Coding Theory 2008 Answers

1. (a) Σ_q^n: ordered n-tuples from Σ_q. (1 marks)

$\{0,1\}^3 = \{000, ..., 111\}$ (using $(i,j,k) \mapsto ijk$). (1 marks)

$P(\Sigma_q^n) = 2^{q^n}$ (also accept count excluding empty set). (2 marks)

Roughly determine an n and a q such that this entire document is a codeword in some $C \subset \Sigma_q^n$.
ANSWER: Estimate $q = 100$ (62 alphanumeric + punctuation + symbols), $n = 80 \times 25 \times 5 = 10,000$.
OR: $n = q = 1$ where the entire paper is a single 'symbol', and Σ_q just contains this symbol.
OR: any correctly argued variant. (1 marks)

(b) i. Hamming distance $d(x,y) = \#\{i | x_i \neq y_i\}$ (1 marks)

ii. Suppose $d(x,z) + d(z,y) = d(x,y)$. Then

$$D(x,y) := \{i | x_i \neq y_i\}$$

clearly obeys

$$D(x,y) = D(x,z) \cup D(z,y)$$

and $I = D(x,z) \cap D(z,y) = \emptyset$.
Otherwise $I \neq \emptyset$ and for $i \in I$ then $z_i \neq x_i = y_i \neq z_i$ is possible.
Either way $d(x,y) \leq |I| \leq d(x,z) + d(z,y)$.
OR EQUIVALENT. (4 marks)

iii. Weight 0/1: Vectors of form $(0,0,...,0,X,0,...,0)$. There are $n \times (q-1) + 1$ of these.
Weight 2: Vectors of form $(0,0,...,0,X,0,..,Y,0,...,0)$ with $X, Y \neq 0$. There are $\frac{n(n-1)}{2} \times (q-1)^2$ of these. (2 marks)

iv. minimum distance $d(C) = min\{d(x,y)|x,y \in C,\ x \neq y\}$ (2 marks)

v. $d(C) \geq 15$. (1 marks)

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

$$M \sum_{r=0}^{t} \binom{n}{r} (q-1)^r \leq q^n$$

where t such that $d \geq 2t + 1$. (2 marks)

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

$$(6,2,6) \qquad (3,8,1) \qquad (4,8,2) \qquad (8,40,3)$$

If no such code exists, then prove it, stating any theorems used.
ANSWER: (6,2,6): $\{000000, 111111\}$.
(3,8,1): $\{000, 001, 010, 011, 100, 101, 110, 111\}$
(4,8,2): $\{0000, 0011, 0101, 0110, 1001, 1010, 1100, 1111\}$
(8,40,3): fails the BP bound:

$$40(1+8) = 360 \not\leq 2^8 = 256$$

(5 marks)

(d) $p(x\ transmitted) = (1-p)^{n-d(x,w)}p^{d(x,w)}$ so $(1-p) > p$ implies $p(x)$ maximal when $d(x,w)$ minimal. $\square$ (3 marks)

(/25 marks)

2. (a) $|\mathcal{M}_{n,m}(F)| = q^{nm}$ (1 marks)

(b) M generator if rows linearly independent (which implies $n \leq m$), and F finite.
Then M generates a $|F|$-ary [m,n]-code (dimension n, length m code over F). (2 marks)

(c)

$$C_1 = \{000, 101, 011, 110\}$$

(1 marks)

(d) S_1 not closed under $+$.
S_2 is closed under linear combinations, so linear code.
S_3 is closed under linear combinations, so linear code.
S_4 is not closed under $+$. (4 marks)

(e) minimum weight $w(C) = min\{w(x)|x \in C \setminus \{0\}\}$ (where w is weight, and 0 denotes the zero vector).
Prove that, for a linear code, the minimum distance $d(C)$ is equal to $w(C)$.:
$d(x, y) = d(x - y, 0) = w(x - y)$ $\square$ (5 marks)

