Cryptography

Hill-ciphers

an application of
Linear Algebra

This project for my Linear Algebra class is about cryptography. I will discuss a simple method of enciphering and deciphering a message using matrix transformations and modular arithmetic, and show how *elementary row operations* can sometimes be used to break an opponent's code.

The ciphers I will discuss are called ***Hill ciphers*** after Lester S. Hill who introduced them in two papers: "Cryptography in an Algebraic Alphabet," *American Mathematical Monthly*, 36, June-July 1929, pp. 306-312; and "Concerning Certain Linear Transformation Apparatus of Cryptography," *American Mathematical Monthly*, 38, March 1931, pp. 135-154.

I will show an example of how a message is enciphered and deciphered using *Hill ciphers,* I will also briefly discuss how to break a *Hill cipher* using elementary row operations by giving an example from "Elementary Linear Algebra, Applications version, edition 6".

...What is cryptography?...

Cryptography is the study of encoding and decoding secret messages. In the language of cryptography, codes are called the *ciphers*, uncoded messages are called *plaintext*, and coded messages are called *ciphertext*.

Cryptography has for long been an important issue in the realm of computers. It was mainly used for the security needed for passwords but now cryptography is very important due to the Internet's flow of sensitive information such as credit card information and other sensitive information which is fairly easy to monitor by unintended third hand parties.

The idea behind enciphering a message is to make it worthless to everyone except for the party with the deciphering "key".

...Procedure for enciphering and deciphering plaintext using a simple Hill-cipher...

For *Hill ciphers* I assign numerical values to each plaintext and ciphertext letter so that A=1, B=2, C=2 and so on. If I wanted to I could have assigned numerical values for all the other characters on a keyboard, but for simplicity I will only assign numerical values to the letters in the alphabet in this project.

The following procedure shows the simplest Hill ciphers (Hill 2-cipher), successive pairs of plaintext that are transformed into ciphertext by a 2 x 2 matrix ***A***.

NOTE: I will impose an additional condition on matrix *A* later.

Here I have assigned numerical values to the alphabet:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

Enciphering Step 1.

Choose a 2 x 2 matrix ***A*** with integer entries to perform the encoding.

(The matrix has to be invertible modulo *m,* but I will discuss this later)

Enciphering Step 2.

Group successive *plaintext* letters into pairs. If we end up with one single letter at the end, simply add an arbitrary "dummy" letter to fill out the last pair of letters.

Enciphering Step 3.

Convert each plaintext pair *p1p2* into a column vector **p**.

$$\mathbf{p} = \begin{pmatrix} p1 \\ p2 \end{pmatrix}$$

Then form the plaintext matrix ***P*** of all our plaintext column vectors.

$$\mathbf{P} = (p1 \;\; p2 \;\; p3 \;\; \ldots \;\; pn)$$

To encipher the message we multiply our plaintext matrix ***P*** by our transformation matrix ***A*** to form the product ***AP***.

$$\mathbf{C} = \mathbf{AP}$$

The product of our matrix multiplication is the *ciphertext* matrix ***C.***

$$\mathbf{C} = (c1 \;\; c2 \;\; c3 \;\; \ldots \;\; cn)$$

Enciphering Step 4.

Now we convert each ***ciphertext vector*** into its alphabetical equivalent and write out our enciphered message.

This was the encoding procedure, pretty simple, huh:) Let's see how we decipher our enciphered message.

Deciphering Step 1.

Now we group the successive *ciphertext* letters into pairs and convert each *ciphertext* pair *c1c2* into a column vector **c**. Then form the *ciphertext* matrix ***C*** of all our *ciphertext* column vectors.

Deciphering Step 2.

Multiply the *ciphertext* matrix ***C*** with the inverse of our enciphering matrix ***A*** to obtain the deciphered message. Not too difficult, huh:)

NOTE: To use this procedure we have to understand the concept of *modular arithmetic*. In the 6 steps I showed you above, I chose not to include the *modular arithmetic* in the steps for simplicity. However, modular arithmetic is important for this procedure to work. Keep reading and I'll show you why this is so important:)

...The concept of Modular Arithmetic...

