Communication Theory (Part II)
Information Theory (Part IIB)
Coding and Cryptography (Part II)
Information and Coding (Part III)

Math Tripos Questions and Solutions 1990–2006

YMS

2006

2006, II104G Coding and Cryptography

Define a linear feedback shift register. Explain the Berlekamp–Massey method for "breaking" a key stream produced by a linear feedback shift register of unknown length. Use it to find the feedback polynomial of a linear feedback shift register with output sequence

$$0\ 1\ 0\ 1\ 1\ 1\ 1\ 0\ 0\ 0\ 1\ 0\ \ldots.$$

Solution. An LFSR of length d is a function $f\colon \mathbb{F}_2^d \to \mathbb{F}_2^d$ of the form

$$f(x_0, \ldots, x_{d-1}) = \big(x_1, \ldots, x_{d-1}, C(x_0, \ldots, x_{d-1})\big),$$

for some linear function $\mathcal{C}\colon \mathbb{F}_2^d \to \mathbb{F}_2$:

$$\mathcal{C}(x_0, \ldots, x_{d-1}) = \sum_{i=0}^{d-1} a_i x_i.$$

The initial fill $(x_0, \ldots, x_{d-1})$ provides output stream $(x_n)_{n\geq 0}$ with $x_{n+d} =$

$\sum_{i=0}^{d-1} a_i x_{n+i}$. This is equivalent to the equation

$$A_d \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{d-1} \\ 1 \end{pmatrix} = 0, \quad \text{where } A_d = \begin{pmatrix} x_0 & x_1 & \dots x_d \\ x_1 & x_2 & \dots x_{d+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_d & x_{d+1} & \dots x_{2d} \end{pmatrix}. \qquad (*)$$

The Berlecamp–Massey method is as follows:

Start at A_r if we know that $d \geq r$.

For each A_i compute det (A_i):

if det $(A_i) \neq 0$ then $d \neq i$;

if det $(A_i) = 0$ then we solve $(*)$ by the Gaussian elimination method and test whether the solutiion works for the rest of the sequence. If so, we know that $d > i$ and start again with A_{i+1}.

In the Example we have:

$$\det(A_1) = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} \neq 0, \det(A_2) = \begin{vmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{vmatrix} \neq 0, \det(A_3) = \begin{vmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{vmatrix} \neq 0$$

and

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

This yields the solution: $d = 4$, $x_{n+4} = x_n + x_{n+1}$. The linear recurrence relation is satisfied by every term of the output sequence given. The feedback polynomial is then $X^4 + X + 1$.

2006, II204G Coding and Cryptography

Let Σ_1 and Σ_2 be alphabets of sizes m and a. What does it mean to say that an a-ary code $f : \Sigma_1 \to \Sigma_2^*$ is decipherable? Show that if f is decipherable then the word lengths $s_1, \dots, s_m$ satisfy

$$\sum_{i=1}^{m} a^{-s_i} \leq 1 \ .$$

Find a decipherable binary code consisting of codewords 011, 0111, 01111, 11111, and three further codewords of length 2. How do you know the example you have given is decipherable?

Solution. The notation: $\Sigma^* = \cup_{n \geq 0} \Sigma^n$. then $f\colon \Sigma_1 \to \Sigma_2^*$ induces $f^*\colon \Sigma_1^* \to \Sigma_2^*$ by

$$x_1 \cdots x_k \mapsto f(x-1) \cdots f(x_k).$$

f is called decipherable if f^* is injective.

Theorem. (McMillan) *Any decipherable code satisfies the Kraft inequality.*

Proof let $f\colon \Sigma_1 \to \Sigma_2^*$ be adecipherable code with word lengths $s_1, \ldots, s_m$ and $\bar{s} = \max_{1 \leq i \leq m} S_i$. Let r be an integer. Then

$$\left(\sum_{i=1}^{m} a^{-s_i} \right)^r \leq \sum_{l=1}^{rs} b_l a^{-l},$$

where b_l is the number of ways of choosing r codewords of total length l.

If f is decipherable then $b_l \leq \left| \Sigma_2^l \right| = a^l$, and hence,

$$\left(\sum_{i=1}^{m} a^{-s_i} \right)^r \leq rs.$$

Thus,

$$\sum_{i=1}^{m} a^{-s_i} \leq (rs)^{1/r}.$$

the RHS ghoes to 1 as $r \to \infty$. So,

$$\sum_{i=1}^{m} a^{-s_i} \leq 1.$$

In the example, the three extra codewords must be 00, 01, 10 (we can't take 11, as then a sequence of ten 1's is ambiguous). Reversing the order in every codeword gives a prefix-free code. But prefix-free codes are decipherable. Hence, the code is decipherable.

2006, II212G Coding and Cryptography

Define a cyclic code. Show that there is a bijection between the cyclic codes of length n, and the factors of $X^n - 1$ in $\mathbb{F}_2[X]$.

If n is an odd integer then we can find a finite extension K of $\mathbb{F}_2$ that contains a primitive nth root of unity α. Show that a cyclic code of length n with defining set $\{\alpha, \alpha^2, \dots, \alpha^{\delta-1}\}$ has minimum distance at least δ. Show that if $n = 7$ and $\delta = 3$ then we obtain Hamming's original code.
[*You may quote a formula for the Vandermonde determinant without proof.*]

Solution. A subset $\mathcal{C} \subset \mathbb{F}_2^n$ is a cyclic code if

(a) $\mathcal{C}$ is linear, i.e., (i) $0 \in \mathcal{C}$, (ii), $x, y \in \mathcal{C}$ implies that $x + y \in \mathcal{C}$;

(b) $(x_1, \dots, x_n) \in \mathcal{C}$ implies that $(x_2, \dots, x_n, x_1) \in \mathcal{C}$. With the identification

$$\frac{\mathbb{F}_2[X]}{(X^n - 1)} = \big\{ f \in \mathbb{F}_2[X];\ \deg\,(f) < n \big\} = \mathbb{F}_2^n$$

the cyclic codes become ideals in the polynomial ring $\dfrac{\mathbb{F}_2[X]}{(X^n - 1)}$. [They are in a 1–1 correspondence with the ideals in $\mathbb{F}_2[X]$ containing polynomial $X^n - 1$. Because $\mathbb{F}_2[X]$ is a Euclidean domain, all ideals in $\mathbb{F}_2[X]$ are principal, i.e., of the form $\{f(X)g(X) :\ f(X) \in \mathbb{F}_2[X]\}$. The lemma below states that ideals in $\mathbb{F}_2[X]/(X^n - 1)$ are also principal ideals.

Lemma 1. *Let* $\mathcal{C} \subseteq \dfrac{\mathbb{F}_2[X]}{(X^n - 1)}$ *be a cyclic code. then there exists a unique polynomial* $g(X) \in \mathbb{F}_2[X]$ *such that*

(i) $\mathcal{C} = \big\{ f(X)g(X)\ (\mathrm{mod}\ (X^n - 1)) :\ f(X) \in \mathbb{F}_2[X] \big\}$.

(ii) $g(x) \big| (X^n - 1)$.

Proof. Take $g(X) \in \mathbb{F}_2[X]$ a non-zero polynomial of least degree in $\mathcal{C}$. Take $p(X) \in \mathbb{F}_2[X]$ with $p(X)\big(\mathrm{mod}\ (X^n - 1)\big)$ in $\mathcal{C}$. Write

$$p(X) = q(X)g(X) + r(X), \quad \text{with } \deg\, r(X) < \deg\, g(X).$$

Then $r(X)\big(\mathrm{mod}\ (X^n - 1)\big)$ belongs to $\mathcal{C}$. This contradicts the choice of $g(X)$ unless $r(X) = 0$. Therefore, $g(X)|p(x)$ which proves (i).

