techspace

Hill Ciphers

Jonaki B Ghosh

Introduction

Cryptography is the science of making and breaking codes. It is the practice and study of techniques for secure communication. Modern cryptography intersects the disciplines of mathematics, computer science, and electrical engineering. Applications of cryptography include ATM cards, computer passwords, and e-commerce. In this article we explore an interesting cryptography method known as the *Hill Cipher*, based on matrices. We explore the method using the spreadsheet MS Excel to perform operations on matrices.

Hill Ciphers

Hill ciphers are an application of matrices to cryptography. Ciphers are methods for transforming a given message, the *plaintext*, into a new form unintelligible to anyone who does not know the key – the transformation used to convert the plaintext to the *ciphertext*. The inverse key is required to reverse the transformation to recover the original message. To use the key to transform plaintext into ciphertext is to *encipher* the plaintext. To use the inverse key to transform the ciphertext back into plaintext is to *decipher* the ciphertext.

In order to understand Hill ciphers, we must first understand *modular arithmetic.*

Definition 1: A *Hill n-cipher* has for its key a given $n \times n$ matrix whose entries are non-negative integers from the set $\{0, 1, 2, 3, \ldots, m-1\}$, where m is the number of characters used for the encoding process. Suppose we wish to use all 26 alphabets from A to Z and three more characters, say ‘.’, ‘-‘ and ‘?’.

This means we have 29 characters with which to write our plaintext. These have been shown in Table 1, where the 29 characters have been numbered from 0 to 28.

A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Q	**R**	**S**	**T**	**U**	**V**	**W**	**X**	**Y**	**Z**	**.**	**_**	**?**			
16	17	18	19	20	21	22	23	24	25	26	27	28			

Table 1: The substitution table for the Hill Cipher

Let us apply this to an example of a Hill 2-cipher corresponding to the substitution scheme shown in Table 1 with 29 characters. Let the key be the 2 × 2 matrix

$$E = \begin{bmatrix} 1 & 4 \\ 2 & 9 \end{bmatrix}$$

We can also refer to E as the encoding matrix. We will use E to encipher groups of two consecutive characters. Suppose we have to encipher the word **GO**. The alphabets G and O correspond to the numbers 6 and 14, respectively from our substitution table. We represent it by a 2 × 1 matrix.

$$\begin{bmatrix} 6 \\ 14 \end{bmatrix}$$

To encipher **GO**, we pre-multiply this matrix by the encoding matrix E.

$$\begin{bmatrix} 1 & 4 \\ 2 & 9 \end{bmatrix} \begin{bmatrix} 6 \\ 14 \end{bmatrix} = \begin{bmatrix} 62 \\ 138 \end{bmatrix}$$

The product is a 2 × 1 matrix with entries 62 and 138. But what characters do the numbers 62 and 138 represent? These are not in our substitution table! What we do is as follows:

We divide these numbers by 29 and consider their respective remainders after the division process is done. Thus when we divide 62 by 29, the remainder is 4 and when we divide 138 by 29, the remainder is 22.

To express this in the language of *modular arithmetic*, we write:

$$62 \equiv 4 \ (mod\ 29),$$
$$138 \equiv 22 \ (mod\ 29).$$

Equivalence and residues modulo an integer

Definition 2: Given an integer $m > 1$, called the *modulus*, we say that the two integers a and b are *congruent* to one another *modulo* m if $a - b$ is an integral multiple of m. To denote this, we write: $a \equiv b\ (mod\ m)$.

We read this as: 'a is congruent to b modulo m'.

In other words, $a \equiv b\ (mod\ m)$ means that $a = b + km$ for some integer k which could be positive, negative or zero.

For our cipher, we have: $62 \equiv 4\ mod\ \ 29$ and $138 \equiv 22\ mod\ 29$. Note: $62 - 4 = 2 \times 29$, or $62 = 4 + 2 \times 29$, and $138 - 22 = 4 \times 29$, or $138 = 22 + 4 \times 29$.

The numbers 4 and 22 correspond to the letters E and W respectively from our substitution table. Thus, the word **GO** is enciphered to **EW**!

