CSE 533: Error-Correcting Codes **Autumn 2006**

Problem Set 2

Due on Friday, December 1

There are problems for a total of 80 points. **It is enough to turn in any subset of (whole) problems worth 60 points.** If you turn in more problems, the score will be scaled down to 60.

Homework policy: Students are encouraged to work on the problems in small groups (of up to 3 people); however, all writeups should be done individually and must clearly cite any collaborators. You are strongly urged to try and solve the problems without consulting any reference material besides what we cover in class (such as any textbooks or material on the web where the solution may appear either fully or in part). If for some reason you feel the need to consult some source, please acknowledge the source and try to explain what difficulty you couldn't overcome before consulting the source and how it helped you overcome that difficulty.

1. (10 points) **Computing a linear code on GV bound in singly exponential time.**
   Give a deterministic algorithm that given positive integers $n \geq d \geq 2$, runs in $2^{O(n)}$ time and outputs an $[n, k, d]_2$ binary linear code with

   $$2^k \geq \frac{2^{n-1}}{\mathrm{Vol}_2(n-1, d-2)} \ .$$

   Conclude that one can compute a linear code meeting the GV bound in singly exponential time.

2. (10 points) **Dual of Reed-Muller codes.**
   Denote by $R(r, m)$ the binary m-variate Reed-Muller code of order r (recall that it is a $[2^m, k, 2^{m-r}]_2$ code where $k = \sum_{i=0}^{r} \binom{m}{i}$).

   Prove that the dual of $R(r, m)$ is the Reed-Muller code $R(m - r - 1, m)$.

3. (20 points) **Parity Check view of Reed-Solomon codes.**
   Consider the Reed-Solomon code over a field $\mathbb{F}$ of size q and block length $n = q - 1$ defined as

   $$\mathrm{RS}_{\mathbb{F}}[n, k, n-k+1] = \{(p(1), p(\alpha), \ldots, p(\alpha^{n-1})) \ \mid \ p(X) \in \mathbb{F}[X] \text{ has degree } \leq k-1\}$$

   where α is the generator of the multiplicative group $\mathbb{F}^*$ of $\mathbb{F}$.

   (a) Prove that

   $$\begin{aligned} \mathrm{RS}_{\mathbb{F}}[n, k, n-k+1] \ &= \ \{(c_0, c_1, \ldots, c_{n-1}) \in \mathbb{F}^n \mid c(\alpha^\ell) = 0 \text{ for } 1 \leq \ell \leq n-k \ , \\ & \quad \text{where } c(X) = c_0 + c_1 X + \cdots + c_{n-1} X^{n-1}\} \ . \end{aligned} \tag{1}$$

   Hint: Prove that the identity $\sum_{i=0}^{n-1} \alpha^{ji} = 0$ holds for all j, $1 \leq j \leq n-1$, and then make use of it.

   (b) Now suppose that $\mathbb{F} = \mathbb{F}_{2^m}$. Consider the binary code C_{BCH} defined as $\mathrm{RS}_{\mathbb{F}}[n, k, n-k+1] \cap \mathbb{F}_2^n$.[1]

[1] The subscript BCH stands for Bose-Chaudhuri-Hocquenghem, the discoverers of this family of codes.

i. Prove that C_{BCH} is a binary linear code of distance at least $d = n - k + 1$ and dimension at least $n - (d-1)\log_2(n+1)$.
<u>Hint</u>: Use the characterization (1) of the Reed-Solomon code.

ii. Prove a better lower bound of $n - \left\lceil \frac{d-1}{2} \right\rceil \log_2(n+1)$ on the dimension of C_{BCH}.
<u>Hint</u>: Try to find redundant checks amongst the "natural" parity checks defining C_{BCH}.

iii. For $d = 3$, C_{BCH} is the same as another code we have seen. What is that code?

iv. For constant d (and growing n), prove that C_{BCH} have nearly optimal dimension for distance d, in that the dimension cannot be $n - t\log_2(n+1)$ for $t < \frac{d-1}{2}$.

4. (20 points) **BCH-like codes that meet the Gilbert-Varshamov bound.**
In this problem, we continue in the theme of Problem 3b and look at the intersection of a Reed-Solomon code with $\mathbb{F}_2^n$ to get a binary code. Let $\mathbb{F} = \mathbb{F}_{2^m}$. Fix positive integers d, n with $(d-1)m < n < 2^m$, and a set $S = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ of n distinct nonzero elements of $\mathbb{F}$. For a vector $\mathbf{v} = (v_1, \ldots, v_n) \in (\mathbb{F}^*)^n$ of n not necessarily distinct nonzero elements from $\mathbb{F}$, define the *Generalized Reed-Solomon code* $\mathrm{GRS}_{S,\mathbf{v},d}$ as follows:

$$\mathrm{GRS}_{S,\mathbf{v},d} = \{(v_1 p(\alpha_1), v_2 p(\alpha_2), \ldots, v_n p(\alpha_n)) \mid p(X) \in \mathbb{F}[X] \text{ has degree } \leq n-d\} .$$

(a) (Gentle warm-up) Prove that $\mathrm{GRS}_{S,\mathbf{v},d}$ is an $[n, n-d+1, d]_{\mathbb{F}}$ linear code.

(b) (Some more warm-up) Argue that $\mathrm{GRS}_{S,\mathbf{v},d} \cap \mathbb{F}_2^n$ is a binary linear code of rate at least $1 - \frac{(d-1)m}{n}$.

