Carnegie Mellon University Instructor: Venkatesan Guruswami
15-859Y: Coding Theory Term: Fall 2018

PROBLEM SET 1
Due date: Friday, September 28

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 (see the last bullet item for some "credit" for doing so).

- 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).
- You should not search for solutions on the web. More generally, you should try and solve the problems without consulting any reference material other than the course notes and what we cover in class. If for some reason you feel the need to consult some source, *please acknowledge the source* and try to articulate the difficulty you couldn't overcome before consulting the source and how it helped you overcome that difficulty. Alternatively, before turning to any such material, we encourage you to ask the instructor for hints or clarifications.
- Please start work on the problem set early. The problem set has **seven** problems and is worth a total of 120 points. As a rough estimate, a score around 105, or 95 if you work by yourself, might correspond to an A level performance on this problem set.

---

1. (15 points) Let $G = (V, E)$ be any undirected graph (assume no loops or multiple edges). A *cut* in the graph is the subset of all edges that connect a vertex in S to vertex in $V \setminus S$, for some subset $S \subseteq V$. Let Cuts$(G) \subseteq \{0, 1\}^E$ consist of the characteristic vectors of all cuts of G.

   (a) Prove that Cuts(G) is an $[|E|, |V| - 1]_2$ binary linear code. What parameter of G equals the distance of Cuts(G)?

   (b) Describe the dual code Cuts$(G)^\perp$ of Cuts(G). What is its dimension? What parameter of G equals the distance of Cuts$(G)^\perp$?

2. (20 points) In this problem you will need to come up with some ways of constructing new codes from existing ones, and prove the following statements. Recall that $[n, k, d]_q$ stands for an length n *linear code* over $\mathbb{F}_q$ of dimension k.

   (a) If there exists an $[n, k, d]_q$ code ($d \geq 2$), then there also exists an $[n - 1, k, d' \geq d - 1]_q$ code.

(b) If there exists an $[n, k, d]_2$ code with d *odd*, then there also exists an $[n+1, k, d+1]_2$ code. What is the code that you get when you apply this transformation to the $[2^r - 1, 2^r - 1 - r, 3]_2$ Hamming code? What is the dual of this code?

(c) If there exists an $[n, k, d]_q$ code, then there also exists an $[n - d, k - 1, d' \geq \lceil d/q \rceil]_q$ code. (Hint: Drop the d positions corresponding to the support of a minimum weight codeword.)

(d) If there exists an $[n, k, d]_2$ code ($0 < d < n/2$), then for every $m \geq 1$, there also exists an $\left[n^m, k, \frac{n^m - (n-2d)^m}{2}\right]_2$ code.

(Hint: Given an $n \times k$ generator matrix G for the code, consider the $n^m \times k$ generator matrix whose $(i_1, i_2, \ldots, i_m)$'th row is the sum of rows $i_1, i_2, \ldots, i_m$ of G. It is also more slick to use a ± 1 notation for binary alphabet via the translation $b \mapsto (-1)^b$ from $\{0, 1\}$ to $\{1, -1\}$, and track the bias $\mathbf{E}_{i \in \{1,\ldots,N\}}[x_i]$ of a string $x \in \{-1, 1\}^N$ as a proxy for its relative Hamming weight.)

3. (15 points) A set of vectors $S \subseteq \mathbb{F}_q^n$ is called t-wise independent if for every set of positions I with $|I| = t$, the set S projected to I has each of the vectors in $\mathbb{F}_q^t$ appear the same number of times. (In other words, if one picks a vector $(s_1, \ldots, s_n)$ from S at random then any of the t random variables are uniformly and independently random over $\mathbb{F}_q$).

   Prove that any linear code C whose dual $C^\perp$ has distance $d^\perp$ is $(d^\perp - 1)$-wise independent.

4. (15 points) Let C_1 be an $[n_1, k_1, d_1]_2$ binary linear code, and C_2 an $[n_2, k_2, d_2]$ binary linear code. Let $C \subseteq \mathbb{F}_2^{n_2 \times n_1}$ be the subset of $n_2 \times n_1$ matrices whose rows belong to C_1 and whose columns belong to C_2 (view elements of C as binary vectors of length $n_1 n_2$ in some canonical way).

   Prove that C is an $[n_1 n_2, k_1 k_2, d_1 d_2]_2$ binary linear code.

5. (15 points) A $n \times k$ *Toeplitz Matrix* $A = \{A_{i,j}\}_{i=1,\ j=1}^{k\ ;\ n}$ satisfies the property that $A_{i,j} = A_{i-1,j-1}$. In other words, any diagonal has the same value. For example, the following is a 6×4 Toeplitz matrix:

$$\begin{pmatrix} 1 & 7 & 8 & 9 \\ 2 & 1 & 7 & 8 \\ 3 & 2 & 1 & 7 \\ 4 & 3 & 2 & 1 \\ 5 & 4 & 3 & 2 \\ 6 & 5 & 4 & 3 \end{pmatrix}$$

   A *random* $n \times k$ Toeplitz matrix $T \in \mathbb{F}_q^{n \times k}$ is chosen by picking the entries in the first row and column uniformly (and independently) at random.

