PROJECT 2 – ERROR CORRECTING CODES

Modular Arithmetic. Let p be a prime number, such as 2, 3, 5, 7, 11, ... We write $\mathbb{F}_p$ for integers with operations $+$, $-$ and $\times$ considered only up to remainder when we divide by p. For example, here are addition and multiplication tables for $\mathbb{F}_5$.

$a+b$	$a=0$	1	2	3	4
$b=0$	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

$a-b$	$a=0$	1	2	3	4
$b=0$	0	1	2	3	4
1	4	0	1	2	3
2	3	4	0	1	2
3	2	3	4	0	1
4	1	2	3	4	0

$a\times b$	$a=0$	1	2	3	4
$b=0$	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

.

To be clear, $\mathbb{F}_5$ refers to a set with 5 elements in it, $\{0, 1, 2, 3, 4\}$, which add and multiply by the table above. For example, we have $3 \times 4 = 2$ because the remainder when we divide 3×4 by 5 is 2.

The great thing when p is prime is that we can also divide![1] Here is a table of $a \div b$ for a and b in $\mathbb{F}_5$. As usual, we can not divide by 0.

$a\div b$	$a=0$	1	2	3	4
$b=1$	0	1	2	3	4
2	0	3	1	4	2
3	0	2	4	1	3
4	0	4	3	2	1

.

For example, $2 \div 3 = 4$ because $3 \times 4 = 2$.

Since we have the basic operations $+$, $-$, $\times$ and $\div$, we can perform row reduction and other computations of linear algebra. Of course, you may wonder whether these algorithms still work; they answer is yes (at least for anything in the first three chapters of Bretscher's book.) You may take this for granted throughout this assignment.

Problem 1 By row reduction, parametrize all solutions to the linear equations

$$\begin{array}{rcrcrcl} x & + & y & + & 2z & = & 4 \\ & & 4y & + & 3z & = & 2 \\ -2x & + & 2y & - & z & = & 4 \end{array}$$

in $\mathbb{F}_5$. You should get a one parameter family of solutions.

Problem 2 Give bases for the kernel and image of

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 3 \\ 0 & 2 & 1 \end{bmatrix}$$

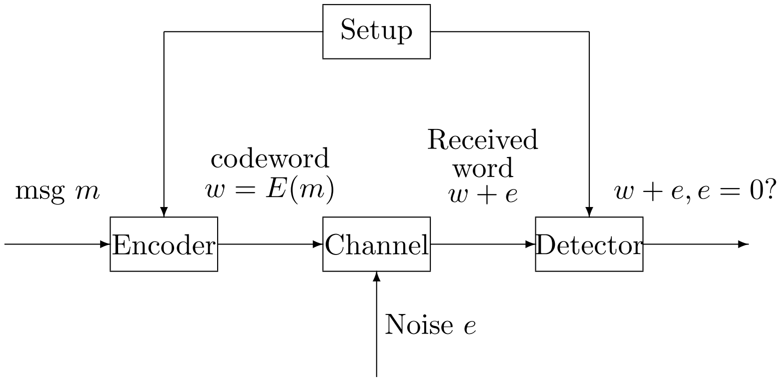
over $\mathbb{F}_5$. You should find that the image is two dimensional and the kernel is one dimensional.

Error-Correcting Codes. We now discuss basic concepts of error correcting codes, and then make the connection to finite fields and linear algebra. Alice and Bob anticipate that they are

[1] For more, beyond the scope of this course, look up *modular arithmetic* and the *extended Euclidean algorithm*, or take Math 312, Applied Modern Algebra.

going to want to communicate through an unreliable method where some part of the message may be garbled. Precisely, Alice expects that she will want to send one of N possible messages, which she will do using n characters from an alphabet of size p. She is concerned that as many as $d-1$ of the characters may be received wrong. She will therefore try to choose N of the possible p^n strings, chosen such that no one can be turned into any other by changing fewer than d characters. That is, any $d-1$ errors should be detectable.

For reference, below is a block diagram of a channel with Encoder and Detector.



Problem 3 Where in Computer Science, IOE or your major are error-correcting codes used? Discuss.