(f) Define $C^{\perp}$, the dual code to a linear code C.
$C \subset F_q^n$, $C^{\perp} = \{v \in F_q^n | v.x = 0 \forall x \in C\}$ where $v.x = \sum_i v_i x_i$ (over F).
Prove that $C^{\perp}$ is also a linear code:
$v.x = 0$ is a linear constraint on $\{v_i\}$ for any given x. (5 marks)

(g) Compute the dual of C_1 above, and hence or otherwise determine if it is self-dual.
$(x, y, z).(1, 0, 1) = x + z = 0$
$(x, y, z).(0, 1, 1) = y + z = 0$
$(x, y, z).(1, 1, 0) = x + y = 0$

Take $x = 1$ (WLOG), then these imply $x = y = z = 1$, so $C^\perp = \{000, 111\}$. So $C^\perp \neq C$, so not self-dual. (4 marks)

(h) Give an example of an error correcting linear code used by humans in everyday life:
ANY SENSIBLE ANSWER IS OK. EXAMPLE:
Repetition code is linear. Let M be the number of message words required. Choose $q = p^e$ such that $q \geq M$ and assign each messageword m to a $\psi(m) \in F_q$. Then $C \subset F_q^n$ ($n = 4$, say) has $G = (1, 1, 1, 1)$. Thus C a q-ary [n,1]-code.
In practice this code is constructed on the fly by deaf or hearing impaired people, who routinely force others to transmit using it by repeatedly using their retransmission signal "beg pardon?!", and assembling the result into a codeword, until d is big enough for the channel. (3 marks)

3. SEE ALSO HANDWRITTEN SOLUTIONS.

(a) Explain a way to construct a field of order 4.
ANSWER: Consider degree 2 polynomials over $\mathbb{Z}_2$. Quadratics are x^2+1, x^2+x+1, x^2, x^2+x. Only x^2+x+1 irreducible, so extend $\mathbb{Z}_2$ by x obeying this. (4 marks)

Write down the addition and multiplication tables for this field.

$+$	0	1	x	$1+x$
0	0	1	x	$1+x$
1	1	0	$1+x$	x
x	x	$1+x$	0	1
$1+x$	$1+x$	x	1	0

$\times$	0	1	x	$1+x$
0	0	0	0	0
1	0	1	x	$1+x$
x	0	x	$1+x$	1
$1+x$	0	$1+x$	1	x

(4 marks)

Construct the table of multiplicative inverses for the field $\mathbb{Z}_7$.
$\mathbb{Z}_7$: $1^{-1}=1$, $2^{-1}=4$, $3^{-1}=5$, $6^{-1}=6$. (2 marks)

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

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

i. Write down a parity check matrix H for C.

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

(2 marks)

ii. Compute the matrix $G.H^t$ (where H^t is the transpose of H). Interpret your result.
$GH^t = 0$ (show calculations)
Columns of H^t are rows of H so GH^t assembles the various inner product calculations for C and $^\perp$, which must all be zero by definition. (2 marks)

iii. Show that $d(C) = 3$.
H has no zero or parallel columns, but $w(G_3) = 3$ so $d(C) \leq 3$.
So $d(C) = 3$. (3 marks)

iv. How many of the coset leaders of C have weight 1?
There are 7^2 coset leaders. There are $5 \times 6 = 30$ weight 1 vectors, none of which lie in C, and no distint pair of which have $x - y \in C$. So number = 30. (3 marks)

v. 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?
$Hy^t = \begin{pmatrix} 5 \\ 0 \end{pmatrix}$. Now find coset leader: $H \begin{pmatrix} 0 \\ 0 \\ 0 \\ 5 \\ 0 \end{pmatrix} = \begin{pmatrix} 5 \\ 0 \end{pmatrix}$
(3 marks)
so $x = 11254 - 00050 = 11204$. (2 marks)

4. (a) i. a standard array for $C = \{0000, 1010, 0101, 1111\}$:
0000 1010 0101 1111
1000 0010 1101 0111
0100 1110 0001 1011
1100 0110 1001 0011

(5 marks)

ii. Decode the received message 1101 using your array:
the coset leader is 1000, so 1101-1000=0101 is the decoding.

