Chapter 6
The Hill Cipher

6.1 Matrices

The next cipher we will look at, known as the *Hill cipher*, is based on matrices. It is a form of a substitution cipher, except that it doesn't just substitute one *letter* for another but rather one *block* of letters for another. For example, it might swap some three-letter block with another three-letter block.

Let us start with an overview of matrices. We will just state the facts that we need, without proof, since it would take too long to give all the proofs.

Definition 6.1 Let m and n be positive integers. An $m \times n$ *matrix* is an m by n array of numbers, put into a box. Here, m denotes the number of *rows*, and n denotes the number of *columns*.

Example

$$\begin{pmatrix} 1 & 4 \\ 0 & -7 \\ 19 & 8\pi^3 \end{pmatrix}$$

is an example of a 3×2 matrix.

Typically, we use either round brackets, as above, or square brackets to delimit a matrix. Generally, we fill our matrices with numbers, for example, integers or real numbers. But we may also fill them with other things, as long as we are able to add and multiply those things. In particular, we might want to fill them with elements of $\mathbb{Z}/m\mathbb{Z}$, for some m. And that is what we will need to do for the Hill cipher: our key will be a matrix whose elements lie in $\mathbb{Z}/26\mathbb{Z}$. We will write them as numbers from 0 to 25, but we must recall that their addition and multiplication must be taken modulo 26.

Given two matrices, there is *sometimes* a way of multiplying them to get a new matrix. In particular, we can multiply an $m \times n$ matrix by an $n \times p$ matrix, and the result will be an $m \times p$ matrix. However, it is *not* possible to multiply an $m \times n$

S. Rubinstein-Salzedo, *Cryptography,* Springer Undergraduate Mathematics Series, https://doi.org/10.1007/978-3-319-94818-8_6

matrix by an $n' \times p$ matrix if $n \neq n'$. The matrix multiplication operation is a bit strange-looking, but it turns out to be a good idea for reasons that we will see in Chapter 19.

Notation Given a matrix A, we often write a_{ij} for the entry in the ith row and jth column.

Definition 6.2 Let A be an $m \times n$ matrix and B an $n \times p$ matrix. Then, we define AB to be an $m \times p$ matrix whose entry in the ith row and jth column is $\sum_{k=1}^{n} a_{ik}b_{kj}$.

Example

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} 1 \times 5 + 2 \times 7 & 1 \times 6 + 2 \times 8 \\ 3 \times 5 + 4 \times 7 & 3 \times 6 + 4 \times 8 \end{pmatrix} = \begin{pmatrix} 19 & 22 \\ 43 & 50 \end{pmatrix}.$$

If we assume that these matrices have entries in $\mathbb{Z}/26\mathbb{Z}$, then we would get $\begin{pmatrix} 19 & 22 \\ 17 & 24 \end{pmatrix}$.

Note that sometimes we can form the product AB, but *not* the product BA the other way around. And even when we can, these two products might not be equal.

Example Let

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Then

$$AB = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}, \qquad BA = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

One particularly important special case of matrix multiplication is *matrix-vector multiplication*. If we have a matrix with only one column, we call that matrix a *vector*, or perhaps a *column vector*. (By contrast, there are also *row vectors*, which have only one row. But we will not need those at the moment.) If we have an $m \times n$ matrix, then we can multiply it by a (column) vector of length n, and then result will be a vector of length m.

Example

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 2 \\ 5 \end{pmatrix} = \begin{pmatrix} 12 \\ 26 \\ 40 \end{pmatrix}.$$

There is a special matrix that does nothing when we multiply it by another matrix. This is the *identity matrix*. In fact, there is an identity matrix for each n: an $n \times n$ square matrix whose *diagonal* entries a_{ii} are 1, and everything else is 0. Here is the 4×4 identity matrix:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

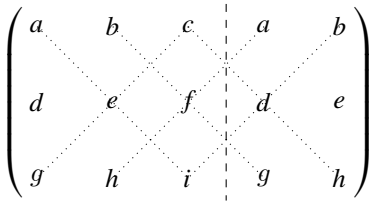
If we let I_n denote the $n \times n$ identity matrix, and A is any $m \times n$ matrix, then $AI_n = A$, and if B is any $n \times m$ matrix, then $I_n B = B$.

