EE720: Home Assignment

Virendra Sule
Dept. of EE, IIT Bombay

September 29,2022

1 Instructions

1. Create a group of upto 5 students and choose a group leader for correspondence.
2. The home assignment will be a single document for the group. One evaluation (out of 20 marks) will be awarded for all the group members.
3. There should be only one submission from the group by the group leader. Make sure that your names and roll numbers are written in the common submission. Addition of your name to a group after submission will not be allowed. Please do not submit the assignment individually just make sure that your group leader has submitted and you are included as member.
4. Submissions which have parts copied will be penalized in evaluation. Hence dont copy same numbers, sage code and figures.
5. Read and understand the background and the problem tasks explained below.
6. The assignment is being uploaded in the assignment channel of the Team EE720. Deadline is sufficiently relaxed. Late submissions or zip file submissions will not be evaluated. Make sure that the group leader includes your name on the paper. No names will be added after submission.

2 Linear Complexity and Minimal Polynomial of recursive sequences defined by maps

This background has been discussed in class during the discussion on minimal polynomial computation for binary outputs streams of LFSRs. The concept of linear complexity (LC) and minimal polynomial is now extended over finite fields $\mathbb{F}_q$. Consider a finite field $\mathbb{F}_q$ and a map $F : \mathbb{F}_q^n \to \mathbb{F}_q^n$ which generates value $y = F(x)$ in $\mathbb{F}_q^n$ for a given x in $\mathbb{F}_q^n$. The map F itself may be given as an algorithm or a computer code or an algorithmic pseudocode or a polynomial function which allows us to compute y efficiently given x. A *Recursive Sequence* is defined by F in $\mathbb{F}_q^n$ for a given y as follows

$$y(k+1) = F(y(k)), k = 0, 1, 2, \ldots \tag{1}$$

where $y(0)$ is given. For $y(0) = y$ the sequence is denoted as

$$S(F, y) = \{y, F(y), F^{(2)}(y), \ldots, F^{(k)}(y), \ldots\} \tag{2}$$

where $F^{(2)}(y) = F(F(y))$ and $F^{(k)}(y) = F(F^{(k-1)}(y))$.

Now consider a finite sequence $\hat{s} = \{s_0, s_1, s_2, \ldots, s_{(2M-1)}\}$ of $2M$ numbers and define the $m \times m$ matrices (called a Hankel matrices)

$$H(m) = \begin{bmatrix} s_0 & s_1 & \ldots & s_{(m-1)} \\ s_1 & s_2 & \ldots & s_{(m)} \\ \vdots & \vdots & \ldots & \vdots \\ s_{(m-1)} & s_{(m)} & \ldots & s_{(2m-2)} \end{bmatrix}$$

This leaves the last element of the sequence $s_{(2M-1)}$ to be included in the larger matrix $H(M+1)$ whose last $(M+1) \times (M+1)$-th element is $s_{(2M-1)}$.

Definition 1. ***Linear Complexity (LC).*** *Given the finite sequence $\hat{s}$ the Linear Complexity of $\hat{s}$ over $\mathbb{F}_q$ is the largest rank of $H(m)$ in $\mathbb{F}_q$ for $m \leq M$.*

From definition it follows that the largest LC of a sequence $\hat{s}$ of even length $2M$ is $m = M$.

2.1 Minimal polynomial

Let $0 < m \leq M$ such that the Hankel matrix $H(m)$ has maximum rank (rank $H(m+j) =$ rank $H(m)$ for $j = 01, 2, \ldots, M-m$). Then the sequence $\{s_0, s_1, \ldots, s_{(2m-1)}$ satisfies the linear recurrence relations

$$s_{(m+j)} = a_0 s_j + a_1 s_{(j+1)} + \ldots + a_{(m-1)} s_{(j+m-1)} \tag{3}$$

for $j = 0, 1, 2, \ldots, (2M-m-1)$ for unique co-efficients $(a_0, a_1, \ldots, a_{(m-1)})$ in $\mathbb{F}_q$ because the linear equation

$$H(m)\alpha = h(m+1) \tag{4}$$

has the unique solution (due to the full rank of $H(m)$ and being maximal) where

$$\alpha = (a_0, a_1, \ldots, a_{(m-1)})^T$$

and $h(m+1)$ is the m-tuple vector obtained by taking the last $(m+1)$-th column of $H(m+1)$ after dropping the last component (which is $s_{(2m-1)}$).

