Queen Mary, University of London

B. Sc. Examination by course unit 2007

MAS309 Coding Theory

16th May, 2007

14:30

Duration: 2 hours.

You may attempt as many questions as you wish and all questions carry equal marks. Except for the award of a bare pass, only the best FOUR questions will be counted.
Electronic calculators are not permitted.

Question 1. (a) Define the constants $A_3(n, d)$. [3]

(b) Prove that $A_3(n, d) \leq A_3(n-1, d-1)$, for all $n \geq d \geq 2$. [6]

(c) Consider the ternary "parity check code"

$$\mathcal{C} = \left\{ v_1 v_2 v_3 v_4 \in \mathbb{F}_3^4 : v_1 + v_2 + v_3 + v_4 = 0 \right\}$$

of length 4. Demonstrate that $\mathcal{C}$ has 27 codewords and minimum distance 2. [7]

(d) Write down the Hamming ("sphere packing") bound as it applies to ternary, 1-error-correcting codes. (This is the special case $t = 1$ and $q = 3$.) Hence show that a ternary 1-error-correcting code of length 5 has at most 22 codewords. [4]

(e) Prove that if a code has minimum distance 3, then it is 1-error-correcting. [3]

(f) Hence deduce that $A_3(5, 3) < A_3(4, 2)$. [2]

Question 2. (a) What is meant by a *binary* (n, M, d)*-code*? [3]

(b) Suppose $\mathcal{C}$ is a binary (n, M, d)-code. Regard the codewords as vectors over $\mathbb{F}_2$, and define a $\binom{M}{2} \times n$ matrix D as follows: The rows of D correspond to all (unordered) pair of codewords in $\mathcal{C}$. The row corresponding to codewords u and v is simply the vector sum of u and v. (The ordering of the rows of D is not significant.) Write down the array D for the particular code

$$\mathcal{C} = \{\mathtt{000000}, \mathtt{001111}, \mathtt{111001}, \mathtt{110110}\}.$$

[3]

(c) Now suppose $\mathcal{C}$ is an arbitrary (n, M, d)-code. Prove that the number of $\mathtt{1}$s in D is at least $\binom{M}{2} d$. (Hint: consider D row-wise.) [5]

(d) Prove that the number of 1s in D is at most $nM^2/4$. (Hint: consider D columnwise.) [5]

(e) Deduce that $M \leq 2d/(2d - n)$, provided $2d > n$. [3]

(f) State, without proof, a bound relating $A_2(n, d)$ and $A_2(n-1, d)$. [3]

(g) Deduce that $A_2(2d, d) \leq 4d$, for all $d \geq 1$. [3]

Question 3. (a) Suppose $\mathcal{C}$ is a code of length n over the alphabet A. What is meant by a *decoding process* for $\mathcal{C}$? What does it mean for a decoding process to be *nearest-neighbour*? [4]

(b) For this part of the question, let $\mathcal{C}$ be the binary code

$$\{\mathtt{000}, \mathtt{001}, \mathtt{110}, \mathtt{111}\}.$$

i. Construct a nearest-neighbour decoding process for $\mathcal{C}$ with the following additional property of "balance": Let $N(v)$ the number of input words that result in codeword $v \in \mathcal{C}$ being output. Then the decoding process is *balanced* if $N(v)$ is independent of $v \in \mathcal{C}$. [5]

ii. Suppose the word $\mathtt{000}$ is transmitted through a noisy channel with error probability $p = \frac{1}{4}$. What is the probability that $\mathtt{000}$ is correctly decoded, assuming your balanced decoding process for $\mathcal{C}$ is used? [5]

(c) Let q denote the size of A, and suppose $x \in A^n$ is any word. Define the (Hamming) sphere $S(x,t)$ of radius t with centre x. Write down and briefly justify a formula for V, the number of words contained in $S(x,t)$. [5]

(d) Recall that a t-error-correcting code $\mathcal{C}$ is said to be *perfect* if $MV = q^n$, where $M = |\mathcal{C}|$. Prove that the nearest-neighbour decoding process for a perfect code $\mathcal{C}$ is unique. [6]