Given an $n \times n$ square matrix A, we might wonder if there is another $n \times n$ matrix B so that $AB = I_n$ and $BA = I_n$; this is the matrix version of a *unit*, as discussed in Definition 4.10. When such a matrix B exists, we say that A is *invertible*. We usually write A^{-1} for the matrix B, and it turns out that this inverse matrix, when it exists, is unique.

It turns out that there is a way to check if a matrix is invertible without actually finding the inverse matrix B, just as we saw that $a \in \mathbb{Z}/m\mathbb{Z}$ is a unit if and only if $\gcd(a, m) = 1$, without having to find the inverse b. This relies on the *determinant*. The determinant is slightly complicated to define in general, but we can write it down in the case of a 2×2 or 3×3 matrix. It is only defined for square matrices.

Definition 6.3 If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is an 2×2 matrix, then its determinant is defined to be $\det(A) = ad - bc$. If $B = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$ is a 3×3 matrix, then its determinant is defined to be $\det(B) = aei + bfg + cdh - ceg - afh - bdi$.

The 4×4 determinant already contains 24 terms, so the formula is cumbersome, and there are often better ways of computing it. There is a picture that helps in remembering the formula for a 3×3 determinant:



Just add up the products that go from northwest to southeast, and subtract the products that go from northeast to southwest.

It turns out (see, for instance, [Bre12, Theorem 6.2.4] for the case of matrices with entries in $\mathbb{R}$, or [DF04, Theorem 30, §11.4] for the general version) that a matrix A is invertible if and only if its determinant is a unit. Thus, if the entries are in $\mathbb{R}$, it is invertible if and only if its determinant is nonzero. If the entries are integers and we want an inverse that also has integer entries, then this happens if and only if its determinant is ± 1. If the entries are in $\mathbb{Z}/m\mathbb{Z}$, then it has an inverse if and only if $\gcd(\det(A), m) = 1$.

Although the above statement is nonconstructive in that it doesn't tell us how to find the inverse matrix A^{-1}, there is a way of writing down A^{-1} explicitly. We will only do so in the 2×2 and 3×3 cases.

Proposition 6.4 • *If* $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ *is an invertible* 2×2 *matrix, then its inverse is*

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

• *If* $A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$ *is an invertible* 3×3 *matrix, then its inverse is*

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} ei - fh & ch - bi & bf - ce \\ fg - di & ai - cg & cd - af \\ dh - eg & bg - ah & ae - bd \end{pmatrix}.$$

Here $\frac{1}{\det(A)}$ *means the reciprocal if our matrix entries are in* $\mathbb{Z}$ *or* $\mathbb{R}$ *or something similar, and it's the inverse in* $\mathbb{Z}/m\mathbb{Z}$ *if our matrix entries are in* $\mathbb{Z}/m\mathbb{Z}$.

This can be proved, in a rather unenlightening manner, simply by multiplying A by its claimed inverse and checking that the result is in fact the desired identity matrix. Based on Proposition 6.4, we can see why we need the determinant to be a unit: the formula for the inverse involves dividing by the determinant. Note that the formula for matrix inversion is a bit unwieldy to use by hand, but Sage is happy to help us out: if we type

```
A = matrix(Integers(26),[[0,1,2],[3,4,6],[8,1,17]])
A^-1
```

then we get the result

```
[22 19  6]
[ 9 22  8]
[ 9  2  9]
```

And we can check that, modulo 26, it is indeed true that

$$\begin{pmatrix} 0 & 1 & 2 \\ 3 & 4 & 6 \\ 8 & 1 & 17 \end{pmatrix}^{-1} = \begin{pmatrix} 22 & 19 & 6 \\ 9 & 22 & 8 \\ 9 & 2 & 9 \end{pmatrix}.$$

