Algebra II 189-251B **Instructor: Dr. E. Goren.**

Assignment 1 – Corrections

6. **Rudiments of Coding Theory I.**[1] In this exercise $\mathbb{F} = \mathbb{Z}/2\mathbb{Z}$.[2] Let $V = \mathbb{F}^n$. Thus, an element of V is just an n-tuple $(x_1, \dots, x_n)$ where each coordinate x_i is either 0 or 1. Define a distance function $d(x, y)$ on V as follows. If x and y are vectors and $z = x - y = (z_1, \dots, z_n)$ then

$$d(x, y) = z_1 + \cdots + z_n.$$

For example: if $n = 6$, $x = (1, 1, 0, 0, 1, 0)$ and $y = (1, 1, 1, 0, 0, 0)$ then $z = (0, 0, 1, 0, 1, 0)$ and $d(x, y) = 0+0+1+0+1+0 = 2$ (it is considered as an integer and not as an element of $\mathbb{F}$). Namely, *the distance between x and y is precisely the number of coordinates in which x and y differ.* This distance is called the *Hamming distance*[3]. Prove that:

1. $d(x, y) \geq 0$ with equality holding if and only if $x = y$;
2. $d(x, y) \leq d(x, z) + d(z, y)$ for every $x, y, z \in V$ (the *Triangle Inequality*).

We also call $d(x, 0)$, where 0 is the zero vector, the *Hamming weight of x*, it is equal to the number of 1's in x.

Coding theory has nothing to do with concealing information. It is rather the science of transmitting information over noisy, or defective, channels. This could either your telephone line when you connect to the internet, or this poor rover transmitting to NASA from Mars (or MARS). The purpose in each case is to find some means to ensure that the receiving side either receives the correct information or is able to reconstruct it from the information it does receive. Assume that the original message, that consists of "words" (or chunks of information) of some fixed length, is written as a string of zeros and ones. For example we might be interested in sending the following information

00101 11001 00010...

(This might mean "all is well, tell mum I'll be back for supper"). To do that we have a "code". A code is like a dictionary that substitutes for each original word a longer word and it is that longer word that is being transmitted. The receiving side has the same code (or dictionary) and has no problem translating longer words to the original words. The logic is, in a sense, that longer words are more "robust" and can be recognized even if distorted.

Definition 0.1. An (n, k) linear code C is a subspace of $\mathbb{F}^n$ having 2^k elements.

Show that the minimal distance between two distinct elements of a code C is the minimal weight of a non-zero vector. Namely: Note Correction

$$\min\{d(x, y) : x \neq y,\ x \in C, y \in C\} = \min\{d(x, 0) : x \in C, x \neq 0\}.$$

The procedure of coding continues as follows. The transmitting side is sending words that belong to an (n, k) code C that is known to the receiving side and sends only such words. The receiving side receives vectors of $\mathbb{F}^n$. Each such vector may be in C (i.e., if no errors occurred, or if errors did occur but the erroneous vector happens to belong to C as well). In case it isn't, the receiving side looks for the word in C that is closest to the vector that was received.

We say that a linear code *corrects t errors* if every code word that is transmitted with t or less errors has a unique element of C which is nearest to it and is correctly decoded by assigning to it that nearest element. We say that a linear code *detects t errors* if every received word with at least Note Correction

[1] I believe that the text provided here and the exercises are completely self contained and the best way to understand this material is reading the text carefully and attempting to solve the questions. If, however, you find that you cannot follow my explanation a reference is Hungerford's book of the first semester.

[2] Note carefully, in this semester we shall use the notation $\mathbb{Z}/n\mathbb{Z}$ for the ring of integers $\pmod n$ that we denoted in the first semester by $\mathbb{Z}_n$. The notation $\mathbb{Z}/n\mathbb{Z}$ *is* the customary one and you should use only this notation from now on. The reason for sticking to the "bad" notation in the first semester was that this is the notation used in the text book (Hungerford) of that course.

[3] After the scientist Richard W. Hamming.

one, but no more that t errors, is not a code word. Prove the following Theorem

Theorem 0.2. *A linear code C corrects t errors if and only if the Hamming distance of every two elements of C is at least $2t+1$.*

A linear code C detects t errors if and only if the Hamming distance between any two elements of C is at least $t+1$.

Find the linear $(7,4)$ code which is the subspace of $\mathbb{F}^7$ spanned by the following vectors:

$$(1,0,0,0,0,1,1)$$
$$(0,1,0,0,1,0,1)$$
$$(0,0,1,0,1,1,0)$$
$$(0,0,0,1,1,1,1)$$

Prove that this code corrects single errors and detects double errors.