Problem 1 (5 pts x 2) A key $K$ is involutory if $E_K(x) = D_K(x)$ for any plaintext $x$. In the case of the permutation cipher, this relation holds if the permutation $\pi$ satisfies $\pi(\pi(i)) = i$ for each $i$. $\pi(\pi(i)) = i$ holds under two cases: first, if $\pi(i) = i$; and second, if $\pi(i) = j$ for some $j \neq i$ and $\pi(j) = i$.

(a) The proof requires two direction. First, suppose there is a permutation function $\pi$ such that $\pi(\pi(i)) = j$, and then $\pi(j) = i$. Then it is clear that $\pi(\pi(i)) = i$, hence the forward direction completes. On the other hand, suppose the permutation function is an involutory key, such that $\pi(\pi(r)) = r$, for some number $r \in (1, \ldots, m)$, then suppose $\pi(r) = q$ for some number $q \in (1, \ldots, m)$, then it is clear that $\pi(q) = r$, which satisfies the requirement and hence completes proof.

(b) This problem can be solved through a recursive approach. Let $P_m$ denote the number of involutory keys of block size $m$. We have $P_1 = 1$ and $P_2 = 2$. Now, suppose that we have computed the values of $P_k$ up to $k = m - 1$. Consider an involutory permutation $\pi$ of size $m$, and in particular consider the value of $\pi(m)$. By the paragraph above, there are two cases on $\pi(m)$. First, if $\pi(m) = m$, then it suffices to choose the $\pi(i)$ for all $i \neq m$ such that an involutory permutation is formed. This is equal to the number of involutory permutations of the integers up to $(m - 1)$, which is equal to $P_{m-1}$.

Now, in the second case, $\pi(m) = i$ and $\pi(i) = m$ for some $i \neq m$. The permutation $\pi$ will be involutory if and only if the remaining $(m - 2)$ integers is an involutory permutation. There are $P_{m-2}$ such permutations for each of the $(m - 1)$ possible values of $i$, implying that there are $(m - 1)P_{m-2}$ involutory permutations of this form. Hence the total number of involutory permutations of size $m$ is equal to $P_m = P_{m-1} + (m-1)P_{m-2}$.

For $m$ up to 6, we have that $P_1 = 1$, $P_2 = 2$. $P_3$ is given by $P_3 = P_2 + 2P_1 = 4$. $P_4$ is given by $P_4 = P_3 + 3P_2 = 10$. $P_5$ is given by $P_5 = P_4 + 4P_3 = 10 + 16 = 26$. Finally $P_6$ is given by $P_6 = P_5 + 5P_4 = 26 + 50 = 76$.

Problem 2 (10 pts) The keystream for 8 sets of initial values is given in Table 1. For each keystream, the period is 8 except for (0, 0, 0).
<table>
<thead>
<tr>
<th>((z_0, z_1, z_2))</th>
<th>Key stream</th>
</tr>
</thead>
<tbody>
<tr>
<td>((0, 0, 0))</td>
<td>00000000</td>
</tr>
<tr>
<td>((0, 0, 1))</td>
<td>11201001</td>
</tr>
<tr>
<td>((0, 0, 2))</td>
<td>22102002</td>
</tr>
<tr>
<td>((0, 1, 0))</td>
<td>01112010</td>
</tr>
<tr>
<td>((0, 2, 0))</td>
<td>02221020</td>
</tr>
<tr>
<td>((1, 0, 0))</td>
<td>11120100</td>
</tr>
<tr>
<td>((2, 0, 0))</td>
<td>22210200</td>
</tr>
<tr>
<td>((1, 1, 0))</td>
<td>12202110</td>
</tr>
</tbody>
</table>

Table 1: Key streams generated for Exercise 1.19

**Problem 3 (10 pts)** Based on the sequence 1, 1, 0, 2, 2, 0, 1, 1, we have the equations

\[
c_0 \cdot 1 + c_1 \cdot 1 = 0 \mod 3
\]

\[
c_0 \cdot 1 + c_1 \cdot 0 = 2 \mod 3
\]

Combining these equations yields \(c_0 = 2\), \(c_1 = 1\). Hence the next four elements are 0, 2, 2, 0.

**Problem 4 (20 pts)** (See the attached Matlab code `hw2_prob4.m`) The first step in solving this exercise is to compute the key length, \(m\). Its procedures are listed as follows: first trying the key lengths, \(m\), from 1 to 10, we divide the text into \(m\) vector, so that each vector contains ciphertext shifted by a constant value. After that, for each vector, we need to find out the frequency of each alphabet, and compute the autocorrelation of the frequency distribution using autocorrelation function. Then the autocorrelation is compared with the standard alphabet distribution in the handout, we see that when \(m = 6\), the autocorrelation value is the highest. The reasonable guess is that \(m = 6\). By trial and error and examination of the frequency distribution (see `Ex21_b.m`), we find that G maps to e (in \(y_1\)), I maps to t (in \(y_2\)), C maps to e (in \(y_3\)), T maps to e (in \(y_4\)), X maps to e (in \(y_5\)), and S maps to e (in \(y_6\)). The key is \((2, 17, 24, 15, 19, 14)\). Instead of trial and error, you can use cyclic cross correlation for finding out the key length, the idea behind it is, each vector has some distorted English word frequency, correction by shifting them can cancel the distortion, but in order to see the difference, a cross correlation with standard English frequency can make it more obvious, so in case that’s the right shift, it will give a peak at 0.0656. The plaintext is given by

I learned how to calculate the amount of paper needed when I was at school. You multiply the square footage of the wall by the cubic contents of the floor and ceiling, combine, and double it. You then allow have the total for openings, such as windows and doors, and then you allow the other half for matching the pattern. Then you double the whole thing again to give a margin of error and then you order the paper.
Problem 5 (10 pts) This problem can be solved by doing an exhaustive search on \( m \). The attached codes `compModuloInverse.m` and `compModuloAdjoint.m` can help calculate modulo matrix inverse. The correct value of \( m = 3 \). The plaintext `breathtaking` is converted to numbers \((1, 17, 4, 0, 19, 7, 19, 0, 10, 8, 13, 6)\), and the ciphertext is converted to numbers \((17, 20, 15, 14, 9, 13, 19, 14, 8, 5, 21)\).

