Coding Theory 24 Aug, 2016

Problem Set 1

- Due Date: **12 Sep (Mon), 2016**
- 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.
- Be clear in your writing.
- Problems 1,2,4,5 and 6 are from the book "Essential Coding Theory" (Guruswami, Rudra and Sudan) while Problem 7 is due to Zeev Dvir.

1. (2+4+2+7) There are n people in a room, each of whom is given a black/white hat chosen uniformly at random (and independent of the choices of all other people). Each person can see the hat color of all other people, but not their own. Each person is asked if they wish to guess their own hat color. They can either guess, or abstain. Each person makes their choice without knowledge of what the other people are doing. They either win collectively, or lose collectively. They win if all the people who don't abstain guess their hat color correctly and at least one person does not abstain. They lose if all people abstain, or if some person guesses their color incorrectly. Your goal below is to come up with a strategy that will allow the n people to win with pretty high probability. We begin with a simple warmup:

   (a) Argue that the n people can win with probability at least $\frac{1}{2}$.
   Next we will see how one can really bump up the probability of success with some careful modeling, and some knowledge of Hamming codes.

   (b) Lets say that a directed graph G is a subgraph of the n-dimensional hypercube if its vertex set is $\{0,1\}^n$ and if $u \to v$ is an edge in G ,then u and v differ in at most one coordinate. Let $K(G)$ be the number of vertices of G with in-degree at least one, and out-degree zero. Show that the probability of winning the hat problem equals the maximum, over directed subgraphs G of the n-dimensional hypercube, of $K(G)/2^n$.

   (c) Using the fact that the out-degree of any vertex is at most n, show that $K(G)/2^n$ is at most $\frac{n}{n+1}$ for any directed subgraph G of the n-dimensional hypercube.

(d) Show that if $n = 2^r - 1$, then there exists a directed subgraph G of the n-dimensional hypercube with $K(G)/2^n = \frac{n}{n+1}$.
Hint: This is where the Hamming code comes in

2. (1+2+3+4+5) 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 block-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.

   (c) If there exists an $[n,k,d]_q$ code, there 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 $[n,k_1,d_1]_q$ code and an $[n,k_2,d_2]_q$ code, then there also exists a $[2n,k_1+k_2,\min(2d_1,d_2)]_q$ code.

   (e) 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 $b \to (-1)^b$ from $\{0,1\}$ to $\{1,-1\}$ and track the bias $\mathbb{E}_{i \in 1,\ldots,N}[x_i]$ of a string $x \in \{-1,1\}^N$ as a proxy for its relative Hamming weight.

3. (5+5+5) Given a $(n_1,k_1,d_1)_q$ code C_1 and a $(n_2,k_2,d_2)_q$ code C_2, the direct product of C_1 and C_2, denoted $C_1 \otimes C_2$, is an $(n_1n_2, k_1k_2, d)_q$ code constructed as follows. View a message of $C_1 \otimes C_2$ as a k_2 by k_1 matrix M. Encode each row of M by the code C_1 to obtain an k_2 by n_1 intermediary matrix. Encode each column of this intermediary matrix with the C_2 code to get an n_2 by n_1 matrix representing the codeword encoding M.

   In this problem, we first show that the resulting code has distance at least d_1d_2 in either case. Then we show that if C_1 and C_2 are linear, then the resulting code is also linear, and furthermore is the same as the code that would be obtained by encoding the columns with C_2 first and then encoding the rows with C_1.

   (a) Prove that the distance of the code $C_1 \otimes C_2$ is at least d_1d_2.

   (b) Suppose C_1 and C_2 are linear codes. Let $G_1 \in \mathbb{F}_q^{n_1 \times k_1}$ be a generator matrix for the code C_1 and $G_2 \in \mathbb{F}_q^{n_2 \times k_2}$ be a generator matrix for the code C_2. Show that the direct product code $C_1 \otimes C_2$ is a linear code that has as its codewords

$$\{G_2 M G_1^T \mid M \in \mathbb{F}_q^{k_2 \times k_1}\}.$$

Conclude that the code $C_1 \otimes C_2$ is linear if C_1 and C_2 are. Also, that the same code is obtained by encoding the columns with C_2 first and then encoding the rows in the intermediate matrix with C_1.

(c) Suppose C_1 and C_2 are linear codes. Show that the code $C_1 \otimes C_2$ is equivalent to the following code whose codewords are all $n_2 \times n_1$ matrices whose rows are codewords of C_1 and columns are codewords of C_2.

4. (7+7+1) In this exercise we will prove the following q-ary version of the Plotkin bound via a purely combinatorial proof.

If $C \subseteq [q]^n$ is a code with distance d and if $d > \left(1 - \frac{1}{q}\right) n$, then $|C| \leq \frac{qd}{qd-(q-1)n}$.