6.2 Encrypting with Matrices

The idea of the Hill cipher is to encrypt blocks of characters using matrices, replacing a block of some small number of letters with a different block of the same size. For instance, suppose we choose to encrypt blocks of length 3 by blocks of length 3. (The

plaintext and ciphertext blocks must be the same size for the Hill cipher to work.) To do this, we choose some invertible 3×3 matrix A with coefficients in $\mathbb{Z}/26\mathbb{Z}$ as our key. A common way of doing this is to start with a 9-letter word and then convert it into numbers. We will use the key `TANGERINE`. (I had to try a bunch of different 9-letter fruits before I was able to find one that was usable as a key.) To turn this into a key, write the letters of `TANGERINE` in a 3×3 matrix, as

$$\begin{pmatrix} \mathrm{T} & \mathrm{A} & \mathrm{N} \\ \mathrm{G} & \mathrm{E} & \mathrm{R} \\ \mathrm{I} & \mathrm{N} & \mathrm{E} \end{pmatrix}.$$

Next, convert each letter to a number, by replacing `A` with 0, `B` with 1, `C` with 2, and so forth, up to `Z` with 25. We get

$$\begin{pmatrix} 19 & 0 & 13 \\ 6 & 4 & 17 \\ 8 & 13 & 4 \end{pmatrix}.$$

Its determinant is 5, which is a unit modulo 26, so we're safe. (The reason I had to try a bunch of possible words for keys is that when I tried several other fruit names, such as `PERSIMMON` and `RASPBERRY`, I ended up with a determinant that is not a unit, and hence isn't usable.)

To encrypt a message, we must first convert the plaintext into numbers, using the same scheme: replacing `A` with 0, `B` with 1, and so forth. Let us suppose that we wish to encrypt the plaintext message `rhinoceros`. In order for our encryption scheme to work, we need the message length to be a multiple of the block size, which is 3. Since our message length is 10, we need to append a few random characters at the end so that the message length becomes the next multiple of 3, namely, 12. Once the message is decrypted on the other side, it will be obvious that the extra characters are simply junk and should be discarded. So, let us append `jc` at the end, making our plaintext message `rhinocerosjc`.

We now replace each letter with a number, giving us

```
17 7 8   13 14 2   4 17 14   18 9 2.
```

To form the ciphertext, we turn each block into a column vector of length 3 and multiply it by our key matrix. For instance, with the first block, we get

$$\begin{pmatrix} 19 & 0 & 13 \\ 6 & 4 & 17 \\ 8 & 13 & 4 \end{pmatrix} \begin{pmatrix} 17 \\ 7 \\ 8 \end{pmatrix} = \begin{pmatrix} 11 \\ 6 \\ 25 \end{pmatrix}.$$

Converting that back to letters, we get `LGZ`. Similarly, the other blocks encrypt to `NMI`, `YSX`, and `EWJ`. Putting all these blocks together and rearranging them into our standard length-5 blocks, we get our complete ciphertext:

```
LGZNM IYSXE WJ.
```

Knowing the key and the block length, decryption is very similar to encryption, but instead of multiplying each length-3 block by the key matrix, we multiply it by the *inverse* of the key matrix. In this case, the inverse of the key matrix is $\begin{pmatrix} 11 & 13 & 0 \\ 12 & 10 & 3 \\ 4 & 13 & 10 \end{pmatrix}$. Multiplying this by the vector $\begin{pmatrix} 11 \\ 6 \\ 25 \end{pmatrix}$ corresponding to the first block `LGZ` gives us back the first block of the plaintext, which is $\begin{pmatrix} 17 \\ 7 \\ 8 \end{pmatrix}$, or `rhi`, just as expected.

6.3 Attacking the Hill Cipher

Unfortunately, the Hill cipher is quite weak. Intuitively, this makes sense: it is defined in a very structured way, based on the operation of matrix multiplication. A common theme in cryptography is that the more mathematical structure our cryptosystem has, the weaker it tends to be, as an eavesdropper can use that structure to attack the system.