In order to use this method of sending secret messages, the sender has to encrypt the plaintext **GO** and send the encrypted form. This means the sender sends **EW** instead. The receiver gets the message **EW**. The secret key, that is, the encoding matrix E is known only to the sender and the receiver. Now let us see how the receiver deciphers what **EW** stands for.

In order to decipher the message **EW**, we begin by looking for the numbers corresponding to E and W in our substitution table. These are 4 and 22 respectively. We represent this in the form of a 2 × 1 matrix

$$\begin{bmatrix} 4 \\ 22 \end{bmatrix}$$

We pre-multiply this matrix by the inverse of the matrix E, that is, by $E^{-1} = \begin{bmatrix} 9 & -4 \\ -2 & 1 \end{bmatrix}$. So:

$$E^{-1}W = \begin{bmatrix} 9 & -4 \\ -2 & 1 \end{bmatrix}\begin{bmatrix} 4 \\ 22 \end{bmatrix} = \begin{bmatrix} -52 \\ 14 \end{bmatrix}$$

We need another definition. Note that any arbitrary integer a can be divided by m to yield a quotient q and remainder r; that is, $a = q\,m + r$. Then we say $a \equiv r\ (mod\ m)$.

Definition 3: Let m be any integer exceeding 1. For an arbitrary integer a, the *residue of a modulo m* is the unique integer r in the set $\{0,\ 1,\ 2,\ 3,\ \ldots,\ m-1\}$ such that $a \equiv r\ (mod\ m)$.

Thus: $23 \equiv 7\ (mod\ 8)$, since $23 = 2\times 8 + 7$. Here $a = 23, m = 8$ and $r = 7$. What about negative integers? For example, what is r when $a = -18$ and $m = 8$? Clearly r has to be an integer between 0 and 7. Note that $-18 = -3 \times 8 + 6$. Thus $r = 6$ and we can write $-18 \equiv 6\ (mod\ 8)$.

Now coming back to deciphering our Hill cipher, we need to find r for -52 and 14 for $m = 29$. Note that $-52 = -2 \times 29 + 6$ and $14 = 0 \times 29 + 14$.Thus: $-52 \equiv 6\ (mod\ 29)$ and $14 \equiv 14\ (mod\ 29)$.

Thus we may write $\begin{bmatrix} -52 \\ 14 \end{bmatrix} \equiv \begin{bmatrix} 6 \\ 14 \end{bmatrix} (mod\ \ 29)$

Isn't this great! 6 represents G and 14 represents O from our substitution table. Hence we have deciphered **EW** to obtain **GO** the original characters or the plaintext!

Let us now see how to encipher a longer message or plaintext using the key $\begin{bmatrix} 1 & 4 \\ 2 & 9 \end{bmatrix}$.

We shall encipher the plaintext **MATH_IS_FUN.** The steps are indicated below. For matrix computations we use MS Excel. In Excel, the commands for multiplying matrices and finding the inverse of a matrix are MMULT and MINVERSE respectively. For reducing a number modulo a divisor the required command is MOD.

Encoding or enciphering the plaintext

Step 1: Convert the plaintext **MATH_IS_FUN.** to the corresponding substitution values from the substitution table. The values are

12 0 19 7 27 8 18 27 5 20 13 26

We need to make a 2 × n matrix using these values

Step 2: Form pairs of these numbers as follows

12 0   19 7   27 8   18 27   5 20   13 26

Note: In case the message has an odd number of characters, a full stop or underscore (a ‘.’ ‘ or a ‘_’) may be added at the end to complete the pair. For example, if the plaintext is “LET_US_GO”, then we have 9 characters. So we add a ‘.’ at the end to make the message “LET_US_GO.”

Each pair will form a column of a 2 × 6 matrix (as there are 6 pairs). Let us call this matrix P (the plaintext matrix)

