Name:

Points: /96pt

EXAM SOLUTIONS - ACCQ204 - JANUARY 2025

1 Warm-up

(3pt) Consider the parity check matrix

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

What are the parameters of this code?

Solution:

$n = 7, k = 4, d = 3$

(3pt) Consider the received vector $y = (0, 1, 1, 0, 1, 1, 0)$ and suppose there is at most one error. Tell whether this is a codeword and if not correct the error.

Solution:

$Hy^T = (010)$. There is an error in the second position, hence the sent codeword is (0010110).

(3pt) Using this code, explain how errors are detected and how many errors can be detected.

Solution:

If $Hy^T \neq 0$ an error is detected, and if $Hy^T = 0$ we declare no error. $d - 1 = 2$ errors can be detected.

2 MDS codes

(9pt) Prove that a linear code $\mathcal{C}$ is MDS if and only if its dual code $\mathcal{C}^\perp$ is MDS.

Solution:

We prove that if $\mathcal{C}$ is an MDS code, then its dual code $\mathcal{C}^\perp$ is also MDS. The converse is then immediate by swapping the roles of $\mathcal{C}$ and $\mathcal{C}^\perp$ since $\mathcal{C}^{\perp\perp} = \mathcal{C}$.

Suppose $\mathcal{C}$ is an MDS code with parameters $[n, k, d]$ and generator and parity check matrices G and H, respectively. Its dual code $\mathcal{C}^\perp$ has parameters $[n, k^\perp = n - k, d^\perp]$ and we need to show that $d^\perp = n - k^\perp + 1 = k + 1$. Since the Singleton bound already gives us $d^\perp \leq k + 1$, it suffices to show that $d^\perp \geq k + 1$.

To prove this we rely on the property that G is the parity matrix of $\mathcal{C}^\perp$, i.e., $c \in \mathcal{C}^\perp$ if and only if $Gc^T = 0$. Hence, to any $c \in \mathcal{C}^\perp$ corresponds $wt(c)$ linearly dependent columns of G. Since G has rank k, we know that $wt(c)$ should be greater or equal to $k + 1$. Therefore $d^\perp = \min_{c \neq 0, c \in \mathcal{C}^\perp} wt(c) \geq k + 1$.

3 Group testing (this is new and hopefully interesting)

Suppose $x = (x_1, x_2, \ldots, x_N) \in \{0,1\}^N$ represents N individuals. If the i-th individual is infected by Covid, $x_i = 1$, and if not $x_i = 0$. The number of infected individuals is known to be no more than d, that is

$$wt(x) \leq d$$

for some d between 1 and N (as usual, $wt(x)$ denotes the number of ones in x).

To identify the infected individuals, we perform tests described as follows. A test is speficied by a subset $S \subseteq [N]$.[1] The Answer to a test S, denoted $A(S)$, is defined as

$$A(S) = \begin{cases} 1 & \text{if } \vee_{i \in S} x_i = 1 \\ 0 & \text{otherwise,} \end{cases}$$

where "$\vee$" denotes the logical OR binary operation.[2]

The goal is to successfully determine x with a minimum number of tests knowing only N and d.

There are two types of test procedures, adaptive and non-adaptive. Non-adaptive procedures describe the situation where all tests are set before even we collect the result of the first test. Adaptive procedures do not have this restriction and tests may depend on the results of previous tests.

We define $t^a(d, N)$ as the minimum number of tests needed to identify any x when adaptive procedures are allowed, and $t^{na}(d, N)$ as the minimum number of tests with non-adaptive procedures.

1. (*6pt*) Show that
$$1 \leq t^a(d, N) \leq t^{na}(d, N) \leq N.$$

   Solution:

   If $t^a(d, N) = 0$ then this means that without any test we can exactly infer about x. This would be possible only if $d = 0$ in which case there is only one possible x, known in advance, which we don't need to test. If $d \geq 1$, the number of possible x's is greater than one and we need at least one test—for otherwise the decision is independent of x, and is therefore unreliable.

   The second inequality is obvious since non-adaptive procedures are a subset of the adaptive procedures.

   The third inequality holds since we can always test all individuals non-adaptively to retrieve x.

2. (*3pt*) For a given (adaptive or nonadaptive) procedure, let $r(x)$ denote the binary vector that gives the results obtained from x (the i-th entry of $r(x)$ refers to the i-th test outcome). Show that a successful procedure cannot result in $r(x) = r(y)$ for different but valid x, y.

