Error Correcting Codes – Questions Pool

Amnon Ta-Shma and Dean Doron

January 3, 2018

General guidelines

The questions fall into several categories:

(Know).	Make sure you know how to solve. Do not submit.
(Mandatory).	Mandatory questions.
(Bonus).	Bonus questions.

On the submission date we will collect your answers. We will then go over the questions and solve them in class. After that you have a week to write the solutions and submit to us, as long as

1. You write the solutions alone,
2. You give credit to any source (or any person) you consulted with.

You have to submit solutions to, at least, the mandatory questions. We give the same grade to solutions that were submitted before or after we solved the question in class.

HW 1 – Basic error-correcting codes.

Out: 29.10.2017
Due: 12.11.2017

1. (Know). What is the expected number of roots of a random univariate polynomial of degree d over $\mathbb{F}_q$?

2. (Know). Let $\mathbb{F}_q$ be a finite field of odd characteristic. An element $x \in \mathbb{F}_q^\star$ is a *quadratic residue* if there exists $y \in \mathbb{F}_q$ such that $x = y^2$. What is the number of quadratic residues in $\mathbb{F}_q$? Prove that the set of quadratic residues is a multiplicative group.

3. (Know). Prove that the number of distinct generator matrices of $[n,k]_q$ codes is $\prod_{i=0}^{k-1}(q^k - q^i)$.

4. (Know). Let $\mathcal{C}$ be an $[n,k,d]_q$ code. We can *extend* $\mathcal{C}$ to $\mathcal{C}^\star$ by adding a parity-check bit. That is,
$$\mathcal{C}^\star \;=\; \left\{x \in \mathbb{F}_q^{n+1} \mid x_1 \cdots x_n \in \mathcal{C}, x_1 + \ldots x_{n+1} = 0\right\}$$
Prove that $\mathcal{C}^\star$ is a linear code. Given a parity-check matrix for $\mathcal{C}$, show how to obtain a parity-check matrix for $\mathcal{C}^\star$.

5. (Know). Let $\mathcal{C}$ be an $[n,k,d]_q$ code. We can *puncture* $\mathcal{C}$ by deleting the same coordinate in each codeword. Fix some $i \in [n]$ and let $\mathcal{C}^\star$ be the code $\mathcal{C}$ punctured on the i-th coordinate. Prove:

   (a) $\mathcal{C}^\star$ is a linear code.

   (b) If $d > 1$, $\mathcal{C}^\star$ is an $[n-1,k,d^\star]_q$ code where $d^\star = d-1$ if $\mathcal{C}$ has minimum weight codeword with a nonzero i-th coordinate and $d^\star = d$ otherwise.

6. (Mandatory). Let $\mathcal{C}$ be an $[n,k,d]_2$ code.

   (a) Prove that all the codewords of $\mathcal{C}$ have even weight, or half the codewords have even weight and half of them have odd weight.

   (b) Assume d is odd. Show that there exists a linear code $\mathcal{C}'$ that is a $[n,k-1,d+1]_2$ code.

7. (Know). For integers $k \le n$ we write $B(k,n) = |\{x \in \{0,1\}^n \mid w(x) \le k\}|$. For $p \in [0, \frac{1}{2}]$, prove that $B(pn,n) \le 2^{H(p)\cdot n}$ and $\lim_{n\to\infty} \frac{1}{n} \log B(pn,n) = H(p)$.

   $H(x) = -x\log_2 x - (1-x)\log_2(1-x)$ is the binary entropy function, and $H(0) = H(1) = 0$.

8. (Know). Prove that for every $[n,k,d]_q$ code, $d \le n-k+1$.

9. (Know). For every k, give an explicit construction of a $[2^k, k+1, 2^{k-1}]_2$ code.

10. (Mandatory). Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be cyclic codes over $\mathbb{F}_q$ with generator polynomials $g_1(x)$ and $g_2(x)$ respectively.

    (a) Prove that $\mathcal{C}_1 \cap \mathcal{C}_2$ and $\mathcal{C}_1 + \mathcal{C}_2$ are also cyclic and find their generating polynomial.

    (b) Prove that $\mathcal{C}_1 \subseteq \mathcal{C}_2$ if and only if $g_2(x) \mid g_1(x)$.

11. (Know). A local decoder for a code $\mathcal{C} : \mathcal{C}^k \to \{0,1\}^n$ handling d errors is an algorithm that given an access to a string y such that $\Delta(y, \mathcal{C}(x)) \le d$ for some unknown x and an index $j \in [k]$, runs in time $\mathrm{poly}(\log n)$ and outputs x_j with high probability.