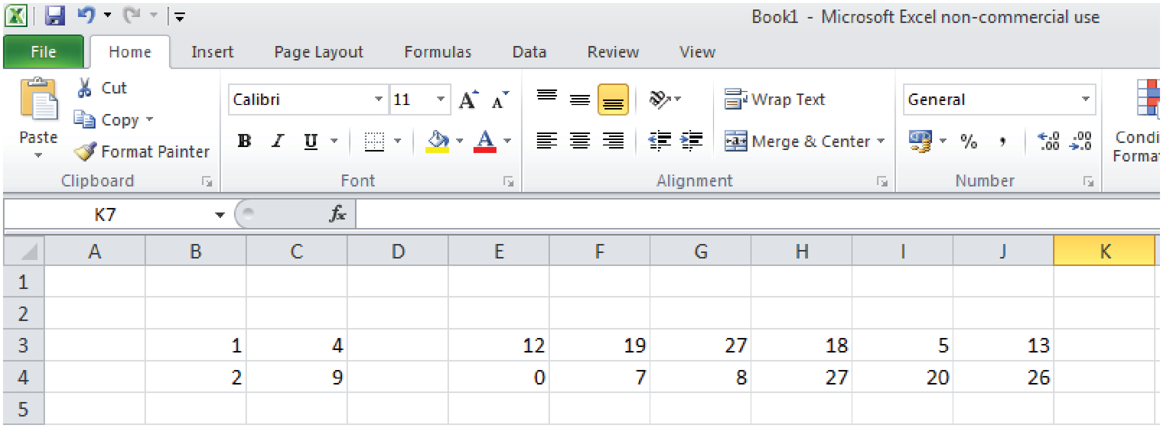
$$P = \begin{bmatrix} 12 & 19 & 27 & 18 & 5 & 13 \\ 0 & 7 & 8 & 27 & 20 & 26 \end{bmatrix}$$

Step 3: Compute the product EP

$$EP = \begin{bmatrix} 1 & 4 \\ 2 & 9 \end{bmatrix} \begin{bmatrix} 12 & 19 & 27 & 18 & 5 & 13 \\ 0 & 7 & 8 & 27 & 20 & 26 \end{bmatrix} = \begin{bmatrix} 12 & 47 & 59 & 126 & 85 & 117 \\ 24 & 101 & 126 & 279 & 190 & 260 \end{bmatrix}$$

In order to perform this computation in Excel we proceed as follows

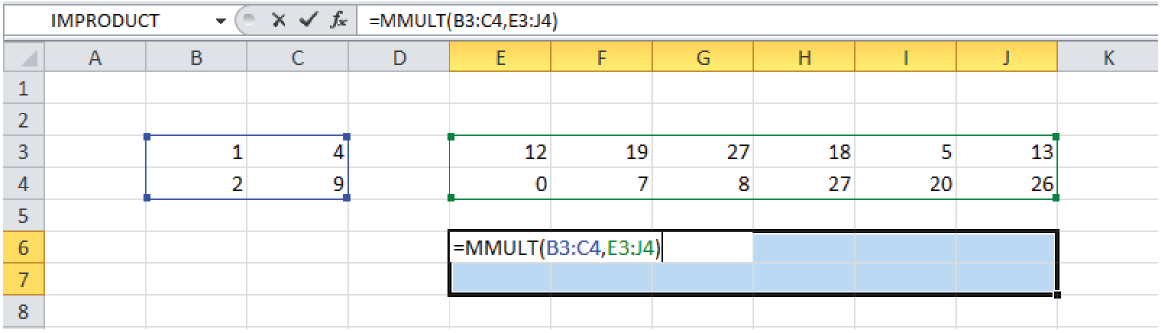
Enter the 2 × 2 matrix E and the 2 × 6 matrix P as separate arrays as shown. Each entry of a matrix may be entered by typing a number in a cell and pressing Enter. The arrow keys may be used to move to the next appropriate cell.



K7

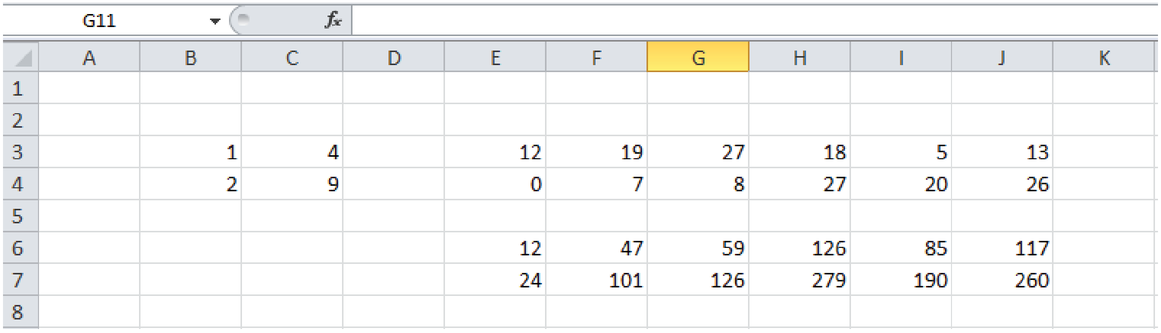
	A	B	C	D	E	F	G	H	I	J	K
1											
2											
3		1	4		12	19	27	18	5	13	
4		2	9		0	7	8	27	20	26	
5											

