Coding and cryptography

Assignment 1 (2022-2023)

Solutions.

1. (a) Clearly, if v is in C then v is in $B_v(d-1)$ and in C, hence in $B_v(d-1) \cap C$. Suppose $v' \neq v$ is another word in $B_v(d-1) \cap C$. Then v' is in the code C, and $d(v,v') \leq d-1$. But this is impossible because $d \leq d(C) = \min\{d(c,c') \text{ with } c \neq c' \text{ both in } C\}$.

   (b) We look at $S = \cup_{v \in C} B_v(d-1) \subseteq K^n$. Each $B_w(d-1)$ has the same number of elements, namely $m = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{d-1}$. So in total $|S| \leq |C| \cdot m$. By assumption then, $|S| < 2^n = |K^n|$, so S is not the whole of K^n. Therefore there is a word w in K^n that is not in S. Note that $d(v,w) \geq d$ for all v in C, otherwise w would be in some $B_v(d-1)$.

   Now let $C' = C \cup \{w\}$. Then $d(C') \geq d$. Namely, the distinct (unordered) pairs in C' are of the form $\{v,v'\}$ with $v \neq v'$ both in C, or $\{w,v\}$ with v in C. But $d(v,v') \geq d(C) \geq d$ for the former, and $d(w,v) \geq d$ for the latter. So $d(C') \geq d$.

   (c) We can start with the code $\{00\ldots0, 11\ldots1\}$, which has distance $n \geq d$. Assume we have a code D with $d(D) \geq d$ and $|D| \cdot m < 2^n$ where m is as before. Then we saw in the previous part that we can enlarge D to a code D' by adding one new codeword and still have $d(D') \geq d$. We can keep doing this until we hit a code C for which $d(C) \geq d$ but $|C| \cdot m < 2^n$ fails, i.e., $|C| \cdot m \geq 2^n$.

2. (a) Note that the Singleton bound means that $5-1 \leq 12-6$. This is correct, so such a code *may* exist according to this bound (i.e., it does not rule out its existence, but does not prove its existence either). However, the Hamming bound gives, for a code C with length 12 and distance 5, that $|C| \leq \frac{2^{12}}{\binom{12}{0}+\binom{12}{1}+\binom{12}{2}} = \frac{2^{12}}{1+12+66} = \frac{2^{12}}{79} = 51.8\ldots$, so there is no such code with $2^6 = 64$ elements.

   (Because we are looking for a linear code, we could round down $\frac{2^{12}}{79}$ to a power of 2 by increasing 79 to the next power of 2, namely $2^7 = 128$, to find that $|C| \leq 2^5 = 32$ for a *linear* code.)

   (b) In this case the Singleton bound reads $3-1 \leq 63-57$. This is correct, so such a code *may* exist. The Hamming bound now reads $|C| \leq \frac{2^{63}}{\binom{63}{0}+\binom{63}{1}} = \frac{2^{63}}{1+63} = \frac{2^{63}}{2^6} = 2^{57}$, so again the code might exist (and it would have to be perfect).

   Note that the Gilbert-Varshamov bound here gives that there is a linear code $|C|$ with $n = 63$ and $d = 3$ with $|C| \geq \frac{2^{62}}{\binom{62}{0}+\binom{62}{1}} = \frac{2^{62}}{63} > \frac{2^{62}}{2^6} = 2^{56}$, so with dimension at least 57 (and by the Hamming bound above the dimension would be exactly 57). But this does not produce an actual such code without going trough the construction of the parity check matrix in the proof of that bound. Given the size of the matrix, this would be quite painful.

   The simplest solution is to notice that the Hamming codes (which are linear) for $r \geq 2$ have parameters $(2^r-1, 2^r-r-1, 3)$, and that we get the correct parameters for $r = 6$. So we can use one of those, and get a 63×6 parity check matrix by using all the non-zero elements of K^6 as its rows.

3. (a) It is clear that $\mathrm{wt}(r_1) = 6$ as all rows of X have exactly six 1s. Note that the positions of the 0s in two distinct rows of X are disjoint. So if we add them we compute $8-4 = 4$ times $1+1 = 0$ and 4 times $1+0 = 1$ or $0+1 = 1$. So there will be four 1s and four 0s in the result, and the sum has weight 4. If we do this for three distinct rows than in six positions we compute $1+1+0$, $1+0+1$ or $0+1+1$, giving 0, and in the remaining two positions $1+1+1 = 1$. So the sum will have weight 2.

   (b) For this we use Theorem 2.9.1.

   The first three rows of X and the first two rows of I_8 add up to the zero vector, so $d(C) \leq 5$.

   We still have check that $d(C) \geq 5$ in order to obtain $d(C) = 5$, which we do as follows:

- $d(C) \geq 1$ because H contains no zero rows (i.e., every single row of H is linearly independent);
- $d(C) \geq 2$ because all rows of H are different (i.e., every two rows of H are now linearly independent);
- $d(C) \geq 3$: we have to check that no three rows of H add up to the zero vector. If we use three rows of X and zero rows of I_8 then this follows from (a). If we use two rows of X and one row of I_8 then this follows from (a) as the weight of the sum of the first two is 4 and the row of I_8 has weight 1, so the three rows cannot add up to the zero vector. If we use one row of X and two rows of I_8 then this is not possible because the row of X has weight 6 and the sum of the two rows of I_8 has weight 2. No three rows of I_8 are linearly dependent (in fact, no subset of rows of I_8 is linearly dependent);
- $d(C) \geq 4$. We proceed as in the previous case, considering a sum of i rows of X and $4-i$ rows of I_8 that gives the zero vector. For $i = 4$ there is only one sum, which is 11111111. For $i = 3$ we use (a): the three rows of X sum up to a vector of weight 2, and this is not a row of I_8. For $i = 1$ the row of X has weight 4 and the sum of three rows in I_8 has weight 3, so they are not equal. And for $i = 0$ we have four rows of I_8, which are always linearly independent.

(c) Note that H really is a parity check matrix, i.e., its columns are linearly independent (because of the 8×8 identity matrix in it). So C has dimension $k = 12 - 8 = 4$. The Singleton bound now reads $5 - 1 \leq 12 - 4$, which is correct. This is not an equality, so C is not an MDS-code.

(d)
- 110000000011 has syndrome $00001111 + 00000011 = 00001100$. This is non-zero, the given word is not in C. The syndrome is the sum of the 9th and 10th rows of H, so we correct the word to a codeword by changing the 9th and 10th positions i.e., we decode it to 110000001111. Because $d(C) = 5$, this is the unique codeword with distance at most 2 to the given (received) word.
- 101010101010 has syndrome $00110011 + 10101010 = 10011001$ so the given word is not in C. The syndrome is not a row of H (which has weight 6 or 1). It is also not the sum of two rows of I_8 (which has weight 2), or of a row of X and a row of I_8 (which has weight 5 or 7), or of two rows of X (which has weight 4 by (a)). So this word is not at distance at most two from a codeword, and we do not decode it under the given instructions.
- 111111111111 has syndrome 00000000 (note that the rows of X and the rows of I_8 both add up to 11111111), so it is in C and decodes to itself.