All right, everything so far was pretty basic stuff, right? Before moving on to my example, we need to understand the important concept of Hill-ciphers, *modular arithmetic*:

Let's consider the enciphering of the letters "TH" who form the column vector P.

$$\mathbf{P} = \begin{pmatrix} p1 \\ p2 \end{pmatrix} = \begin{pmatrix} T \\ H \end{pmatrix} = \begin{pmatrix} 20 \\ 8 \end{pmatrix}$$

We have the transformation matrix ***A***

$$\mathbf{A} = \begin{pmatrix} 5 & 6 \\ 2 & 3 \end{pmatrix}$$

When we multiply this vector by our transformation matrix ***A***, we get the enciphered column vector

$$\mathbf{C} = \mathbf{AP} = \begin{pmatrix} 5 & 6 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} 20 \\ 8 \end{pmatrix} = \begin{pmatrix} 148 \\ 64 \end{pmatrix}$$

Uh...hmmm, what letters correspond to the integers 148 and 64? This is where Modular Arithmetic comes in handy.

Our alphabet is given by non negative integers from 1, 2, , ..., m, where m is the length of our alphabet (in this case m = 26).

What we do when we have over 26, is simply "wrapping around" the numbers from 27 to 52 to represent the 26 letters again, then we do the same thing from 53 to 78 etc. We can do the same with negative integers (in this case Z=0, Y=-1, X=-2 etc.).

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52
53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78

The procedure of "wrapping" is quite general. It is the same procedure we use every noon and midnight when we begin again to number the hours 1, 2, etc. In a 24 hour system, 18:00 is the same as 6:00 (pm) and 13:00 is 1:00 (pm).

How we do this mathematically is as follows:
When we have integers greater than 26, we replace it by the remainder that results when this integer is divided by 26. So if we have the number 148 from the example above, we divide 148 by 26 and the remainder is 18.

148 - (5 * 26) = 18

Here are a couple examples for some different *modulus*:

7 = 2 (mod 5) because the remainder is 2 after dividing 7 by 5

19 = 3 (mod 2) because the remainder is 3 after dividing 19 by 2

-1 = 25 (mod 26) because the remainder is 25 after dividing -1 by 26

The formal definitions:

If *m* is a positive integer and *a* and *b* are any integers, then we say that *a* is ***equivalent*** to *b* modulo *m*, written

a = b (mod *m*)

if *a-b* is an integer multiple of *m*.

Now to the most important part of the concept of Modular Arithmetic for *Hill ciphers*. As mentioned in the procedure for enciphering and deciphering plaintext using a simple Hill-cipher above, we have to impose an additional condition for our transformation matrix ***A***:

The transformation matrix ***A*** must be *invertible modulo m* for this procedure to work.

So when finding the inverse of our transformation matrix ***A*** we have to take (mod *m*) into consideration.

We have our set of ***residues*** modulo *m* denoted by Z*m* = { 1, 2, 3, ..., *m* } and if *a* is a number in Z*m*, then a number a^-1 in Z*m* is called a ***reciprocal*** or ***multiplicative inverse*** of a modulo m if (a * a^-1) = (a^-1 * a) = 1 (mod *m*).

However, since this project is about Linear Algebra, I chose to skip the details about the modular arithmetic here, and provide a table of the reciprocols of modulo 26 instead. The inmportant thing is to keep in mind when checking our transformation matrix to see if it is invertile it ha to be invertible modulo *m,* you see how this is done in the example provided below the table of reciprocals modulo 26:.

a	1	3	5	7	9	11	15	17	19	21	23	25
a^{-1}	1	9	21	15	3	19	7	23	11	5	17	25

$$A^{-1} = \det^{-1} \begin{pmatrix} 3 & -6 \\ -2 & 5 \end{pmatrix} \qquad \det^{-1} = 3^{-1} = 9 \pmod{26}$$

The inverse of the 2 x2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ $A^{-1} = 9 \begin{pmatrix} 5 & -6 \\ -2 & 5 \end{pmatrix} = \begin{pmatrix} 45 & -54 \\ -18 & 45 \end{pmatrix} = \begin{pmatrix} 19 & 24 \\ 8 & 19 \end{pmatrix} \pmod{26}$

is defined by

$$A^{-1} = \frac{1}{\det} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \text{ where } \det = ad - bc$$

...Simple example of enciphering and deciphering a message using Hill ciphers...