Question 4. (a) Define the notions of $[n,k]$- and $[n,k,d]$-*code* over a field $\mathbb{F}_q$. [4]

(b) Suppose $\mathcal{C}$ is an $[n,k]$-code over $\mathbb{F}_q$. Explain what it means for a matrix G over $\mathbb{F}_q$ to be a *generator matrix* for $\mathcal{C}$. What are the dimensions of G? [3]

(c) Consider the following column operations on a generator matrix:

- Add a multiple of one column to another.
- Let $f : \mathbb{F}_q \to \mathbb{F}_q$ be an arbitrary permutation of the field elements. Apply f to all the elements in some column of G.

Suppose G' is a matrix that results from applying one or other operation to G. The matrices G and G' do not in general generate equivalent codes. Illustrate this fact by presenting two counterexamples (one for each operation) based on the generator matrix

$$\begin{pmatrix} \mathtt{101} \\ \mathtt{011} \end{pmatrix}$$

of the binary parity-check code of length 3. [7]

(Hint: Use an invariant of equivalent codes, such as minimum distance.)

(d) Explain how to restrict the permutation f in the second operation of part (c) so that only equivalent codes are produced. [2]

(e) Prove that there is a unique ternary $[4,3,2]$-code (over $\mathbb{F}_3$), up to equivalence. Provide a simple description of this unique code. [9]

(Hint. Start with a generator matrix in normal form.)

Question 5. (a) Suppose $\mathcal{C}$ is an $[n, k]$-code over $\mathbb{F}_q$, with generator matrix G. Define the *dual code* $\mathcal{C}^\perp$ of $\mathcal{C}$. Prove that $\mathcal{C}^\perp$ is an $[n, n-k]$-code. [6]

(Note: you may use standard results from linear algebra, provided they are correctly stated.)

(b) Write down conditions involving matrices G and H that express the situation that H is a *parity-check matrix* for $\mathcal{C}$. What is the relationship between H and the dual code $\mathcal{C}^\perp$? [3]

(c) Explain what it means for a generator matrix G to be in *standard form*. Given a matrix G in standard form, show how to write down a parity-check matrix H for the code $\mathcal{C}$ generated by G. [4]

(d) Write down a parity-check matrix H for the binary $[5, 2]$-code $\mathcal{C}$ with generator matrix

$$G = \begin{pmatrix} \texttt{10101} \\ \texttt{01011} \end{pmatrix}.$$

(Note that G is in standard form.) [3]

(e) Construct a syndrome look-up table for H. Explain how the syndrome look-up table determines a decoding process for $\mathcal{C}$, and illustrate this process by decoding the word `11101`. [9]

Question 6. (a) Define the *redundancy* of a linear code. [2]

(b) Describe the construction of the Hamming code $\mathrm{Ham}(r, q)$, of redundancy r over the alphabet $\mathbb{F}_q$. [7]

(c) Illustrate your answer by writing down the parity-check matrix H for $\mathrm{Ham}(2, 5)$, and deriving the associated generator matrix. [5]

(d) State, without proof, the minimum distance of the code $\mathrm{Ham}(r, q)$. [1]

(e) Prove that $\mathrm{Ham}(r, q)$ is an $[n, n-r]$-code, where $n = (q^r - 1)/(q - 1)$. [5]

(f) Show that there is no Hamming code whose codewords have length 16. [5]

Solutions

Question 1. (a) A (n, M, d)-code $\mathcal{C}$ is one with M codewords, all of length n, such that $d(u, v) \geq d$ for all distinct $u, v \in \mathcal{C}$. Then

$$A_3(n, d) = \max\{M : \text{there exists a } (n, M, d)\text{-code over } \mathbb{F}_3\}.$$

(b) [Bookwork.] Let $\mathcal{C}$ be a (n, M, d)-code with $M = A_3(n, d)$. For any $v = v_1 \dots v_n \in \mathcal{C}$, let $\hat{v} = v_1 \dots v_{n-1}$. Consider $\mathcal{C} = \{\hat{v} : v \in \mathcal{C}\}$. Since $d(\hat{u}, \hat{v}) \geq d(u, v) - 1$, $\mathcal{C}'$ is an $(n-1, M, d-1)$-code. (Since $d \geq 2$ the construction preserves the number of codewords.) Thus $A_3(n-1, d-1) \geq M = A_3(n, d)$.