Definition 2. ***Minimal polynomial of*** $\hat{s}$ *If $m = rank\, H(m)$ is the maximal rank of the Hankel matrices constructed from $\hat{s}$, then the polynomial defined by the unique solution α of the linear system (4)*

$$f(X) = X^m - (a_0 + a_1 X + a_2 X^2 + \ldots + a_{(m-1)} X^{(m-1)}) \tag{5}$$

is called the minimal polynomial of the sequence $\hat{s}$.

Minimal polynomial thus determines the shortest linear recurrence (3) possible in the sequence $\hat{s}$. LC of the sequence is the degree on the minimal polynomial.

2.2 Algorithm to compute the LC of a finite sequence and a solution x of the local inverse

Assume that the finite sequence $\hat{s}$ be given over a finite field $\mathbb{F}_q$ with the last element $s_{(2M-1)}$. Hence the sequence has even length $2M$. An algorithm to find the LC of the sequence and the Hankel matrix size at which the LC is achieved is computed by the following algorithm. It is required to find the Hankel matrix size k at which maximal rank is achieved and is automatically equal to k. The matrix $H(k)$ at which the rank is maximal and equal to k is required for the purpose of actually inverting the map F. See the paper [**?**] for details of the formula for solving the minimal polynomial and inverting the map F. An algorithm for computing the LC and the minimal polynomial is described below.

2.2.1 Computation of the local inverse of $y = F(x)$

If $f(X)$ is the minimal polynomial of the limited sequence

$$\{y, F(y), \ldots, F^{(2M-1)}(y)\}$$

then it is known that $a_0 \neq 0$ and a possible inverse in the periodic sequence $S(F, y)$ whose first $2M$ terms are given above is computed as

$$x = (1/a_0)[F^{(m-1)}(y) - (\sum_{i=1}^{(m-1)} a_i F^{(i-1)}(y))] \tag{6}$$

However since we have a limited sequence the data may not be sufficient to compute the correct minimal polynomial of $S(F, y)$. Hence we may have to discard a false positive solution x. A correct solution is obtained by verification of $y = F(x)$. If the above solution does not satisfy this equation we discard the solution and declare that the data is insufficient to find a local inverse. Following algorithm describes all these steps.

Algorithm. Input: $\hat{s} = \{s_1, s_2, \ldots, s_{(2M-1)}\}$.

1. Set $m = M$, construct the Hankel matrix $H(m)$.
2. Repeat
3. While $m > 1$
4. Compute $r = \operatorname{rank} H(m)$.
5. If $r = m$
6. Compute the unique solution α of (4).
7. Compute the minimal polynomial (5) $f(X)$.
8. Compute x from (6).
9. If $y = F(x)$ is satisfied
10. Return local inverse x, $LC = m$, $f(X)$.
11. EndIf (9)
12. Else Return "insufficient data to compute local inverse".

13. EndIf (5)
14. Else $m \leftarrow r$.
15. EndWhile
16. Until $m = 1$
17. End

2.2.2 Computation by the Berlekamp Massey algorithm

Fortunately the computation of the minimal polynomial of scalar sequences over $\mathbb{F}_q$ for small q and for small length sequences can be done practically by using the Berlekamp Massey algorithm whose implementation is available in SAGE. The function Berlekamp_Massey($\hat{s}, \mathbb{F}_q$) (BM) returns the minimal polynomial when the scalar sequence $\hat{s}$ of even length and the field $\mathbb{F}_q$ are specified. Hence the steps 4, 5, 6 in the above algorithm for computing the maximal rank and solution of the linear system (4) and the minimal polynomial are not needed. The algorithm is modified as follows

Algorithm using BM. Input: $\hat{s} = \{s_1, s_2, \ldots, s_{(2M-1)}\}$.