(c) Let $\mathbf{c} \in \mathbb{F}_2^n$ be a nonzero binary vector. Prove that (for every choice of d, S) there are at most $(2^m - 1)^{n-d+1}$ choices of the vector $\mathbf{v}$ for which $\mathbf{c} \in \mathrm{GRS}_{S,\mathbf{v},d}$.

(d) Using the above, prove that if the integer D satisfies $\mathrm{Vol}_2(n, D-1) \leq (2^m-1)^{d-1}$ (where $\mathrm{Vol}_2(n, D-1) = \sum_{i=0}^{D-1} \binom{n}{i}$), then there exists a vector $\mathbf{v} \in (\mathbb{F}^*)^n$ such that the minimum distance of the binary code $\mathrm{GRS}_{S,\mathbf{v},d} \cap \mathbb{F}_2^n$ is at least D.

(e) Using Parts (4b) and (4d) above, argue that the family of codes $\mathrm{GRS}_{S,\mathbf{v},d} \cap \mathbb{F}_2^n$ contains binary linear codes that meet the Gilbert-Varshamov bound.

5. (10 points) **Rate of linear list-decodable codes.** For $0 < p < 1$ and a positive integer L, call a code $C \subset \Sigma^n$ to be (p, L)-list decodable if every Hamming ball of radius pn (in the space Σ^n) has at most L codewords of C. Prove that for every finite field $\mathbb{F}_q$, $0 < p < 1 - 1/q$, integer $L \geq 1$, and large enough n, there is a (p, L)-list decodable linear code $C \subseteq \mathbb{F}_q^n$ that has rate at least $1 - H_q(p)\left(1 + \frac{1}{\log_q(L+1)}\right) - o(1)$.
<u>Hint</u>: Apply the usual random coding method of picking a generator matrix at random. In estimating the probability that L nonzero messages all get mapped into a ball of radius pn, these L events are not all independent (and this is the difference compared to picking a general random code). But at least how many of these events are independent of one another?

6. (10 points) **Chinese Remainder codes.** In this problem, we will consider the number-theoretic counterpart of Reed-Solomon codes. Let $1 \leq k < n$ be integers and let $p_1 < p_2 < \cdots < p_n$ be n distinct primes. Denote $K = \prod_{i=1}^{k} p_i$ and $N = \prod_{i=1}^{n} p_i$. The notation $\mathbb{Z}_M$ stands for integers modulo M, i.e., the set $\{0, 1, \ldots, M-1\}$. Consider the *Chinese Remainder code* defined by the encoding map $E : \mathbb{Z}_K \to \mathbb{Z}_{p_1} \times \mathbb{Z}_{p_2} \times \cdots \times \mathbb{Z}_{p_n}$ defined by:

$$E(m) = (m \mod p_1,\ m \mod p_2,\ \cdots,\ m \mod p_n) .$$

(Note that this is not a code in the usual sense we have been studying since the symbols at different positions belong to different alphabets. Still notions such as distance of this code make sense and are studied in the questions below.)

(a) Suppose that $m_1 \neq m_2$. For $1 \leq i \leq n$, define the indicator variable $b_i = 1$ if $E(m_1)_i \neq E(m_2)_i$ and $b_i = 0$ otherwise. Prove that $\prod_{i=1}^{n} p_i^{b_i} > N/K$.

Use the above to deduce that when $m_1 \neq m_2$, the encodings $E(m_1)$ and $E(m_2)$ differ in at least $n-k+1$ locations.

(b) This exercise examines how the idea behind the Welch-Berlekamp decoder for Reed-Solomon codes can be used to decode Chinese Remainder codes.

Suppose $\mathbf{r} = (r_1, r_2, \ldots, r_n)$ is the received word where $r_i \in \mathbb{Z}_{p_i}$. By Part (a), we know there can be at most one $m \in \mathbb{Z}_K$ such that

$$\prod_{i:E(m)_i \neq r_i} p_i^{b_i} \leq \sqrt{N/K}\ . \tag{2}$$

(Be sure you see why this is the case.)

The exercises below develop a method to find the unique such m, assuming one exists.

In what follows, let r be the unique integer in $\mathbb{Z}_N$ such that $r \mod p_i = r_i$ for every $i = 1, 2, \ldots, n$ (note that the Chinese Remainder theorem guarantees that there is a unique such r).

i. Assuming an m satisfying (2) exists, prove that there exist integers y, z with $0 \leq y < \sqrt{NK}$ and $1 \leq z \leq \sqrt{N/K}$ such that $y \equiv rz \pmod{N}$.

ii. Prove also that if y, z are any integers satisfying the above conditions, then in fact $m = y/z$.

(**Remark:** A pair of integers (y, z) satisfying above can be found by solving the integer linear program with integer variables y, z, t and linear constraints: $0 < z \leq \sqrt{N/K}$; and $0 \leq z \cdot r - t \cdot N < \sqrt{NK}$. This is an integer program in a fixed number of dimensions and can be solved in polynomial time. Faster, easier methods are also known for this special problem.)

(c) **(For your cognitive pleasure only; no need to turn this part in)** Instead of condition (2) what if we want to decode under the more natural condition for Hamming metric, that is $|\{i : E(m)_i \neq r_i\}| \leq \frac{n-k}{2}$? Using ideas similar to GMD decoding, show how this can be done by calling the above decoder many times with different sets of erasures.