    Show that for every $\delta < \frac{1}{4}$, the Hadamard code has a local decoder handling δ-fraction of errors. How many queries to you need in order to succeed with probability at least $1 - \varepsilon$?

12. (Bonus). Let $\mathcal{C}_{\mathrm{RS}}$ be the $[n, k, n-k+1]_{q=n+1}$ RS code defined by evaluating degree $k-1$ polynomials over $\alpha^0, \dots, \alpha^{n-1}$, where α is a generator of $\mathbb{F}_q^\star$. Suppose that $q = 2^m$ for some integer m and consider $\mathcal{C} = \mathcal{C}_{\mathrm{RS}} \cap \mathbb{F}_2^n$.

    (a) Prove that $\mathcal{C}$ is a binary linear code of distance at least $d = n - k + 1$ and dimension at least $n - (d-1)\log(n+1)$.

    (b) Prove a better bound of $n - \left\lceil \frac{d-1}{2} \right\rceil \log(n+1)$ on the dimension of $\mathcal{C}$ by finding redundant checks amongst the "natural" checks defining $\mathcal{C}$.

    (c) For constant d (and growing n), prove that $\mathcal{C}$ has nearly optimal dimension, in the sense that the dimension cannot be $n - t\log(n+1)$ for $t < \frac{d-1}{2}$.

13. Alice holds $x \in \{0,1\}^n$ and Bob holds $y \in \{0,1\}^n$ and they wish to verify that $x = y$. The goal is to use as few communication bits as possible. We allow Alice and Bob to use private random coins, and allow one-sided error, i.e., if $x = y$ Alice and Bob should always accept, whereas whenever $x \ne y$ Alice and Bob should reject with probability (over their coins) at least $1 - \varepsilon$.

    (a) (Mandatory). Show a protocol with ε one-sided error and $2\log\frac{n}{\varepsilon}$ communication bits.

    (b) (Bonus). A bipartite graph $G = (V_1 = [N], V_2 = [M], E)$ is a (K, α)-disperser if for every $X \subseteq V_1$ of cardinality K, $|\Gamma(X)| > (1-\alpha)M$ (that is, every large enough set in V_1 misses less than an α-fraction of the vertices in V_2).

    Prove that for every $M \le K \le N$ and constant $0 < \alpha < 1$ there exists a (K, α)-disperser $G = (V_1 = [N], V_2 = [M], E)$ with degree $D = O(\log \frac{N}{K})$.

    (c) (Bonus). Show a protocol with ε one-sided error and $\log n + O(\log \frac{1}{\varepsilon})$ communication bits.

14. (Mandatory). Write explicitly a parity check matrix for the Hamming $[7,4,3]_2$ code that gives a cyclic code.

15. (Know). Prove Lemma 10 and 11 of Lecture 1.

HW 2 – Decoding and List Decoding.

Out: 26.11.2017
Due: 10.12.2017

1. (Know). Show that the ring of multivariate polynomials is not a Euclidean ring (i.e., it does not satisfy the division-with-remainder property) but it is a unique factorization domain (i.e., every non-zero non-unit element can be written as a product of irreducible polynomials, uniquely up to order scalar multiplication).

2. (Know). Let $\mathcal{C}$ be an $[n, k, d = n - k + 1]_q$ RS code.

   (a) Give a polynomial time algorithm that decodes $\mathcal{C}$ from up to $d - 1$ *erasures*.[1]

   (b) Use the Berlekamp-Welch algorithm to give a polynomial time algorithm that decodes $\mathcal{C}$ from e errors and s erasures as long as $d > 2e + s$.

3. (Mandatory). To uniquely decode concatenated code, we first considered uniquely decoding the inner code and got bad parameters. In the GMD algorithm, we applied "soft decoding" on the inner code and reached half the distance. Instead, we can consider *list-decoding* the inner code (say even by brute force).

   Analyze this approach on the code that is obtained by concatenating Reed-Solomon with Hadamard. What is the decoding radius that you obtain?

4. (Mandatory). Prove the correctness of the following algorithm for unique decoding RS codes.

