On Linear Complexity of Finite Sequences: Coding Theory and Applications to Cryptography

Edoardo Persichetti(✉) and Tovohery H. Randrianarisoa

Florida Atlantic University, Boca Raton, FL 33434, USA
{epersichetti,trandrianarisoa}@fau.edu

Abstract. We define two metrics on vector spaces over a finite field using the linear complexity of finite sequences. We then develop coding theory notions for these metrics and study their properties. We give a Singleton-like bound as well as constructions of subspaces achieving this bound. We also provide an asymptotic Gilbert-Varshamov-like bound for random subspaces. We show how to reduce the problem of finding codewords with given Hamming weight into a problem of finding a vector of a given linear complexity. This implies that our new metric can be used for cryptography in a similar way to what is currently done in the code-based setting.

Keywords: Linear code · Linear complexity · Periodic linear complexity · Gilbert-Varshamov · Signature scheme

1 Introduction

Code-based Cryptography was informally born in 1978, when Robert J. McEliece proposed a new cryptosystem based on the hardness of decoding linear codes (binary Goppa codes) in the Hamming metric [20]. The advantage of this approach is that cryptosystems of this kind are considered safe against adversaries with access to quantum computers. More precisely, there is no known quantum algorithm that can decode a random linear code in polynomial time. After 40 years of cryptanalysis, the cryptosystem is still considered to be secure, as a general framework. However, the protocol requires the use of relatively large public keys, which may be undesirable in certain applications.

To address the key size issue, it was initially suggested to use different families of linear codes, as well as "structured" linear codes (e.g. [10,19,24]). After several years and various unsuccessful attempts, the field has stabilized, and one can say that code-based encryption/key-establishment protocols are going to be crystallized (also thanks to NIST's Standardization effort [23]) into one of two main categories: the original McEliece framework (with only minor improvements that do not affect security, e.g. [22]) or protocols based on structured parity-check codes such as QC-MDPC [21]. The former, although with its well-known limitations, provides a safe choice relying on 40 years of security history [1]. The latter,

C.-M. Cheng and M. Akiyama (Eds.): IWSEC 2022, LNCS 13504, pp. 24–44, 2022.
https://doi.org/10.1007/978-3-031-15255-9_2

instead, represents the opposite trend, namely a choice aimed at a performance advantage, which however fails to fully explore some security aspects [3].

The situation is different for code-based signature schemes, for which a satisfactory solution has yet to be found; it is worth noting that the few code-based signature schemes submitted to NIST's process were all either broken, or withdrawn. This has prompted a large body of work in recent years, trying to circumvent the traditional issues by either relying on a different coding problem [7,11] or leveraging innovative frameworks [14–16] in the Hamming metric. As we will show, the notion of weight for vectors is closely related to the notion of linear complexity for sequences. This motivates us to study the linear complexity of sequences as a new metric for coding theory, with an eye towards cryptographic applications.

1.1 Overview

Let $\mathbb{F}_q$ be the finite field of size q. We recall some notions from coding theory in the Hamming metric. Let $\mathbf{x} = (x_1, \cdots, x_n) \in \mathbb{F}_q^n$. The *Hamming weight* $w_H(\mathbf{x})$ of $\mathbf{x}$ is the number of non-zero entries of $\mathbf{x}$. If $\mathbf{x}$ and $\mathbf{y}$ are two elements of $\mathbb{F}_q^n$, we define the *Hamming distance* between $\mathbf{x}$ and $\mathbf{y}$ as $d_H(\mathbf{x}, \mathbf{y}) = w_H(\mathbf{x} - \mathbf{y})$. A *linear code* $\mathcal{C}$ of length n over $\mathbb{F}_q$ is a subspace of $\mathbb{F}_q^n$ paired with the distance d_H. The *minimum distance* of a linear code $\mathcal{C}$ is the smallest value of $d_H(\mathbf{x}, \mathbf{y})$ for any two distinct codewords $\mathbf{x}, \mathbf{y} \in \mathcal{C}$.

The most important parameters for a linear code $\mathcal{C}$ are the size q of the base field, the length n, the dimension k and minimum distance d of the code. We denote such code by $[n, k, d]$ and the field is assumed to be understood. One has to optimize the choice of these parameters for applications. For example, one typically wants to construct codes that have simultaneously large dimension and large minimum distance, while the base field should preferably be as small as possible (binary field for example). A trade-off between the minimum distance and the dimension should be considered, as captured by the Singleton bound.

Theorem 1 (Singleton bound). *Let $\mathcal{C}$ be an $[n, k, d]$ linear code over $\mathbb{F}_q$. Suppose that d is the minimum distance of $\mathcal{C}$. Then,*

$$d \leq n - k + 1.$$

Due to this, we want to have codes which maximize both the dimension and the minimum distance of the codes. Thus we want to have codes for which the inequality in the above definition is an equality. Such codes are defined as follows.

Definition 1. *An $[n, k, d]$ linear code $\mathcal{C}$ which attains the Singleton bound i.e. $d = n - k + 1$, is called a* Maximum Distance Separable (MDS) *code.*

There exist explicit instances of maximum distance separable codes. One easy construction is given by the following. Let $n = q - 1$ and let $\alpha = (\alpha_1, \cdots, \alpha_n)$

be a vector having as entries all the distinct non-zero elements of $\mathbb{F}_q$. We define the evaluation map as

$$\begin{aligned} ev_\alpha : \mathbb{F}_q[x] &\to \mathbb{F}_q^n \\ f(x) &\mapsto (f(\alpha_1), \cdots, f(\alpha_n)) \end{aligned}$$

Let $\mathbb{F}_q[x]_{<k}$ be the vector space of all polynomials of degree at most $k-1$. Then the image $\mathcal{C} = ev_\alpha\left(\mathbb{F}_q[x]_{<k}\right)$ is an MDS code. This comes from the fact that a polynomial of degree at most $k-1$ can have at most $k-1$ roots. The code we just described is called *Reed-Solomon* code.

It is this relation between the property of the roots of polynomials and the weights of vectors which is interesting for us. The following theorem is a consequence of the König-Rados Theorem [18, Chap. 6].

Theorem 2. *Let $\mathbb{F}_q^* = \{\alpha_1, \ldots, \alpha_{q-1}\}$ and let $f(x) = a_0 + a_1 x + \cdots + a_{q-2}x^{q-2}$ be a polynomial over $\mathbb{F}_q$. If $f(x)$ has $q-1-r$ roots, then $(f(\alpha_1), \cdots, f(\alpha_{q-1}))$ has (Hamming) weight r and the periodic sequence $\overline{(a_0, \cdots, a_{q-2})}$ has linear complexity r i.e. there exist $c_0, \ldots, c_{r-1} \in \mathbb{F}_q$ such that*

$$a_{i+r \bmod (q-1)} = \sum_{j=0}^{r-1} c_j a_{i+j \bmod (q-1)}, \quad \forall i \in \mathbb{N}$$

and r is the smallest for such integer.

Through Theorem 2, we can relate the linear complexity of a periodic sequence with the Hamming weight of a vector. However, we have only periodic sequences with period $q-1$. This raises the following question: what happens if we consider sequences $(a_0, \ldots, a_n)$ of any length not necessarily equal to $q-1$? Even more generally, what is the situation with any type of sequences which are not necessarily assumed to be periodic? We will answer these questions in the next sections. Our goal is to provide a theory of the linear complexity of subspaces of sequences. Such a theory can in fact be used as a basis to consider new code-based cryptosystems based on the linear complexity of sequences.