(3 marks)

ALSO SEE HANDWRITTEN FOR OTHER ANSWERS.

iii. 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. 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.
ANSWER:

$$P_{corr}(C) =$$

$$P(e = 0000) + P(e = 1000) + P(e = 0100) + P(e = 1100) =$$
$$(1-p)^4 + 2p(1-p)^3 + p^2(1-p)^2 = 0.9801$$
$$P_{err}(C) = 1 - P_{corr}(C) = 0.0199$$

(3 marks)

$$P_{undetec}(C)|_{p=0.01} = (0.99)^4 + 2\times 0.01(0.99)^3 + 0.0001(0.99)^2 = 0.039...$$

(2 marks)

iv. 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.

$$P_{retrans} = 1 - P(no\ error\ detected) = 1 - P(no\ error) - P(undetected\ error)$$

$$P_{undetec} = 2p^2(1-p)^2 + p^4$$

(3 marks)

so

$$P_{retrans}(C) = 1 - (1-p)^4 - 2p^2(1-p)^2 - p^4$$

so

$$P_{retrans}(C)|_{p=0.01} = etc.$$

(3 marks)

(b) Give the definition of the syndrome of a received word. Prove that two words have the same syndrome iff they lie in the same coset of the code C.
Syndrome $S(y) = yH^t$ where H is the PCM of C. (2 marks)

Proof of Lemma: $y_1H^t = y_2H^t$ if and only if $(y_1 - y_2)H^t = 0$ iff $y_1 - y_2 \in C$ iff $C + y_1 = C + y_2$. $\square$
(or equivalent). (4 marks)

5. (a) Confirm that G is a generator matrix for C:
   1. rows linearly independent
   2. $GH^t = ...calculation... = 0$
   3. # rows = 6-3

   All ok. (3 marks)

   (b) Compute the encoded form of the letter E:
   E is 5th letter, so rep is 012 and encoding is 220112 (3 marks)

   (c) What is $d(C)$?
   Clearly $d(C) \leq 3$, but no column of H is "parallel" to another, so $d(C) = 3$. (2 marks)

   This implies no y with $w(y) = 1$ or 2 lies in C.
   Now suppose x, y of wt 1 lie in $C + x$. Then $y - x$ lies in C. But $w(y - x) \leq 2$, so wt.1 vectors lie in distinct cosets.
   There are $6 \times 2 = 12$ of them. Syndromes:
   $S(000000) = 000$
   $S(100000) = 100$
   $S(200000) = 200$
   etc
   $S(000002) = 202$

   (8 marks)

   (d) Message:
   212012 012212 220112 112100 220112 000000
   200021 112000 220112 000000 022021 221000
   022200 002000 022021 221000 111112 022022
   Now:
   $212012.H^t = 000$ so we have 202, which gives T;
   $012212.H^t = 220$ so we have 012212-002000=010212, which gives 022, which gives H;
   we already encoded E to get the next word;

and so on, until

$$(022021)\begin{pmatrix} 100 \\ 010 \\ 110 \\ 200 \\ 001 \\ 101 \end{pmatrix} = (010) = S(010000)$$

giving 201, and hence S;

and so on, until the last vector. To this point we have:

THERE ARE SIX SIN[?]

The syndrome for this last vector is not among the computed ones. By linearity (say) one can find a couple of weight 2 possibilities, from which (by context!) one might guess that S is the transmitted symbol.

(Alternatively, full marks are available for suggesting RETRANSMISSION at this point.)

Altogether we get:

THERE ARE SIX SINS

(9 marks)