To obtain the product, select a blank 2 × 6 array and type **=MMULT(** in the top leftmost cell of the closed array. Within the parentheses, first select the array for matrix E and then the array for matrix P separated by a comma. Press **Crtl + Shift** followed by **Enter** to obtain the product. (Note that you need to press Crtl and Shift simultaneously and then press Enter.)



IMPRODUCT =MMULT(B3:C4,E3:J4)

	A	B	C	D	E	F	G	H	I	J	K
1											
2											
3		1	4		12	19	27	18	5	13	
4		2	9		0	7	8	27	20	26	
5											
6					=MMULT(B3:C4,E3:J4)						
7											
8											



G11 *fx*

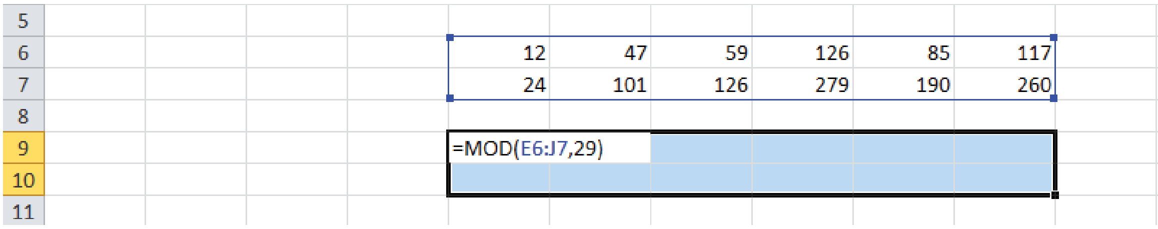
	A	B	C	D	E	F	G	H	I	J	K
1											
2											
3		1	4		12	19	27	18	5	13	
4		2	9		0	7	8	27	20	26	
5											
6					12	47	59	126	85	117	
7					24	101	126	279	190	260	
8											

Step 4: Reduce the product modulo 29 to obtain the Hill 2-cipher values. This means we have to divide each number by 29 and find the remainder. In Excel we can reduce the entire matrix modulo 29 in one go!

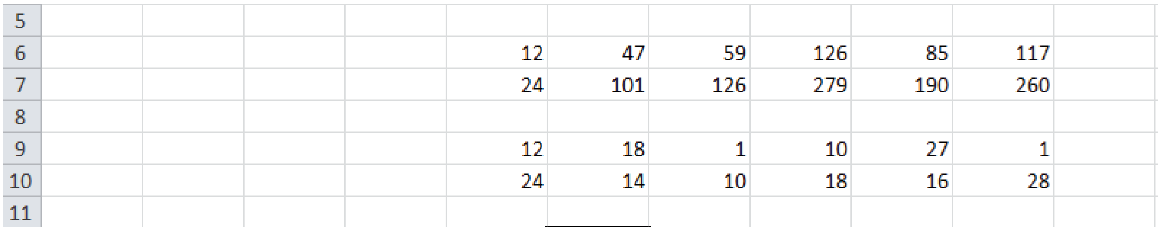
$$EP = \begin{bmatrix} 12 & 47 & 59 & 126 & 85 & 117 \\ 24 & 101 & 126 & 279 & 190 & 260 \end{bmatrix} \equiv \begin{bmatrix} 12 & 18 & 1 & 10 & 27 & 1 \\ 24 & 14 & 10 & 18 & 16 & 28 \end{bmatrix} (mod\ 29)$$

To do this in Excel proceed as follows

Select a blank 2 × 6 array and type **=MOD(** in the top leftmost cell of the array. Within the parentheses, select the array of the product matrix EP and type 29 for the divisor.