1. Compute minimal polynomial $f(X)$ of $\hat{s}$ using the BM function. (After using the scalar components of $\hat{s}$ as described below).
2. Compute x from (6).
3. If $y = F(x)$ is satisfied
4. Return local inverse x, $LC = \deg f$, $f(X)$.
5. EndIf
6. Else Return "insufficient data to compute local inverse".
7. End

(Note that if an odd length sequence is input to the BM algorithm it extends it to even length by repeating the sequence. This may give a wrong result because the rank of the Hankel matrix may increase after periodic extension of the sequence). One more important aspect of the problems you will be doing in this home paper is that the sequences will be initial parts of periodic sequences. Hence if the minimal polynomial of degree $\leq m$ exists for the given sequence $\hat{s}$ of length $2M$ then the co-efficient a_0 of the minimal polynomial will be nonzero.

2.3 Computation of minimal polynomial of vector sequences over binary field

In this project you need to face another difficulty in using the BM algorithm. The function F is not acting in $\mathbb{F}_q$ and the sequence $S(F, y)$ is not a scalar sequence over $\mathbb{F}_q$. But $S(F, y)$ can be represented by a vector sequence over $\mathbb{F}_2$. In order to use the BM algorithm we need to reduce the problem of finding the minimal polynomial of the vector sequence $S(F, y)$ in terms of scalar sequences. Following procedure does this computation. The sequences $S(F, y)$ for the functions defined below are sequences of numbers in $[0, n-1]$. Hence every

element of the sequence can be represented by its $0, 1$ binary co-ordinates after binary expansion. Hence we can consider $S(F, y)$ to be a sequence over $\mathbb{F}_2^l$ where l is the bit length of n. For an elements s_k of the sequence $S(F, y)$ there is a vector

$$s_k = [s_k(1), s_k(2), \ldots, s_k(l)]^T$$

where $s_k(i)$ is the i-th binary co-ordinate of s_k. Hence the sequence $S(F, y)$ is a collection of l sequences for each of the i-th co-ordinates. Let these co-ordinate sequences be denoted as $S(i)$. Then $S(i)$ can be given as input to BM with field $\mathbb{F}_2$ and the LC and minimal polynomial of $S(i)$ can be computed. To find the LC and minimal polynomial of $S(F, y)$ carry out following procedure (this algorithm is needed to execute item 1 of algorithm using BM described above).

2.3.1 Minimal polynomial computation for vector sequence $\hat{s}$

1. compute minimal polynomial $f_1(X)$ of $S(1)$.
2. Minimal polynomial $f \leftarrow f_1$
3. for $j = 2$ to l Verify: Boolean

   $$b = f \text{ satisfies the recurrence relation for } S(j)$$

4. alternatively compute minimal polynomial of $S(j)$ by BM and verify whether $f_j | f$
5. If $b = 0$ compute minimal polynomial f_j of $S(j)$. (Alternatively if f_j does not divide f_i)
6. $f(X) \leftarrow \text{lcm}\,(f, f_j)$.
7. return Minimal polynomial of $\hat{s}$: $f(X)$
8. return LC of $S(F, y)$: $LC = \deg f(X)$.