(c) [Bookwork, at least for $q = 2$, adapted to $q = 3$.] Choose $v_1 v_2 v_3 \in \mathbb{F}_3^3$ freely: 3^3 choices. Now v_4 is forced by the equation $v_4 = -(v_1 + v_2 + v_3)$. So $\mathcal{C}$ has 27 words. Suppose $\mathcal{C}$ had two words u, v at Hamming distance 1. Then u, v agree in three positions, say $1, 2, 3$, and so $u_1 = v_1, \dots, u_3 = v_3$. But now $u_4 = v_4$, and $u = v$, a contradiction. So $\mathcal{C}$ has minimum distance 2.

(d) [Routine application of bookwork.] A 1-error correcting ternary code of length n has at most $\lfloor 3^n/(1 + 2n) \rfloor$ codewords. (This is the total number of words of length n divided by the number of words in any "sphere" of radius 1.) When $n = 5$, this bound evaluates to $\lfloor 3^5/(1 + 2 \times 5) \rfloor = \lfloor 243/11 \rfloor = 22$.

(e) [Bookwork, adapted to $t = 1$.] Let $x \in \mathbb{F}_q^n$ be a word at distance at most 1 from some codeword v. The word x cannot be at distance 1 from any other codeword v' otherwise, by the triangle inequality, $d(v, v') \leq 2$.

(f) [Routine synthesis.] A code with minimum distance 3 is 1-error-correcting. So, from part (d), $A_3(5, 3) \leq 22$. On the other hand, from part (c), $A_3(4, 2) \geq 27$.

Question 2. (a) A binary (n, M, d)-code $\mathcal{C}$ is a code over the alphabet $\{0, 1\}$, having M codewords, all of length n, such that $d(u, v) \geq d$ for all distinct $u, v \in \mathcal{C}$.

(b) [Parts (b)-(e) lead the student through a proof of a simplified version of the Plotkin bound. The more precise version of the bound was proved in the course using a similar approach.]

$$D = \begin{pmatrix} \mathtt{001111} \\ \mathtt{111001} \\ \mathtt{110110} \\ \mathtt{110110} \\ \mathtt{111001} \\ \mathtt{001111} \end{pmatrix}$$

(c) The row corresponding to codewords u and v is $u+v$ (with addition in $\mathbb{F}_2$). Now $\mathrm{weight}(u+v) = d(u, v) \geq d$. There are $\binom{M}{2}$ rows, so the total number of $\mathtt{1}s$ in D is at least $\binom{M}{2} d$.

(d) Consider any column of D, say column 1. There is a $\mathtt{1}$ in position 1 of the row containing $u + v$ iff $u_1 \neq v_1$, i.e., u and v differ in position 1. Suppose j codewords start with $\mathtt{1}$, so that $M - j$ start with $\mathtt{0}$. The number of $\mathtt{1}$s in column 1 of D is then $j(M - j) \leq \frac{1}{4}M^2$. (Maximise a quadratic.) There are n columns, so the total number of $\mathtt{1}$s in D is at most $nM^2/4$.

(e) From (c) and (d), $\binom{M}{2}d \le nM^2/4$. Thus, $2(M-1)d \le nM$ and $M \le 2d/(2d-n)$.

(f) [Bookwork, specialised to $q = 2$.] $A_2(n,d) \le 2A_2(n-1,d)$.

(g) [Unseen.] $A_2(2d,d) \le 2A_2(2d-1,d) \le 4d/(2d-(2d-1)) = 4d$, where the inequalities are from parts (f) and (e), resp.

Question 3. (a) A decoding process is a function $f : A^n \to \mathcal{C}$. It is nearest neighbour if $d(w, f(w)) \le d(w,v)$ for all $w \in A^n$ and $v \in \mathcal{C}$.