5											
6					12	47	59	126	85	117	
7					24	101	126	279	190	260	
8											
9					=MOD(E6:J7,29)						
10											
11											



5											
6					12	47	59	126	85	117	
7					24	101	126	279	190	260	
8											
9					12	18	1	10	27	1	
10					24	14	10	18	16	28	
11											

Step 5: Write out the columns of the matrix in a sequence

$$\begin{bmatrix} 12 & 18 & 1 & 10 & 27 & 1 \\ 24 & 14 & 10 & 18 & 16 & 28 \end{bmatrix}$$

These are

12 24 18 14 1 10 10 18 27 16 1 28

Replace these values by the characters from the substitution table to which these values correspond.

The encrypted message or ciphertext is **MYSOBKKS_QB?**

Decoding or deciphering the ciphertext

In this section we will try to decipher the ciphertext **MYSOBKKS_QB?**

Step 1: Convert the characters to their respective Hill-2-cipher values from the substitution table

12 24 18 14 1 10 10 18 27 16 1 28

Form a 2 × 6 matrix using these values. Make pairs of these numbers as follows

12 24 18 14 1 10 10 18 27 16 1 28

Each pair will form a column of a 2 × 6 matrix (since there are 6 pairs). Let us call this matrix C (the ciphertext matrix)

$$C = \begin{bmatrix} 12 & 18 & 1 & 10 & 27 & 1 \\ 24 & 14 & 10 & 18 & 16 & 28 \end{bmatrix}$$

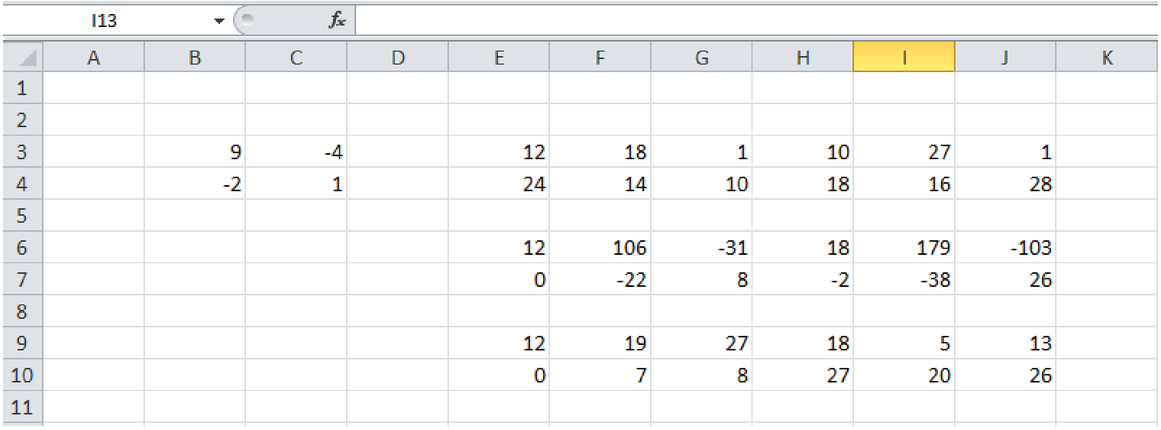
Step 2: Compute the product E^{-1} C

$$E^{-1}C = \begin{bmatrix} 9 & -4 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 12 & 18 & 1 & 10 & 27 & 1 \\ 24 & 14 & 10 & 18 & 16 & 28 \end{bmatrix} = \begin{bmatrix} 12 & 106 & -31 & 18 & 179 & -103 \\ 0 & -22 & 8 & -2 & -38 & 26 \end{bmatrix}$$

Step 3: Reduce the product modulo 29 to obtain the substitution values.

$$\begin{bmatrix} 12 & 106 & -31 & 18 & 179 & -103 \\ 0 & -22 & 8 & -2 & -38 & 26 \end{bmatrix} \equiv \begin{bmatrix} 12 & 19 & 27 & 18 & 5 & 13 \\ 0 & 7 & 8 & 27 & 20 & 26 \end{bmatrix} (mod\ 29)$$

The reader may perform these computations using Excel. The screenshot of the Excel sheet is as follows.