Given an $(n,k,d)_q$ code C with $d > \left(1 - \frac{1}{q}\right) n$, define

$$S = \sum_{c_1 \neq c_2 \in C} \Delta(c_1, c_2).$$

For the rest of the problem, think of C as an $|C| \times n$ matrix where each row corresponds to a codeword in C . Now consider the following:

(a) Looking at the contribution of each column in the matrix above , argue that

$$S \leq \left(1 - \frac{1}{q}\right) n|C|^2.$$

(b) Looking at the contribution of the rows in the matrix above, argue that

$$S \geq |C|(|C|-1)d.$$

(c) Conclude the q-ary version of Plotkin's bound.

5. (6+9) For integers $1 \leq k \leq n$, call a (multi)set $S \subseteq \{0,1\}^n$ to be k-wise independent if for every $1 \leq i_1 < i_2 < \cdots < i_k \leq n$ and $(a_1, a_2, \ldots, a_k) \in \{0,1\}^k$,

$$\Pr_{x \in S}\left[x_{i_1} = a_1 \wedge x_{i_2} = a_2 \wedge \cdots \wedge x_{i_k} = a_k\right] = \frac{1}{2^k}$$

where the probability is over an element x chosen uniformly at random from S . Small sample spaces of k-wise independent sets are of fundamental importance in derandomization. In this problem, you will see how codes can be used to construct k-wise independent sets of near-optimal size.

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

(b) Using BCH codes and the previous part, show how one can construct a $2t$ -wise independent subset of $\{0,1\}^n$ of size at most $(n+1)^t$ when n is of the form $2^m - 1$.

(c) [Extra credit] Prove an almost matching lower bound, namely any k-wise independent set $S \subseteq \{0,1\}^n$ satisfies

$$|S| \geq \sum_{i=0}^{\lfloor \frac{k}{2} \rfloor} \binom{n}{i}. \tag{1}$$

Hint: Find a set of linearly independent vectors in $\mathbb{R}^{|S|}$ of cardinality at least the R.H.S of (**??**). Specifically, for $T \subseteq \{1,2,\ldots,n\}$ of size $\leq \lfloor k/2 \rfloor$, consider the $\langle \chi_T(x) \rangle_{x \in S}$ where $\chi_T(x) = (-1)^{\sum_{i \in T} x_i}$.

6. (15) For this problem, consider the following problem:

   Input Instance: A set $S = \{\alpha_1, \ldots, \alpha_n\} \subseteq \mathbb{F}_{2^m}$, an element $\beta \in \mathbb{F}_{2^m}$, and an integer $1 \leq k < n$.

   Question: Is there a nonempty subset $T \subseteq \{1,2,\ldots n\}$ with $|T| = k+1$ such that $\sum_{i \in T} \alpha_i = \beta$.

   [Note: It can be shown that this problem is NP-hard via a reduction from subset sum.]

   Consider the $[n, k, n-k+1]_{2^m}$ Reed-Solomon code $RS_{n,k,S}$ over $\mathbb{F}_{2^m}$ obtained by evaluating polynomials of degree at most $k-1$ at points in S. Define $y \in (\mathbb{F}_{2^m})^n$ as follows: $y_i = \alpha_i^{k+1} - \beta \alpha_i^k$ for $i = 1, 2, \ldots, n$.

   Prove that there is a codeword of $RS_{n,k,S}$ at Hamming distance at most $n-k-1$ from y if and only if there is a set T as above of size $k+1$ satisfying $\sum_{i \in T} \alpha_i = \beta$.

   This implies that finding the nearest codeword in a Reed-Solomon code over exponentially large fields is NP-hard. (Proving this for polynomial sized fields remains an embarrassing open question.)

7. (4+2+2+2) Let $\mathbb{F}$ be a finite field of size q. A *Kakeya set* in $\mathbb{F}^m$ is a set $K \subseteq \mathbb{F}^n$ such that K contains a line in every direction. More precisely, K is a Kakeya set if for every $y \in \mathbb{F}^m$ there exists a $z \in \mathbb{F}^m$ such that the line

   $$L_{z,y} = \{z + t \cdot y | t \in \mathbb{F}\}$$

   is contained in K.

   A trivial upper bound on th size of K is q^m and this can be improved to $q^m/2^{m-1}$. In this problem, we will use the polynomial method to show a lower bound of $q^m/m!$. More precisely, we will show that

   $$K \geq \binom{q+m-1}{m}.$$

   Suppose, for contradiction that this is not the case.

   (a) Show that there exists a m-variate non-zero polynomial g of degree $d \leq q-1$ such that $g(x) = 0$ for all $x \in K$.

Let g_d be the homogenous part of degree d of g so that g_d is non-zero and homogenous.

For any $y \in \mathbb{F}^m$, we know that there exists a $z \in \mathbb{F}^m$ such that the line $L_{z,y}$ is contained in K. Consider the following univariate polynomial

$$P_{y,z}(t) := g(z + t \cdot y).$$

(b) Argue that $P_{y,z}$ is identically zero and hence the coefficient of t^d in $P_{y,z}(t)$ is zero.

(c) Show that the coefficient of t^d in $P_{y,z}(t)$ is exactly $g_d(y)$.

(d) Conclude that g_d is identically zero, a contradiction.