Hill Substitution Ciphers

Text Reference: Section 4.1, p. 223

In this set of exercises, using matrices to encode and decode messages is examined.

Perhaps the simplest way to encode a message is to simply replace each letter of the alphabet with another letter. This is the method used in the "Cryptograms" often found in puzzle books or newspapers, and is called a **substitution cipher**.

Consider this **cipher array** for a substitution cipher:

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
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
Y	C	W	O	G	R	D	B	P	I	Z	X	K	L	N	M	T	S	E	F	H	J	A	V	U	Q

To encode a message like VECTOR SPACE, encode each letter in turn and get

JGWFNSEMYWG

Notice that the space between the words hasn't been encoded; this character could be added (as could any other punctuation or symbol) to the cipher array and encoded also. To decode a message, the **decipher array** is needed, which for our substitution cipher is

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
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
W	H	B	G	S	T	E	U	J	V	M	N	P	O	D	I	Z	F	R	Q	Y	X	C	L	A	K

Thus the secret message FBGSNNEFGSWSNAEYFOYAL is easily revealed as

THE ROOSTER CROWS AT DAWN

A major drawback of the substitution cipher is that it is very easy for a person to "crack" the code; that is, to determine the decipher array from an encoded message. If you have done a crytogram before, you know how this is done: the relative frequency of letters in English is known, as are the frequencies of certain groups of letters like TH or ST. See Reference 3 (p.16 and p.19) for sample tables. Common short words like THE and OF also help the code cracker. To make the code harder to crack, groups of letters can be encoded at the same time. For example, consider splitting the above message into units of three letters each, although any length of block would be allowable. The message

THE ROOSTER CROWS AT DAWN

is thus converted to

THE ROO STE RCR OWS ATD AWN

If the number of letters is not a multiple of three, the final set of letters can be padded with random letters; thus VECTOR SPACE could be split up as VEC TOR SPA CEX.

Since linear algebra will come in handy for the encoding process, replace each letter by its position in the alphabet. The positions are numbered from 0 to 25 for reasons which will shortly become apparent.

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
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
0	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

The batches of three letters can be converted into batches of three numbers, which can be thought of as **vectors**. For example, THE ROOSTER CROWS AT DAWN has become

$$\begin{matrix} T \\ H \\ E \end{matrix}\begin{bmatrix} 19 \\ 7 \\ 4 \end{bmatrix} \quad \begin{matrix} R \\ O \\ O \end{matrix}\begin{bmatrix} 17 \\ 14 \\ 14 \end{bmatrix} \quad \begin{matrix} S \\ T \\ E \end{matrix}\begin{bmatrix} 18 \\ 19 \\ 4 \end{bmatrix} \quad \begin{matrix} R \\ C \\ R \end{matrix}\begin{bmatrix} 17 \\ 2 \\ 17 \end{bmatrix} \quad \begin{matrix} O \\ W \\ S \end{matrix}\begin{bmatrix} 14 \\ 22 \\ 18 \end{bmatrix} \quad \begin{matrix} A \\ T \\ D \end{matrix}\begin{bmatrix} 0 \\ 19 \\ 3 \end{bmatrix} \quad \begin{matrix} A \\ W \\ N \end{matrix}\begin{bmatrix} 0 \\ 22 \\ 13 \end{bmatrix}$$

In order to encode messages using the numbers instead of letters, an easy way of calculating using only the numbers 0 through 25 is needed. A way to do this is called **modular arithmetic**. Sums, differences, and products are calculated as normal, but if the result is larger than 25 or less than 0, it is replaced by the remainder left when the result is divided by 26. Thus only numbers from 0 to 25 result from this arithmetic. This form of arithmetic is denoted by stating that the calculations are to be done **modulo 26**, and by placing the symbol $(\mathrm{mod}\, 26)$ after the calculation.

