This question paper consists of 6 printed pages, each of which is identified by the reference MATH315301.

All calculators must carry an approval sticker issued by the School of Mathematics.

School of Mathematics

January 2019

MATH315301

Coding Theory

Time Allowed: 2.5 hours

You must attempt to answer 4 questions.
If you answer more than 4 questions, only your best 4 answers will be counted towards your final mark for this exam.
All questions carry equal marks.

1. Let F be a field. We write $\mathcal{M}_{n,m}(F)$ for the set of $n \times m$ matrices with entries in field F. For example

$$\begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \in \mathcal{M}_{3,5}(\mathbb{Z}_2)$$

Associated to each $M \in \mathcal{M}_{n,m}(F)$ is the row space $R(M)$ of M. This is the vector space over F spanned by the rows of M (regarded as vectors).

(a) What does it mean to say that a set of vectors such as the row vectors of matrix $M \in \mathcal{M}_{n,m}(F)$ above are linearly independent?

(b) If F is a field of order q, what is the size of $\mathcal{M}_{n,m}(F)$?

(c) Under what conditions is $M \in \mathcal{M}_{n,m}(F)$ a generator matrix for a linear code; and what are the block length and dimension of the code it then generates?

(d) The function $\pi_n : F^n \to F^{n-1}$ is defined by $(x_1, x_2, ..., x_{n-1}, x_n) \mapsto (x_1, x_2, ..., x_{n-1})$. We define the action of this function on a code $C \subseteq F^n$ by restriction of the action on F^n. Similarly we may define an action of π_n on $\mathcal{M}_{m,n}(F)$, specifically $\pi_n : \mathcal{M}_{m,n}(F) \to \mathcal{M}_{m,n-1}(F)$, by deleting the final column of the matrix. Construct an example where M is a generator matrix and $\pi_n(M)$ is not.

(e) Write down the binary linear code C_1 with generator matrix

$$G_1 = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{pmatrix}$$

(f) Consider the following sets: $S_1 = \{000000, 000100, 200000\} \subset \mathbb{Z}_3^6$;

$$S_2 = \{0000, 0001, 1000, 1001\} \subset \mathbb{Z}_2^4;$$

$$S_3 = \{00000, 01110, 01000, 00110\} \subset \mathbb{Z}_2^5;$$

$$S_4 = \{0000, 0001, 1000, 1002\} \subset \mathbb{Z}_5^4.$$

Determine which of these are linear codes (giving the reasons for your answers).

(g) Define the minimum weight $w(C)$ for a code C over a symbol set that is a field. Prove that, for a linear code, the minimum distance $d(C)$ is equal to $w(C)$.

(h) Define $C^{\perp}$, the dual code to a linear code C. Prove that $C^{\perp}$ is also a linear code.

(i) A linear code is self-dual if $C^{\perp} = C$. Compute the dual of C_1 in part (e) above, and hence or otherwise determine if it is self-dual.

2. (a) Explain why the following two matrices generate equivalent linear codes in $\mathbb{Z}_7^5$

$$A = \begin{pmatrix} 1 & 1 & 0 & 6 & 4 \\ 1 & 0 & 0 & 2 & 1 \\ 0 & 0 & 1 & 6 & 5 \end{pmatrix}, \qquad G'' = \begin{pmatrix} 1 & 1 & 0 & 4 & 6 \\ 1 & 0 & 0 & 1 & 2 \\ 0 & 0 & 1 & 5 & 6 \end{pmatrix}$$

(b) Explain why the following two matrices generate the same linear code $C \subset \mathbb{Z}_7^5$

$$G'' = \begin{pmatrix} 1 & 1 & 0 & 4 & 6 \\ 1 & 0 & 0 & 1 & 2 \\ 0 & 0 & 1 & 5 & 6 \end{pmatrix}, \qquad G' = \begin{pmatrix} 0 & 1 & 0 & 3 & 4 \\ 1 & 0 & 0 & 1 & 2 \\ 0 & 0 & 1 & 5 & 6 \end{pmatrix}$$

(c) Let $C \subset \mathbb{Z}_7^5$ be the linear code with generator matrix

$$G' = \begin{pmatrix} 0 & 1 & 0 & 3 & 4 \\ 1 & 0 & 0 & 1 & 2 \\ 0 & 0 & 1 & 5 & 6 \end{pmatrix}$$

i. By a suitable row permutation, bring this matrix G' into a standard form G. Hence write down a parity check matrix H for C.

ii. Compute the matrix $G.H^t$ (where H^t is the transpose of H). Interpret your result.

iii. What are the weights of the three row vectors of G'? What does this tell us about the minimum distance $d(C)$?

iv. Show that $d(C) = 3$.

v. How many of the coset leaders of C have weight 1?

vi. Codeword x is transmitted down a noisy channel, so that $y = 11254$ is received, with exactly one error having occured. What was the transmitted codeword x?

(d) Let $\mathbb{Z}_4$ denote the set of integers modulo 4, together with the associated mod.4 arithmetic. Give the addition and multiplication tables for $\mathbb{Z}_4$. Explain why this number system of modulo 4 arithmetic does *not* form a field.

(e) Explain a way to construct a field of order 4. Write down the addition and multiplication tables for this field.

3. Let Σ_q denote a set of symbols (an 'alphabet') of size q. That is, $|\Sigma_q| = q$. We shall assume that there is a 'zero' element $0 \in \Sigma_q$.