1.2 Our Contribution

Using rank metric in lieu of the Hamming metric is a popular trend in code-based cryptography, occasionally leading to interesting results [2,4]. While this approach is not always completely satisfactory and its security is still not fully explored [5,6,25], it does hint at the possibility of using other metrics for building protocols, which provides additional motivation for our work. In this paper, we strive to show that the metric connected to the linear complexity of finite sequences is viable to build cryptographic schemes. To do that, we first carefully develop the necessary coding theory notions, beginning in Sect. 2 by describing linear-feedback shift registers and some of their properties. We then present the definition of linear complexity for both arbitrary finite sequences and sequences

with a fixed period. Accordingly, in Sect. 3, we define two new metrics on $\mathbb{F}_q^n$ by considering the linear complexity of finite sequences and periodic sequences with a fixed period. We give a Singleton-like bound with respect to the new metrics and we construct optimal subspaces i.e. subspaces that achieve the bound. In the interest of space, these subspaces and their applications are described in Appendix B. We then move on to studying hard problems in this metric, which is a fundamental step to apply the metric to cryptography. Thus, in Sect. 4, we show that, given a subspace of $\mathbb{F}_q^n$, the problem of finding codewords with a given linear complexity is NP-complete. We do this for both finite and periodic sequences. The result is achieved by reducing the problem of finding a codeword with given Hamming weight to a problem of finding vectors with given linear complexity. In Sect. 5, we describe further properties of the linear complexity of sequences. We give an asymptotic Gilbert-Varshamov-like bound, which shows that most subspaces have large minimum distance with respect to the linear complexity. Furthermore, we describe techniques for solving the hard problems introduced earlier, which effectively constitute attack techniques for the schemes, and analyze their complexity. Finally, in Sect. 6 we describe a sample application to the cryptographic setting, by adapting a construction of Feneuil et al. [15] and explaining why its formulation in terms of linear complexity provides a computational advantage.

2 Linear-Feedback Shift Registers

We fix a finite field $\mathbb{F}_q$ where q is a power of a prime.

Definition 2. *A* Linear-Feedback Shift Register (LFSR) *of order l over $\mathbb{F}_q$ is an infinite sequence (a_i) over $\mathbb{F}_q$ such that there are fixed $c_j \in \mathbb{F}_q$, $j = 0, \ldots, l-1$ with,*

$$a_{i+l} = \sum_{j=0}^{l-1} c_j a_{i+j}, \quad \forall i \in \mathbb{N}.$$

The feedback polynomial *associated to (a_i) is $f(z) = z^l - \sum_{j=0}^{l-1} c_j z^j$.*

Definition 3. *Let (a_i) be an LFSR over $\mathbb{F}_q$. The* generating function *$A(z)$ associated to (a_i) is the formal power series*

$$A(z) = \sum_{i=0}^{\infty} a_i z^i.$$

Given an LFSR over $\mathbb{F}_q$ with feedback polynomial $f(z)$ and generating function $A(z)$, one can show [18, Chap. 8] that for some polynomial $g(z)$ of degree $l-1$ at most, we have

$$A(z) = \frac{g(z)}{f^*(z)},$$

where f^* is the reciprocal polynomial given by $f^*(z) = z^l f\left(\frac{1}{z}\right)$.

Definition 4 (Linear Complexity). *Given a non-zero finite sequence* $(a_i) = (a_0, \cdots, a_{n-1}) \in \mathbb{F}_q^n$*, the* linear complexity $\mathfrak{L}(a_i)$ *of the sequence is the smallest* l *such that*

$$a_{i+l} = \sum_{j=0}^{l-1} c_j a_{i+j}, \quad \forall i,\ 0 \leq i \leq n-l-1,$$

for some fixed $c_j \in \mathbb{F}_q$*. We set* $\mathfrak{L}(\mathbf{0}) = 0$*, where* $\mathbf{0} = (0, \ldots, 0)$*.*

Another family of sequences are periodic sequences.

Definition 5 (n-periodic Linear Complexity). *Let* n *be a positive integer. An infinite sequence* (a_i) *is called* n*-periodic if for all* $i \geq 0$*,* $a_{i+n} = a_i$*. Such sequences are written as* $\overline{(a_0, \ldots, a_{n-1})}$*. The linear complexity of the sequence is defined as*

$$\mathfrak{L}_p\overline{(a_0, \ldots, a_{n-1})} = \mathfrak{L}(a_0, \ldots, a_{n-1}, a_0, \ldots, a_{n-1}).$$

Remark 1. It is possible that an n-periodic sequence is l-periodic for some $l < n$. The context tells us what period we consider for our sequences.

Given a finite sequence, it is possible to compute the shortest LFSR that produces it. This can be done using the Berlekamp-Massey algorithm in $\mathcal{O}(n^2)$ field operations in $\mathbb{F}_q$ [18, Chap. 8]. Furthermore, if the linear complexity of the sequence is $n/2$, then n successive terms of the sequence are enough to uniquely find the shortest shift register.
We have the following property for the linear complexity of finite sequences.

Proposition 1. *Let* $(a_i) = (a_0, \cdots, a_{n-1})$ *be a finite sequence over* $\mathbb{F}_q$*. Then* $\mathfrak{L}(a_i) \leq n$*. Furthermore the only sequences attaining the bound* n *are of the form* $(0, \cdots, 0, a)$*, with* $a \in \mathbb{F}_q^*$*.*

Proof. We can just use an LFSR with (a_i) as initial state so that the maximum linear complexity is at most n. It is obvious that $(0, \cdots, 0, a)$ has linear complexity n. Finally, if $(a_i) = (a_0, \cdots, a_{n-1})$ is such that $a_j \neq 0$ for some j with $0 \leq j \leq n-2$, then by taking $c_i = 0$ except when $i = j$, where $c_j = a_{n-1}/a_j$, we prove that $a_{n-1} = \sum_{j=0}^{n-2} c_j a_j$ so that the linear complexity is at most $n-1$. □

The corresponding property for periodic sequences is given in the following proposition.

Proposition 2. *Let* $\overline{(a_i)} := \overline{(a_0, \cdots, a_{n-1})}$ *be an* n*-periodic sequence over* $\mathbb{F}_q$*. Then* $\mathfrak{L}_p\overline{(a_i)} \leq n$*.*

Proof. It is enough to show that the LFSR defined by the coefficients in $\mathbb{F}_q$, $(c_0, \ldots, c_{n-1}) = (1, 0, \ldots, 0)$ generates the periodic sequence with initial input $(a_0, \ldots, a_{n-1})$. Thus the linear complexity is smaller or equal to n. □

Remark 2. Unlike the case of finite sequences, there can be periodic sequences other than $\overline{(0, \cdots, 0, a)}$, with $a \in \mathbb{F}_q^*$ that attain the bound in the proposition.

As an example, we can use Theorem 2. Start with a codeword in $\mathbb{F}_q^n$ with Hamming weight $n = q-1$: then, the corresponding polynomial will have coefficients which form an n-periodic sequence of linear complexity n.

The key property of the linear complexity of sequences which will be used later is the following.

Theorem 3. *Let (a_i) and (b_i) be two finite sequences of the same length. If $(c_i) = (a_i) + (b_i)$, then*

$$\mathfrak{L}(c_i) \leq \mathfrak{L}(a_i) + \mathfrak{L}(b_i).$$

Proof. With a slight abuse of notation, we denote by (a_i) (resp. (b_i)) the LFSR generating the finite sequence (a_i) (resp. (b_i)). Suppose that these LFSR have generating functions

$$\frac{g_a(z)}{f_a^*(z)} \text{ and } \frac{g_b(z)}{f_b^*(z)},$$

respectively. Then the generating function of the LFSR generating (c_i) is

$$\frac{g_a(z)f_b^*(z) + g_b(z)f_a^*(z)}{f_a^*(z)f_b^*(z)}.$$

Thus, (c_i) can be generated by an LFSR with the feedback polynomial $f_a(z)f_b(z)$. It follows that the linear complexity of the sequence is at most $\mathfrak{L}(a_i) + \mathfrak{L}(b_i)$. □

Corollary 1. *Let $\overline{(a_i)}$ and $\overline{(b_i)}$ be two finite periodic sequences of the same period. If $\overline{(c_i)} = \overline{(a_i)} + \overline{(b_i)}$, then*

$$\mathfrak{L}_p\overline{(c_i)} \leq \mathfrak{L}_p\overline{(a_i)} + \mathfrak{L}_p\overline{(b_i)}.$$

3 Coding Theory Using Linear Complexity

Let $\mathbb{F}_q$ be a finite field and let n be a positive integer. We will consider vectors in $\mathbb{F}_q^n$ and embed them with two different metrics using the linear complexity of finite (resp. periodic) sequences.

