Mathematics 2130
Project 3A
Cryptography

The necessity of keeping information secret is an old one. Julius Caesar is said to have employed a code based on shifting letters — for example, replacing the word `HELLO` with `IFMMP`, where each letter of `HELLO` is shifted ahead a letter. (`I` follows `H`, `F` follows `E`, and so on.) It is widely believed that the computer HAL, in the classic film *2001: A Space Odyssey*, got his name by shifting the letters of computer pioneers IBM backwards, making HAL one step ahead of IBM.

Unfortunately, such codes can easily be deciphered by unintended recipients. In fact, every single possibility could be tried, since there are only 25 of them! Instead, we need more complex methods to encrypt information. In this lab, we will examine the **Hill Cipher**, which was invented in 1929 by Lester Hill and employs some elementary linear algebra.

Here's what two parties — say, Alice and Bob — have to do when using the Hill cipher. They first agree to represent the letters of the alphabet numerically, so that `A` is 1, `B` is 2, ..., and `Z` is 26. Punctuation and capitalization is to be ignored, but spaces will still count and will be represented with the number 0. But because spaces are hard to see, a period will be printed whenever a space is intended. Alice and Bob also pick a square n by n matrix whose entries are all integers from 0 to 26, inclusive. They call this matrix K and keep it a secret, since anybody who knows K will be able to decipher their messages.

For Alice to send Bob a message, she takes her English text and converts it into a sequence of numbers. She then divides this sequence into groups of n consecutive numbers, so that each group can be treated as a vector of length n. If the last group of numbers doesn't have n numbers, she pads in a few numbers at random. For each vector p, Alice now computes

$$c = p \cdot K$$

and reduces the entries of the vector c modulo 27. She then sends c to Bob. Since c is encrypted, it is hoped that anybody else who might have access to c will be unable to decipher the message. However Bob, who knows K, can retrieve p from c by computing

$$p = c \cdot K^{-1}$$

where K^{-1} denotes the inverse, modulo 27, of the matrix K.

The principal object of this project is to decipher the text in the files `cipher2.txt` (where $n = 2$) and `cipher3.txt` (where $n = 3$), found on the course website, by determining the respective matrices K.

A good report will contain several elements. First, it will contain the two matrices K and the deciphered texts (or at least a suitable excerpt to demonstrate the success of the decryption), as well as the methodology used to obtain them. Second, since K is needed to encode a text but K^{-1} is needed to decode a text, it would be good to consider the conditions under which a matrix is invertible modulo 27. Finally, your analysis should contain a critique of the Hill cipher. In particular, you should try to illustrate strengths and weaknesses of the current implementation, and discuss ways to overcome some of these issues.

Note: Students who successfully decrypt both of the given ciphers, and who would like to undertake a further challenge for additional credit, should contact me.