Taking $p(X) = X^n - 1$ proves (ii).

Finally, if $g(X)$ and $\widetilde{g}(X)$ both satisfy (i) and (ii) then $g(X)|\widetilde{g}(X)$ and $\widetilde{g}(X)|g(X)$, implying $g(X) = \widetilde{g}(X)$.

Polynomial $g(X)$ is called teh generator polynomial of $\mathcal{C}$.

Corollary. *The map*

$$\begin{array}{rcl} \{\text{cyclic codes of length } n\} & \to & \{\text{ divisors of } X^n-1\} \\ C & \mapsto & g(X) \end{array}$$

is a bijection.

Lemma 2. (The Vandermonde determinant)

$$\begin{vmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \dots & x_n^{n-1} \end{vmatrix} = \prod_{i>j}(x_i - x_j).$$

Proof. This is an identity in $\mathbb{Z}[x_1, \dots, x_n]$. The LHS vanishes when we specify $x_i = x_j$ for $i \neq j$. So, $(x_i - x_j)|$LHS and hence RHS|LHS. But both sides have the total degree $\binom{n}{2}$ and the coefficient of $x_2 x_3^2 \cdots x_n^{n-1}$ is one. This implies that LHS=RHS.

The BCH code, $\mathcal{C}^{\rm BCH}_{n,\delta}$, of length n and design distance δ, is constructed by indicating zeros of polynomials $f(X) \in \mathcal{C}^{\rm BCH}_{n,\delta}$:

$$\mathcal{C}^{\rm BCH}_{n,\delta} = \big\{ f(X)\big(\text{mod } (X^n - 1)\big) : \ f(X) \in \mathbb{F}_2[X] \text{ and } f(\alpha^i) = 0, \ 1 \leq i \leq \delta - 1 \big\}.$$

Here α is a primitive root of unity of degree n, i.e., a root of the polynomial $X^n - 1$. (These roots form a commutative multiplicative group which is cyclic, of order n, and α is a generator of this group.)

Lemma 3. *The BCH code $\mathcal{C}^{\rm BCH}_{n,\delta}$ has minimum distance $\geq \delta$.*

Proof. Set the parity-check $(\delta - 1) \times n$ matrix

$$H = \begin{pmatrix} 1 & \alpha & \alpha^2 & \dots & \alpha^{n-1} \\ 1 & \alpha^2 & \alpha^4 & \dots & \alpha^{2(n-1)} \\ \vdots & \vdots & \ddots & \vdots & \\ 1 & \alpha^{\delta-1} & \alpha^{2(\delta-1)} & \dots & \alpha^{(\delta-1)(n-1)} \end{pmatrix}.$$

The codewords of $\mathcal{C}$ are linear dependence relations between the columns of H. Then Lemma 2 implies that any $\delta - 1$ columns of H are linearly independent. In fact, select columns with top (row 1) entries α^{k_1}, $\dots$, $\alpha^{k_{\delta-1}}$ where $0 \leq k_1 < \dots < k_{\delta-1} \leq n-1$. They form a square $(\delta - 1) \times (\delta - 1)$

matrix

$$D = \begin{pmatrix} \alpha^{k_1} \cdot 1 & \alpha^{k_2} \cdot 1 & \dots & \alpha^{k_{\delta-1}} \cdot 1 \\ \alpha^{k_1} \cdot \alpha^{k_1} & \alpha^{k_2} \cdot \alpha^{k_2} & \dots & \alpha^{k_{\delta-1}} \cdot \alpha^{k_{\delta-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha^{k_1} \cdot \alpha^{k_1(\delta-2)} & \alpha^{k_2} \cdot \alpha^{k_2(\delta-2)} & \dots & \alpha^{k_{\delta-1}} \cdot \alpha^{k_{\delta-1}(\delta-2)} \end{pmatrix}$$

that differs from the Vandermonde matrix by factors α^{k_s} in front of the sth column. Then the determinant of D is the product

$$\begin{aligned} \det D &= \left(\prod_{s=1}^{\delta-1} \alpha_j^{k_s} \right) \begin{vmatrix} 1 & 1 & \dots & 1 \\ \alpha^{k_1} & \alpha^{k_2} & \dots & \alpha^{k_{\delta-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha^{k_1(\delta-2)} & \alpha^{k_2(\delta-2)} & \dots & \alpha^{k_{\delta-1}(\delta-2)} \end{vmatrix} \\ &= \left(\prod_{s=1}^{\delta-1} \alpha_j^{k_s} \right) \times \left(\prod_{i>j} \left(\alpha^{k_i} - \alpha^{k_j} \right) \right) \neq 0. \end{aligned}$$

Thus, indeed any $\delta - 1$ columns of H are linearly independent. In turn, this means that any non-zero codeword in $\mathcal{C}$ has weight at least δ. Thus, $\mathcal{C}$ has minimum distance $\geq \delta$.

Factorising in $\mathbb{F}_2^7$ we find

$$X^7 - 1 = (X-1)(X^3 + X + 1)(X^3 + X^2 + 1) := (X-1) f_1(X) f_2(X).$$

Suppose α is a root of $f_1(X)$. Since $f_1(X)^2 = f_1(X^2)$ in $\mathbb{F}_2[X]$ we have

$$f_1(\alpha) = f_1(\alpha^2) = 0.$$

It follows that $\mathcal{C}$ has the generator polynomial $f_1(X)$ and the check polynomial $(X-1)f_2(X) = X^4 + X^2 + X + 1$. Then the parity check matrix is

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

The columns of H run over $\mathbb{F}_2^3 \setminus \{0\}$. This property characterises Hamming's original code (up to equivalence).

The case where α is a root of $f_2(X)$ is similar (in fact, we just reverse every codeword).

2006, II304G Coding and Cryptography

What does it mean to say that a binary code $\mathcal{C}$ has length n, size m and minimum distance d? Let $A(n,d)$ be the largest value of m for which there exists an $[n,m,d]$-code. Prove that

$$\frac{2^n}{V(n,d-1)} \leq A(n,d) \leq \frac{2^n}{V(n, \lfloor \frac{1}{2}(d-1) \rfloor)}$$

where $V(n,r) = \sum_{j=0}^{r} \binom{n}{j}$.

Solution. A code of length n and size m is a subset $\mathcal{C} \subset \mathbb{F}_2^n$ with m elements (called codewords).The Hamming distance between $x, y \in \mathbb{F}_2^n$ is

$$d(x,y) = \#\{1 \leq i \leq n : \ x_i \neq y_i\}.$$

The minimum distance of $\mathcal{C}$ is the minimum Hamming distance between distinct codewords.

Let $B(x,r) = \{y \in \mathbb{F}_2^n : \ d(x,y) \leq r\}$ be the closed Hamming ball centred at x and of radius r. It hax $V(n,r)$ elements. let $\mathcal{C}$ be an $[n,m,d]$ code. Then it is $e = \left\lfloor \frac{d-1}{2} \right\rfloor$ error correcting. Hence, $\cup_{c\in\mathcal{C}} B(c,e)$ is a disjoint union, which implies that

$$\sum_{c\in\mathcal{C}} \left|B(c,e)\right| \leq \left|\mathbb{F}_2^n\right| = 2^n.$$

Therefore,

$$m = |C| \leq \frac{2^n}{V(n,e)}, \quad \text{i.e.,} \quad A(n,d) \leq \frac{2^n}{V(n,e)}.$$

Now let $m = A(n,d)$ and let $\mathcal{C}$ be an $[n,m,d]$ code. Then $\nexists\, x \in \mathbb{F}_2^n$ with $d(x,c) \geq d$ for all $c \in \mathcal{C}$, otherwise we could add x to $\mathcal{C}$. therefore,

$$\cup_{c\in\mathcal{C}} B(c,d-1) = \mathbb{F}_2^n, \quad \text{implying that} \quad A(n,d) = |C| \geq \frac{2^n}{V(n,d-1)}.$$

2006, II312G Coding and Cryptography

Describe the RSA system with public key (N, e) and private key (N, d). Briefly discuss the possible advantages or disadvantages of taking (i) $e = 2^{16} + 1$ or (ii) $d = 2^{16} + 1$.