In the rest of this problem we will take $p = 2$. This is particular natural for applications to computer science, because computers like to communicate using an alphabet with 2 symbols. Other primes also occur in applications but would take us a bit too far afield.[2]

The Hamming code. For example, take $N = 8$, $p = 2$, $n = 7$. Alice could choose the following N codewords to represent the messages she might send:

Code Word	Message
`0000000`	Hello.
`1001101`	Goodbye.
`0101011`	The water is fine.
`1100110`	I have met an unusual animal.
`0010111`	Bring an umbrella.
`1011010`	I just saw a rainbow.
`0111100`	I'm ready to return.
`1110001`	Send help immediately!

A code is called *linear* when the list of code words is a linear subspace of $\mathbb{F}_p^n$. Let V_{Hamming} be the list of codewords above.

Problem 4 Give a basis of V_{Hamming}, and give a matrix G_{Hamming} whose image is V_{Hamming}.

The matrix G is called the *generator matrix.*

Let V be a linear code and let d be the smallest number of nonzero bits of any nonzero vector $\vec{v}$ in V. For V_{Hamming}, we have $d = 4$. The number d is called the *distance* of the code.

Problem 5 Explain why, if $\vec{u}$ and $\vec{v}$ are different code words in V, then there are at least d bits where $\vec{u}$ and $\vec{v}$ differ.

[2] Also, there are other mathematical structures where we can add, subtract, multiply and divide besides $\mathbb{R}$ and $\mathbb{F}_p$. These are called *fields.* For example, music CD's are encoded in an error correcting code based on a field of size 256.

Problem 6 Let Alice send Bob a message using a linear code V, and suppose e of the bits are garbled. Show that, if $e < d$, then Bob will be able to detect that an error has occurred and, if $e < d/2$, then Bob will be able to determine Alice's intended message.

We now return to the example of V_{Hamming}.

Problem 7 Give linear equations defining $V_{\text{Hamming}}^{\perp}$ (see Definition 5.1.7 in your textbook). Compute a basis for $V_{\text{Hamming}}^{\perp}$.

Problem 8 Give a matrix H_{Hamming} whose kernel is V_{Hamming}

A matrix H with kernel V is called a **parity check matrix**. Applying any row operation to H will not change the kernel, so the parity check matrix is not unique.

A larger code. We now consider a code with $N = 32$, $p = 2$ and $n = 9$. This code has generator matrix and parity check matrix:

$$G_{\text{big}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 \end{bmatrix} \qquad H_{\text{big}} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Problem 9 Explain how to use H_{big} to check whether or not a vector is in the image of G_{big}. For example, of the three vectors `100110101`, `101100010` and `110111101`, exactly one is in the image of G_{big}. Which one and why?

Problem 10 Let $\vec{y}$ be the vector which you just discovered is in the image of G_{big}. Find the vector $\vec{x}$ for which $\vec{y} = G_{\text{big}}\vec{x}$. (Hint: Look at the top 5 rows of G_{big}.)

Problem 11 Suppose that Alice meant to send the vector $G_{\text{big}}\vec{x}$. However, the first bit of her message got flipped, so Bob received $\vec{y}$ instead. What will $H_{\text{big}}\vec{y}$ be? What would $H_{\text{big}}\vec{y}$ be if the ninth bit had been flipped instead?

Alice encodes her message, which is 50 characters long, into 5 bit vectors in the following manner:

space	00000
A	00001
B	00010
C	00011
...	...
Z	11010

She then encodes these vectors with G_{big}, making them 9 bits long, and sends them to Bob. The message Bob receives is on the next page, and can be downloaded from `http://www.math.lsa.umich.edu/courses/214/project_files/message.csv`. Each vector is either the one Alice sent, or else differs from what she sent by a single bit flip.

Problem 12 Recover Alice's message. You are encouraged to use software for this computation: `MATLAB`, `Mathematica`, even `Excel` if you like.

The message Bob receives reads down the first column, then down the second column, etc.

```
101101001
011001100
001011111
010000000
011110000
011101101
011001010
110011010
000000000
101110010
000011011
110011010
000000010
100001001
011110000
000000000
011001011
001111111
000011001
110101100
011100001
000000000
011010011
000011001
101001001
010101100
001011111
011010010
000011001
101001011
010010101
000111010
100110001
000010000
010010101
100110100
000000010
101001001
010110000
000000000
001000110
011110000
000001000
011010011
000011001
101101001
000001100
001011111
011010011
000011001
101001001
110010101
000111011
100110101
000000000
100001111
000011000
101010001
011001010
000000000
010001110
000011001
011001010
011010011
011110000
100110111
```