Let us first suppose that our eavesdropper has managed to acquire a ciphertext encrypted by a Hill cipher, together with a couple extra pieces of information. In particular, she knows that the block size is 2, and she knows that the block `th` encrypts to `BP`, and `at` encrypts to `YL`. It turns out that this alone is enough to calculate the entire key, and thus decrypt the ciphertext.

How can that be? Let us suppose that the 2×2 key matrix is $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Based on the two plaintext–ciphertext pairs, Eve knows that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 19 \\ 7 \end{pmatrix} = \begin{pmatrix} 1 \\ 15 \end{pmatrix}, \qquad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 0 \\ 19 \end{pmatrix} = \begin{pmatrix} 24 \\ 11 \end{pmatrix}. \tag{6.1}$$

Writing that out in terms of equations, we get the following system of four linear equations in four variables:

$$\begin{aligned} 19a + 7b &= 1 \\ 19c + 7d &= 15 \\ 0a + 19b &= 24 \\ 0c + 19d &= 11. \end{aligned}$$

(And remember, all these equations are modulo 26.) So, to find a, b, c, d, we simply solve the system of equations. In fact, this system is easier than a usual system of four equations in four variables, because it splits up into two separate systems of

equations, each of which consists of two equations in two variables. When we solve it, we find that

$$a = 15, \qquad b = 4, \qquad c = 0, \qquad d = 17.$$

Converting this back to letters, we see that the key is PEAR.

We can rephrase this process in matrix form. The pair of equation (6.1) is equivalent to the single matrix identity

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 19 & 0 \\ 7 & 19 \end{pmatrix} = \begin{pmatrix} 1 & 24 \\ 15 & 11 \end{pmatrix}.$$

We wish to solve for a, b, c, d, which means inverting the matrix $\begin{pmatrix} 19 & 0 \\ 7 & 19 \end{pmatrix}$. Its inverse is $\begin{pmatrix} 11 & 0 \\ 11 & 11 \end{pmatrix}$, which Sage is happy to tell us with the following commands:

```
plaintext = matrix(Integers(26),[[19,0],[7,19]])
plaintext^-1
```

Now that we have the inverse, we can solve for $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 24 \\ 15 & 11 \end{pmatrix} \begin{pmatrix} 11 & 0 \\ 11 & 11 \end{pmatrix} = \begin{pmatrix} 15 & 4 \\ 0 & 17 \end{pmatrix}.$$

This is the same answer as before.

More generally, if Eve knows that the block size is n, and she can somehow acquire n different plaintext–ciphertext blocks, then she can determine the key and use it to decrypt any ciphertext message. But how is Eve to get these plaintext–ciphertext blocks? Ordinarily, Alice and Bob will not be eager to provide them to Eve. So, one possibility for Eve is to get them from the ciphertext itself, if it is long enough. She may be able to do this using frequency analysis. If she decrypts a very long message and knows that the block size is 2, then it is very likely that the most common block is th in the plaintext, giving her one block. She can try various other common bigrams like he, in, and er for other common bigrams in the ciphertext. With a bit of time and patience, she should be able to find her necessary two bigrams to decrypt the ciphertext.

There is a bit of a pitfall here though: it's not enough just to have *any* n different plaintext–ciphertext blocks. Rather, they have to be sufficiently special in order for this attack to be completely successful. In the above example, the plaintext matrix $\begin{pmatrix} 19 & 0 \\ 7 & 19 \end{pmatrix}$ is *invertible*: it has an inverse, considered as a matrix with entries in $\mathbb{Z}/26\mathbb{Z}$, because its determinant is relatively prime to 26. This won't always happen: the determinant could be divisible by 2 or 13. When that happens, these plaintext–ciphertext blocks won't be sufficient to work out the key directly, but it can lead to partial information about the key. We encourage you to try problem 3 to see how to get this partial information, and to determine how it might be used.