	A	B	C	D	E	F	G	H	I	J	K
1											
2											
3		9	-4		12	18	1	10	27	1	
4		-2	1		24	14	10	18	16	28	
5											
6					12	106	-31	18	179	-103	
7					0	-22	8	-2	-38	26	
8											
9					12	19	27	18	5	13	
10					0	7	8	27	20	26	
11											

Step 4: Write out the columns of the matrix in a sequence

$$\begin{bmatrix} 12 & 19 & 27 & 18 & 5 & 13 \\ 0 & 7 & 8 & 27 & 20 & 26 \end{bmatrix}$$

These are

12 0 19 7 27 8 18 27 5 20 13 26

Replace these values by the characters from the substitution table to which these values correspond.

The decrypted message or plaintext is **MATH_IS_FUN.**

The method works! Observe that **MATH_IS_FUN** has been encrypted as **MYSOBKKS_QB?**

Note that there are two '_' (underscores) in the original plaintext message. But these do not get encrypted to the same characters. The first one is encrypted to B and the second one gets encrypted to S. Can you explain why?

So far we have learnt how to encrypt a plaintext using a Hill 2-cipher. This means that our encoding matrix is a 2×2 matrix. If we choose a 3×3 matrix, the plaintext will have to be converted to a $3 \times n$ matrix (here the number of columns ‘n’ depends on the length of the message).

The reader is urged to try to decode the messages in the next few exercises to practice the method. All computations may be done on Excel. Note that the substitution table remains the same as before.

EXERCISES

1. Decode the secret message **FY O. KI ZT WA QC** which was encrypted using the encoding matrix

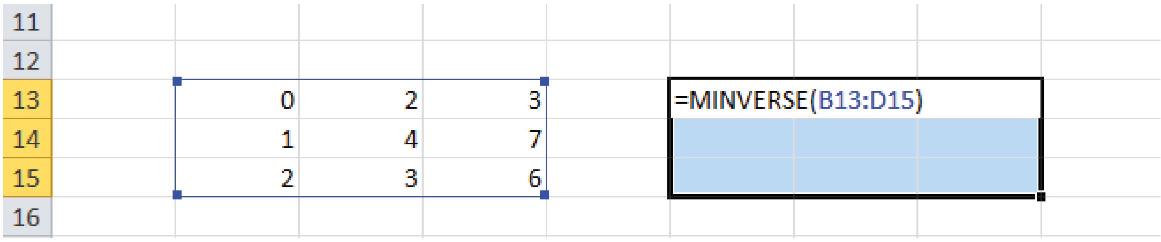
$$E = \begin{bmatrix} 1 & 4 \\ 2 & 9 \end{bmatrix}$$

2. Decode the secret message **ITS DGN STX SJK DVO JHE TCB** which was encrypted using the encoding matrix

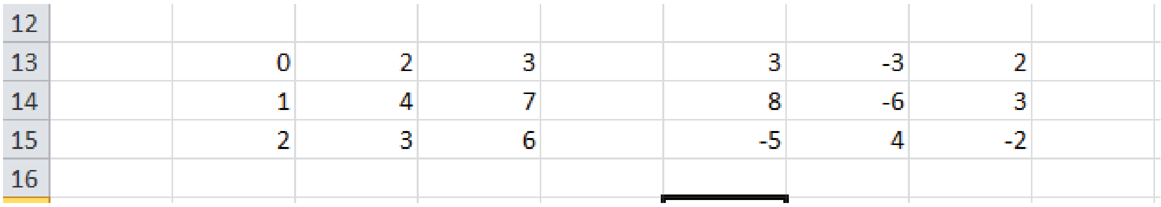
$$E = \begin{bmatrix} 0 & 2 & 3 \\ 1 & 4 & 7 \\ 2 & 3 & 6 \end{bmatrix}$$

This is an example of a Hill 3-cipher.