Examples:

- $1 + 2 = 3 (\mathrm{mod}\, 26)$
- $13 \times 2 = 0 (\mathrm{mod}\, 26)$, since $13 \times 2 = 26 = 26(1) + 0$
- $7 + 24 = 5 (\mathrm{mod}\, 26)$, since $7 + 24 = 31 = 26(1) + 5$
- $10 - 15 = 21 (\mathrm{mod}\, 26)$, since $10 - 15 = -5 = 26(-1) + 21$

Just as numbers can be added and multiplied, vectors may be added and scalar multiplication may be performed:

Examples:

- $\begin{bmatrix} 19 \\ 7 \\ 4 \end{bmatrix} + \begin{bmatrix} 14 \\ 22 \\ 18 \end{bmatrix} = \begin{bmatrix} 7 \\ 3 \\ 22 \end{bmatrix} (\mathrm{mod}\, 26)$
- $10 \times \begin{bmatrix} 14 \\ 22 \\ 18 \end{bmatrix} = \begin{bmatrix} 10 \\ 12 \\ 24 \end{bmatrix} (\mathrm{mod}\, 26)$

The set $\mathbb{Z}_{26}^3$ is the set of all vectors with three elements drawn from the numbers 0 through 25. Addition of vectors and multiplication by scalars are defined as has been done above. The space $\mathbb{Z}_{26}^3$ will be very useful for coding and decoding messages.

To encode a message a **key matrix** A is used, which in our case will be a 3×3 matrix. Multiplying a vector $\mathbf{v}$ in $\mathbb{Z}_{26}^3$ by A will produce another vector $A\mathbf{v}$ which is also in $\mathbb{Z}_{26}^3$, which may be interpreted as the encoded version of $\mathbf{v}$.

Example: Let our key matrix be

$$A = \begin{bmatrix} 3 & 10 & 20 \\ 20 & 9 & 17 \\ 9 & 4 & 17 \end{bmatrix}$$

The message THE ROOSTER CROWS AT DAWN may be encoded by multiplying each message vector in turn by the key matrix; for example, the first message vector becomes

$$\begin{bmatrix} 3 & 10 & 20 \\ 20 & 9 & 17 \\ 9 & 4 & 17 \end{bmatrix} \begin{bmatrix} 19 \\ 7 \\ 4 \end{bmatrix} = \begin{bmatrix} 25 \\ 17 \\ 7 \end{bmatrix}$$

So the triple of letters THE becomes encoded as ZRH. To save effort, combine all of the message vectors into one matrix M and compute AM, because the columns of AM are just the result of applying A to the individual columns of M:

$$AM = \begin{bmatrix} 3 & 10 & 20 \\ 20 & 9 & 17 \\ 9 & 4 & 17 \end{bmatrix} \begin{bmatrix} 19 & 17 & 18 & 17 & 14 & 0 & 0 \\ 7 & 14 & 19 & 2 & 22 & 19 & 22 \\ 4 & 14 & 4 & 17 & 18 & 3 & 13 \end{bmatrix} = \begin{bmatrix} 25 & 3 & 12 & 21 & 24 & 16 & 12 \\ 17 & 2 & 1 & 23 & 4 & 14 & 3 \\ 7 & 5 & 20 & 8 & 0 & 23 & 23 \end{bmatrix}$$

The encoded message would be

ZRHDCFMBUVXIYEAQOXMDX

Now this method of encoding would be useless if there were not a way to decode the encoded messages. Since the key matrix A does the encoding, it stands to reason that if A has an inverse A^{-1}, then A^{-1} should be the decoding matrix. While this is true, the inversion of A must be done in modular arithmetic. So for example the inverse of our key matrix A is

$$A^{-1} = \begin{bmatrix} 11 & 22 & 14 \\ 7 & 9 & 21 \\ 17 & 0 & 3 \end{bmatrix}$$

since (see Question 1d) below)

$$A^{-1}A = \begin{bmatrix} 11 & 22 & 14 \\ 7 & 9 & 21 \\ 17 & 0 & 3 \end{bmatrix} \begin{bmatrix} 3 & 10 & 20 \\ 20 & 9 & 17 \\ 9 & 4 & 17 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \pmod{26}$$