   Solution:

   If $r(x) = r(y)$ then we cannot distinguish x from y.

3. (*4pt*) Deduce that a successful (adaptive or non-adaptive) procedure performs a number of tests at least equal to $\log_2(|Ball(N, d)|)$, where $Ball(N, d)$ refers to the Hamming ball of radius d in dimension N.

   Solution:

   From the previous question,
   $$|\{r(x) : wt(x) \leq d\}| \geq |Ball(N, d)|.$$
   The result then follows by noting that a procedure with t tests produce at most 2^t results.

[1] $[N] := \{1, 2, \ldots, N\}$

[2] For $a, b \in \{0, 1\}$, $a \vee b$ is equal to 0 if $a = b = 0$ and is equal to 1 otherwise.

4. (*3pt*) Using that $\binom{a}{b} \geq (\frac{a}{b})^b$ deduce that

$$t^a(d, N) \geq d \log_2 \left(\frac{N}{d} \right)$$

Solution:

We have

$$|Ball(N, d)| = \sum_{i=0}^{d} \binom{N}{i} \geq \binom{N}{d} \geq (N/d)^d$$

and the result follows from 3.

5. (*4pt*) Show that without any assumption on d, with $O(\log_2 N)$ adaptive tests it is possible to identify one i such that $x_i = 1$, or to identify the situation where nobody is infected (Hint: use binary search).

Solution:

Assuming $wt(x) \geq 1$, a binary search where at each step we split the individuals in two sets of (roughly) equal size identifies an i such that $x_i = 1$ in $O(\log_2 N)$ steps. To accommodate for the case where $wt(x) = 0$, it suffices to check each of the two-halfs of the coordinates in the first step. The number of tests is therefore $O(\log_2 N)$.

6. (*5pt*) Deduce that one can compute any x with $O(wt(x) \cdot \log_2 N)$ adaptive tests.

Solution:

We proceed according to the search procedure of 5. and when we identify an index i such that $x_i = 1$, we record this index, set $x_i = 0$, and repeat the procedure.

7. (*5pt*) Argue that any scheme that computes x with $O(wt(x) \cdot \log_2 N)$ tests can be used to compute x with $O(d \cdot \log_2(N/d))$ adaptive tests with $wt(x) \leq d$ (Hint: Divide the components of x into d groups. How many tests are needed per group as a function of the weight of the group?).

Solution:

Divide the components into d groups of size N/d. According to 6. we can compute the j-th group with $O(w_j \log_2(N/d))$ tests, where w_j denotes the weight of the j-th group of components. Summing over j and using the fact that $\sum_j w_j \leq d$ gives the desired result.

8. (*3pt*) Deduce that $t^a(d, N) = \Theta(d \log_2(N/d))$.[3]

Solution:

Directly follows from 4. and 7.

9. (*4pt*) Argue that a set of t nonadaptive tests $S_1, S_2, \ldots, S_t$ on x can be written as

$$M \odot x^T$$

where

[3]Recall that $f(n) = \Theta(n)$ if there exist two strictly positive constants c_1, c_2 such that

$$c_1 \cdot n \leq f(n) \leq c_2 \cdot n$$

for n large enough.

- $\rightsquigarrow$ M is a $t \times N$ binary matrix where the (i,j)-coordinate is equal to one if $j \in S_i$ and zero otherwise—for instance for $N = 3$ if $S_1 = \{1, 3\}$ then the first row of M is equal to (101).
- $\rightsquigarrow$ $\odot$ represents the standard matrix multiplications over GF(2), except that for the sum we replace the logical exclusive-OR with OR.

Solution:

We have $A(S_j) = 1$ if and only if $\sum_{i \in S_j} \mathbb{1}\{x_i = 1\} \geq 1$ if and only if $\vee_{i \in S_j} x_i = 1$

10. (*6pt*) Show that
$$t(1, N) \leq \lceil \log(N + 1) \rceil + 1$$
(Hint: Consider $x = e_i$ (e_i here represents the all-zero vector with only a one at position i). Deduce a property that the columns of a measurement matrix M should have. Does this remind you about the generator/parity matrix of a well-known code? Distinguish the case where N satisfies $N = 2^r - 1$ for some integer r, from the case where N does not have this property.)