Let's say we want to encipher the following sentence,"**THE PROFESSOR IS EVIL**", into ciphertext.

The first thing we do is to group the letters into pairs of 2 letters. If we would do a Hill 3-cipher, we would group the letters in groups of 3 letters and use a 3 x 3 transformation matrix, but in this example we're using a Hill 2-cipher.

For a Hill n-cipher, use n x n transformation matrix.

So, I have grouped the letters like this:

T H	E P	R O	F E	S S	O R	I S	E V	I L
20 8	5 16	18 15	6 5	19 19	15 18	9 19	5 22	9 12

Voila, this doesn't look like our original message but most 5 year olds could break this simple code.

This leads us to step 3 of the procedure, convert each pair into a column to form the plaintext matrix P.

$$P = \begin{pmatrix} 20 & 5 & 18 & 6 & 19 & 15 & 9 & 5 & 9 \\ 8 & 16 & 15 & 5 & 19 & 18 & 19 & 22 & 12 \end{pmatrix}$$

Form the ciphertext matrix E=AP

Oooops, most of these numbers in E are over 26, but by using the trick we learned from modular arithmetic we easily convert into nicer numbers, remember this is in modulo 26.

$$E = AP = \begin{pmatrix} 148 & 121 & 180 & 60 & 209 & 183 & 159 & 157 & 117 \\ 64 & 58 & 81 & 27 & 95 & 84 & 75 & 76 & 54 \end{pmatrix}$$

Then we assign letters to the numerical values by using our table and this is what we get:

$$E = AP = \begin{pmatrix} 18 & 17 & 24 & 8 & 1 & 1 & 3 & 1 & 13 \\ 12 & 6 & 3 & 1 & 17 & 6 & 23 & 24 & 2 \end{pmatrix}$$

RLQFXCHAAQAFCWAXMB

Yeah, we enciphered the message, let's hope the professor can't break it. I'll show you later how Hill-ciphers can be broken by using row reduction.

$$E = AP = \begin{pmatrix} R & Q & X & H & A & A & C & A & M \\ L & F & C & A & Q & F & W & X & B \end{pmatrix}$$

All right, time to decipher the messages.

Let's imagine we just received this message from one of our classmates, we know the matrix A he/she used to encipher the message with, so what do we do?

Now we work backwards, once again grouping the ciphertext into pairs of 2 letters and assigning numerical values for the letters. We make each pair into a column vector in a matrix E.

Then we simply multiply the matrix E by the inverse of A, but we have to remember our modular arithmetic from the example above.

$$D = A^{-1} E = \begin{pmatrix} 20 & 5 & 18 & 6 & 19 & 15 & 9 & 5 & 9 \\ 8 & 16 & 15 & 5 & 19 & 18 & 19 & 22 & 12 \end{pmatrix}$$

"THE PROFESSOR IS EVIL"

Nice, we just deciphered the message.

...Breaking a *Hill cipher...*

If we are able to obtain a small amount of corresponding plaintext and ciphertext from a secret message, it is possible to determine the deciphering matrix A and then again decipher the entire message. We have learned in class that a *linear transformation* is determined by its values at a basis. This means that if we have a Hill *n*-cipher, and if

p1, p2, ... , pn

are linear independent plaintext vectors whose corresponding ciphertext vectors

Ap1, Ap2, ..., Apn

are known, then we have enough information to determine the matrix A and later A-1 (mod m).

To illustrate this I found an example from "Elementary Linear Algebra, Applications version, edition 6".