Definition 6. *Let $\mathbf{a} = (a_0, \cdots, a_{n-1}) \in \mathbb{F}_q^n$ and $\mathbf{b} = (b_0, \cdots, b_{n-1}) \in \mathbb{F}_q^n$ be two finite sequences of n elements of $\mathbb{F}_q$ each. Then we define two distances on $\mathbb{F}_q^n$ as*

$$d_1(\mathbf{a}, \mathbf{b}) = \mathfrak{L}((a_i) - (b_i)) \text{ and } d_2(\mathbf{a}, \mathbf{b}) = \mathfrak{L}_p(\overline{(a_i)} - \overline{(b_i)}).$$

It is easy to see that both maps define a distance. We only show it for d_1, but the proof for d_2 is similar.

(i) By definition $d_1(\mathbf{a}, \mathbf{b}) = 0 \Leftrightarrow \mathbf{a} = \mathbf{b}$.
(ii) By definition of $\mathfrak{L}$, $\mathfrak{L}(a_i) \geq 0$.
(iii) The symmetry is obvious, i.e. $d_1(\mathbf{a}, \mathbf{b}) = d_1(\mathbf{b}, \mathbf{a})$.

(iv) For the triangular inequality,

$$\begin{aligned} d_1(\mathbf{a},\mathbf{b}) &= \mathfrak{L}((a_i)-(b_i)) \\ &= \mathfrak{L}((a_i)-(c_i)+(c_i)-(b_i)) \\ &\leq \mathfrak{L}((a_i)-(c_i))+\mathfrak{L}((c_i)-(b_i)), \text{ by Theorem 3} \\ &= d_1(\mathbf{a},\mathbf{c})+d_1(\mathbf{c},\mathbf{b}). \end{aligned}$$

Thus, d_1 indeed defines a distance of $\mathbb{F}_q^n$. In a similar fashion, d_2 also defines a distance.

As in traditional coding theory, we can define a subset of $\mathbb{F}_q^n$ and fix a metric d_j, $j = 1, 2$ on this set. We will derive basic coding results for this context.

Definition 7. *Let $\mathcal{S}$ be a subset of $\mathbb{F}_q^n$ together with a distance $d \in \{d_1, d_2\}$. The minimum distance d of $\mathcal{S}$ is the minimum of $d(\mathbf{a},\mathbf{b})$ for distinct $\mathbf{a},\mathbf{b} \in \mathcal{S}$. We describe the parameters of $\mathcal{S}$ as $[n, |\mathcal{S}|, d]$. In case $\mathcal{S}$ is a k-dimensional subspace of $\mathbb{F}_q^n$, then d is the minimum linear complexity of the non-zero sequences in $\mathcal{S}$ and we say $\mathcal{S}$ is an $[n, k, d]$* code *with this metric.*

Next, we inspect the bounds on a $[n, |\mathcal{S}|, d]$-subset of $\mathbb{F}_q^n$.

Theorem 4 (Singleton Bound). *Let $\mathbb{F}_q$ be a finite field of size q. Let $\mathcal{S} \subset \mathbb{F}_q^n$ be a set of elements of $\mathbb{F}_q^n$ together with a distance $d \in \{d_1, d_2\}$, with minimum distance d with respect to the metric. Then $|\mathcal{S}| \leq q^{n-d+1}$.*

Proof. It is clear that for any finite sequence (a_i), $\mathfrak{L}(a_i) \leq \mathfrak{L}_p(\overline{(a_i)})$ and therefore, for any $\mathbf{a}$ and $\mathbf{b}$ in $\mathbb{F}_q^n$, $d_1(\mathbf{a},\mathbf{b}) \leq d_2(\mathbf{a},\mathbf{b})$. Thus it is enough to show the thesis for the distance d_2. We define the linear map ϕ as

$$\begin{aligned} \phi : \mathbb{F}_q^n &\to \mathbb{F}_q^{n-d+1} \\ (a_0, \cdots, a_{n-1}) &\to \begin{pmatrix} 1 \dots 1 \end{pmatrix} \begin{pmatrix} a_0 & \dots & a_{n-d} & a_{n-d+1} & \dots & a_{n-1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{d-1} & \dots & a_{n-1} & a_0 & \dots & a_{d-2} \end{pmatrix} \end{aligned}$$

This map is constrained to be injective on $\mathcal{S}$, otherwise (if two sequences $\mathbf{a}$ and $\mathbf{b}$ were mapped to the same image) then $\mathbf{a}-\mathbf{b}$ would be mapped to zero. In this case, if we write $\mathbf{a}-\mathbf{b} = (c_0, \dots, c_{n-1})$, then

$$\begin{pmatrix} 1 \cdots 1 \end{pmatrix} \begin{pmatrix} c_0 & \dots & c_{n-d} & c_{n-d+1} & \dots & c_{n-1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ c_{d-1} & \dots & c_{n-1} & c_0 & \dots & c_{d-2} \end{pmatrix} = \begin{pmatrix} 0 \cdots 0 \end{pmatrix}.$$

Thus the last row is a linear combination of the previous rows. But this would imply that $\mathfrak{L}_p\left(\overline{(a_i)} - \overline{(b_i)}\right) \leq d-1$ i.e. $d_2(\mathbf{a},\mathbf{b}) \leq d-1$. This contradicts the minimum distance of $\mathcal{S}$. By injectivity, we must have that $|\mathcal{S}| \leq |(\mathbb{F}_q^{n-d+1})|$. □

Note that in this proof, instead of using $(1 \cdots 1)$, we can use any vector with 1 as last entry. These operations are equivalent to the puncturing operation on codes. Namely, using $(0 \cdots 0\ 1)$ is analogue to puncturing at the first $d-1$ positions.

Remark 3. In case $\mathcal{S}$ is linear of dimension k over $\mathbb{F}_q$, then the Singleton bound is $k \leq n - d + 1$.

To conclude this section, we mention the existence of structures that achieve the Singleton bound. We call these *Optimal Sets of Sequences*, and describe them briefly in Appendix B.

4 Linear Complexity Coset Weight Problems

Given that our initial motivation was the possibility of an application to cryptography, in this section we show that the problem of decoding random linear codes with respect to the linear complexity metrics d_1 and d_2 is a difficult problem. Namely, we show that some problems related to the linear complexity are NP-complete. Recall that a decisional problem $\mathcal{P}$ is said to be in NP if, for any instance of $\mathcal{P}$ with a positive answer, there is an algorithm which can verify the solution in polynomial time. A problem $\mathcal{P}$ is called NP-hard if any problem in NP can be reduced to $\mathcal{P}$ in polynomial time. If a problem is both NP and NP-hard, then it is called NP-complete. NP-complete problems are considered to be intractable. One example of an NP-complete problem, which is relevant for us, is the following (where we indicate (I) for Input and (Q) for Question).

Coset Weight Problem (CWP):

(I) A matrix $\mathbf{H} \in \mathbb{F}_q^{r \times n}$, a vector $\mathbf{b} \in \mathbb{F}_q^r$ and a non negative integer ω.
(Q) Is there a vector $\mathbf{a} \in \mathbb{F}_q^n$ such that $w_H(\mathbf{a}) \leq \omega$ and $\mathbf{a}\mathbf{H}^\top = \mathbf{b}$?