Explain how to factor N when both the private key and public key are known.

Describe the bit commitment problem, and briefly indicate how RSA can be used to solve it.

Solution. The RSA system is as follows. We pick a number $N = pq$ where p, q are distinct large primes (which remain secret). Write

$$\phi(N) = \left|\left(\mathbb{Z}/N\mathbb{Z}\right)^*\right| = (p-1)(q-1).$$

Next, we choose e with gcd $(e, \phi(N)) = 1$ and solve for d such that

$$de \equiv 1 \pmod{\phi(N)}.$$

Then form:

private key: (N, d),
public key: (N, e).

A message $m \in \{1, \ldots, N-1\}$ is then encrypted as ciphertext $c = m^e \pmod{N}$.

The ciphertext c is decrypted as $m_1 = c^d \pmod{N}$.

The Euler-Fermat theorem ensures that $m_1 \equiv m^{de} \equiv m \pmod{N}$ (we neglect the possibility gcd $(m, N) > 1$, assuming that m and N are co-prime).

(i) With $e = 2^{16} + 1$ we can encrypt messages quickly (need just 17 multi[plications using repeated squareing).

(ii) With $d = 2^{16} + 1$ we can decrypt messages quickly . However, an eavesdropper can easily guess it.

We show that if we know a multiple m of $\phi(N)$ then it is easy to factor N.

Denote by $o_p(x)$ the order of x in $\mathbb{F}_p^* = \left(\mathbb{Z}/p\mathbb{Z}\right)^*$. Let $m = 2^a b$ where $a \geq 0$, b odd. Set

$$X = \left\{x \in \left(\mathbb{Z}/p\mathbb{Z}\right)^* : \ o_p(x^b) \neq o_q(x^b)\right\}.$$

Lemma. (i) *If $x \in X$ then $\exists\ 0 \leq t < a$ such that gcd $\left(x^{2^t b} - 1, N\right)$ is a non-trivial factor of N.*

(ii) $|X| \geq \frac{1}{2}\left|\left(\mathbb{Z}/N\mathbb{Z}\right)^*\right| = \frac{\phi(N)}{2}$.

Proof. Put $y = x^b \pmod{N}$. The Euler-Fermat theorem implies that $x^{\phi(N)} \equiv 1 \pmod{N}$ and hence $y^{2^a} \equiv 1 \pmod{N}$. Then

$$o_p(x^b) \text{ and } o_q(x^b) \text{ are powers of } 2.$$

We are given $o_p(x^b) \neq o_q(x^b)$; wlog, $o_p(x^b) < o_q(x^b)$. Then $\exists\ 0 \leq t, a$ such that

$$\begin{aligned} y^{2^t} &\equiv 1 \pmod{p}, \\ y^{2^t} &\not\equiv 1 \pmod{q}. \end{aligned}$$

So, gcd $(\left(y^{2^t} - 1, N\right) = p$, as required.

(ii) By the Chinese remainder theorem, there is a bijection

$$\begin{array}{rcl} \left(\mathbb{Z}/N\mathbb{Z}\right)^* & \leftrightarrow & \left(\mathbb{Z}/p\mathbb{Z}\right)^* \times \left(\mathbb{Z}/q\mathbb{Z}\right)^* \\ x & \mapsto & \left(x \bmod p, x \bmod q\right). \end{array}$$

It suffices to show that if we partition $\left(\mathbb{Z}/p\mathbb{Z}\right)^*$ into subsets according to the value of $o_p(x^b)$ then each subset has size $\leq \frac{1}{2}\left|\left(\mathbb{Z}/p\mathbb{Z}\right)^*\right|$. We will do this by exhibiting such a subset of size $\frac{1}{2}\left|\left(\mathbb{Z}/p\mathbb{Z}\right)^*\right|$.

Recall, $\left(\mathbb{Z}/p\mathbb{Z}\right)^* = \left\{1, g, g^2, \ldots, g^{p-2}\right\}$. Then

$$\phi(N)\big|2^a b \text{ implies that } o_p(g^b) \text{ is a power of } 2.$$

In turn, this imp;lies that

$$o_p(g^{\delta b}) \begin{cases} = o_p(g^b), & \delta \text{ odd}, \\ < o_p(g^b), & \delta \text{ even}. \end{cases}$$

Therefore, $\left\{g^\delta \bmod p : \ \delta \ \text{ odd}\right\}$ is the required subset.

Now, given both (N, e) and (N, d) we put $m = de - 1$ and use the above lemma to factor N. The probability of failure after r random choices of $x \in X$ is $< \frac{1}{2^r}$.

Finally, the bit commitment problem:

Alice sends a message to Bob in such a way that

(i) Bob can't read the message until Alice sends further information.

(i) Alice can't change the message.

One solution is to use a public key system (e.g., RSA): Bob cannot read the message until Alice (later) reveals her private key.

2006, II404G Coding and Cryptography

A binary erasure channel with erasure probability p is a discrete memoryless channel with channel matrix

$$\begin{pmatrix} 1-p & p & 0 \\ 0 & p & 1-p \end{pmatrix}.$$

State Shannon's second coding theorem, and use it to compute the capacity of this channel.

Solution. Shannon's Second coding theorem states that for a discrete memoryless channel (DMC),

$$\text{(capacity)} = \text{(maximum information transmitted per letter)}.$$

Here the capacity is understood as the supremum over all reliable information rates while the RHS is defined as

$$\max_X \ I(X:Y)$$

where X is a random variable input and Y the corresponding output.

The binary erasure channel keeps an input letter 0 or 1 intact with probability $1-p$ and turns it to a splodge $*$ with probability p. An input random variable X is 0 with probability α and 1 with probability $1-\alpha$. Then the output variable Y takes three values:

$$\begin{aligned}
&\mathbb{P}(Y=0) = (1-p)\alpha, \\
&\mathbb{P}(Y=1) = (1-p)(1-\alpha), \\
&\mathbb{P}(Y=*) = p.
\end{aligned}$$

Thus, conditional on the value of Y, we have:

$$\left.\begin{aligned}
&H(X|Y=0) = 0, \\
&H(X|Y=1) = 0, \\
&H(X|Y=*) = H(\alpha)
\end{aligned}\right\}, \quad \text{implying that} \ \ H(X|Y) = pH(\alpha).$$

Therefore,

$$\begin{aligned} \text{capacity} &= \max_\alpha \; I(X,Y) \\ &= \max_\alpha \left[H(X) - H(X|Y)\right] \\ &= \max_\alpha \left[H(\alpha) - pH(\alpha)\right] \\ &= (1-p)\max_\alpha \; H(\alpha) = 1-p. \end{aligned}$$

It is useful to note that $H(\alpha) = -\alpha \log\; \alpha - (1-\alpha)\log\;(1-\alpha)$ attains it maximum value at $\alpha = 1/2$.

