**Hill Ciphers:**

**An Application of Linear Algebra**

Julia Vonessen

Gabrielle Legaspi

Saleema Qazi

**Introduction**

The hill cipher is a method of encryption invented in 1929 by Lester S. Hill. When they were invented they were the most practical polygraphic substitution cipher because the process is simple and can be used on more than three symbols, which was a unique attribute at the time. However, they never gained much popularity because even though they were the most efficient polygraphic substitution cipher, they are still fairly complex.

        Hill ciphers use modular and linear algebra to encrypt and decrypt messages. First, each letter of an alphabet is given a numerical value. Modular algebra is used to keep calculations within the range of values that represents letters. Any message can be converted into a matrix of numbers. Then this matrix is multiplied by a key matrix. It is important that the key matrix be invertible because its inverse is essential to decrypting the message. The rest of this paper will further detail the process of encryption, decryption, and how to break the cipher using sample pairings.

**Ciphering Process**

First we choose a key matrix of any size that is invertible, square, and contains nonnegative integers. For this example, we have chosen this 3x3 matrix:

A= $\begin{matrix}2&1&1\\3&2&1\\2&1&2\end{matrix}$

We want to encipher the message MATHISFUN. We assign numbers to the corresponding letters of the alphabet as shown below. We use this to create a matrix with the letters of the message.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |

$\begin{matrix}M&H&F\\A&I&U\\T&S&N\end{matrix}$ = $\begin{matrix}13&8&6\\1&9&21\\20&19&14\end{matrix}$

Then we multiply the message matrix by the key matrix.

$\begin{matrix}2&1&1\\3&2&1\\2&1&2\end{matrix}$ \*$ \begin{matrix}13&8&6\\1&9&21\\20&19&14\end{matrix}$ = $\begin{matrix}47&44&47\\61&61&74\\67&63&61\end{matrix}$

Then modulo every value of the resulting matrix by 26 (number of characters in the set). This means you divide the matrix by 26 and put the remainders into a matrix.

$\begin{matrix}47&44&47\\61&61&74\\67&63&61\end{matrix}$ = $\begin{matrix}21&18&21\\9&9&22\\15&11&9\end{matrix}$ mod26

The remainder matrix contains the code word. Use the alphabet character and number assignments to solve for the code word.

$\begin{matrix}21&18&21\\9&9&22\\15&11&9\end{matrix}$ mod26 = $\begin{matrix}U&R&U\\I&I&V\\O&K&I\end{matrix}$

MATHISFUN translates to UIORIKUVI

## Decryption of the Hill Cipher

Let’s say we receive the encrypted text UIORIKUVI and we wish to decrypt it. First, we split this message up into separate nx1 matrices, where n is the rank of our encryption matrix. Then, we replace each letter by its corresponding number.

$$B\_{1}=\left[\begin{matrix}U\\I\\O\end{matrix}\right]=\left[\begin{matrix}21\\9\\15\end{matrix}\right]$$

$$B\_{2}=\left[\begin{matrix}R\\I\\K\end{matrix}\right]=\left[\begin{matrix}18\\9\\11\end{matrix}\right]$$

$$B\_{2}=\left[\begin{matrix}U\\V\\I\end{matrix}\right]=\left[\begin{matrix}21\\22\\9\end{matrix}\right]$$

For convenience, we can combine these nx1 matrices into one larger matrix.

$$B=\left[\begin{matrix}U&R&U\\I&I&V\\O&K&I\end{matrix}\right]=\left[\begin{matrix}21&18&21\\9&9&22\\15&11&9\end{matrix}\right]$$

We know our encryption matrix, K, but for this part, we’ll need K-1. That’s why it was important that the encryption matrix was invertible. Our next step, then, is to calculate K-1 mod 26.

$$K^{-1}=\left[\begin{matrix}3&-1&-1\\-4&2&1\\-1&0&1\end{matrix}\right]\left(mod 26\right)=\left[\begin{matrix}3&25&25\\22&2&1\\25&0&1\end{matrix}\right]$$

Once we have K-1, the process of decryption is very like that of encryption. We multiply K-1 and B, find the modulus of the matrix, and transform the matrix back into letters to discover our message.

$$K^{-1}∙B=\left[\begin{matrix}3&25&25\\22&2&1\\25&0&1\end{matrix}\right]∙\left[\begin{matrix}21&18&21\\9&9&22\\15&11&9\end{matrix}\right]=\left[\begin{matrix}663&554&838\\495&425&515\\540&461&534\end{matrix}\right]$$

$$\left[\begin{matrix}663&554&838\\495&425&515\\540&461&534\end{matrix}\right](mod 26)=\left[\begin{matrix}13&8&6\\1&9&21\\20&19&14\end{matrix}\right]=\left[\begin{matrix}M&H&F\\A&I&U\\T&S&N\end{matrix}\right]$$

Thus, our decrypted message is MATHISFUN.

What if our encrypted text doesn’t fit exactly into an nxm matrix? This could be the case if we only intercepted part of a message. It turns out that it’s impossible to completely decrypt a block of text whose total number of characters isn’t divisible by n. For example, if a message consisted of 24 characters and was encrypted by a 5x5 encryption matrix, it would only be possible to decrypt the first 20 characters (Lyons).