Solution:

For any M we have that $M \odot e_i^T$ gives the i-th column of M. By 2. we deduce that all the columns of M must be distinct if we want to retrieve any e_i. Conversely, if M is such that all its columns are distinct then such an M allows to recover any e_i and also allows to recover the all-zero sequence x when $M \odot x^T = 0$.

There are lots of choices for such a matrix. One corresponds to the parity matrix of a Hamming code. Recall that the order-r matrix has $2^r - 1$ columns, with the i-th column gives the binary expansion of the i. If $N = 2^r - 1$ for some integer $r \geq 1$ then we immediately get that $t(1, N) \leq \log_2(N + 1)$. If not, let r^* be the smallest integer r such that $N \leq 2^r - 1$, that is $r^* = \lceil \log_2(N+1) \rceil$. And let $N' = 2^{r^*} - 1$. By considering the order-r^* Hamming parity matrix from which we deleted the last $N' - N$ columns gives the desired result.

4 Concatenated Codes (a small break)

(2pt) Give a non-trivial lower bound on the number of errors that can be corrected with a concatenated code whose outer code and inner code have parameters $[n_1 = 7, k_1 = 5, d_1 = 3]$ and $[n_2 = 7, k_2 = 4, d_2 = 3]$, respectively.

Solution:

The minimum distance of the concatenated code is $d = 3 \cdot 3 = 9$. We saw in class that $\lfloor d/4 \rfloor = 2$ errors can be corrected using the "simple" decoding rule (decode first the inner code then the outer code), but we know that using more complex decoding rules (nearest neighbor decoding) it is possible to correct $\lfloor d/2 \rfloor = 4$ errors.

5 From transmission to compression (this should also be interesting)

We consider communication over a BSC(p) channel (the channel's probability law is $Q(y|x) = P_Z(y - x)$, $x, y \in \{0, 1\}$, where P_Z is the Bernoulli(p) distribution) with a linear code $\mathcal{C}$ with parity matrix H.

(6pt) Show that, given a received vector $y \in \{0, 1\}^n$, the maximum likelihood (ML) decoder[4]
$$\text{ML}(y) := \arg\max_{c \in \mathcal{C}} Q^{(n)}(y|c)$$
is equal to
$$y - g_{\text{synd.}}(Hy^T)$$
where $g_{\text{synd}}(\cdot)$ is a function that depends only on the syndrome Hy^T. Your argument should explicit $g_{\text{synd.}}(Hy^T)$. (Hint: $Q(y|c) = P_Z(y - c)$, and observe that $\{c : Hc^T = 0\} = \{y - z : Hz^T = Hy^T\}$)

[4] $Q^{(n)}(y|c) := \prod_{i=1}^n Q(y_i|c_i) = \prod_{i=1}^n P_Z(y_i - c_i)$

Solution:

We have

$$\begin{aligned}
\arg\max_{c\in\mathcal{C}} Q^{(n)}(y|c) &= \arg \max_{c:Hc^T=0} Q^{(n)}(y|c) \\
&= \arg \max_{c:Hc^T=0} P_Z^{(n)}(y-c) \\
&= \arg \max_{y-z:Hz^T=Hy^T} P_Z^{(n)}(z) \\
&= y - \arg \max_{z:H(z)=H(y)} P_Z^{(n)}(z) \\
&= y - \arg \max_{z:H(z)=H(y)} p^{wt(z)}(1-p)^{n-wt(z)} \\
&= y - \underbrace{\arg \max_{z:H(z)=H(y)} \left(\frac{p}{1-p}\right)^{wt(z)}}_{=g_{\text{synd.}}(Hy^T)} \\
&= y - \underbrace{\arg \min_{z:H(z)=H(y)} wt(z)}_{=g_{\text{synd.}}(Hy^T)}
\end{aligned}$$

where the last equality holds if $0 \le p \le 1/2$—replace the min by max if $1/2 < p \le 1$.

(4pt) Denote by $\bar{c}$ the sent codeword. Argue that the ML decoder outputs $\bar{c}$ if and only if

$$g_{\text{synd.}}(Hy^T) = g_{\text{synd.}}(Hz^T) = z$$

where $z = y - \bar{c}$ is the noise binary vector caused by the channel.

Solution:

From the previous question, the ML decoder outputs $\bar{c}$ if and only if $g_{\text{synd.}}(Hy^T) = z$. Since $Hy^T = H\bar{c}^T + Hz^T = Hz^T$ the result follows.