CWP was proven to be NP-complete in [13]. However, the statement in [13] is proved only for the binary field. A more general statement with arbitrary field size is proved in [8]. For our theory, we want to show that the following problems related to the linear complexity are NP-complete.

Linear Complexity Coset Weight Problem (LCCWP):

(I) A matrix $\mathbf{H} \in \mathbb{F}_q^{r \times n}$, a vector $\mathbf{b} \in \mathbb{F}_q^r$ and a non-negative integer ω.
(Q) Is there a vector $\mathbf{a} \in \mathbb{F}_q^n$ such that $\mathfrak{L}(\mathbf{a}) \leq \omega$ and $\mathbf{a}\mathbf{H}^\top = \mathbf{b}$?

Periodic Linear Complexity Coset Weight Problem (PLCCWP):

(I) A matrix $\mathbf{H} \in \mathbb{F}_q^{r \times n}$, a vector $\mathbf{b} \in \mathbb{F}_q^r$ and a non-negative integer ω.
(Q) Is there a vector $\mathbf{a} \in \mathbb{F}_q^n$ such that $\mathfrak{L}_p(\mathbf{a}) \leq \omega$ and $\mathbf{a}\mathbf{H}^\top = \mathbf{b}$?

To show that these decision problems are NP-complete, we first show that CWP can be reduced to PLCCWP. Then we show that PLCCWP can be reduced to LCCWP. To begin, we show a reduction from CWP to a more specialized problem, which we state below. Its difference with CWP is that the size of the field $\mathbb{F}_q$ is not arbitrary but it is fixed to be $q = n + 1$.

Fixed-Field Coset Weight Problem (FFCWP):

(I) A matrix $\mathbf{H} \in \mathbb{F}_q^{r\times(q-1)}$, a vector $\mathbf{b} \in \mathbb{F}_q^r$ and a non negative integer ω.
(Q) Is there a vector $\mathbf{a} \in \mathbb{F}_q^{q-1}$ such that $w_H(\mathbf{a}) \leq \omega$ and $\mathbf{a}\mathbf{H}^\top = \mathbf{b}$?

The result is proven in the following theorem.

Theorem 5. *FFCWP is NP-complete.*

Proof. The fact that FFCWP is NP is easy to see. Next, we transform an instance of CWP to an instance of FFCWP. Let $\mathbf{H} \in \mathbb{F}_q^{r\times n}$, $\mathbf{b} \in \mathbb{F}_q^r$ and ω a non-negative integer from an instance of CWP. Let $Q = q^{\lceil \log_q(n+1)\rceil}$. It is clear that $Q \geq n+1$. If $Q > n+1$, then construct the matrix $\mathbf{H}_1 \in \mathbb{F}_Q^{r\times(Q-1)}$ by appending columns of zeros to the matrix $\mathbf{H}$. Finding $\mathbf{a} \in \mathbb{F}_q^n$ such that $w_H(\mathbf{a}) \leq \omega$ and $\mathbf{a}\mathbf{H}^\top = \mathbf{b}$ is reduced to finding $(\mathbf{a}|\mathbf{0}) \in \mathbb{F}_q^{Q-1}$ such that $w_H(\mathbf{a}) \leq \omega$ and $(\mathbf{a}|\mathbf{0})\mathbf{H}_1^\top = \mathbf{b}$. Now, if there was a polynomial-time algorithm which solves FFCWP, we could use it to find $\mathbf{a}_1$ such that $\mathbf{a}_1\mathbf{H}_1^\top = \mathbf{b}$. Note that $\mathbf{a}_1$ can still be a vector over $\mathbb{F}_Q$. However, we show that we can use this to get a solution over $\mathbb{F}_q$. Due to the form of $\mathbf{H}_1$, we may assume that $\mathbf{a}_1 = (a_1, \ldots, a_n, 0, \ldots, 0)$. Now, let $Tr_{Q/q}$ be the trace function corresponding to the finite extension $\mathbb{F}_Q/\mathbb{F}_q$. We also denote by $Tr_{Q/q}(\mathbf{x})$, for any vector $\mathbf{x}$ over $\mathbb{F}_Q$, where the trace map is applied individually on the entries of $\mathbf{x}$. Then, since the matrix $\mathbf{H}_1$ and the vector $\mathbf{b}$ have entries in $\mathbb{F}_q$, we have that

$$(Tr_{Q/q}(a_1), \ldots, Tr_{Q/q}(a_n), 0, \ldots, 0)\mathbf{H}_1^\top = Tr_{Q/q}\left((\mathbf{a}|\mathbf{0})\mathbf{H}_1^\top\right) = Tr_{Q/q}(\mathbf{b}) = \mathbf{b}.$$

This gives us $(Tr_{Q/q}(a_1), \ldots, Tr_{Q/q}(a_n))\mathbf{H}^\top = \mathbf{b}$ where the vector given by $(Tr_{Q/q}(a_1), \ldots, Tr_{Q/q}(a_n))$ has entries over $\mathbb{F}_q$. Notice that the trace over $\mathbb{F}_q/\mathbb{F}_q$ can be computed in polynomial time. Therefore a polynomial-time algorithm solving FFCWP also solves CWP in polynomial time. Since CWP is NP-complete, it is NP-hard, from which it follows that FFCWP must also be NP-hard, and hence NP-complete. □

Remark 4. Switching from the field $\mathbb{F}_q$ of size q to the field $\mathbb{F}_Q$ with $Q = q^{\lceil \log_q(n)\rceil}$ does not increase the difficulty of the problem exponentially. Indeed, instead of working over the field $\mathbb{F}_q$, we just work on a field of size $Q \sim n$.

We now use the result in Theorem 5 to show that PLCCWP is also NP-complete. First of all, note that it is easy to see that PLCCWP is in NP. Next, we will need to translate the notion of Hamming distance into the notion of linear complexity. For that we recall the results from Sect. 1.

Let q be a power of a prime. Theorem 2 says that if $f(x) = f_0 + f_1x + \cdots + f_{q-2}x^{q-2}$ is a polynomial over a finite field $\mathbb{F}_q$ of size q, then the number of roots of $f(x)$ in $\mathbb{F}_q^*$ is given by $q - 1 - \omega$, where $\mathfrak{L}_p(f_0, f_1, \ldots, f_{q-2}) = \omega \leq q-1$.

Another tool that we need is how to convert a vector into a polynomial. That is done via the interpolation using a Vandermonde matrix. Suppose that

we have a finite field with q elements $\mathbb{F}_q = \{0, b_1, \ldots, b_{q-1}\}$. Then the following Vandermonde matrix is invertible.

$$V = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ b_1 & b_2 & \ldots & b_{q-1} \\ b_1^2 & b_2^2 & \ldots & b_{q-1}^2 \\ \vdots & \vdots & \ddots & \vdots \\ b_1^{q-2} & b_2^{q-2} & \ldots & b_{q-1}^{q-2} \end{pmatrix}. \tag{1}$$

Thus for any $(c_1, \ldots, c_{q-1}) \in \mathbb{F}_q^{q-1}$, there is a unique polynomial $f_0 + f_1 x + \cdots + f_{q-2}x^{q-2}$ such that $f(b_i) = c_i$. This can be computed via $(f_0, \ldots, f_{q-2}) = (c_1, \ldots, q_{q-1})V^{-1}$. We denote the map by

$$\begin{aligned} \phi : \quad \mathbb{F}_q^{q-1} &\to \mathbb{F}_q^{q-1} \\ (c_1, \ldots, c_{q-1}) &\mapsto (f_0, \ldots, f_{q-2}) \end{aligned} \tag{2}$$