   (a) Prove the following claim: For any non-zero $\mathbf{m} \in \mathbb{F}_q^k$, the vector $T \cdot \mathbf{m}$ is uniformly distributed over $\mathbb{F}_q^n$.

   (b) Briefly argue why the claim above implies that a random code defined by picking its generator matrix as a random Toeplitz matrix with high probability achieves the Gilbert-Varshamov bound.

   (c) Conclude that an $[n, k]_q$ code on the Gilbert-Varshamov bound can be constructed in time $q^{O(n)}$.

6. (20 points) In this problem, we will re-derive the Gilbert-Varshamov bound in a graph-theoretic view, and then proceed to improve it (in the lower-order terms in the asymptotic view). We will restrict ourselves to binary codes for simplicity. For integers $1 \leq d \leq n$, consider the graph $G_{n,d}$ whose vertex set is $\{0, 1\}^n$ and two vertices are adjacent if their Hamming distance is *less than* d.

(a) Argue that any independent set of the graph $G_{n,d}$ is a code of distance at least d. (Recall that an independent set in a graph G is a subset of vertices no two of which are adjacent.)

(b) It is a well-known and easy fact that any graph G on N vertices and maximum degree Δ has an independent set of size at least $N/(\Delta+1)$. Using this fact, argue a lower bound on the size of the maximum independent set in $G_{n,d}$, and then deduce the Gilbert-Varshamov bound, namely the existence of binary codes with distance at least d and size at least $\frac{2^n}{\sum_{i=0}^{d-1}\binom{n}{i}}$.

We will now slightly improve the above bound using a more sophisticated argument, via counting triangles in the graph $G_{n,d}$. (A triangle in a graph $G=(V,E)$ is a set $\{u,v,w\} \subset V$ of three distinct vertices such that all three vertices are adjancent, i.e., $(u,v),(v,w),(w,u) \in E$.)

(a) Argue that a graph on N vertices of maximum degree Δ has at most $O(N\Delta^2)$ triangles.

(b) Prove that the number of triangle in graph $G_{n,d}$ is at most

$$2^n \cdot \sum_{0 \le \ell \le 3d/2} \binom{n}{\ell} \cdot 3^\ell.$$

Hint: Fix u and let ℓ count the number of coordinates where at least one of v or w disagree with u. Prove that ℓ is at most $3d/2$.

(c) For simplicity, we will restrict to the case $d = n/5$ below. Simplify the above expression in the case where $d = n/5$ to show that the number of triangles in $G_{n,n/5}$ is $O(N \cdot \Delta^{2-\eta})$ for some $\eta > 0$.

(d) A famous result in the "probabilistic method" shows (and you don't have to prove this), that if a graph on N vertices of maximum degree Δ has at most $O(N \cdot \Delta^{2-\eta})$ triangles, then it has an independent set of size $\Omega(\frac{N}{\Delta}\log\Delta)$. Use this result to conclude that there is a binary code of block length n and distance $n/5$ of size $\Omega\left(n\frac{2^n}{\sum_{i=0}^{n/5-1}\binom{n}{i}}\right)$. (Note that this improves over the GV-bound by an $\Omega(n)$ factor.)

7. (20 points) Let C be $[n,k]_2$ linear code with a_j denoting the number of codewords of C of Hamming weight j, for $0 \le j \le n$. (So $a_0 = 1$ and $\sum_{j=0}^n a_j = 2^k$.) Let $A(X) = \sum_{j=1}^n a_j X^j$ be the "weight-enumerator" polynomial of C (leaving out the all-zeroes codeword).

Suppose C is used for transmission on a discrete memoryless channel ($\mathcal{X} = \{0,1\}, \mathcal{Y}, W$) with maximum likelihood decoding at the receiver. That is, if $\mathbf{y} \in \mathcal{Y}^n$ is received, the decoding rule outputs a codeword $\mathbf{c} \in C$ for which $p(\mathbf{y}|\mathbf{c}) = \prod_{i=1}^n W(y_i|c_i)$ is maximum (ties broken arbitrarily).

Prove that regardless of which codeword was transmitted, the resulting error probability P_{err} is at most $P_{\text{err}} \le A(Z(W))$ where $Z(W) := \sum_{y\in\mathcal{Y}} \sqrt{W(y|0)W(y|1)}$.

For W corresponding to BSC_p, what is the value of $Z(W)$? Using this and the above bound on P_{err}, conclude that for every fixed $p < 1/2$, any asymptotically good binary can be used to communicate on the BSC_p with positive rate and decoding error probability at most $2^{-\Omega(n)}$.