2006, III 301 Information and Coding

(a) Consider two discrete random variables X and Y. Define the conditional entropy $h(X|Y)$, and show that it satisfies

$$h(X|Y) \le h(X),$$

giving necessary and sufficient conditions for equality. You may assume the Gibbs inequality, provided that you state it carefully.

(b) Consider two discrete random variables U and V with corresponding probability mass functions p_U and p_V. For each $\alpha \in [0,1]$, define the mixture random variable $W(\alpha)$ by its mass function

$$p_{W(\alpha)}(x) = \alpha p_U(x) + (1-\alpha)p_V(x).$$

Prove that for all α the entropy of $W(\alpha)$ satisfies:

$$h(W(\alpha)) \ge \alpha h(U) + (1-\alpha)h(V).$$

(c) Define $F(\lambda)$ to be the entropy of a Poisson random variable with mean $\lambda > 0$. Show that $F(\lambda)$ is a non-decreasing function of $\lambda > 0$.

Solution. By definition,

$$\begin{aligned} h(X|Y) &= h(X,Y) - h(Y) \\ &= -\sum_{x,y} \mathbb{P}(X=x,Y=y)\log\;\mathbb{P}(X=x,Y=y) \\ &\quad + \sum_y \mathbb{P}(Y=y)\log\;\mathbb{P}(Y=y) \end{aligned}$$

The inequality is then equivalent to

$$h(X,Y) \le h(X) + h(Y)$$

and is proved via the Gibbs inequality $\sum\limits_i p_i \log \dfrac{p_i}{q_i} \geq 0$. In fact, take $i \sim (x,y)$ and

$$p_i = \mathbb{P}(X=x, Y=y), \;\; q_i = \mathbb{P}(X=x)\mathbb{P}(Y=y).$$

Then

$$\begin{aligned} 0 \;\; &\leq \sum_{x,y} \mathbb{P}(X=x,Y=y) \log \frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(X=x)\mathbb{P}(Y=y)} \\ &= \sum_{x,y} \mathbb{P}(X=x,Y=y) \log \; \mathbb{P}(X=x,Y=y) \\ &\quad - \sum_{x,y} \mathbb{P}(X=x,Y=y) \big[\log \; \mathbb{P}(X=x) + \log \; \mathbb{P}(Y=y) \big] \\ &-h(X,Y) + h(X) + h(Y). \end{aligned}$$

Equality here occurs iff X and Y are independent.

(b) Define a random variable T equal to 0 with probqability α and 1 with probqability $1-\alpha$. Then set

$$Z = \begin{cases} U, & \text{if } \; T=0, \\ V, & \text{if } \; T=1. \end{cases}$$

The variable Z has the distribution $W(\alpha)$. By part (a),

$$h(Z|T) \leq h(Z),$$

with the LHS $= \alpha h(U) + (1-\alpha) h(V)$, the RHS $= h(W(\alpha))$.

(c) the key fact is that for X and Y independent,

$$h(X+Y) \geq h(X).$$

Proof: Since only probabilitiews matter, not values,

$$h(X+Y|X) = h(Y|X) = h(Y).$$

Hence, again by part (a),

$$h(X+Y) \geq h(X+Y|X) = h(Y).$$

Using this fact, $\forall \; \lambda_1 < \lambda_2$, take $X \sim$Po(λ_1), $Y \sim$Po$(\lambda_2 - \lambda_1)$, independently. Then

$$h(X+Y) \geq h(X) \;\text{ implies }\; h\big[\text{Po}(\lambda_2)\big] \geq h\big[\text{Po}(\lambda_1)\big], \;\; \text{i.e.,} \;\; F(\lambda_2) \geq F(\lambda_2).$$

2006, III 302 Information and Coding

State the Entropy Power Inequality for n-dimensional random vectors.

Let X be a real-valued random variable with a density and finite differential entropy, and let function $g : \mathbb{R} \to \mathbb{R}$ have strictly positive derivative g' everywhere. Prove that the random variable $g(X)$ has differential entropy satisfying

$$h(g(X)) = h(X) + \mathbb{E} \log_2 g'(X),$$

assuming that $\mathbb{E} \log_2 g'(X)$ is finite.

Let Y_1 and Y_2 be independent, strictly positive random variables with densities. Show that the differential entropy of the product $Y_1 Y_2$ satisfies

$$2^{2h(Y_1 Y_2)} \geq \alpha_1 2^{2h(Y_1)} + \alpha_2 2^{2h(Y_2)},$$

where $\log_2(\alpha_1) = 2\mathbb{E} \log_2 Y_2$ and $\log_2(\alpha_2) = 2\mathbb{E} \log_2 Y_1$.

Solution.

2006, III 303 Information and Coding

(a) Prove the Hamming and Gilbert–Varshamov bounds on the size of a binary code of length N and minimum distance δ, in terms of $v_N(d)$, the volume of an N-dimensional Hamming ball of radius d.

Suppose that the minimum distance is $\lfloor \lambda N \rfloor$ for some fixed $\lambda \in (0, 1/2)$. Describe the asymptotic behaviour of both of the above bounds as $N \to \infty$.

[*You may assume that*

$$\lim_{N \to \infty} \frac{1}{N} \log v_N(\lfloor \lambda N \rfloor) = h(\lambda),$$

where h is the binary entropy function].

(b) State Shannon's Second Coding Theorem, giving the capacity of a general memoryless channel. Use this to calculate the capacity of a memoryless binary symmetric channel with error probability p.

(c) Fix $R \in (0, 1)$ and suppose we want to send one of a collection U_N of messages of length N, where the size $|U_N| = 2^{NR}$. The message is transmitted through a memoryless binary symmetric channel with error probability $p < 1/2$, so that we expect about pN errors. According to the asymptotic Gilbert–Varshamov bound of part (a), for which values of p can we correct

pN errors, for large N? Why does this give a different answer to the Shannon capacity of part (b)?

Solution.

2006, III 304 Information and Coding

Prove that the binary code of length 23 generated by the polynomial $g(X) = 1 + X + X^5 + X^6 + X^7 + X^9 + X^{11}$ has minimum distance 7, and is perfect.

[*You may use the BCH theorem without proof provided it is clearly stated, and you may assume that* $X^{23} + 1 \equiv (X+1)g(X)g^{\mathrm{rev}}(X) \bmod 2$, *where* $g^{\mathrm{rev}}(X) = X^{11}g(1/X)$ *is the reversal of* $g(X)$.]

Solution. First, show that the code is BCH, of designed distance 5.

Lemma 1. *If* α *is a root of a polynomial* $f(X) \in \mathbb{F}_2[X]$ *then so is* α^2.

Proof. Write:

$$0 = (f(\alpha))^2 = \left(\sum_i f_i \alpha^i\right)^2$$
$$= \sum_i f_i^2 \alpha^{2i} + 2\sum_{i<j} f_i f_j \alpha^{i+j} = \sum_i f_i \alpha^{2i} = f(\alpha^2).$$

thus, if α is a root of $g(X) = 1 + X + X^5 + X^6 + X^7 + X^9 + X^{11}$ then so are α, α^2, α^4, α^8, α^{16}, α^9, α^{18}, α^{13}, α^3, α^6, α^{12}. This yields the design sequence $\{\alpha, \alpha^2, \alpha^3, \alpha^4\}$. The BCH theorem states: *if a generator polynomial of a cyclic code has roots* $\{\beta, \beta^2, \ldots, \beta^{\delta-1}\}$ *then the code has distance* $\geq \delta$.

Hence, the code in question, $\mathcal{C} = \langle g(X) \rangle$, has distance ≥ 5.

