USS: Introduction to mathematical cryptography
Code-based cryptography

This is a self-study homework if you wish to learn more about code-based cryptography. The notation and vocabulary is taken from Trappe and Washington's book *Introduction to Cryptography with Coding Theory*, which is available at the PCMI library. These problems cover Sections 16.1, 16.2, 16.4, 16.5, and 16.10 of the first edition of the book, or Sections 18.1, 18.2, 18.4, 18.5, and 18.10 of the second edition.

1. Let C be a code over a field and d its Hamming distance. Prove that d is a metric. In other words, prove that if c_1, c_2 and c_3 are code words, then
   (a) $d(c_1, c_2) \geq 0$, with $d(c_1, c_2) = 0$ if and only if $c_1 = c_2$;
   (b) $d(c_1, c_2) = d(c_2, c_1)$; and
   (c) $d(c_1, c_2) \leq d(c_1, c_3) + d(c_3, c_2)$.
2. Let C be a code with minimum distance $d(C)$. Show that
   (a) C can detect up to s errors if $d(C) \geq s + 1$; and
   (b) C can correct up to t errors if $d(C) \geq 2t + 1$.
3. (adapted from Trappe and Washington Section 18.12, problem 5)
   Let $C = \{(0, 0, 1), (1, 1, 1), (1, 0, 0), (0, 1, 0)\}$ be a code over $\mathbb{F}_2$.
   (a) Show that C is not a linear code.
   (b) Compute $d(C)$, the minimum distance of C.
4. Let C be a **linear** code. Prove that $d(C)$, the minimum distance of C, is equal to the smallest Hamming weight of nonzero code words:
$$d(C) = \min\{\mathrm{wt}(c) : 0 \neq c \in C\}.$$
5. Consider the linear binary code C given by the generating matrix
$$G = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix}.$$
   (a) Please enumerate all of the elements of C.
   (b) Using the notation from the book, what are n and k for this code?
   (c) Give a parity check matrix H for this code.
   (d) What is $d(C)$, the minimum distance?
   (e) How many errors can C detect? How many errors can C correct? You can use the results from problem 2 to answer this question.

6. Consider the n-repetition code C_n over $\mathbb{F}_2$, which encodes the 1-bit message $m = 0$ with the n-bit codeword $0000\ldots0$ and the 1-bit message $m = 1$ with the n-bit codeword $1111\ldots1$.

   (a) Prove that C_n is a linear code.

   (b) Using the notation from the book, what are n and k for this code?

   (c) What is $d(C_n)$, the minimal distance?

   (d) How many errors can C_n correct?

7. Consider the linear binary code C given by the generating matrix

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

   (This is not important to this question, but it is an example of a Goppa code.)

   (a) Compute a parity check matrix H for this code.

   (b) For each of the following vectors, compute the syndrome. Which vectors below are codewords?

      i. $v_1 = [0, 1, 0, 1, 1, 0, 1, 1, 1, 0, 0, 0]$
      ii. $v_2 = [1, 1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0]$
      iii. $v_3 = [0, 1, 1, 1, 1, 0, 0, 1, 1, 0, 0, 0]$
      iv. $v_4 = [0, 1, 0, 0, 1, 0, 1, 1, 1, 0, 0, 0]$

8. Please read Example 4 of Section 18.1 of Trappe and Washington's book, on the Hamming $[7, 4]$ code. The "mysterious" decoding algorithm is simply this: Given a received message v, the vector vH^T will always be a row of H^T, or zero. Since H is a parity matrix, if vH^T is zero then v is a codeword and does not need to be corrected (it is already decoded). Otherwise, if vH^T is the ith row of H^T, then v is a codeword c but with an error in the ith entry. To correct v to a codeword it suffices then to flip the ith entry of v.

   For this problem, please use the Hamming $[7, 4]$ code to decode the following received messages:

   (a) $v_1 = [1, 1, 1, 0, 1, 0, 1]$

   (b) $v_2 = [1, 0, 1, 0, 0, 0, 1]$

   (c) $v_3 = [0, 0, 1, 1, 1, 0, 0]$

   (d) $v_4 = [1, 0, 1, 0, 0, 1, 1]$

9. Suppose that Person B publishes the "scrambled" generating matrix

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

based on a code that can correct $t = 2$ errors. Encrypt the following messages to send to Person B. Don't forget to introduce a random error!

(a) $m_1 = [1, 0, 1, 1]$

(b) $m_2 = [0, 0, 1, 1]$

10. Now suppose that you are Person B, and you have set up a McEliece cryptosystem based on the following data: You are using the Hamming $[7, 4]$ code from above, with generating matrix

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

You have chosen the invertible matrix

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

and the permutation matrix

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

(a) Given that this Hamming code can correct 1 error, what would be your public key (the information that you publish)?

(b) Decrypt the following ciphertexts:

i. $c_1 = [0, 0, 1, 0, 0, 1, 0]$

ii. $c_2 = [1, 0, 1, 0, 0, 1, 1]$

iii. $c_3 = [0, 0, 1, 1, 1, 0, 1]$

iv. $c_4 = [0, 1, 1, 1, 1, 0, 0]$