Now, let us see how we can convert a linear code into a subspace of periodic sequences. Suppose that we have a finite field $\mathbb{F}_q$ with q elements. Assume that $\mathcal{C} \subset \mathbb{F}_q^{q-1}$. Let $\mathbf{c} = (c_1, \cdots, c_{q-1}) \in \mathbb{F}_q^{q-1}$. If we assume that $\{a_1, \cdots, a_{q-1}\} = \mathbb{F}_q^*$, then, via the map in Eq. (2), any $\mathbf{c}$ can be written as

$$\mathbf{c} = (f(a_1), \cdots, f(a_{q-1})),$$

for some polynomial $f(x)$ of degree $q-2$ over $\mathbb{F}_q$. Using Theorem 2 and the above discussion, we see that the Hamming weight of $\mathbf{c}$ is the same as the periodic linear complexity $\mathfrak{L}_p(\overline{f_0, f_1, \ldots, f_{q-2}})$. Therefore, we have the following correspondence.

$$\left\{ \begin{array}{c} \text{Codewords } \mathbf{c} \text{ in } \mathbb{F}_q^{q-1} \\ \text{using the Hamming weight} \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{c} \text{Finite sequences } \mathbf{c}V^{-1} \text{ in } \mathbb{F}_q^{q-1} \\ \text{using the linear complexity} \end{array} \right\} \tag{3}$$

Now, with FFCWP, we have a parity-check matrix $\mathbf{H} \in \mathbb{F}_q^{r \times (q-1)}$ and a vector $\mathbf{b} \in \mathbb{F}_q^r$. We want to find $\mathbf{c} \in \mathbb{F}_q^{q-1}$ such that $\mathbf{c}\mathbf{H}^\top = \mathbf{b}$ and $w_H(\mathbf{c}) \leq \omega$. Using the Vandermonde matrix in Eq. (1) and the correspondence (3), we can write $\mathbf{c} = \mathbf{a}V$. Thus $\mathbf{a}V\mathbf{H}^\top = \mathbf{b}$. So if we set $\mathbf{H}_1 = V\mathbf{H}^\top$, then the problem is equivalent to finding $\mathbf{a} \in \mathbb{F}_q^{q-1}$ such that $\mathbf{a}\mathbf{H}_1^\top = \mathbf{b}$ and $\mathfrak{L}_p(\mathbf{c}) \leq \omega$. In other words, solving FFCWP over $\mathbb{F}_q^{q-1}$, can be reduced to solving PLCCWP over $\mathbb{F}_q$.

Theorem 6. *Solving PLCCWP is at least as hard as solving FFCWP.*

Since by Theorem 5, solving a general instance of FFCWP is NP-complete, we can also conclude that solving PLCCWP is NP-hard. Thus, we have the following theorem.

Theorem 7. *PLCCWP is NP-complete.*

Now, because we know that $\mathfrak{L}_p(a_1, \ldots, a_n) = \mathfrak{L}(a_1, \ldots, a_n, a_1, \ldots, a_n)$, we can reduce an instance of PLCCWP to an instance of LCCWP in the following manner. Suppose we are looking for $\mathbf{a}$ such that $\mathfrak{L}_p(\mathbf{a}) = \omega \leq n$ and $\mathbf{a}\mathbf{H}^\top = \mathbf{b}$. This can be interpreted as looking for $(\mathbf{a}|\mathbf{a})$ such that $\mathfrak{L}(\mathbf{a}|\mathbf{a}) = \omega \leq n$ and $(\mathbf{a}|\mathbf{a})\mathbf{H}_1^\top = (\mathbf{b}|\mathbf{0})$, where

$$\mathbf{H}_1 = \left[\begin{array}{c|c} \mathbf{H} & \mathbf{0} \\ \hline \mathbf{I}_n & -\mathbf{I}_n \end{array}\right]$$

If there were an algorithm solving LCCWP, then we could use it with the parity-check matrix $\mathbf{H}_1$ and syndrome $(\mathbf{b}|\mathbf{0})$ to find a solution $(\mathbf{a}_1|\mathbf{a}_2)$. The identity matrix $\mathbf{I}_n$ in $\mathbf{H}_1$ ensures that $\mathbf{a}_1 = \mathbf{a}_2$, so we find a solution of the form $(\mathbf{a}|\mathbf{a})$ and therefore we get a solution to PLCCWP. Thus, we have the following theorem.

Theorem 8. *LCCWP is NP-complete.*

From Theorem 2, we have seen that there is a correspondence between linear complexity and Hamming weight. As we have seen in this section, the problem of decoding in the Hamming metric can be translated into a problem of decoding with linear complexity, where the period of the sequences is fixed. It is therefore natural to ask if we can do the converse. It is not straightforward to use the previous results. Namely, when we start with a finite field $\mathbb{F}_q$ with the Hamming metric, we end up with the field $\mathbb{F}_Q$ with the linear complexity, for $Q = q^l$, and the period of the finite sequence is fixed to be $n = Q^l - 1$. Thus, for the converse, if we start with periodic sequences with period n such that $n+1$ is not a power of a prime, we cannot use the above correspondence. However, we are going to show that, with a more general version of Theorem 2, we are still able to switch from periodic linear complexity to Hamming metric. We begin with the following.

Proposition 3 ([12])**.** *Let $\mathbb{F}_q$ be a finite field and let w be a primitive n-th root of unity lying in $\mathbb{F}_{q^m}$ for some m. The linear complexity of $\mathfrak{L}_p(\overline{a_0, \ldots, a_{n-1}})$ with $a_i \in \mathbb{F}_q$ is equal to the Hamming weight of $(c_0, \ldots, c_{n-1})$, where $c_i = \sum_{j=0}^{n-1} w^{-ij} a_j$.*

By the previous proposition, if one starts with a subspace of $\mathbb{F}_q^n$ embedded with the n-periodic linear complexity, then one can transform the problem to a Hamming-metric version over the field $\mathbb{F}_{q^m}$ in a straightforward way. In this case, the problem can be easily translated to a problem with Hamming metric, the theoretic results from Hamming metric can be translated into results in the periodic linear complexity metric. Hence, from here on, we focus on finite sequences that are not restricted to a fixed period and are measured with the distance d_1.

5 Properties of Linear Complexity

As we have seen, one can compute the linear complexity of a sequence using the Berlekamp-Massey algorithm. Thus, if a sequence has small linear complexity,

one can easily find an LFSR generating this sequence. Due to this fact, we usually want to have sequences with large linear complexity. Therefore, one important question is to know how many finite sequences have large linear complexity. Another motivation for this section is that knowing the number of sequences with a given linear complexity is important for the security aspect of a code-based cryptosystem using linear complexity as metric. In the vast majority of the traditional code-based cryptosystems, in fact, one has to randomly generate error vectors with a fixed Hamming weight. We may think of the same by replacing the Hamming weight by linear complexity. In order to parametrize the security of such scheme, one again needs to know the number of sequences with a given linear complexity. When we consider finite sequences, there is already an answer to this question [17]. As mentioned in the last paragraph of the previous section, we are only interested in finite sequences without fixed periods. Thus, we will only use the linear complexity $\mathfrak{L}$ and the distance $d = d_1$.

Theorem 9 ([17]). *Let $\omega \leq n$ be positive integers. Then, the number of sequences $(a_i) = (a_1, \ldots, a_n)$ having length n and linear complexity $\mathfrak{L}(a_i) = \omega$ over a finite field $\mathbb{F}_q$ of size q is given by*

$$\begin{cases} 1 & \text{if } \omega = 0, \\ q^{2\omega-1}(q-1) & \text{if } \omega \leq \lfloor \frac{n}{2} \rfloor, \\ q^{2(n-\omega)}(q-1) & \text{if } \omega > \lfloor \frac{n}{2} \rfloor. \end{cases}$$

Theorem 10. *Given two integers $\omega \leq n$, the number $b(n, \omega)$ of finite sequences $(a_i) = (a_1, \ldots, a_n)$ having length n and linear complexity $\mathfrak{L}(a_i)$ at most ω over a finite field $\mathbb{F}_q$ of size q is*

$$\begin{cases} 1 & \text{if } \omega = 0, \\ \dfrac{q^{2\omega+1}+1}{q+1} & \text{if } \omega + 1 \leq n - \omega, \\ \dfrac{1-q^{2(n-\omega)}}{1+q} + q^n & \text{if } n - \omega \leq \omega. \end{cases}$$

Proof. Direct computation from Theorem 9. □

Since we also know the size of balls with respect to the linear complexity from Theorem 10, we can give a formula for the sphere packing bound.