Note that the command **=MINVERSE()** may be used in Excel to find the inverse of the given matrix E. Enter the matrix in a 3 × 3 array. Then select a blank 3 × 3 array and type the command as shown in the screenshot

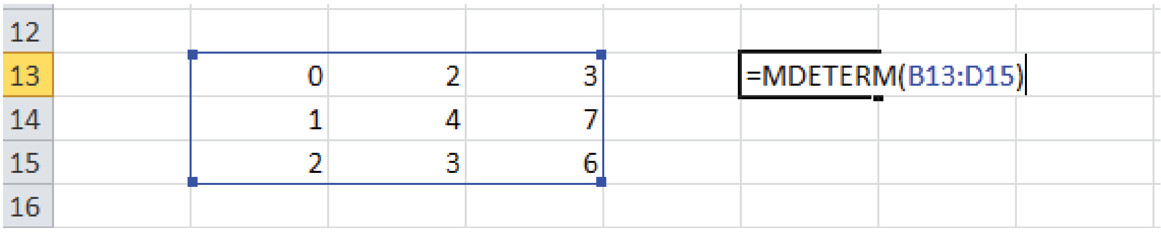


11								
12								
13		0	2	3		=MINVERSE(B13:D15)		
14		1	4	7				
15		2	3	6				
16								

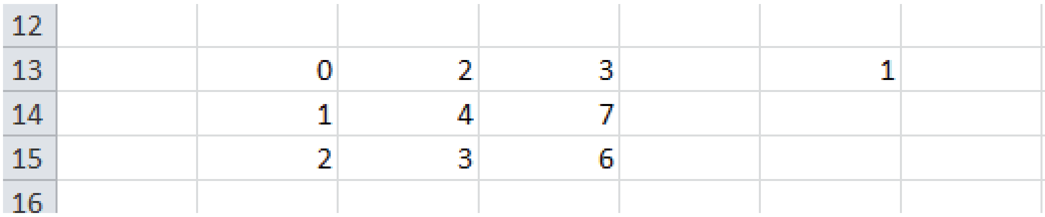


12								
13		0	2	3		3	-3	2
14		1	4	7		8	-6	3
15		2	3	6		-5	4	-2
16								

Also if you want to compute the determinant of the matrix, then select any cell and type the command **=MDETERM()** and press Enter.



12								
13		0	2	3		=MDETERM(B13:D15)		
14		1	4	7				
15		2	3	6				
16								



12							
13		0	2	3		1	
14		1	4	7			
15		2	3	6			
16							

Note that the given 3 × 3 matrix has determinant 1.

3. Decode the secret message **UAR CR? WBQ BYW WBL LCD RWD** which was encrypted using the encoding matrix

$$E = \begin{bmatrix} 0 & 2 & 3 \\ 1 & 4 & 7 \\ 2 & 3 & 6 \end{bmatrix}$$

4. Find a 4 × 4 matrix whose determinant is equal to 1 and use it to encrypt your own message.

Solutions to these exercises are given on page 74.

Conclusion

The Hill Cipher presents an interesting application of matrices and number theory to cryptography. It is open to exploration and students find it exciting to use this method. This is an example of a practical situation where performing matrix operations such as matrix multiplication and finding the inverse are actually required. It helps the student to understand the need and importance of matrix operations and also explore the method by using different keys (that is, encoding matrices). Any computing tool which can perform matrix operations will be helpful, as the computations may be tedious and time consuming (especially when the plaintext or ciphertext are lengthy). In this article we have chosen square matrices whose determinant is equal to 1 as the key. However this is not necessary. Any invertible matrix may be chosen. In case the determinant of the encoding matrix is anything other than 1, the computations are slightly different, and this case will be discussed in another article.

References

1. http://en.wikipedia.org/wiki/Hill_cipher
2. http://practicalcryptography.com/ciphers/hill-cipher/

JONAKI GHOSH is an Assistant Professor in the Dept. of Elementary Education, Lady Sri Ram College, University of Delhi where she teaches courses related to math education. She obtained her Ph.D. in Applied Mathematics from Jamia Milia Islamia University, New Delhi, and her M.Sc. from IIT Kanpur. She has taught mathematics at the Delhi Public School, R K Puram, where she set up the Math Laboratory & Technology Centre. She has started a Foundation through which she conducts professional development programmes for math teachers. Her primary area of research interest is in the use of technology in mathematics instruction. She is a member of the Indo Swedish Working Group on Mathematics Education. She regularly participates in national and international conferences. She has published articles in journals and authored books for school students. She may be contacted at jonakibghosh@gmail.com.