Telecom ParisTech
Teacher: Aslan Tchamkerten

ACCQ204, Coding Theory
November 2017

ASSIGNMENT 1 - SOLUTIONS

Exercise 1 (RAID, distributed storage)**.** Redundant Arrays of Independent Disks consist of a set of disks such that any subset of s disks can be disabled and the others are still able to reconstruct any requested file (the system can tell which disks are disabled).The rate of a RAID system corresponds to the rate at which data is stored.

1. Design a RAID system for 7 disks and $s = 2$. To do this you may want to consider the $(7, 4)$ Hamming code.

2. What happens if we use this code and try correct 3 erasures?

Solution. 1. Encode each successive 4 bits of data into the corresponding seven bit codeword of a $(7, 4)$ Hamming code, and write each bit of the codeword on a different disk. If s disks are disabled, this means that we get to observe codewords with erasures at some specific s positions. To reconstruct the original codeword x from a corrupted (i.e., x with 2 erased positions) vector y, observe that because the minimum distance is 3, all except codeword x will differ from y in at least one of the non-erased positions. Therefore only x will be consistent with y, and erasure decoding will be error-free.

2. Things might go wrong. Consider two codewords whose distance is 3 and suppose one of them is stored. If the 3 erased positions correspond to the positions where the 2 codewords differ then it won't be possible to (fully) recover the original data.

□

Exercise 2. Let C be a code with minimum distance d. Prove that C can correct any pattern of e_1 errors and e_2 erasures provided that $2e_1 + e_2 + 1 \leq d$. (Hint: given an erasure pattern, consider the code obtained by the deleting the erasure positions.)

Solution. Consider a pattern of $e_2 \leq d - 1$ erasures, and the code obtained by deleting the erasure positions of the code. The resulting code C' has a minimum distance at least $d - e_2$ and thus can be corrected as long as $2e_1 \leq (d - e_2) - 1$. Once C' has been error corrected, C can be erasure corrected since $e_2 \leq d - 1$ (see Ex.1).

□

Exercise 3 (Perfect codes)**.** A code is a perfect $t-$error correcting code if the set of $t-$spheres centered on the codewords fill the Hamming space $\{0, 1\}^n$ without overlapping. Here we will show that such codes do not, in general, achieve the capacity of the BSC.

Consider a set of three codewords of length n. Let un denote the number of positions where the first codeword differs from both the second and the third codewords, let vn denote the number of positions where the second codeword differs from both the first and the third codewords, let wn

denote the number of positions where the third codeword differs from both the first and the second codewords, and finally let zn denote the number of positions where the three codewords agree.

1. Argue that we can assume, without loss of generality, that one of them is the all-zero codeword.
2. Assuming that the code is $f \cdot n$-error correcting, give necessary conditions on u, v, w.
3. Show that for a certain range of f we must have $u + v + w > 1$ which is impossible.
4. Conclude that, for a certain range of f, perfect codes do not exist.
5. Reconcile this result with the Shannon's result which says that 'with high probability it is possibe to correct $f \cdot n$ errors with exponentially many codewords'.

Solution.

1. From a given code, pick any codeword, and XOR it with each codeword in the code. The new code is simply a translated version of the original code, with one codeword being the all-zero codeword. As such it achieves the exact same performance as the original code, both in terms of rate and error probability.
2. $$u + v > 2f \qquad u + w > 2f \qquad u + w > 2f$$
3. Summing the three inequalities and dividing by two we get $u + v + w > 3f$. So if $f > 1/3$ the sum exceeds 1, a contradiction.
4. See 3.
5. Hamming error correction requires to be able to correct error patterns exactly. This gives rise to a worst case scenario where the minimium distance plays the central role. In fact, beyond half the minimum distance we cannot guarantee that all error patterns will be corrected. By contrast, Shannon requires to correct error patterns with high probability which, in turn, allows to correct many error patterns above the minimum distance. More specifically, since Shannon's theorem says that (for the BSC) capacity is $1 - H(f)$ (where f is the crossover probability of the channel), it is indeed possible to correct with arbitrarily high probability up to a fraction fn of the codeword.

□

Exercise 4 ($A(n, d, w), A(n, d)$)**.** For any integers n, d, w with $d \leq 2w \leq n$, let $A(n, d, w)$ be the largest possible size of a set of binary vectors of length n and weight w whose minimum distance is at least d, and let $A(n, d)$ be the largest possible size of a set of length n binary vectors whose minimum distance is at least d. Prove that

$$A(n, d) \leq \sum_{w=0}^{n} A(n, d, w)$$

Solution. Consider a code which achieves $A(n, d)$. This code is the disjoint union of classes of codewords of different weight w. Since each class has a minimum distance at least equal to d it has at most $A(n, d, w)$ elements. □