(b) [The "balance" condition is an novelty element.] E.g.,

$$\begin{aligned} f(\mathtt{000}) &= \mathtt{000} \\ f(\mathtt{001}) &= \mathtt{001} \\ f(\mathtt{010}) &= \mathtt{000} \\ f(\mathtt{011}) &= \mathtt{111} \\ f(\mathtt{100}) &= \mathtt{110} \\ f(\mathtt{101}) &= \mathtt{001} \\ f(\mathtt{110}) &= \mathtt{110} \\ f(\mathtt{111}) &= \mathtt{111}. \end{aligned}$$

(There are three other possibilities.)

(c) [Similar to calculations in coursework.] Let w be the (possibly) corrupted word leaving the channel. For $\mathtt{000}$ to be correctly decoded, either $w = \mathtt{000}$ or $w = \mathtt{010}$. Now $\Pr(w = \mathtt{000}) = (1-p)^3 = \frac{27}{64}$ and $\Pr(w = \mathtt{010}) = p(1-p)^2 = \frac{9}{64}$. So the probability of correct decoding is $\frac{27}{64} + \frac{9}{64} = \frac{36}{64} = \frac{9}{16}$.

(d) [Bookwork] $S(x,t) = \{v \in A^n : d(v,x) \le t\}$. The number of words at distance exactly j from x is $\binom{n}{j}(q-1)^j$. (Choose j positions to be changed; for each position there are $q-1$ choices for the new symbol and all choices are independent.) Thus

$$V = \binom{n}{0} + \binom{n}{1}(q-1) + \cdots + \binom{n}{t}(q-1)^t.$$

(e) [Unseen, but touched on in the course.] Consider the set $\{S(v,t) : v \in \mathcal{C}\}$ of all spheres centered at codewords of $\mathcal{C}$. All spheres are disjoint, otherwise there would exist words that are not uniquely decodable. On the other hand, the spheres must cover all words in A^n since $MV = q^n$. Thus the spheres partition the space of all words. The only possible choice for $f(w)$ in a nearest neighbour decoding process is the centre of the sphere containing w.

Question 4. (a) An $[n,k]$-code over $\mathbb{F}_q$ is a vector subspace of $\mathbb{F}_q^n$ of dimension k. An $[n,k,d]$-code in addition has minimum distance d: no two codewords are closer than d in Hamming distance.

(b) G is a generator matrix for $\mathcal{C}$ if the rows of G form a basis for $\mathcal{C}$. Thus G has k rows and n columns.

(c) [That these operations fail to preserve equivalence was discussed in the course. The counterexample is different.] First add column 2 to column 3 to obtain

$$G' = \begin{pmatrix} 101 \\ 010 \end{pmatrix}.$$

Clearly, G' is the generator matrix of a code $\mathcal{C}'$ containing the word 010 of weight 1. But minimum distance is equal to the weight of a minimum weight non-zero codeword, which for $\mathcal{C}'$ is 1. So $\mathcal{C}'$ cannot be equivalent to $\mathcal{C}$, which has minimum weight 2.
Now apply the transposition $\binom{01}{10}$ to column 2 to obtain

$$G'' = \begin{pmatrix} 111 \\ 001 \end{pmatrix}.$$

G'' is the generator matrix of a code $\mathcal{C}''$ of minimum distance 1, which cannot be equivalent to $\mathcal{C}$.

(d) [Bookwork.] Restrict permutations of $\mathbb{F}_q$ to ones of the form $\sigma(x) = ax$ for some $a \in \mathbb{F}_q \setminus \{0\}$.

(e) [Unseen, but similar to examples from the lectures or exercises.] We know that any $[4, 3, 2]$-code $\mathcal{C}$ is equivalent to one in standard form:

$$\begin{pmatrix} 100a \\ 010b \\ 001c \end{pmatrix},$$

with $a, b, c \in \mathbb{F}_3$. Since $\mathcal{C}$ has minimum distance 2, $a, b, c \neq 0$. If $a = 1$ then multiply row 1 by 2 and column 1 by 2. Repeat for b and c. The generator matrix is now

$$G = \begin{pmatrix} 1002 \\ 0102 \\ 0012 \end{pmatrix},$$

which is the generator matrix for the "parity-check" code over $\mathbb{F}_3$, which has minimum distance 2. So any $[4, 3, 2]$-code is equivalent to one with generator matrix G.