Next, the parity-check extension, $\mathcal{C}^+$, is self-orthogonal. To check this, we need only to show that any two rows of the generating matrix of $\mathcal{C}^+$ are orthogonal. These are represented by

$$(X^i g(X)|1) \text{ and } (X^j g(X)|1)$$

and their dot-product is

$$\begin{aligned}
1 + (X^i g(X))(X^j g(X)) &= 1 + \sum_r g_{i+r} g_{j+r} \\
&= 1 + \sum_r g_{i+r} g^{\mathrm{rev}}_{11-j-r} \\
&= 1 + \text{coefficient of } X^{11+i-j} \text{ in } \underbrace{g(X) \times g^{\mathrm{rev}}(X)}_{\substack{\| \\ 1+\ldots+X^{22}}} \\
&= 1 + 1 = 0.
\end{aligned}$$

So,

$$\text{any two words in } \mathcal{C}^+ \text{ are dot-orthogonal.} \qquad (*)$$

Lemma 2. *All words in $\mathcal{C}^+$ have weight divisible by 4.*

Proof. By inspection, all rows $(X^i g(X)|1)$ of the generating matrix of $\mathcal{C}^+$ have weight 8. Then, by induction on the number of rows involved in the sum, if $c \in \mathcal{C}^+$ and $g^{(i)} \sim (X^i g(X)|1)$ is a row of the generating matrix of $\mathcal{C}^+$ then

$$w\big(g^{(i)} + c\big) = w\big(g^{(i)}\big) + w(c) - 2w\big(g^{(i)} \wedge c\big),$$

where $\big(g^{(i)} \wedge c\big)_l = \min\ \big[\big(g^{(i)}\big)_l, c_l\big]$, $l = 1, \ldots, 24$. We know that $8|w\big(g^{(i)}\big)$ and by the induction hypothesis, $4|w(c)$. Next, by $(*)$, $w\big(g^{(i)} \wedge c\big)$ is even, so $2w\big(g^{(i)} \wedge c\big)$ is divisible by 4. Then LHS, $w\big(g^{(i)} + c\big)$, is divisible by 4.

Therefore, the distance of $\mathcal{C}^+$ is 8, as it is ≥ 5 and is divisible by 4. (It is easy to see that it cannot be > 8 as then it would be 12.) Then the distance of the original code, $\mathcal{C}$, equals 7.

Then code $\mathcal{C}$ is perfect 3-error correcting, since the volum of the 3-ball in $\mathbb{F}_2^{23}$ equals

$$\binom{23}{0} + \binom{23}{1} + \binom{23}{2} + \binom{23}{3} = 1 + 23 + 253 + 1771 = 2048 = 2^{11},$$

and

$$2^{11} \times 2^{12} = 2^{23}.$$

Here, obviously, 12 represents the rank and 23 the length.

2005

2005, II104J Coding and Cryptography

Briefly describe the methods of Shannon-Fano and Huffman for economical coding. Illustrate both methods by finding decipherable binary codings in the case where messages $\mu_1, \ldots, \mu_5$ are emitted with probabilities $0.45, 0.25, 0.2, 0.05, 0.05$. Compute the expected word length in each case.

Solution. Let messages μ_1, ..., μ_m be emitted with probabilities p_1, ..., p_m. We code in an alphabet Σ with $|\Sigma| = \alpha$.

The Shannon-Fano encoding is as follows. Set: $s_i = \lceil -\log_a p_i \rceil$. We construct a prefix-free a-ary code with codeword-lengths s_i, by taking the s_i's in the increasing order and ensuring that previous codewords are not prefixes. (The Kraft inequality guarantees there is enough room.)

The Huffman encoding, with $a = 2$, is as follows.Assume, wlog, that $p_1 \geq p_2 \geq \ldots \geq p_m$. If $m = 2$, we simply assign codewords 0 and 1 to μ)1 and ν_2. If $m > 2$, we first find a Huffman coding for messages μ_1, ..., μ_{m-2}, $\mu_{m-1} \vee \mu_m$, with probabilities p_1, ..., p_{m-2}, $p_{m-1} + p_m$. Then append 0, respectively, 1, to the codeword for $\mu_{m-1} \vee \mu_m$ to give a codeword for μ_{m-1}, respectively, μ_m. The Huffman e=code is optimal as it minimises the expected codeword-length.

In the example given (with $a = 2$), we have:

Shannon–Fano:

p_i	$\lceil -\log_2 p_i \rceil$	codeword
.45	2	00
.25	2	01
.2	3	100
.05	5	11100
.05	5	11111

with $\mathbb{E}\big(\text{codeword-length}\big) = .9 + .5 + .6 + .25 + .25 = 2.5$, and

Huffman:

p_i	codeword
.45	1
.25	01
.2	000
.05	0010
.05	0011

with $\mathbb{E}\big(\text{codeword-length}\big) = .45 + .5 + .6 + .2 + .2 = 1.95$.

2005, II204J Coding and Cryptography

What is a *linear binary code*? What is the *weight* $w(C)$ of a linear binary code $\mathcal{C}$? Define the bar product $\mathcal{C}_1|\mathcal{C}_2$ of two binary linear codes $\mathcal{C}_1$ and $\mathcal{C}_2$, stating the conditions that $\mathcal{C}_1$ and $\mathcal{C}_2$ must satisfy. Under t hese conditions show that

$$w(\mathcal{C}_1|\mathcal{C}_2) \geq \min(2w(\mathcal{C}_1), w(\mathcal{C}_2)).$$

Solution. A linear binary code of length n is a vector subspace of $\mathbb{F}_2^n$.

The weight $w(x)$ of a vector $x \in \mathbb{F}_2^n$ is

$$w(x) = d(0, x) = \#\{1 \leq i \leq n : \ x_i \neq 0\}.$$

The weight $w(C)$ of a code $\mathcal{C} \subset \mathbb{F}_2^n$ is the minimum weight of a non-zero codeword in $\mathcal{C}$.

To define the bar-product code $\mathcal{C}_1|\mathcal{C}_2$, $\mathcal{C}_1$ and $\mathcal{C}_2$ must be two linear codes of the same length (say n). Then $\mathcal{C}_1|\mathcal{C}_2$ is the code of length $2n$ given by

$$\mathcal{C}_1|\mathcal{C}_2 = \{(x, x + y) : \ x \in \mathcal{C}_1, \ y \in \mathcal{C}_2\}.$$

The construction is used when $\mathcal{C}_2 \subseteq \mathcal{C}_1$; in this case the weight $w\big(\mathcal{C}_1|\mathcal{C}_2\big)$ can be estimated.

Thus, assume that $\mathcal{C}_2 \subseteq \mathcal{C}_1$. Let $x \in \mathcal{C}_1$ and $y \in \mathcal{C}_2$. Then

(i) If $y \neq 0$ then $w(x, x + y) = w(x) + w(x + y) \geq w(y) \geq w(\mathcal{C}_2)$.

(ii) If $y = 0$ and $x \neq 0$ then $w(x, x) = 2w(x) \geq 2w(\mathcal{C}_2)$.

Therefore,

$$w\big(\mathcal{C}_1|\mathcal{C}_2\big) \geq \min \ \big[2w(\mathcal{C}_1), w(\mathcal{C}_2)\big].$$

2005, II212J Coding and Cryptography

What does it means to say that $f : \mathbb{F}_2^d \to \mathbb{F}_2^d$ is a linear feedback shift register? Let $(x_n)_{n\geq 0}$ be a stream produced by such a register. Show that there exist N, M with $N + M \leq 2^d - 1$ such that $x_{r+N} = x_r$ for all $r \geq M$.