   - Input: q, n, k where $k < n \leq q$,$e < \frac{n-k+1}{2}$. $\{(\alpha_i, y_i)\}_{i=1}^{n}$, where α_i are distinct elements of $\mathbb{F} = \mathbb{F}_q$ and $y_i \in \mathbb{F}$.
   - Output: A polynomial $p \in \mathbb{F}[x]$ of degree at most $k - 1$ with distance at most e from the given word.
   - Algorithm: Find a polynomial $Q : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$ such that
     - $\deg_Y(Q) = 1$, i.e., $Q(x, y) = A(x) + yB(x)$.
     - $\deg(A) \leq k + e - 1$, $\deg(B) \leq e$,
     - For every $i \in [n]$, $Q(\alpha_i, y_i) = 0$.

     You have to prove that such a Q exists and can be efficiently found.

     Then, find $g \in \mathbb{F}[x]$ such that $y - g | Q$ and output it. You have to show that such a g exists, has small degree and is correct.

5. (Mandatory). In this question we give a decoding algorithm for *binary* $\mathcal{C} = [n = 2^m, k = \binom{m}{r}, 2^{m-r}]_2$ RM codes (of evaluations over $\mathbb{F}_2^m$ of all degree r multi-linear polynomials) from errors up to half the minimum distance. The running time of the algoritm will be polynomial in n, and therefore is allowed to be exponential in m.

   For a subset $S \subseteq [m]$, define the polynomial $\chi_S : \mathbb{F}_2^m \to \mathbb{F}_2$ defined by $\chi_S(x) = \prod_{i \in S} x_i$. Any degree r polynomial P over $\mathbb{F}_2^m$ can be uniquely expressed as $P(x) = \sum_{S \subseteq [m], |S| \leq r} c_S \cdot \chi_S(x)$ with $c_S \in \mathbb{F}_2$ (Why?).

[1] Recall that an erasure is marked with a special symbol '?', so we *know* where an error has occurred, unlike the adversarial error model.

We are given some word $g \in [n] \to \mathbb{F}_2$ and we think of it as $g : \mathbb{F}_2^m \to \mathbb{F}_2$. We are promised that g has distance $e < 2^{m-r}/2$ from a correct codeword $p = \sum_S p_S \chi_S$.

(a) Prove that for all $b \in \mathbb{F}_2^{m-r}$ and a degree r polynomial $P(x) = \sum_{S \subseteq [m], |S| \le r} c_S \cdot \chi_S(x)$, for any $S \subseteq [m]$, $|S| = r$,

$$\sum_{a \in \mathbb{F}_2^m, a_{\bar{S}} = b} P(a) = c_S.$$

(b) For a given $S \subset [m]$, $|S| = r$, the algorithm goes over all $b \in \mathbb{F}_2^{m-r}$. For each such b it computes $z_b = \sum_{a \in F_2^m, a_{\bar{S}} = b} g(a)$. It outputs $Maj_b(z_b)$. Prove that the algorithm outputs p_s.

(c) Give a polynomial (in n) time decoding algorithm for $\mathcal{C}$ from any number of errors smaller than $\frac{1}{2} \cdot 2^{m-r}$.

6. (Mandatory). Let $\mathcal{C}$ be the $[n = 2^m, m, \frac{1}{2}]_2$ Hadamard code. Let $w \in \mathbb{F}_2^n$ be an arbitrary word. Prove that the number of codewords in radius $\frac{1-\varepsilon}{2}$ around w is $O(\frac{1}{\varepsilon^2})$, and in particular is independent of n.

7. (Mandatory).

(a) Prove that there exists a mapping $\phi : \mathbb{F}_q \to \mathbb{R}^q$ such that:

- For every $a \in \mathbb{F}_q$, $\|\phi(a)\| = 1$.
- For every $a \ne b \in \mathbb{F}_q$, $\langle \phi(a), \phi(b) \rangle = -\frac{1}{q-1}$.

(b) Let C be a q-ary code of length n and minimum distance at least d. Prove that if $\frac{d}{n} > 1 - \frac{1}{q}$ then $|C| \le \frac{qd}{qd-(q-1)n}$.

8. (Bonus). Let $1 \le k \le n$ be integers and let $p_1 < \ldots < p_n$ be n distinct primes. Denote $K = \prod_{i=1}^k p_i$ and $N = \prod_{i=1}^n p_i$. Consider the mapping $E : \mathbb{Z}_K \to \mathbb{Z}_{p_1} \times \ldots \times \mathbb{Z}_{p_n}$ defined by:

$$E(m) \;=\; (m \bmod p_1, \ldots, m \bmod p_n).$$

(a) Suppose that $m_1 \ne m_2$. For $i \in [n]$, define the indicator b_i such that $b_i = 1$ iff $E(m_1)_i \ne E(m_2)_i$. Prove that $\prod_{i=1}^n p_i^{b_i} > N/K$.
Deduce that when $m_1 \ne m_2$, $\Delta(E(m_1), E(m_2)) \ge n - k + 1$.