Theorem 11 (Sphere Packing Bound). *Let $\mathcal{S}$ be a set of sequences of length n and with minimum distance d. Then*

$$|\mathcal{S}| \leq \begin{cases} \dfrac{q^n(q+1)}{q^{2\lfloor \frac{d-1}{2} \rfloor}+1} & \text{if } 2\lfloor \frac{d-1}{2} \rfloor \leq n-1, \\ \dfrac{q^n(q+1)}{1-q^{2(n-\lfloor \frac{d-1}{2} \rfloor)}+(1+q)q^n} & \text{if } 2\lfloor \frac{d-1}{2} \rfloor > n-1. \end{cases}$$

Proof. This is a direct consequence of Theorem 10 and uses the fact that the union of the spheres of radius $\lfloor \frac{d-1}{2} \rfloor$ centered at the sequences in $\mathcal{S}$ is a disjoint union. □

Our next theorem is the analogue to the Gilbert-Varshamov bound.

Theorem 12 (Gilbert-Varshamov Bound for Linear Complexity). *Let $d \leq n$ be positive integers. Let $A_q(n,d)$ be the size of the largest possible subset $\mathcal{S}$ of $\mathbb{F}_q^n$ with minimum distance d with respect to the metric d_1. Then*

$$\begin{cases} A_q(n,d) = q^n & \text{if } d = 1, \\ A_q(n,d) \geq \dfrac{q^n(q+1)}{q^{2d-1}+1} & \text{if } d \leq n-d+1, \\ A_q(n,d) \geq \dfrac{q^n(q+1)}{1+q^n(q+1)-q^{2(n-d-1)}} & \text{if } d \geq n-d+2. \end{cases}$$

Proof. We follow the proof in the classical Hamming metric. When $d = 1$, the result is trivial. Suppose that $|\mathcal{C}| = A_q(n,d)$. Because of the maximality of $\mathcal{S}$, any elements of $\mathbb{F}_q^n$ should be contained in a ball $B(\mathbf{x}, d-1)$, with center $\mathbf{x}$ and radius $d-1$, for some $\mathbf{x} \in \mathcal{S}$. Thus $\mathbb{F}_q^n = \cup_{\mathbf{x}\in\mathcal{S}} B(\mathbf{x}, d-1)$. Thus, we have $|\mathbb{F}_q^n| \leq |\mathcal{S}| b(n, d-1)$. The results follow from Theorem 10. □

The following is a version of the Gilbert-Varshamov bound for linear spaces of sequences.

Theorem 13 (Gilbert-Varshamov Bound for Linear Spaces). *Let $d \leq n$ be positive integers. Let $\mathcal{D}_q(n,d)$ be the dimension of the largest possible subspace $\mathcal{S}$ of $\mathbb{F}_q^n$ with minimum distance d with respect to the metric d_1. Then*

$$\begin{cases} \mathcal{D}_q(n,d) = n & \text{if } d = 1, \\ \mathcal{D}_q(n,d) \geq \log_q\left(\dfrac{q^n(q+1)}{q^{2d-1}+1}\right) & \text{if } d \leq n-d+1, \\ \mathcal{D}_q(n,d) \geq \log_q\left(\dfrac{q^n(q+1)}{1+q^n(q+1)-q^{2(n-d-1)}}\right) & \text{if } d \geq n-d+2. \end{cases}$$

Proof. Again, the case $d = 1$ is trivial. For a non-zero vector $\mathbf{x} \in \mathbb{F}_q^n$, we denote by $\langle \mathbf{x} \rangle$, the one-dimensional $\mathbb{F}_q$-space generated by $\mathbf{x}$. Now, if $\mathcal{S}$ has maximal dimension, say k, then for any element $\mathbf{x} \in \mathbb{F}_q^n \backslash \mathcal{S}$, the space $\mathcal{S} +_{\mathbb{F}_q} \langle \mathbf{x} \rangle$ should contain an elements of linear complexity smaller than d. Thus, there is $\mathbf{a} \in \mathcal{S}$ and $b \in \mathbb{F}_q^*$ such that $\mathbf{a} + b\mathbf{x}$ has linear complexity at most $d-1$. Thus $\mathbf{x} \in B(\mathbf{a}, d-1)$. On the other hand, if $\mathbf{x} \in \mathcal{S}$ then $\mathbf{x} \in B(\mathbf{x}, d-1)$. Thus we get to the same proof of the previous theorem: $\mathbb{F}_q^n = \cup_{\mathbf{x}\in\mathcal{S}} B(\mathbf{x}, d-1)$. The results follow. □

The bounds in Theorems 11, 12 and 13 were given for the reader to compare to the case of linear codes equipped with the Hamming metric. However, we have already seen a bound on the maximum size of set of sequences with a given minimum distance (see Theorem 4) and we have shown that the bound is attained for any parameters and without restriction on the base field. Now, we want to give a criteria for the optimal subspaces of sequences. To do this, for any integer $t < n$ and a vector $\mathbf{b} = (b_1, \ldots, b_t)$, we define the matrix $\mathbf{M_b} \in \mathbb{F}_q^{n\times(n-t)}$ by

$$\mathbf{M_b} = \begin{pmatrix} b_1 & 0 & \dots & 0 \\ \vdots & b_1 & \dots & 0 \\ b_t & \vdots & \ddots & \vdots \\ -1 & b_t & \ddots & b_1 \\ 0 & -1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & b_t \\ 0 & \dots & 0 & -1 \end{pmatrix}$$

Theorem 14. *Let $\mathcal{S}$ be an $[n,k]$ subspace of sequences over $\mathbb{F}_q$ and let $k \leq n-d+1$. Let $\mathbf{G}$ be a generator matrix of $\mathcal{S}$. Then the following statements are equivalent:*

(i) The minimum distance of $\mathcal{S}$ is d.
(ii) There exists a vector $\mathbf{c} = (c_1, \dots, c_d) \in \mathbb{F}_q^d$ such that $\mathbf{GM_c}$ has rank strictly smaller than k. Furthermore, $\mathbf{GM_b}$ has full rank k for any vector $\mathbf{b} = (b_1, \dots, b_{d-1}) \in \mathbb{F}_q^d$.

Proof. Suppose that the minimum distance is d. Because no element of $\mathcal{S}$ has linear complexity smaller than d, then if $\mathbf{a} = (a_1, \dots, a_n) \in \mathcal{S}$ no coefficients $\mathbf{b} = (b_1, \dots, b_{d-1})$ can generate $\mathbf{a}$ with initial state $a_1, \dots, a_{d-1}$. Thus we have that $(m_1, \dots, m_k)\mathbf{GM_b} \neq \mathbf{0}$ for any $(m_1, \dots, m_k) \in \mathbb{F}_q^k$. Therefore $\mathbf{GM_b} \in \mathbb{F}_q^{k\times(n-d)}$ has no left kernel i.e. it has full rank k. In a similar fashion, if there is a codeword of linear complexity d, then we can find $\mathbf{c} = (c_1, \dots, c_d)$ such that $\mathbf{GM_c} \in \mathbb{F}_q^{k\times(n-d)}$ has non-empty left kernel and thus its rank is smaller than k. The converse can be proved using the same idea in reverse fashion. □

Corollary 2. *Let $\mathcal{S}$ be an $[n,k,d]$ subspace of sequences over $\mathbb{F}_q$. Then $\mathcal{S}$ is optimal, i.e. $d = n-k+1$, if and only if $\mathbf{GM_b} \in \mathbb{F}_q^{k\times k}$ is invertible for any $\mathbf{b} = (b_1, \dots, b_{n-k}) \in \mathbb{F}_q^{n-k}$. In particular $\mathcal{S}$ has a generator matrix of the form $\mathbf{G} = [\mathbf{X}|\mathbf{I}_k]$.*

Proof. A direct consequence of Theorem 14. □

The previous corollary gives a characterization of optimal subspaces of sequences. Our next step is to give a bound on the minimum distance of random subspaces. This follows a method analogous to the asymptotic Gilbert-Varshamov bound in the Hamming metric case (See [9] for example).