Let's say that we obtain an enciphered message and we are able to deduce that it is a letter starting with "DEAR". With a small amount of such data it may be possible to determine the deciphering matrix of a Hill-cipher and consequently get access to the rest of the message.

Example 8 The following Hill 2-cipher is intercepted:

IOSBTGXESPXHOPDE

Decipher the message, given that it starts with the word *DEAR*.

Solution. From Table 1, the numerical equivalent of the known plaintext is

DE	*AR*
4 5	1 18

and the numerical equivalent of the corresponding ciphertext is

IO	*SB*
9 15	19 2

so the corresponding plaintext and ciphertext vectors are

$$\mathbf{p}_1 = \begin{bmatrix} 4 \\ 5 \end{bmatrix} \leftrightarrow \mathbf{c}_1 = \begin{bmatrix} 9 \\ 15 \end{bmatrix}$$

$$\mathbf{p}_2 = \begin{bmatrix} 1 \\ 18 \end{bmatrix} \leftrightarrow \mathbf{c}_2 = \begin{bmatrix} 19 \\ 2 \end{bmatrix}$$

We want to reduce

$$C = \begin{bmatrix} \mathbf{c}_1^t \\ \mathbf{c}_2^t \end{bmatrix} = \begin{bmatrix} 9 & 15 \\ 19 & 2 \end{bmatrix}$$

to I by elementary row operations and simultaneously apply these operations to

$$P = \begin{bmatrix} \mathbf{p}_1^t \\ \mathbf{p}_2^t \end{bmatrix} = \begin{bmatrix} 4 & 5 \\ 1 & 18 \end{bmatrix}$$

to obtain $(A^{-1})^t$ (the transpose of the deciphering matrix). This can be accomplished by adjoining P to the right of C and applying row operations to the resulting matrix $[C \mid P]$ until the left side is reduced to I. The final matrix will then have the form $[I \mid (A^{-1})^t]$. The computations can be carried out as follows:

$\left[\begin{array}{cc\|cc} 9 & 15 & 4 & 5 \\ 19 & 2 & 1 & 18 \end{array}\right]$	We formed the matrix $[C \mid P]$.
$\left[\begin{array}{cc\|cc} 1 & 45 & 12 & 15 \\ 19 & 2 & 1 & 18 \end{array}\right]$	We multiplied the first row by $9^{-1} = 3$.
$\left[\begin{array}{cc\|cc} 1 & 19 & 12 & 15 \\ 19 & 2 & 1 & 18 \end{array}\right]$	We replaced 45 by its residue modulo 26.
$\left[\begin{array}{cc\|cc} 1 & 19 & 12 & 15 \\ 0 & -359 & -227 & -267 \end{array}\right]$	We added -19 times the first row to the second.
$\left[\begin{array}{cc\|cc} 1 & 19 & 12 & 15 \\ 0 & 5 & 7 & 19 \end{array}\right]$	We replaced the entries in the second row by their residues modulo 26.
$\left[\begin{array}{cc\|cc} 1 & 19 & 12 & 15 \\ 0 & 1 & 147 & 399 \end{array}\right]$	We multiplied the second row by $5^{-1} = 21$.
$\left[\begin{array}{cc\|cc} 1 & 19 & 12 & 15 \\ 0 & 1 & 17 & 9 \end{array}\right]$	We replaced the entries in the second row by their residues modulo 26.
$\left[\begin{array}{cc\|cc} 1 & 0 & -311 & -156 \\ 0 & 1 & 17 & 9 \end{array}\right]$	We added -19 times the second row to the first.
$\left[\begin{array}{cc\|cc} 1 & 0 & 1 & 0 \\ 0 & 1 & 17 & 9 \end{array}\right]$	We replaced the entries in the first row by their residues modulo 26.

Thus,

$$(A^{-1})^t = \begin{bmatrix} 1 & 0 \\ 17 & 9 \end{bmatrix}$$

so the deciphering matrix is

$$A^{-1} = \begin{bmatrix} 1 & 17 \\ 0 & 9 \end{bmatrix}$$