(b) We will now adopt the Welch-Berlekamp algorithm to handle E. Suppose $r = (r_1, \ldots, r_n)$ is the received word, where $r_i \in \mathbb{Z}_{p_i}$.

i. Prove there can be at most one $m \in \mathbb{Z}_K$ such that

$$\prod_{i : E(m)_i \ne r_i} p_i \le \sqrt{N/K}. \qquad (1)$$

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

ii. Assuming such an m exists, prove that there exist integers y, z with $0 \le y < \sqrt{NK}$ and $1 \le z \le \sqrt{N/K}$ such that $y \equiv rz \pmod{N}$.

iii. Prove that if y, z are any integers satisfying the above conditions, then in fact $m = y/z$. Note that a pair of integers (y, z) satisfying the above can be found by integer linear programming in a fixed number of dimensions in polynomial time.

(c) Instead of condition (1), what if we want to decode under the more natural condition: $|\{i \mid E(m)_i \neq r_i\}| \leq \frac{n-k}{2}$? Show how this can be done by calling the above decoder many times and erasing the last i symbols for each choice of $i \in [n]$.

HW 3 – List decoding, What can and cannot be done I.

Out: 11.12.2017
Due: 31.12.2017

1. (Mandatory). Generalize the Guruswami-Sudan list-decoding algorithm to RM codes (the one with the multiplicities).

   For rate R RS codes, the Guruswami-Sudan decodes from relative distance $\delta \leq 1 - \sqrt{R}$. Show that your generalized algorithm achieves $\delta \leq 1 - \sqrt[m+1]{\frac{r}{q}}$ where m is the number of variables, r is the total degree and q is the field size. Note that when $m = 1$ this is exactly the RS bound.

2. (Mandatory). In this question you will generalize the Guruswami-Sudan list-decoding algorithm to handle *list recovery.*

   Design and analyze an algorithm that gets as input q, n, k, L and n pairs $\{(\alpha_i, L_i)\}_{i=1}^{n}$ such that $L_i \subseteq \mathbb{F}_q$ and $\frac{1}{n}\sum_{i=1}^{n} |L_i| \leq L$. It should be efficient, and output a list of polynomials $p(x)$ of degree less than k such that for more than $agr = \sqrt{nLk}$ pairs j, $p(\alpha_j) \in L_j$.

3. (Mandatory). Let $P : \mathbb{F}_q^n \to \mathbb{F}_q^m$ and $Q : \mathbb{F}_q^\ell \to \mathbb{F}_q^n$. Prove that for every $a \in \mathbb{F}_q^\ell$,

$$mult(P \circ Q, a) \;\geq\; mult(P, Q(a)) \cdot mult(Q - Q(a), a).$$

4. (Mandatory). In this question you will prove another bound on the length of a linear code of specific dimension and minimum distance, which when $d \gg q$ strengthens the Singleton bound.

   (a) Let $\mathcal{C}$ be an $[n, k, d]_q$ code. Let $c \in \mathcal{C}$ and consider the code $\mathcal{C}'(c)$ that is obtained by puncturing the codewords of $\mathcal{C}$ at the nonzero indices of c.

   - Prove that if $c \in \mathcal{C}$ has weight $w < \frac{dq}{q-1}$, the code $\mathcal{C}'(c)$ has length $n - w$ and dimension $k - 1$.
   - Prove that if $c \in \mathcal{C}$ has weight $w < \frac{dq}{q-1}$, the code $\mathcal{C}'(c)$ has distance at least $d - w + \lceil \frac{w}{q} \rceil$.

   (b) Prove that $n \geq \sum_{i=0}^{k-1} \lceil \frac{d}{q^i} \rceil$.

5. (Mandatory). Consider the q-ary erasure channel with erasure probability α: the input to this channel is a field element $x \in \mathbb{F}_q$, and the output is x with probability $1 - \alpha$, and an erasure '?' with probability α. For a linear code C generated by an $n \times k$ matrix G over $\mathbb{F}_q$, let $D : (\mathbb{F}_q \cup \{?\})^n \to C \cup \{\text{fail}\}$ be the following decoder:

$$D(y) \;=\; \begin{cases} c & \text{if } y \text{ agrees with exactly one } c \in C \text{ on the unerased entries in } \mathbb{F}_q \\ \text{fail} & \text{otherwise} \end{cases}$$

   (a) For a set $J \subseteq [n]$, let $p_G(J)$ be the probability (over the channel noise and choice of a random message) that D outputs "fail" conditioned on the erasures being indexed by J. Prove that $\mathbb{E}_{G \in \mathbb{F}_q^{n \times k}}[p_G(J)] \leq q^{k-n+|J|}$.