**Breaking the Hill Cipher**

The easiest way to break the Hill Cipher is using a known ciphertext attack. With a four-letter block of text encoded with a 2x2 matrix and corresponding four letters of code, it’s possible to determine the encrypting matrix. Let’s say that we have an encrypted text that begins HHKRBEUYMXUYQNPXNMDUVKDADKWW and that we know that the first four letters correspond to the letters TOTH. We place the first four letters of each into an augmented matrix as follows.

$$\left[\begin{matrix}\begin{matrix}H&H\\K&R\end{matrix}&|&\begin{matrix}T&O\\T&H\end{matrix}\end{matrix}\right]$$

We then replace each letter with its corresponding number. Let’s say that for this message, we know that the letters were encrypted with the scheme A=0, B=1, etc. Then, we perform operations on the matrix mod 26 until we have a matrix of the form [I|(K-1)T] (Schiefloe).

$\left[\begin{matrix}\begin{matrix}7&7\\10&17\end{matrix}&|&\begin{matrix}19&14\\19&7\end{matrix}\end{matrix}\right]$ Original matrix

$\left[\begin{matrix}\begin{matrix}105&105\\10&17\end{matrix}&|&\begin{matrix}285&210\\19&7\end{matrix}\end{matrix}\right]$ mult(1,15)

$\left[\begin{matrix}\begin{matrix}1&1\\10&17\end{matrix}&|&\begin{matrix}25&2\\19&7\end{matrix}\end{matrix}\right]$ mod 26

$\left[\begin{matrix}\begin{matrix}1&1\\0&7\end{matrix}&|&\begin{matrix}25&2\\-231&-13\end{matrix}\end{matrix}\right]$ combo(1,2,-10)

$\left[\begin{matrix}\begin{matrix}1&1\\0&7\end{matrix}&|&\begin{matrix}25&2\\3&13\end{matrix}\end{matrix}\right]$ mod 26

$\left[\begin{matrix}\begin{matrix}1&1\\0&105\end{matrix}&|&\begin{matrix}25&2\\45&195\end{matrix}\end{matrix}\right]$ mult(2,15)

$\left[\begin{matrix}\begin{matrix}1&1\\0&1\end{matrix}&|&\begin{matrix}25&2\\19&13\end{matrix}\end{matrix}\right]$ mod 26

$\left[\begin{matrix}\begin{matrix}1&0\\0&1\end{matrix}&|&\begin{matrix}6&-11\\19&13\end{matrix}\end{matrix}\right]$ combo(2,1,-1)

$\left[\begin{matrix}\begin{matrix}1&0\\0&1\end{matrix}&|&\begin{matrix}6&15\\19&13\end{matrix}\end{matrix}\right]$ mod 26

$$K^{-1}=\left[\begin{matrix}6&19\\15&13\end{matrix}\right]$$

We can see that this matrix is indeed the correct decryption matrix if we use it to decrypt our text. The decrypted message turns out to be “TO THE PEOPLE OF THE UNITED STATES,” the first line of George Washington’s Farewell Address (Washington).

This same process can be used to decrypt a message encrypted by any nxn matrix, as long as we know n2 letters of each (Lyons). For example, for n=3, we need to know nine letters of both encrypted and decrypted text. If we have the text FKQVEIDPF and know that these letters also correspond to the beginning of the Farewell Address, we follow the same process to find the decryption matrix.

$$\left[\begin{matrix}\begin{matrix}F&K&Q\\V&E&I\\D&P&F\end{matrix}&|&\begin{matrix}T&O&T\\H&E&P\\E&O&P\end{matrix}\end{matrix}\right]$$

$$\left[\begin{matrix}\begin{matrix}5&10&16\\21&4&8\\3&15&5\end{matrix}&|&\begin{matrix}19&14&19\\7&4&15\\4&14&15\end{matrix}\end{matrix}\right]$$

$$\left[\begin{matrix}\begin{matrix}1&0&0\\0&1&0\\0&0&1\end{matrix}&|&\begin{matrix}24&3&9\\8&15&18\\21&7&21\end{matrix}\end{matrix}\right]$$

$$K^{-1}=\left[\begin{matrix}24&8&21\\3&15&7\\9&18&21\end{matrix}\right]$$

This is the correct decryption matrix.

One of the main advantages of the Hill cipher is that frequency analysis of single characters does not help break the code. This is because there is not a one-to-one correspondence of letters in the original message and of those in the encoded message. Instead, each block of n letters is transformed into a different block of n letters.

Frequency analysis can still be used to break the cipher, but it must be n-gram frequency, not the frequency of individual letters. For example, one common 2-gram in English is HE. To attempt to break the cipher with this method, a decoder would need to analyze the frequency of all the blocks of 2 letters that occur in the sample text, and then guess a solution. However, this method requires a very long encrypted message to be feasible (Lyons).

## Conclusion

Since the transformations of the Hill Cipher are linear, it is little challenge for modern computers to break the code. On its own, it is extremely vulnerable to known ciphertext attacks. However, because the cipher equalizes letter distribution and prevents frequency analysis of characters, the Hill Cipher and some of its variations are still used today in conjunction with other ciphers.

## References

Lyons, James. "Hill Cipher." *Practical Cryptography*. N.p., n.d. Web. 5 Mar. 2017.

Schiefloe, Paal. "Hill Ciphers: An Application of Linear Algebra." *Department of Mathematics.* University of Washington, 3 Dec. 2001. Web. 09 Apr. 2017.

Washington, George. "Farewell Address." *Archiving Early America*. Varsity Tutors, n.d. Web. 09 Apr. 2017.