Fix a positive integer $1 \leq d \leq n$. Let $\mathbf{G}$ be a matrix in $\mathbb{F}_q^{k\times n}$ chosen uniformly at random. Suppose that $\mathcal{S}_\mathbf{G}$ is the row space of $\mathbf{G}$. Let P be the probability that the minimum distance $d(\mathcal{S}_\mathbf{G})$ of $\mathcal{S}_\mathbf{G}$ is strictly smaller than d i.e.

$$P = Prob\left(d(\mathcal{S}_\mathbf{G}) < d\right) = Prob\left(\exists \mathbf{x} \in \mathbb{F}_q^k \backslash \{\mathbf{0}\} \colon \mathfrak{L}(\mathbf{xG}) < d\right)$$

It is clear that

$$P \leq \sum_{\mathbf{x} \in \mathbb{F}_q^k \setminus \{\mathbf{0}\}} Prob\left(\mathfrak{L}(\mathbf{xG}) < d\right).$$

Now, because $\mathbf{G}$ is a uniformly random variable, so is $\mathbf{xG}$. Thus

$$Prob\left(\mathfrak{L}(\mathbf{xG}) < d\right) = \frac{b(n, d-1)}{q^n}.$$

Thus

$$P \leq (q^k - 1)\frac{b(n, d-1)}{q^n}.$$

Thus we have the following theorem.

Theorem 15. *Let $\mathbf{G}$ be a random $(k \times n)$ matrix over $\mathbb{F}_q$ and let $\mathcal{S}_{\mathbf{G}}$ be the row space of $\mathbf{G}$ over $\mathbb{F}_q$. Let $d < n/2$ be the minimum distance of $\mathcal{S}$ and let $\epsilon > 0$, where $k = n - 2d - \epsilon$. Then $Prob\left(d(\mathcal{S}_{\mathbf{G}}) < d\right) \leq \frac{2}{q^2 q^{\epsilon n}}$.*

Proof. Let $P = Prob\left(d(\mathcal{S}_{\mathbf{G}}) < d\right)$. From the previous paragraph, we have

$$P \leq (q^k - 1)\frac{b(n, d-1)}{q^n}.$$

Because $d/n < 1/2$, then by Theorem 10,

$$P \leq (q^k - 1)\frac{q^{2d-1} + 1}{(q+1)q^n} \leq \frac{2q^k q^{2d-1}}{q^{n+1}}$$

Thus

$$P \leq \frac{2}{q^2 q^{n-k-2d}},$$

and the result follows. □

Now, in Theorem 15, $q^{-\epsilon n}$ decreases exponentially with respect to n. Thus, we can conclude the following.

Corollary 3. *With high probability, a random $k \times n$ matrix over $\mathbb{F}_q^n$ generates a space of sequences with minimum distance at least $\frac{n-k}{2}$.*

6 Cryptographic Applications

In this section, we illustrate one possible application of our theory to cryptography. Namely, we show how a recent signature scheme by Feneuil et al. [15], which uses the popular "MPC-in-the-head" paradigm, can be formulated in terms of linear complexity, and how this leads to an improvement. Due to space constraints, we are not able to describe the signature scheme in full; instead, we summarize the relevant part of the scheme, and present our proposed modification.

Let $\mathbf{H}$ be a parity-check matrix of a random $[n,k]$ code and let $\mathbf{y} \in \mathbb{F}_q^{n-k}$. For the purpose of verification, a prover wants to prove that he knows $\mathbf{x} \in \mathbb{F}_q^n$ such that $\mathbf{x}\mathbf{H}^T = \mathbf{y}$ and $w_H(\mathbf{x}) \leq w$. The prover does not want to reveal information about $\mathbf{x}$. Note that, by taking $\mathbf{H} = [\mathbf{H}'|\mathbf{I}_{n-k}]$, we can write $(\mathbf{x}_A|\mathbf{x}_B)\mathbf{H}^T = \mathbf{y}$ for $\mathbf{x} = (\mathbf{x}_A|\mathbf{x}_B)$. In this case, $\mathbf{x}_A$ uniquely determines $\mathbf{x}$ from $\mathbf{y}$ and $\mathbf{H}$.

Following the notation of [15], let $\mathbb{F}_{\text{poly}}$ be a finite extension of $\mathbb{F}_q$ such that $n \leq |\mathbb{F}_{\text{poly}}|$ and let $\{\gamma_1, \ldots, \gamma_n\}$ be distinct elements of $\mathbb{F}_{\text{poly}}$. Let $S(X) \in \mathbb{F}_{\text{poly}}[X]$ be the polynomial interpolation of the points (γ_i, x_i). It is easily seen that the condition $w_H(\mathbf{x}) \leq r$ is equivalent to $S(x)$ having at least $n - w$ roots in $\{\gamma_1, \cdots, \gamma_n\}$. In [15], it is shown that this is equivalent to the existence of two polynomials $P, Q \in \mathbb{F}_{\text{poly}}[X]$ such that $Q \cdot S - P \cdot F = 0$, where $\deg P \leq w - 1$, $\deg Q = w$ and $F = \prod_{i=1}^{n}(X - \gamma_i)$. In order to prove his knowledge, the prover does the following.

(1) Write $\mathbf{x}_A = \sum_{j=1}^{N} \mathbf{x}_A^{(j)}$. These define $\mathbf{x} = \sum_{j=1}^{N} \mathbf{x}^{(j)}$ and ensures that the syndrome relation $\mathbf{x}\mathbf{H}^T = \mathbf{y}$ is satisfied. The elements of these sums are what we call the shares in the MPC protocol.
(2) Find the interpolation polynomial $S^{(j)}(X)$ using the points $(\gamma_i, x_i^{(j)})$, where $i = 1, \ldots, n$ and $\mathbf{x}^{(j)} = (x_1^{(j)}, \ldots, x_n^{(j)})$. By the linearity of the Lagrange interpolation, $S(X) = \sum_{j=1}^{N} S^{(j)}(X)$.
(3) Write $Q(X) = \sum_{j=1}^{N} Q^{(j)}(X)$.
(4) Write $(P \cdot F)(X) = \sum_{j=1}^{N} (P \cdot F)^{(j)}(X)$.
(5) To verify that $Q(X)S(X) = (P \cdot F)(X)$. One can verify that $Q(r_l)S(r_l) = (P \cdot F)(r_l)$ for $1 \leq l \leq r$ and r_j elements of an extension $\mathbb{F}_{\text{points}}$ of $\mathbb{F}_{\text{poly}}$.
(6) To make this verification without revealing $Q(r_l)$ and $S(r_l)$, one needs to use the decompositions $Q(r_l) = \sum_{j=1}^{N} Q^{(j)}(r_l)$, $S(r_l) = \sum_{j=1}^{N} S^{(j)}(r_l)$ and $(P \cdot F)(r_l) = \sum_{j=1}^{N} (P \cdot F)^{(j)}(r_l)$ in an MPC protocol.

For full details about the usage of these steps in a zero-knowledge protocol for syndrome decoding, we refer the reader to [15].