   (b) Let p_G be the be the probability that D outputs "fail". Show that when $k = Rn$ for $R < 1 - \alpha$, $\mathbb{E}_{G \in \mathbb{F}_q^{n \times k}}[p_G]$ is exponentially small in n.

(c) Conclude that one can reliably communicate on the q-ary erasure channel with erasure probability α at any rate less than $1-\alpha$ using a linear code.

6. (Bonus). Let $p < 1/2$. Prove that there exists a constant $c = c(p)$ such that for every $\varepsilon > 0$ and large enough n, if $C \subseteq \mathbb{F}_2^n$ is a random *linear* code of rate at least $1 - H(p) - \varepsilon$ then it is $(p, \frac{c}{\varepsilon})$ list-decodable with probability $1 - 2^{-\Omega(n)}$.

   Use the following theorem: Let $p < 1/2$. There exists a constant $c = c(p)$ such that for every n and $\ell = o(\sqrt{n})$, if $X_1, \ldots, X_\ell$ are drawn uniformly and independently from $B(0, pn)$, then

$$\Pr[|\operatorname{Span}(\{X_1, \ldots, X_\ell\}) \cap B(0, pn)| \geq c\ell] \;\leq\; 2^{-(6-o(1))n}.$$

HW 4 – List decoding, What can and cannot be done II.

Out: 31.12.2017
Due: 14.1.2018

1. (Know). Give an explicit representation of a Fourier basis (i.e., an orthonormal basis of homomorphisms) for $V = \{f : G \to \mathbb{C}\}$, where:

   (a) $G = \mathbb{Z}_p$ for a prime p.

   (b) $G = \mathbb{Z}_2^n$.

   Prove correctness.

2. (Mandatory). Let $\mathcal{C}$ be an $[n, k = rn, d = \delta n]_q$ code. Prove that $r \leq 1 - \frac{q}{q-1}\delta + o(1)$.

3. (Mandatory). In class we saw that there exists a linear code $\mathcal{C}$ of length n and distance d with $|\mathcal{C}| \geq \frac{2^n}{|B(d-1,n)|}$ and we want to (very slightly) improve upon this.

   Let $v_1, \ldots, v_r \subseteq \mathbb{F}_2^t$ such that no $d-1$ of them are linearly dependent. Prove that if $|B(d-2, r)| < 2^t$ then there exists $w \in \mathbb{F}_2^t$ such that in $v_1, \ldots, v_r, w$ no $d-1$ vectors are linearly dependent.

   Use the above to prove that there exists a linear code C of length n and distance d with $|\mathcal{C}| \geq \frac{2^n}{|B(d-2,n)|}$.

4. (Mandatory). We say $H \subseteq \mathbb{F}_q^n$ is a *hitting set* for n-variate polynomials of total degree at most d over $\mathbb{F}_q$ if for every nonzero such polynomial f there exists $x \in H$ such that $f(x) \neq 0$.

   Prove that if H is a hitting set for the above family of polynomials then $|H| \geq \binom{n+d}{n}$.

5. (Mandatory). Let H be a group and $S \subseteq H$ be a generating set. The Cayley graph $C(H, S)$ is defined as follows: The vertices are labeled with elements of H, and (a, b) is an edge iff $a = bs^{-1}$ for some $s \in S$.

   (a) What is $G = C(\mathbb{Z}_2^n, \{e_1, \ldots, e_n\})$, where e_i has 1 in the i-th coordinate and 0 elsewhere? Calculate the eigenvalues and eigenvectors of the normalized adjacency matrix of G.

   (b) Let C be an $[\bar{n}, n, \frac{1-\varepsilon}{2}\bar{n}]_2$ code where the weight of every codeword is at most $\frac{1+\varepsilon}{2}$. Let $S \subseteq \mathbb{Z}_2^n$ be the rows of a generating matrix for C and denote $G = C(\mathbb{Z}_2^n, S)$.

   - What are the eigenvectors of A_G, the normalized adjacency matrix of G?
   - Let $\lambda_{2^n} \leq \ldots \leq \lambda_2 \leq \lambda_1$ be the eigenvalues of A_G. Prove that $\max\{\lambda_2, -\lambda_{2^n}\} \leq \varepsilon$.