Question 5. (a) [Bookwork. As in the course notes, codewords are row vectors.] $\mathcal{C}^\perp = \{w : Gw^{\mathrm{T}} = 0\}$. Define $\alpha : \mathbb{F}_q^n \to \mathbb{F}_q^k$ by $\alpha(w) = Gw^{\mathrm{T}}$. Then $\mathcal{C}^\perp = \ker \alpha$. By the Rank-nullity Theorem, $\dim \ker \alpha = n - \dim \operatorname{Im} \alpha = n - k$, since G is of rank k.

(b) [Bookwork.] The conditions are: $GH^{\mathrm{T}} = 0$ and H has full rank $(n - k)$. H is a generator matrix for $\mathcal{C}^\perp$.

(c) [Bookwork.] G is in standard form if $G = [I_k \mid A]$, where I_k is the $k \times k$ identity matrix, and A an unrestricted $k \times (n - k)$ matrix. If G has this form then the parity-check matrix is $H = [-A^{\mathrm{T}} \mid I_{n-k}]$.

(d) [Routine application.]

$$H = \begin{pmatrix} 10100 \\ 01010 \\ 11001 \end{pmatrix}$$

(e) [Application of bookwork; similar to examples from class/notes/exercises.] Syndrome decoding table:

$$\begin{aligned}
\mathtt{000} &\to \mathtt{00000}\\
\mathtt{001} &\to \mathtt{00001}\\
\mathtt{010} &\to \mathtt{00010}\\
\mathtt{011} &\to \mathtt{01000}\\
\mathtt{100} &\to \mathtt{00100}\\
\mathtt{101} &\to \mathtt{10000}\\
\mathtt{110} &\to \mathtt{00110}\\
\mathtt{111} &\to \mathtt{01100}
\end{aligned}$$

Given a received word w, compute the syndrome Hw^{T}. Look up the syndrome in the table to find a coset leader u. The decoded codeword is then $w - u$.

If $w = \mathtt{11101}$ is received then the syndrome is $\mathtt{011}$ and the coset leader $u = \mathtt{01000}$. Then the decoded codeword is $\mathtt{11101} - \mathtt{01000} = \mathtt{10101}$.

Question 6. (a) Redundancy $r = n - k$.

(b) [Bookwork.] From each 1-dimensional linear subspace of $\mathbb{F}_q^r$ select one non-zero vector. Suppose there are n such. Form a $r \times n$ matrix H whose columns are the n vectors just selected (in any order). The matrix H is the parity-check matrix of the code $\mathrm{Ham}(r, q)$.

(c) [Routine application of the above.] For $\mathrm{Ham}(2, 5)$ the parity-check matrix (in standard form) is

$$\begin{pmatrix} \mathtt{101111} \\ \mathtt{011234} \end{pmatrix}.$$

The associated generator matrix is

$$\begin{pmatrix} \mathtt{441000} \\ \mathtt{430100} \\ \mathtt{420010} \\ \mathtt{410001} \end{pmatrix}.$$

(d) [Bookwork.] The minimum distance is 3.

(e) [Bookwork.] Each non-zero word in $\mathbb{F}_q^r$ finds itself in one linear subspace together with $q - 1$ other non-zero words. There are $q^r - 1$ non-zero words in all, so $(q^r - 1)/(q - 1)$ different linear subspaces. This is also the number of columns, n, of the parity-check matrix. Since the parity-check matrix has r rows, the generator matrix must have $n - r$.

(f) [Unseen.] From the previous part $n = (q^r - 1)/(q - 1)$ with q a prime power.

- Try $r = 2$: $n = 16 = q + 1$, and $q = 15$ is not a prime power.
- Try $r = 3$: $n = 16 = q^2 + q + 1$, and $3 < q < 4$ is not integer.
- Try $r = 4$: $n = 16 = q^3 + q^2 + q + 1$, and $2 < q < 3$ is not integer.
- For $r > 4$, $q < 2$.

So there is no solution for q and r in integers, with q a prime power.