Explain and justify the Berlekamp–Massey method for ‘breaking’ a cipher stream arising from a linear feedback register of unknown length.

Let x_n, y_n, z_n be three streams produced by linear feedback registers. Set

$$k_n = x_n \text{ if } y_n = z_n$$

$$k_n = y_n \text{ if } y_n \neq z_n \,.$$

Show that k_n is also a stream produced by a linear feedback register. Sketch proofs of any theorems that you use.

Solution. A linear feedback shift register $f : \mathbb{F}_2^d \to \mathbb{F}_2^d$ is given by

$$f(x_0, x_1, \ldots, x_{d-1}) = \big(x_1, \ldots, x_{d-1}, C(x_0, x_1, \ldots, x_{d-1})\big),$$

where $\mathcal{C} : \mathbb{F}_2^d \to \mathbb{F}_2$ is a linear map, say

$$\mathcal{C}(x_0, x_1, \ldots, x_{d-1}) = a_{d-1}x_0 + a_{d-2}x_1 + \ldots + a_0 x_{d-1}$$

for some $a_i \in \mathbb{F}_2$.

Let x_0, x_1, x_+2, …, be a stream produced by f.

Lemma. *There exist M, N with $N + M \leq 2^d - 1$ such that $x_{r+N} = x_r$ for all $r \geq M$.*

Proof. Let $v_i = (x_i, x_{i+1}, \ldots, x_{i+d-1}) \in \mathbb{F}_2^d$. If $v_i = 0$ for some i then the stream is eventually zero, and the lemma holds true. So, we may assume $v_i \neq 0 \ \forall\, i$.

Then v_0, v_1, …, v_{2^d-1} are 2^d elements (vectors) from $\mathbb{F}_2^d \setminus \{0\}$ which has cardinality $2^d - 1$. So, $v_a = v_b$ for some $0 \leq a < b \leq 2^d - 1$. We put $M = a$, $N = b - a$.

We show by induction that $v_{r+N} = v_r \ \forall\, r \geq M$. Indeed, by construction, $v_{M+N} = v_M$ which establishes the initial step of the induction. Next, assume that $v_{r+N} = v_r$ and write

$$v_{r+1+N} = f(v_{r+N}) = f(v_r) = v_{r+1}.$$

This makes the inductive passage from r to $r + 1$.

Next, since v_i has the first component x_i, it follows that $x_{r+N} = x_r \ \forall\, r \geq M$.

Finally, $N + M = b \leq 2^d - 1$.

The Berlekamp–Massey method is as follows. Let x_0, x_1, x_2, $\dots$ be a stream arising from a linear feedback register. We seek to find the unknown d and a_1, …, a_d such that

$$x_n + \sum_{i=1}^{d} a_i x_{n-i} = 0 \ \forall \ na \geq d.$$

Write

$$\begin{pmatrix} x_d & x_{d-1} & \dots & x_1 & x_0 \\ x_{d+1} & x_d & \dots & x_2 & x_1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{2d} & x_{2d-1} & \dots & x_{d+1} & x_d \end{pmatrix} \begin{pmatrix} 1 \\ a_1 \\ \vdots \\ a_d \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{$*$}$$

We look successively at the matrices

$$A_1 = (x_0),\ A_2 = \begin{pmatrix} x_1 & x_0 \\ x_2 & x_1 \end{pmatrix},\ A_3 = \begin{pmatrix} x_2 & x_1 & x_0 \\ x_3 & x_2 & x_1 \\ x_4 & x_3 & x_2 \end{pmatrix},\ \text{ etc.}$$

(We should start at A_r if it is known that $r \geq d$.) For each A_i we evaluate det A_i. If det $A_i \neq 0$ then $i \neq d$. If det $A_i = 0$, we solve $(*)$ on the assumption $d = i$. This results in some a_1, …, a_d (possibly non-unique). We then check a candidate for $(a_1, \dots, a_d)$ over as many terms of the stream as we wish. If it fails the test we then know that $d > i$ and we start again.

Now, we are given three streams x_n, y_n, z_n produced by linear feedback registers. We have

$$k_n = x_n + (x_n + y_n)(y_n + z_n) \quad (\text{in } \mathbb{F}_2).$$

So it suffices to show that (pointwise) sums and products of streams produced by linear feedback registers also yield some streams produced by linear feedback registers.

Lemma. *Let x_n (respectively, y_n) be a stream produced by a linear feedback register with auxiliary polynomial $P(X)$ (respectively, $Q(X)$). Let P have roots α_1, …, α_M and Q have roots β_1, …, β_N, over some field $K \supset \mathbb{F}_2$. Then*

(i) *$(x_n + y_n)$ is produced by a linear feedback register with auxiliary polynomial $P(X)Q(X)$.*

(ii) $(x_n y_n)$ *is produced by a linear feedback register with auxiliary polynomial* $\prod_{i=1}^{M}\prod_{j=1}^{N}(X-\alpha_i\beta_j)$.

Sketch proof. Suppose for simplicity that the α_i and β_j are distinct. Then

$$x_n=\sum_{i=1}^{M}\lambda_i\alpha_i^n,\ \ y_n=\sum_{j=1}^{N}\mu_j\beta_j^n,$$

for some $\lambda_i,\mu_j\in K$.

(i) We have $x_n+y_n=\sum\limits_{i=1}^{M}\lambda_i\alpha_i^n+\sum\limits_{j=1}^{N}\mu_j\beta_j^n$. So (x_n+y_n) satisfies a linear recurrence with the auxiliary polynomial $P(X)Q(X)=\prod\limits_i(X-\alpha_i)\prod\limits_j(X-\beta_j)$.

(ii) We have $x_ny_n=\sum\limits_{i,j}\lambda_i\mu_j(\alpha_i\beta_j)^n$. So (x_ny_n) satisfies a linear recurrence with the auxiliary polynomial $\prod\limits_{i,j}(X-\alpha_i\beta_j)$.

2005, II304J Coding and Cryptography

Briefly explain how and why a signature scheme is used. Describe the el Gamal scheme.

Solution. Consider a message M sent by A to B. In a signature problem, the receiver B (a) must be able to identify that the message was sent by A and not by a third patry (autentication); (b) must gurantee that a third party did not alter or corrupt the message (integrity); and (c) must be able to convince a third party that A was the sender (non-repudiation).

An important element of the signature scheme is a trapdoor function, i.e., a mapping $s:\mathcal{M}\times\mathcal{K}\to\mathcal{S}$ where

$\mathcal{M}$ is the space of possible messages,
$\mathcal{K}$ is the space of keys,
$\mathcal{S}$ is the space of signatures.

A signs message M with $s(m,K)$ where K is A's private key. Then B uses A's public key to verify thta $s(M,K)$ is in fact the signature of M. The trapdoor function should be designed so that noone can produce signed messages without knowledge of K.

Under the el Gamal scheme, A chooses a large prime p, a primitive root g mod p and some other integer u, with $1 < u < p$. Then:

the public key is the triple p, g and $y := g^u \bmod p$,
the private key is u.

A part of the el Gamal scheme is a hash function $h : \mathcal{M} \to \{1, \dots, p-1\}$. To send a message M, A randomly (i.e. arbitrarily) chooses a number $k \in \left(\mathbb{Z}/(p-1)\mathbb{Z}\right)^*$ (i.e., prime to $p-1$) and computes $1 \le r \le p-1$ and $0 \le s \le p-2$ by the rules

$$r = g^k \pmod p \quad \text{and} \quad h(m) \equiv ur + ks \pmod{p-1} \tag{$*$}$$

Then A signs the message M with (r, s); the triple (M, r, s) is sent to B. B accepts the signature if

$$g^{h(m)} \equiv y^r r^s \pmod p \tag{$**$}$$

(checking autenticity). This verification works since, by $(*)$,

$$\begin{aligned} g^{h(m)} &\equiv g^{ur+ks} \pmod p \\ &\equiv \left(g^y\right)^r \left(g^k\right)^s \pmod p \\ &\equiv y^r r^s \pmod p. \end{aligned}$$

It is believed that the only way to forge a signature is to find u from g^u. This is known as the discrete liogarithm problem (computationally expensive to solve).