Thus to decode the message

ZRHMJHHKVVVWEXWHZIFRQ

the coded message is first split into three-letter units, with are then converted to vectors in $\mathbb{Z}_{26}^3$, then these vectors are collected into a coded message matrix

$$C = \begin{bmatrix} 25 & 12 & 7 & 21 & 4 & 7 & 5 \\ 17 & 9 & 10 & 21 & 23 & 25 & 17 \\ 7 & 7 & 21 & 22 & 22 & 8 & 16 \end{bmatrix}$$

The computation

$$\begin{aligned} A^{-1}C &= \begin{bmatrix} 11 & 22 & 14 \\ 7 & 9 & 21 \\ 17 & 0 & 3 \end{bmatrix} \begin{bmatrix} 25 & 12 & 7 & 21 & 4 & 7 & 5 \\ 17 & 9 & 10 & 21 & 23 & 25 & 17 \\ 7 & 7 & 21 & 22 & 22 & 8 & 16 \end{bmatrix} \\ &= \begin{bmatrix} 19 & 12 & 19 & 13 & 0 & 11 & 3 \\ 7 & 0 & 8 & 18 & 21 & 0 & 4 \\ 4 & 17 & 0 & 7 & 4 & 13 & 3 \end{bmatrix} (\mathrm{mod}\, 26), \end{aligned}$$

gives the decoded message, which may be translated into letters as

THEMARTIANSHAVELANDED, or THE MARTIANS HAVE LANDED

The method which has been used to encode messages is called a **Hill substitution cipher**. The method was invented by mathematician Lester Hill and is reviewed in References 1 and 2. This form of cipher is harder to decode than the simple substitution cipher. Each letter is not encoded by the same letter over and over; for example, in the above code for THE MARTIANS HAVE LANDED, the letter A appears four times and is encoded by J, V, O, and Z.

Since the inverse of the key matrix is needed to decode the encoded message, any prospective key matrix must be invertible. As has been seen in Chapter 3, the determinant of the matrix shows whether a given matrix is invertible. However, since calculations are being done in modular arithmetic, the following criterion must be used. It is given in Reference 3, page 114.

Fact: An $n \times n$ matrix A is invertible modulo n if and only if $\det A \neq 0 (\mathrm{mod}\, p)$ for every prime divisor p of n.

Thus a matrix A will be invertible modulo 26 if and only if $\det A \neq 0 (\mathrm{mod}\, 13)$ and $\det A \neq 0 (\mathrm{mod}\, 2)$.

Example: Since

$$\begin{vmatrix} 3 & 10 & 20 \\ 20 & 9 & 17 \\ 9 & 4 & 17 \end{vmatrix} = -1635$$

which is equal to $3 (\mathrm{mod}\, 13)$ and $1 (\mathrm{mod}\, 2)$, this matrix is invertible modulo 26 and may be used as a key matrix for a Hill substitution cipher.

Example: Since

$$\begin{vmatrix} 21 & 24 & 2 \\ 0 & 1 & 3 \\ 21 & 19 & 17 \end{vmatrix} = 630$$

which is equal to $6 (\mathrm{mod}\, 13)$ but $0 (\mathrm{mod}\, 2)$, this matrix is not invertible modulo 26 and cannot be used as a key matrix.

Once it has been determined that a matrix A is invertible modulo 26, it may be used as a key matrix for a Hill substitution cipher. In order to decode this cipher the inverse of the key matrix A modulo 26 must be found. The same algorithm as presented in Section 2.2 of the text (pages 115-116) may be used: the matrix $[A|I]$ is row reduced to $[I|A^{-1}]$, but the reduction is done in modular arithmetic.