We know that \((1, 17, 4)K = (17, 20, 15), (0, 19, 7)K = (14, 19, 4), \) and \((19, 0, 10)K = (13, 19, 14)\). Combining these three equations, we obtain

\[
\begin{pmatrix}
1 & 17 & 4 \\
0 & 19 & 7 \\
19 & 0 & 10
\end{pmatrix}
K =
\begin{pmatrix}
17 & 20 & 15 \\
14 & 19 & 4 \\
13 & 19 & 14
\end{pmatrix}
\] (1)

and \( K \) can be computed as

\[
K = \begin{pmatrix}
1 & 17 & 4 \\
0 & 19 & 7 \\
19 & 0 & 10
\end{pmatrix}^{-1}
\begin{pmatrix}
17 & 20 & 15 \\
14 & 19 & 4 \\
13 & 19 & 14
\end{pmatrix} = K = \begin{pmatrix}
3 & 21 & 20 \\
4 & 15 & 23 \\
6 & 14 & 5
\end{pmatrix}.
\] (2)

Problem 6 (10 pts) First we can obtain the key stream \( z \) by adding the plain and ciphertext modulo 2. We obtain

\[
z = [010111011000111]
\] (3)

From this, we can form a system of equations

\[
c_1 + c_3 + c_4 = 1 \quad (4)
\]
\[
c_0 + c_2 + c_3 + c_4 = 0 \quad (5)
\]
\[
c_1 + c_2 + c_3 = 1 \quad (6)
\]
\[
c_0 + c_1 + c_2 + c_4 = 1 \quad (7)
\]
\[
c_0 + c_1 + c_3 + c_4 = 0 \quad (8)
\]

By adding the first equation with the last equation modulo 2, we obtain that \( c_0 = 1 \). After substituting \( c_0 \) in the system of equations, and by adding equations modulo 2, we obtain the information \( c_1 = c_2 = c_4 \) mod 2. Moreover, from the second equation, we have \( c_2 + c_3 + c_4 = 1 \). Since \( c_2 = c_4 \), we must have \( c_3 = 1 \), and \( c_1 = c_2 = c_4 = 0 \).

Problem 7 (10 pts) This proof uses 2 lemma, stated as follow:

Lemma 1: If a positive integer \( b \) divides \( a \), then \( b \) must be the greatest common divisor of \( a \) and \( b \), i.e. \( \gcd(a, b) = b \). Proof is omitted.

Lemma 2: For non-zero integer \( a \) and \( b \) suppose that

\[
a = bq + r \quad \text{where} \, q, r \in \mathbb{Z}
\] (9)

Then \( \gcd(a, b) = \gcd(b, r) \)

Proof: If \( c \) divides \( a \) and \( c \) divides \( b \) then \( a = cq_1 \) and \( b = cq_2 \) for some integer
$q_1, q_2$. In this case $r = a - bq = cq_1 - cq_2q = c(q_1 - q_2q)$, so that $c$ divides $r$. Thus $c$ is common divisor of $a$ and $b$ implies $c$ is a common divisor of $b$ and $r$. Conversely, if $c$ is a common divisor of $b$ and $r$ then $c$ is common divisor of $a$ and $b$. Hence the common divisors of $a$ and $b$ are the same as the common divisors of $b$ and $r$ and so their greatest common divisors are the same.

Proof of Algorithm 5.1: We first prove the statement $P(m)$, a statement that, given any sequence of positive numbers, $r_0, r_1, ..., r_m$ generated by $m$-1 step of Euclidean algorithm, $\gcd(r_0, r_1) = \gcd(r_{m-1}, r_m)$ by induction on $m$.

Base case: $P(1)$ simply says $\gcd(r_0, r_1) = \gcd(r_0, r_1)$, and it is certainly true.

Inductive step: suppose we have some $k$ greater than 1 the result is true for $m=k$.

Then, given a sequence of numbers $r_0, r_1, ..., r_k$ generated by $k$ steps of Euclidean Algorithm, $r_k = r_{k-1}q_k + r_{k+1}$, so by Lemma 2 proved above, we know $\gcd(r_{k-1}, r_k) = \gcd(r_k, r_{k+1})$. But by inductive hypothesis, $\gcd(r_0, r_1) = \gcd(r_{k-1}, r_k)$ and so $\gcd(r_0, r_1) = \gcd(r_k, r_{k+1})$. Hence, by induction the statement $P(m)$ is true for all positive integers $m$.

Now complete the proof, using the notation of theorem it follows that $\gcd(a, b) = \gcd(r_0, r_1) = \gcd(r_{m-1}, r_m)$. But since the last reminder is 0 when algorithm finishes, it follows that $r_m$ divides $r_{m-1}$, and so $\gcd(r_{m-1}, r_m) = r_m$, i.e. $\gcd(a, b) = r_m$. 

4