2005, II312J Coding and Cryptography

Define a *cyclic code*. Define the generator and check polynomials of a cyclic code and show that they exist.

Show that Hamming's original code is a cyclic code with check polynomial X^4+X^2+X+1. What is its generator polynomial? Does Hamming's original code contain a subcode equivalent to its dual?

Solution. A linear code $\mathcal{C} \subset \mathbb{F}_2^n$ is cyclic if

$$(a_0, a_1, \dots, a_{n-1}) \in \mathcal{C} \Rightarrow (a_1, a_2, \dots, a_{n-1}, a_0) \in \mathcal{C}.$$

We identify $\mathbb{F}_2^n = \mathbb{F}_2[X]/(X^n - 1)$ via

$$(a_0, a_1, \ldots, a_{n-1}) \mapsto a_0 + a_1 X + \ldots + a_{n-1} X^{n-1} \pmod{X^n - 1}.$$

Then $\mathcal{C}$ is an ideal in $\mathbb{F}_2[X]/(X^n - 1)$.

A generator polynomial for $\mathcal{C}$ is a polynomial $g(X)$ such that

$$\mathcal{C} = \{f(X)g(X) \pmod{X^n - 1} : \ f(X) \in \mathbb{F}_2[X]\}.$$

Lemma. *Every cyclic code of length n has a generator polynomial $g(X)$. Moreover, $g(X)|(X^n - 1)$.*

Proof. Ledt $g(X) \in \mathbb{F}_2[X]$ be a polynomial of least degree such that $g(X) \pmod{X^n - 1}$ belongs to $\mathcal{C}$. let $p(X)$ be another polynomial representing an element of $\mathcal{C}$.

By the division algorithm,

$$p(X) = f(X)g(X) + r(X), \ \text{ where } \deg r(X) < \deg g(X).$$

since $\mathcal{C}$ is an ideal in $\mathbb{F}_2[X]/(X^n - 1)$, it follows that $r(X) = p(X) - f(X)g(X)$ belongs to $\mathcal{C}$. But the choice of $g(X)$ gives $r(X) = 0$. Hence $g(X)$ is a generator polynomial for $\mathcal{C}$. taking $p(X) = X^n - 1$ gives $g(X)|(X^n - 1)$.

The polynomial $h(X)$ with $g(X)h(X) = X^n - 1$ is called the check polynomial for $\mathcal{C}$.

In $\mathbb{F}_2^7$ we have

$$X^7 - 1 = (X^3 + X + 1)(X^4 + X^2 + X + 1).$$

The cyclic code with generator $g(X) = X^3 + X + 1$ has check polynomial $h(X) = X^4 + X^2 + X + 1$. The parity check matrix of the code is

$$\begin{pmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{pmatrix}.$$

The columns of this matrix are the non-zero elements of $\mathbb{F}_2^3$. So, this is equivalent to Hamming's original $(7, 16, 3)$ code.

The dual of Hamming's $(7, 16, 3)$ code has the generator polynomial $X^4 + X^3 + X^2 + 1$ (the reverse of $h(X)$). Since $X^4 + X^3 + X^2 + 1 = (X + 1)g(X)$, it is a subcode of Hamming's $(7, 16, 3)$ code.

2005, II404J Coding and Cryptography

What does it mean to transmit reliably at rate r through a binary symmetric channel (BSC) with error probability p? Assuming Shannon's second coding theorem, compute the supremum of all possible reliable transmission rates of a BSC. What happens if (i) p is very small, (ii) $p = 1/2$, or (iii) $p > 1/2$?

Solution. A BSC can trasmit reliably at rate r if there is a sequence of codes $(\mathcal{C}_n)$, $n = 1, 2, \ldots$, of length $\lfloor 2^{nr} \rfloor$ such that

$$\widehat{e}(\mathcal{C}_n) = \max_{x \in \mathcal{C}_n} \mathbb{P}\big(\text{error}\big|x \text{ sent}\big) \to 0 \text{ as } n \to \infty.$$

Shannon's Second coding therorem states that the supremum sup r over all reliable rates (the operational channel capacity) is equal to $\max_\alpha I(X : Y)$, the maximum information transmitted per input symbol. Here X is a Bernoulli random variable taking values 0 and 1 with probabilities $\alpha \in [0, 1]$ and $1 - \alpha$, and Y is the output random variable for the given input X. Next, $I(X : Y)$ is the mutual entropy (mutual information):

$$I(X : Y) = H(X) - H(X|Y) = H(Y) - H(Y|X).$$

Define the entropy function

$$\begin{array}{rcl} H : [0,1] & \to & [0, \infty); \\ x & \mapsto & -x \log x - (1-x) log\, (1-x). \end{array}$$

Then $H(x)) \leq 1$; equality $X = 1/2$.

The BSC with error probability p has the capacity

$$\begin{aligned} \max_\alpha I(X : Y) &= \max_\alpha \big[H(Y) - H(Y|X)\big] \\ &= \max_\alpha \big[H(\alpha p + (1-\alpha)(1-p)) - H(p)\big] \\ &= 1 - H(p) \\ &= 1 + p \log p + (1-p) \log (1-p). \end{aligned}$$

(i) if p is small, the capacity is only slightly less than 1 (the capacity of a noiseless channel).

(ii) If $p = 1/2$, the capacity is zero (the channel is useless).

(iii) If $p > 1/2$, we may swap the labels on the output alphabet, replacing p by $1 - p$. the channel capacity is non-zero.

2005, III331 Information and Coding

Consider two discrete probability distributions $p(x)$ and $q(x)$. Defining the relative entropy

$$D(p\|q) = \sum_x p(x) \log\left(\frac{p(x)}{q(x)}\right),$$

prove the Gibbs inequality, that is, show that $D(p\|q) \geq 0$, with equality iff $p(x) = q(x)$ for all x.

Using this, show that for any positive functions $f(x)$ and $g(x)$, and for any finite set A:

$$\sum_{x\in A} f(x) \log\left(\frac{f(x)}{g(x)}\right) \geq \left(\sum_{x\in A} f(x)\right) \log\left(\frac{\sum_{x\in A} f(x)}{\sum_{x\in A} g(x)}\right).$$

Assume that for any $0 \leq p, q \leq 1$ then

$$p \log\left(\frac{p}{q}\right) + (1-p)\log\left(\frac{1-p}{1-q}\right) \geq (2\log e)(q-p)^2.$$

Show that for any probability distributions p and q:

$$D(p\|q) \geq \frac{\log e}{2}\left(\sum_x |p(x) - q(x)|\right)^2.$$

Solution. There exists a constant $c > 0$ such that

$$\log\ y \geq c\left(1 - \frac{1}{y}\right), \quad \text{with equality iff}\ \ y = 1.$$

Writing $B = \{x :\ p(x) > 0\}$,

$$\begin{aligned} D(p\|q) &= \sum_{x\in B} p(x) \log \frac{p(x)}{q(x)} \\ &\geq c \sum_{x\in B} p(x) \left(1 - \frac{p(x)}{q(x)}\right) = c\big[1 - q(B)\big] \geq 0. \end{aligned}$$

Equality holds iff $q(x) \equiv p(x)$.