   (a) Let S, T be sets. Explain carefully what is meant by the Cartesian product $S \times T$.

   (b) Given a set S, a closed binary operation on S is a map $\mu : S \times S \to S$. Explain what it means to say that the binary operation μ is associative.

   Give an example of a pair (S, μ) such that μ is associative; and another example such that μ is not associative.

   (c) We write $\Sigma_q^2 = \Sigma_q \times \Sigma_q$ for the Cartesian product of Σ_q with itself. Explain what is meant by the the n-th Cartesian power of Σ_q, denoted Σ_q^n.

   Illustrate your answer by writing out all elements of $\{0, 1\}^3$ explicitly, using a notation in which (a, b) is written as ab, and so on.

   Explain a property of 'symbols' 0 and 1 that allows the above notation to work unambiguously.

   (d) A q-ary code of length n is a subset of Σ_q^n. How many of these are there (as a function of q and n).

   (e) i. Define the Hamming distance d on Σ_q^n.

   ii. Note that $00...0 \in \Sigma_q^n$. How many elements of Σ_q^n have Hamming distance 2 or less from the element 00...0? (Hint: determine how many elements have distance ≤ 1 from 00...0; and how many have distance 2 from 00...0.)

   iii. Define the minimum distance $d(C)$ of a code $C \subset \Sigma_q^n$.

   iv. Given that code C is 7 error correcting, what is the smallest that $d(C)$ could be?

   v. State the ball-packing bound on the size M of a q-ary (n, M, d)-code C.

   (f) For each of the following triples (n, M, d) construct, if possible, a binary (n, M, d)-code:

   $$(19, 2, 19) \qquad (3, 8, 1) \qquad (4, 8, 2) \qquad (8, 89, 3)$$

   If no such code exists, then prove it, stating any theorems used.

4. The 26 letters of the alphabet may be represented in $\mathbb{Z}_3^3$ by $A \mapsto 001$, $B \mapsto 002$, $C \mapsto 010$, $D \mapsto 011$, ..., $Z \mapsto 222$. That is, the k-th letter of the alphabet is represented by $abc \in \mathbb{Z}_3^3$, where $k = a3^2 + b3 + c$. Let us also represent the symbol 'space' by 000.
We are given the parity check matrix

$$H = \begin{pmatrix} 1 & 0 & 1 & 2 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

of a 3-ary $[6, 3, d]$-code C. That is, $w \in C$ if and only if $Hw^t = 0$. (As usual we write simply 0 for the zero vector, where no ambiguity can arise; and write w^t for the transpose of a vector w.)
For example $H(100012)^t = 0$, so $100012 \in C$.

(a) Compute $H(221000)^t$ and hence determine whether $221000 \in C$.

(b) Note that H is not in standard form. Confirm that

$$G = \begin{pmatrix} 2 & 2 & 1 & 0 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 & 0 \\ 2 & 0 & 0 & 0 & 2 & 1 \end{pmatrix}$$

is a generator matrix for C.

(c) Recall that G may be used to encode elements $u = (u_1, u_2, u_3)$ of $\mathbb{Z}_3^3$ by $u \mapsto x = uG$. Thus it may be used to encode letters of the alphabet, via our representation above. Compute the encoded form of the letter D.

(d) What is $d(C)$? How many coset leaders lie within distance 1 of 000000? Compute their syndromes.

(e) Decode as much as possible of the following received message. You may assume that the transmitted message was encoded using C with generator matrix G, and use nearest neighbour decoding. (Marks are available for partial decodings, but all working must be shown.)
Message:
200021 112000 112100 220112 022021 212012
000000 212012 012212 220112 000000 022021
212012 202100 020121 220112 111112 212012

5. (a) Find in standard form the generator matrix for a linear code $D \subseteq Z_{11}^4$ such that D is self-dual.

(b) A block-length n binary linear code is transmitted down a binary symmetric channel with symbol error probability p. Received vectors are decoded by the coset decoding method. Let x denote the sent word, and y the received word, so that $e = y - x$ is the transmission error vector. Then $P(e = 000...0)$ denotes the probability of a codeword being transmitted without error. What is $P(e = 000...0)$? Let S be a set of coset leaders in a standard array for C. Explain why the probability of a codeword being decoded correctly is

$$P_{corr}(C) = \sum_{v \in \mathsf{S}} P(e = v)$$

For the remainder of this question, let C be the binary linear code with generator matrix

$$G = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

(c) Construct a standard array for C.

(d) Decode the received message 1101 using your array.

(e) Code C is transmitted down a binary symmetric channel with symbol error probability $p = 0.01$, with the received vectors being decoded by the coset decoding method, as in (a) above. What is $P(e = 0000)$? Explain why the probability of a transmitted word being decoded correctly can be written as

$$P_{corr}(C) \ = \ P(e = 0000) + P(e = 1000) + P(e = 0100) + P(e = 1100)$$

(Hint: in a binary symmetric channel we have that $P(e = 1100) = P(1010) = P(1001) = P(0110)$ and so on.)
Calculate $P_{err}(C)$, the word error probability of the code; and $P_{undetec}(C)$, the probability of there being an undetected error in a transmitted word.

(f) Code C is again transmitted down a binary symmetric channel with symbol error probability $p = 0.01$, but is now used only for error detection. If an error is detected in a received vector, the receiving device requests retransmission of the codeword. Calculate $P_{retrans}(C)$, the probability that a single codeword transmission will result in a request to retransmit.

 End.