Mathematics 2130
Lab 2012W–4B
Cryptography

The necessity of keeping information secret is an old one. Indeed, one such method is attributed to Julius Caesar. Caesar employed a code based on shifting letters. So, for example, the word "HELLO" might be replaced by "IFMMP", where each letter of HELLO was shifted ahead a letter. (I follows H, F follows E, and so on.) It is widely believed that the evil computer HAL, in *2001: A Space Odyssey*, got his name by shifting the letters of IBM backwards, to make 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 keep this matrix, which they call K, a secret, since anybody who knows K will be able to decipher their messages.

Now, for Alice to send Bob a message, she takes her English and converts it into a sequence of numbers. She then groups her numbers into groups of n, with each group now forming 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 object of this lab is to decipher the text in the files `cipher2.txt` (where $n = 2$) and `cipher3.txt` (where $n = 3$), both found on the class web site. A good report will contain several elements. First, it will contain the deciphered texts, as well as the methodology used to decipher them. Second, since K is needed to encode, but K^{-1} is needed to decode, it would be good to consider when a matrix K is invertible modulo 27. Finally, your analysis should contain a critique of the Hill cipher, as given. In particular, you should illustrate strengths and weaknesses of the current implementation, and perhaps discuss ways to overcome some of these weaknesses.

Note: For students seeking an additional challenge (and bonus marks!), you can also try to decipher the text found in `cipher2hard.txt`. This text exhibits significant deviations from standard linguistic patterns, and is therefore much more difficult to decrypt.