Row reduction can be a bit tricky here, for not all row operations are allowable when using $(\mathrm{mod})\, 26$ arithmetic. Scaling of a row by 13 or 2 is not allowed; since $13 \times 2 = 0 (\mathrm{mod}\, 26)$, this would be equivalent to scaling a row by 0. Similarly, multiplying a row by 13 or 2 for use in a row replacement operation is not allowed.

All other row operations are allowable. Consider using this algorithm to find the inverse of the key matrix A above. The modulo 26 multiplication table at the end of this set of exercises will be helpful in following the arithmetic.

Example: It was shown above that the matrix

$$\begin{bmatrix} 3 & 10 & 20 \\ 20 & 9 & 17 \\ 9 & 4 & 17 \end{bmatrix}$$

is invertible modulo 26. To find its inverse, begin with the matrix

$$\begin{bmatrix} 3 & 10 & 20 & 1 & 0 & 0 \\ 20 & 9 & 17 & 0 & 1 & 0 \\ 9 & 4 & 17 & 0 & 0 & 1 \end{bmatrix}$$

The first row must be scaled so that there is a 1 in the first pivot position. Since $3 \times 9 = 1 (\mathrm{mod}\, 26)$, scale row 1 by 9:

$$\begin{bmatrix} 1 & 12 & 24 & 9 & 0 & 0 \\ 20 & 9 & 17 & 0 & 1 & 0 \\ 9 & 4 & 17 & 0 & 0 & 1 \end{bmatrix}$$

To eliminate the 20 below the first pivot, multiply row 1 by 6 and add it to row 2, since $1 \times 6 + 20 = 0 (\mathrm{mod}\, 26)$; likewise multiply row 1 by 17 and add it to row 3 since $1 \times 17 + 9 = 0 (\mathrm{mod}\, 26)$. The matrix is thus reduced to

$$\begin{bmatrix} 1 & 12 & 24 & 9 & 0 & 0 \\ 0 & 3 & 5 & 2 & 1 & 0 \\ 0 & 0 & 9 & 23 & 0 & 1 \end{bmatrix}$$

Scale rows 2 and 3 by 9 and 3 respectively:

$$\begin{bmatrix} 1 & 12 & 24 & 9 & 0 & 0 \\ 0 & 1 & 19 & 18 & 9 & 0 \\ 0 & 0 & 1 & 17 & 0 & 3 \end{bmatrix}$$

Mulitplying row 3 by 7 and adding to row 2, and multiplying row 3 by 2 and adding to row 1 produce

$$\begin{bmatrix} 1 & 12 & 0 & 17 & 0 & 6 \\ 0 & 1 & 0 & 7 & 9 & 21 \\ 0 & 0 & 1 & 17 & 0 & 3 \end{bmatrix}$$

and finally multiplying row 2 by 14 and adding to row 1 gives

$$\begin{bmatrix} 1 & 0 & 0 & 11 & 22 & 14 \\ 0 & 1 & 0 & 7 & 9 & 21 \\ 0 & 0 & 1 & 17 & 0 & 3 \end{bmatrix}$$

Thus

$$A^{-1} = \begin{bmatrix} 11 & 22 & 14 \\ 7 & 9 & 21 \\ 17 & 0 & 3 \end{bmatrix}$$

as noted above.