Next, write

$$f(A)=\sum_{x\in A}f(x)\quad p(x)=\frac{f(x)}{f(A)}\mathbf{1}(x\in A)$$
$$g(A)=\sum_{x\in A}g(x)\quad q(x)=\frac{g(x)}{g(A)}\mathbf{1}(x\in A).$$

Then

$$\begin{aligned}\sum_{x\in A}f(x)\log\frac{f(x)}{g(x)}&=f(A)\sum_{x\in A}p(x)\log\frac{f(A)p(x)}{g(A)q(x)}\\&=f(A)\underbrace{\sum_{x\in A}p(x)\log\frac{p(x)}{q(x)}}_{\geq 0\text{ by the previous part}}+f(A)\log\frac{f(A)}{g(A)}\\&\geq f(A)\log\frac{f(A)}{g(A)}.\end{aligned}$$

Finally, consider $A=\{x;\ p(x)\leq q(x)\}$. Since

$$\sum_x\big|p(x)-q(x)\big|=2\big[q(A)-p(A)\big]=2\big[q(A^{\mathrm{c}})-p(A^{\mathrm{c}})\big],$$

then

$$\begin{aligned}D(p||q)&=\sum_{x\in A}p(x)\log\frac{p(x)}{q(x)}+\sum_{x\in A^{\mathrm{c}}}p(x)\log\frac{p(x)}{q(x)}\\&\geq p(A)\log\frac{p(A)}{q(A)}+p(A^{\mathrm{c}})\log\frac{p(A^{\mathrm{c}})}{q(A^{\mathrm{c}})}\\&\geq\big(2\log\ e\big)\big[p(A)-q(A)\big]^2\\&=\frac{\log\ e}{2}\left[\sum_x|p(x)-q(x)|\right]^2.\end{aligned}$$

2005, III332 Information and Coding

Define the conditional entropy, and show that for random variables U and V the joint entropy satisfies

$$h(U,V)=h(V|U)+h(U).$$

Given random variables $X_1, \dots X_n$, by induction or otherwise prove the chain rule

$$h(X_1, \dots X_n) = \sum_{i=1}^{n} h(X_i | X_1, \dots X_{i-1}).$$

Define the subset average over subsets of size k to be

$$h_k^{(n)} = \frac{1}{\binom{n}{k}} \sum_{S:|S|=k} \frac{h(X_S)}{k},$$

where if $S = \{s_1, \dots s_k\}$, then $h(X_S) = h(X_{s_1}, \dots X_{s_k})$. Assume that for any i, the $h(X_i|X_S) \leq h(X_i|X_T)$ when $T \subseteq S$, and $i \notin S$.

By considering terms of the form,

$$h(X_1, \dots X_n) - h(X_1, \dots X_{i-1}, X_{i+1}, \dots X_n)$$

show that $h_n^{(n)} \leq h_{n-1}^{(n)}$.

Using the fact that $h_k^{(k)} \leq h_{k-1}^{(k)}$, show that $h_k^{(n)} \leq h_{k-1}^{(n)}$, for $k = 2, \dots n$.

Solution. The definition is

$$\begin{aligned} h(V|U) &= h(Y, V) - h(U) \\ &= \sum_u \mathbb{P}(U = u) h(V|U = u), \end{aligned}$$

where $h(V|U = u)$ is the entropy of the conditional distribution:

$$h(V|U = u) = -\sum_v \mathbb{P}(V = v|U = u) \log\, \mathbb{P}(V = v|U = u).$$

The chain rule is established by induction in n:

$$\sum_{i=1}^{n} h(X_i|X_1, \dots, X_{i-1}) = h(X_1, \dots, X_n).$$

Now assume that

$$h(X_i|X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n) \leq h(X_i|X_1, \dots, X_{i-1}).$$

Next, we have

$$h(X_1, \dots, X_n) = h(X_1, \dots, X_{n-1} + h(X_n|X_1, \dots, X_{n-1}) \tag{1}$$

and in general

$$\begin{aligned} h(X_1,\dots,X_n) = h(X_1,\dots,X_{i-1},X_{i+1},\dots,X_n) \\ +h(X_i|X_1,\dots,X_{i-1},X_{i+1},\dots,X_n) \\ \leq h(X_1,\dots,X_{i-1},X_{i+1},\dots,X_n) + h(X_i|X_1,\dots,X_{i-1},X_{i+1}). \end{aligned} \qquad (i)$$

Then adding Eqns (i) from $i = 1$ to n:

$$\begin{aligned} nh(X_1,\dots,X_n) \ &\leq \sum_{i=1}^n h(X_1,\dots,X_{i-1},X_{i+1},\dots,X_n) \\ &+\sum_{i=1}^n h(X_i|X_1,\dots,X_{i-1}). \end{aligned}$$

The second sum in the RHS equals $h(X_1,\dots,X_n)$ (the chain rule). So,

$$(n-1)h(X_1,\dots,X_n) \leq \sum_{i=1}^n h(X_1,\dots,X_{i-1},X_{i+1},\dots,X_n).$$

This implies that $h_n^{(n)} \leq h_{n-1}^{(n)}$, since

$$\frac{1}{n}h(X_1,\dots,X_n) \leq \frac{1}{n}\sum_{i=1}^n \frac{h(X_1,\dots,X_{i-1},X_{i+1},\dots,X_n)}{n-1}.$$

In general, fis a subset S of size k in $\{1,\dots,n\}$. Writing $S(i)$ for $S\setminus\{i\}$, we obtain

$$\frac{1}{k}h[X(S)] \leq \frac{1}{k}\sum_{i\in S}\frac{h(X[S(i)])}{k-1},$$

by above. This yields

$$h_k^{(n)} = \frac{1}{\binom{n}{k}} \sum_{S\subset\{1,\dots,n\}:\ |S|=k} \frac{h[X(S)]}{k} \leq \frac{1}{\binom{n}{k}} \sum_{S\subset\{1,\dots,n\}:\ |S|=k}\ \sum_{i\in S} \frac{h(X[S(i)])}{k(k-1)}. \qquad (*)$$

Finally, each subset of size $k-1$, $S(i)$, appears $[n-(k-1)]$ times in the sum $(*)$. So, we can re-write it as

$$\begin{aligned} &\frac{1}{\binom{n}{k}}\left[\sum_{T\subset\{1,\dots,n\}:\ |T|=k-1}\frac{h[X(T)]}{k-1}\right]\frac{n-(k-1)}{k} \\ &\qquad \leq \frac{1}{\binom{n}{k-1}}\sum_{T\subset\{1,\dots,n\}:\ |T|=k-1}\frac{h[X(T)]}{k-1} = h_{k-1}^{(n-1)}. \end{aligned}$$

2005, III333 Information and Coding

Explain what is meant by the length, size and distance of a binary code. Define a linear code by both the generator and parity check construction.

Show that the minimum distance of a linear code equals the size of the smallest linearly dependent set of rows of the parity check matrix.

Show that the Hamming code of length $2^l - 1$ is perfect, for any l.

Solution.

2005, III334 Information and Coding

(a) Prove the Plotkin bound, that for a code with size r, length N and minimum distance δ, with $2\delta > N$, the size satisfies

$$r \leq \frac{2\delta}{2\delta - N}.$$

(b) State the MacWilliams identity, connecting the weight enumerator polynomials of a code $\mathcal{X}$ and its dual $\mathcal{X}^{\perp}$.

Give the weight enumerator polynomial of a Hamming code of length $2^l - 1$.