3 Home paper project

The project of this home paper is to find a local inverse of a map F at a value y in the image of the map, to find a value x such that $y = F(x)$. The project is to gather the frequency distribution of output values y for which the sequences of limited length give correct solution x of the inverse. If the whole sequence $S(F, y)$ was given you would have always found the correct solution. The maps F that are being investigated arise from the RSA encryption scheme. To define the maps consider two distinct primes p, q (both greater than 2), $n = pq$, e is a given number in $[1, \phi(n))$ and d is a number such that $ed = 1 \mod \phi(n)$. Two maps are defined as follows.

$$\begin{array}{lll} \text{Encryption map y=E(x)} & c = x^e & \mod n \\ \text{Decryption map y=D(x)} & m = c^x & \mod n \quad \text{for a given } m \end{array} \tag{7}$$

In the map E the choice of y is any value c in $[1, n-1]$. In the map D, y is obtained by choice of m in $[1, n-1]$. In this way a sample of multiple sequences $S(F, y)$ can be constructed by choosing y for each of the functions. The sequences $S(F, y)$ are defined for these functions as follows:

F	y	S(F,y)
E	c	$\{c, c^e \mod n, c^{e^2} \mod n, c^{e^3} \mod n \ldots\}$
D	m	$\{m, c^m \mod n, c^{c^m} \mod n \mod n, \ldots\}$

The algorithm described above is to be used to compute the minimal polynomial for each of these sequences for a sample of values of y (after choosing an input value x) and compute the inverse x using the formula (6) applied on the sequence $\hat{s}$ of length l^2 part of $S(F, y)$. The frequency of values y for which a correct inverse is found is the ratio, defined for specific functions E and D,

$$\nu(E) = \frac{N_1, \text{ the Number of values of } c \text{ for which the algorithm returns correct inverse } x}{N, \text{ Number of sample values of } c \text{ chosen}}$$

Hence the maximum number of sample values of c is $(n-1)$. Since with smallest size of 8-bit primes n will be 16-bits it is suggested in the project task that smaller sample sets be chosen.

Similarly for the function D you have

$$\nu(D) = \frac{N_1, \text{ the Number of values of } m \text{ for which the algorithm returns correct inverse } x}{N, \text{ Number of sample values of } m \text{ chosen}}$$

Hence the maximum number of sample values of m is $n-1$.

3.1 Home paper project task

1. Code the formula for F (both functions E and D) for which y is chosen for a an x, the sequence $S(F, y)$ is constructed and x is computed using the algorithm. Use the binary representation of elements of $S(F, y)$ to represent $S(F, y)$ as a sequence of l-tuple vectors over $\mathbb{F}_2$. (l is the length of n).

2. Choose 5 different sets of parameters p, q, e for functions E and D. p, q primes of minimum 8 bit length and upto 16-bit length.

3. For the algorithm E: choose a sample of $10,000$ to $50,000$ (number increased as bit sizes of p, q are increased) random values of y as input to the algorithm with each fixed parameter set p, q, e. Let this set of values of y be denoted S_y, $N = |S_y|$.

4. For the algorithm D: choose a sample of $10,000$ to $50,000$ (number increased as bit sizes of p, q increased) random values m and compute c (for the inverse d of e). The sequence is given above. Let this set of samples of m be denoted S_y.

5. Fix $M = l^2$ where $l = \log n + 1$ the length of n.

6. $N_1 = 0$

7. For y in S_y,
8. Compute the sequence $\hat{s} = S(F, y)$ for input to the algorithm.
9. Compute the solution x,
10. If the solution is verified $N_1 = N_1 + 1$.
11. Else choose next y in S_y and repeat the computation of $\hat{s}$.
12. Return N_1.
13. Compute the frequency $\nu = N_1/N$ for each of the functions E, D.
14. Generate the table with fields: Parameters p, q, e, Modulus n and size l, Sample size of the set S_y, Functions E, D, density values $\nu(E)$, $\nu(D)$, time taken on SAGEmath by the algorithm for each computation.

The home paper is being uploaded as an assignment in the team EE720. The assignment must be submitted within the deadline. Dont forget to properly submit the assignment. Only single pdf of each team's home paper is to be submitted with names of all team members. Submission should contain the following in a latex PDF file.

1. First page: Title (your choice), Names of group members.
2. Brief description of the work done.
3. Table of results as explained above.