Questions:

1. Compute the following modulo 26.

   a) $17 + 24$

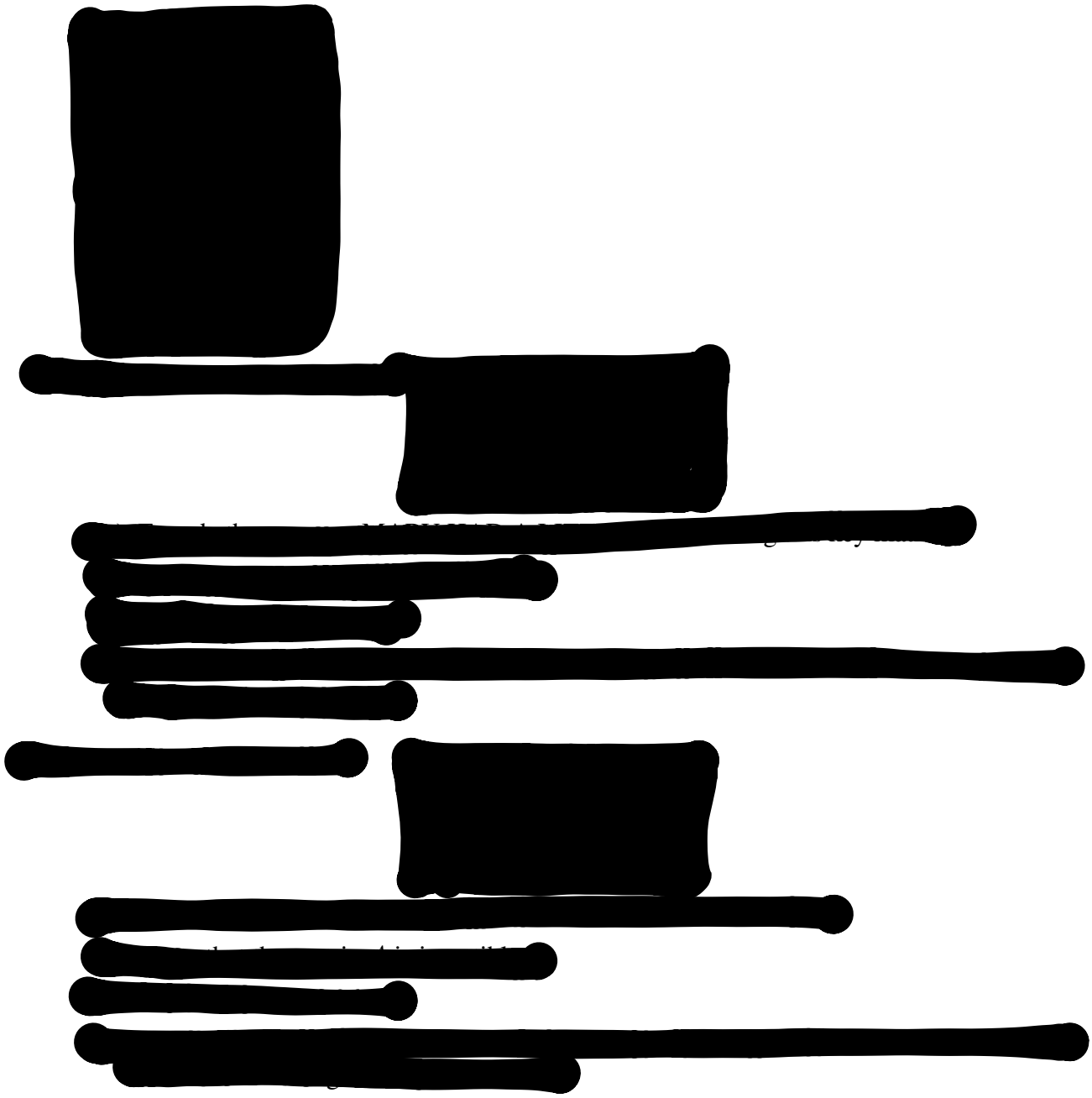
   b) 20×5

   c) $7 \begin{bmatrix} 4 \\ 12 \\ 21 \end{bmatrix} - 3 \begin{bmatrix} 14 \\ 5 \\ 16 \end{bmatrix}$

   d) $\begin{bmatrix} 11 & 22 & 14 \\ 7 & 9 & 21 \\ 17 & 0 & 3 \end{bmatrix} \begin{bmatrix} 3 & 10 & 20 \\ 20 & 9 & 17 \\ 9 & 4 & 17 \end{bmatrix}$