(7pt) Consider the following theorem:

Fix $\varepsilon > 0$ and pick $R < C(p)$. For n large enough there exists a linear code of rate $k/n \ge R$ that, combined with ML decoding, achieves error probability $P_e \le \varepsilon$.

Admiting this theorem, prove the following theorem.

Fix $\varepsilon > 0$, $R > H_b(p)$ and let s be a length n i.i.d. Bernoulli(p) source of information. For n large enough there exists a linear compressor

$$comp : \{0,1\}^n \to \{0,1\}^{nR}$$

which takes s, compressed it to nR bits, and allows correct decompression with probability $\ge 1-\varepsilon$.

Hint: How can we interpret $g_{\text{synd.}}(Hz^T) = z$?

Solution:

Let s be a length n Bernoulli(p) source of information, and consider the previous question (with s playing the role of z). Let us interpret Hs^T as the compression of the source s into $n-k = n(1-R')$ bits (with $R' := k/n$). The compression

rate is therefore $1 - R' := R$. Further, let us intepret $g_{\text{synd.}}(Hs^T)$ as the decompression of s. Now fix $R' < C(p)$ and $\varepsilon > 0$. From the previous theorem, we deduce that for n large enough

$$P(g_{\text{synd.}}(Hs^T) \neq s) \leq \varepsilon.$$

We then deduce that for any compression rate R that satisfies

$$R > 1 - C(p) = 1 - (1 - H_b(p)) = H_b(p)$$

source s can be compressed to nR bits and be correctly decoded with probability $\geq 1 - \varepsilon$.

6 Zero-error capacity and the pentagon channel (... the last interesting problem)

We are interested here in the zero-error capacity of a noisy channel. By zero-error capacity we mean the largest communication rate for which we can communicate without error.

(6pt) For a given channel $Q(y|x)$ with finite input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$, consider the following adjacency graph. The vertices are the elements of $\mathcal{X}$ and two vertices x, x' are connected with an edge if there exists $y \in \mathcal{Y}$ such that $Q(y|x)Q(y|x') > 0$. Show that if a channel $Q(y|x)$ has a fully connected adjacency graph, then it is impossible to communicate at a positive rate with zero probability of error. (Hint: suppose there are only two possible messages to be sent. Show that whatever code we use to encode these two messages the error probability—under optimal decoding—is always bounded away from zero.)

Solution:

To show the claim we show that we cannot achieve zero error probability not even if we have only 2 messages. Suppose we have two equally likely messages m_1 and m_2 that we encode into two length-n codewords $c(m_1)$ and $c(m_2)$. For any pair $(c_i(m_1), c_i(m_2))$ there always exists a "confusing" output symbol y_i^* that satisfies

$$Q(y_i^*|c_i(m_1))Q(y_i^*|c_i(m_2)) > 0.$$

Hence, when the receiver observes as a channel output $Y = y^*$ (which happens with non-zero probability), it makes an error with non-zero probability since this output can be generated with positive probability by both codewords.

(4pt) Consider Shannon's pentagon channel whose adjacency graph is given in Figure 1. Show that the zero-error capacity is at least 1 bit per channel use (Hint: consider a code of length one).

Solution:

Assign message m_1 to codeword $c(m_1) = 1$ and message m_2 to codeword $c(m_2) = 3$ (any non-adjacent pair of symbols can be picked).

(6pt) Through the design of 5 codewords of length 2, show that the zero-error capacity is at least $0.5 \log_2(5) > 1$ (!). (Hint: complement the list of codewords $\{11, 24, 32\}$)[5]

Solution:

We need to find 5 codewords of length 2 that can be distinguished at least in one component. An example would be $\{11, 24, 32, 45, 53\}$.

[5]This was observed by Shannon in 1956 and it is only 23 years after that Lovasz was able to prove that this bound cannot be improved, and is therefore the zero-error capacity of the so-called pentagon channel.

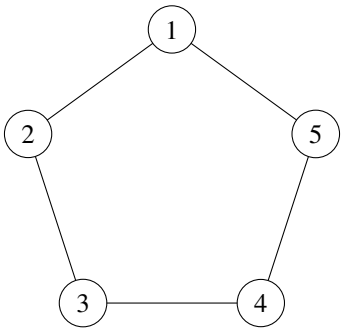

Figure 1: The adjacency graph of Shannon's pentagon channel