{F}| = Rn$) for R bounded away from 0 (as $n \to \infty$) must have minimum distance at most $n^{1/2+o(1)}$.

5. Let $p \in (0, 1/2)$. Suppose $H \in \mathbb{F}_2^{m \times n}$ is a linear compression matrix and Decompress : $\mathbb{F}_2^m \to \mathbb{F}_2^n$ is such that

$$\Pr_{Z \sim \text{Ber}(p)^{\otimes n}} \Big[\text{Decompress}(HZ) \neq Z\Big] \leq 2^{-t} .$$

   Prove that the code with parity check matrix H has minimum distance at least $\frac{t}{\lg(1/p)}$.

   (Thus linear compression schemes for i.i.d Bernoulli sources (equivalently linear codes for BSC) with very good error probability imply codes with good minimum distance.)

   Hint: The optimal decompression algorithm is the maximum likelihood algorithm, that on input $w \in \mathbb{F}_2^m$, outputs $\arg\min_{x:Hx=w} \text{wt}(x)$ (Why?)

6. For a fixed $p \in (0, 1/2)$ and integer $L \geq 2$, we proved in class that there are (p, L)-list decodable binary codes of rate $1-h(p)-1/L$. In this exercise we will prove an upper bound of $\approx 1-h(p)-p^L$ on the rate of any (p, L)-list decodable code. (This shows that for every finite list-size L, the rate has to be bounded away from the capacity $1-h(p)$. However, there is an exponential gap in terms of convergence w.r.t L between the lower and upper bounds. In particular, we can have rate $1-h(p)-\epsilon$ with list size $O(1/\epsilon)$, whereas we only know a lower bound of $\Omega_p(\log(1/\epsilon))$ on the list size needed to be within ϵ of list decoding capacity.)

   (a) Suppose $C \subseteq \{0,1\}^n$ is a $(p.L)$-list decodable code where every codeword has weight exactly μn where $\mu = p + p^L/2$. Prove that $|C| \leq 2L^2/p$.

   Hint: Pick a random L-tuple of codewords from C, and lower bound the expected number of coordinates where they are all 1.

   (b) Using the above, argue that the rate of a (p, L)-list decodable code is at most $1-h(p+p^L/2)+o(1)$. Argue that this is at most $1-h(p)-\frac{(1-2p)}{8} \cdot p^L$.