2. Encode the phrase BUY TEN SHARES TOMORROW using the key matrix A from the example on page 3.

3. Decode the phrase KSKBHXKDYRVTKRZTQE which was encoded using the key matrix A from the example on page 3.



References:

1. Hill, Lester. “Cryptography in an Algebraic Alphabet.” *American Mathematical Monthly*, June-July 1929, pp. 306-312.

2. Hill, Lester. “Concerning Certain Linear Transformation Apparatus of Cryptography.” *American Mathematical Monthly*, March 1931, pp. 135-154.

3. Konheim, Alan G. *Cryptography: A Primer*. New York: John Wiley and Sons, 1981.

MULTIPLICATION TABLE (mod 26)

×	0	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
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	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
2	0	2	4	6	8	10	12	14	16	18	20	22	24	0	2	4	6	8	10	12	14	16	18	20	22	24
3	0	3	6	9	12	15	18	21	24	1	4	7	10	13	16	19	22	25	2	5	8	11	14	17	20	23
4	0	4	8	12	16	20	24	2	6	10	14	18	22	0	4	8	12	16	20	24	2	6	10	14	18	22
5	0	5	10	15	20	25	4	9	14	19	24	3	8	13	18	23	2	7	12	17	22	1	6	11	16	21
6	0	6	12	18	24	4	10	16	22	2	8	14	20	0	6	12	18	24	4	10	16	22	2	8	14	20
7	0	7	14	21	2	9	16	23	4	11	18	25	6	13	20	1	8	15	22	3	10	17	24	5	12	19
8	0	8	16	24	6	14	22	4	12	20	2	10	18	0	8	16	24	6	14	22	4	12	20	2	10	18
9	0	9	18	1	10	19	2	11	20	3	12	21	4	13	22	5	14	23	6	15	24	7	16	25	8	17
10	0	10	20	4	14	24	8	18	2	12	22	6	16	0	10	20	4	14	24	8	18	2	12	22	6	16
11	0	11	22	7	18	3	14	25	10	21	6	17	2	13	24	9	20	5	16	1	12	23	8	19	4	15
12	0	12	24	10	22	8	20	6	18	4	16	2	14	0	12	24	10	22	8	20	6	18	4	16	2	14
13	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13	0	13
14	0	14	2	16	4	18	6	20	8	22	10	24	12	0	14	2	16	4	18	6	20	8	22	10	24	12
15	0	15	4	19	8	23	12	1	16	5	20	9	24	13	2	17	6	21	10	25	14	3	18	7	22	11
16	0	16	6	22	12	2	18	8	24	14	4	20	10	0	16	6	22	12	2	18	8	24	14	4	20	10
17	0	17	8	25	16	7	24	15	6	23	14	5	22	13	4	21	12	3	20	11	2	19	10	1	18	9
18	0	18	10	2	20	12	4	22	14	6	24	16	8	0	18	10	2	20	12	4	22	14	6	24	16	8
19	0	19	12	5	24	17	10	3	22	15	8	1	20	13	6	25	18	11	4	23	16	9	2	21	14	7
20	0	20	14	8	2	22	16	10	4	24	18	12	6	0	20	14	8	2	22	16	10	4	24	18	12	6
21	0	21	16	11	6	1	22	17	12	7	2	23	18	13	8	3	24	19	14	9	4	25	20	15	10	5
22	0	22	18	14	10	6	2	24	20	16	12	8	4	0	22	18	14	10	6	2	24	20	16	12	8	4
23	0	23	20	17	14	11	8	5	2	25	22	19	16	13	10	7	4	1	24	21	18	15	12	9	6	3
24	0	24	22	20	18	16	14	12	10	8	6	4	2	0	24	22	20	18	16	14	12	10	8	6	4	2
25	0	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1