In Step (2), the prover is required to make several of interpolations to find the polynomials $S^{(j)}(X)$. These computations negatively affect the performance of the scheme. In the following, we explain how to use a system with periodic linear complexity as metric, and completely avoid the interpolation steps, thereby considerably speeding up the scheme of [15]. In the remaining part of this section, we set $n = q - 1$ and therefore we can also choose $\mathbb{F}_{\text{poly}} = \mathbb{F}_q$.

Let $\mathbf{H}$ be a parity check matrix of a random $[n,k]$ code and let $\mathbf{y} \in \mathbb{F}_q^{n-k}$. Now, a prover wants to show that he knows $\mathbf{a} \in \mathbb{F}_q^n$ such that $\mathbf{a}\mathbf{H}^T = \mathbf{y}$ and $\mathfrak{L}_p(\mathbf{a}) \leq w$, without revealing information about $\mathbf{a}$. Again, we take $\mathbf{H} = [\mathbf{H}'|\mathbf{I}_{n-k}]$ and we can write $(\mathbf{a}_A|\mathbf{a}_B)\mathbf{H}^T = \mathbf{y}$ for $\mathbf{a} = (\mathbf{a}_A|\mathbf{a}_B)$.

By Theorem 2 and Eq. (3), if $\mathbf{a} = (a_0, \ldots, a_{n-1})$ and $S(X) = \sum_{i=0}^{q-2} a_i X^i$, then $w_H(S(\gamma_0), \ldots, S(\gamma_{q-2})) = \mathfrak{L}_p(\mathbf{a})$, where $\mathbb{F}_q^* = \{\gamma_0, \ldots, \gamma_{n-1}\}$. Using the same method as before, showing that $\mathfrak{L}_p(\mathbf{a}) \leq w$ is therefore the same as showing the existence of two polynomials $P, Q \in \mathbb{F}_{\text{poly}}[X]$ such that $Q \cdot S - P \cdot F = 0$, where $\deg P \leq w-1$, $\deg Q = w$ and $F = \prod_{i=1}^{n}(X - \gamma_i)$. The difference with the scheme in the Hamming metric is that the polynomial $S(X)$ is already defined by $\mathbf{a}$. Thus, no interpolation is needed, as claimed. In general, these are the steps the prover needs to follow.

(1') Write $\mathbf{a}_A = \sum_{j=1}^{N} \mathbf{a}_A^{(j)}$. This defines $\mathbf{a} = \sum_{j=1}^{N} \mathbf{a}^{(j)}$ and ensures that the syndrome relation $\mathbf{a}\mathbf{H}^T = \mathbf{y}$ is satisfied. The elements of these sums are the shares in the MPC protocol.
(2') The coefficients of $\mathbf{a}^{(j)}$ define a polynomial $S^{(j)}(X)$. By linearity, we have $S(X) = \sum_{j=1}^{N} S^{(j)}(X)$.
(3') Write $Q(X) = \sum_{j=1}^{N} Q^{(j)}(X)$.
(4') Write $(P \cdot F)(X) = \sum_{j=1}^{N} (P \cdot F)^{(j)}(X)$.
(5') To verify that $Q(X)S(X) = (P \cdot F)(X)$, one can verify that $Q(r_l)S(r_l) = (P \cdot F)(r_l)$ for $1 \leq l \leq r$ and r_j elements of an extension $\mathbb{F}_{\text{points}}$ of $\mathbb{F}_{\text{poly}}$.
(6') To perform this verification without revealing $Q(r_l)$ and $S(r_l)$, one needs to use the decompositions $Q(r_l) = \sum_{j=1}^{N} Q^{(j)}(r_l)$, $S(r_l) = \sum_{j=1}^{N} S^{(j)}(r_l)$ and $(P \cdot F)(r_l) = \sum_{j=1}^{N} (P \cdot F)^{(j)}(r_l)$ in an MPC protocol.

As mentioned above, since in this setting we have $n = q - 1$ and $\mathbb{F}_{\text{poly}} = \mathbb{F}_q$, Eq. (3) shows that syndrome decoding of the form $\mathbf{x}\mathbf{H}^T = \mathbf{y}$ and $w_H(\mathbf{x}) \leq w$ is equivalent to syndrome decoding of the form $\mathbf{a}\mathbf{H}_1^T = \mathbf{y}$. In this regard, the parameter sets for the Hamming metric are exactly the same parameter sets for the periodic linear complexity metric. In order to find the best parameters for a security of the scheme with the linear complexity, we can therefore use parameters from the Hamming metric. We can for example use a similar set of parameters as in the Variant 3 described in [15], working on a field $\mathbb{F}_q = \mathbb{F}_{\text{poly}} = \mathbb{F}_{256}$, and using a code of length $n = q - 1 = 255$ and dimension $k = 128$. The weight of the secret key $\mathbf{a}$ in this case is $w = 80$. An implementation of the scheme of [15] in this new metric is planned as future work, as well as a translation to the (non-periodic) linear complexity setting.

Acknowledgements. Part of this work was supported by SNF grant no. 169510 when T. Randrianarisoa was at the University of Zurich. The work is also partially funded by the National Science Foundation (NSF) grant CNS-1906360.

A The Berlekamp-Massey Algorithm

Algorithm 1. Berlekamp-Massey

procedure BM($s_0, \cdots, s_{n-1}$)
  $f(z) \leftarrow 1$, $A(z) \leftarrow 1$,
  $L \leftarrow 0$, $m = -1$, $e \leftarrow 1$
  for i from 0 to $n-1$ **do**
    $d \leftarrow s_i + \sum_{j=1}^{L} f_j s_{i-j}$
    if $d \neq 0$ **then**
      $B(z) \leftarrow f(z)$
      $f(z) \leftarrow f(z) - (d/e)A(z)z^{i-m}$
      if $2L \leq i$ **then**
        $L \leftarrow i + 1 - L$
        $m \leftarrow i$
        $A(z) \leftarrow B(z)$
        $e \leftarrow d$
      end if
    end if
  end for
  return L and $f(z)$
end procedure

B Optimal Sets of Sequences

Definition 8 (Optimal Sets of Sequences). *We call a set $\mathcal{S} \subset \mathbb{F}_q^n$ an* Optimal Set of Sequences (OSS) *(resp.* Optimal Set of Periodic Sequences (OSPS)*) if the minimum distance with respect to the metric d_1 (resp. d_2) of $\mathcal{S}$ reaches the bound of the previous theorem i.e. if $\mathcal{S}$ has elements of length n and minimum distance d and $\sharp\mathcal{S} = q^{n-d+1}$.*

Example 1. Let $\mathcal{S}$ be the set of sequences of length n over a finite field $\mathbb{F}_q$ defined by

$$\mathcal{S} = \{(0, \cdots, 0, a_1, \cdots, a_k) : a_i \in \mathbb{F}_q\}.$$

Then, $\mathcal{S}$ is both an OSS an OSPS of dimension k. That is because the sequences cannot be generated by an LFSR of length smaller than $n - k + 1$ except when it is the zero sequence.

The nice property of using the set of sequences with the linear complexity as a metric is that, in opposite to maximum distance separable codes in the Hamming metric, we can have an optimal set of sequences for any parameters. The construction works even for the binary field. Furthermore, the decoding of OSS given in Examples 1 is straightforward. They are similar and we will only describe it for the OSS in Example 1. First let us look at the unique decoding property.

Proposition 4. *Suppose that $\mathcal{S}$ is an $[n, M, d]$ set of sequences. Suppose that $\mathbf{y} \in \mathbb{F}_q$ is equal to $\mathbf{x} + \mathbf{e}$, where $\mathbf{x} \in \mathcal{S}$ and $\mathfrak{L}(\mathbf{e}) < \frac{d}{2}$. Then, the decomposition $\mathbf{x} + \mathbf{e}$ is unique.*

Proof. If $\mathbf{y} = \mathbf{x}_1 + \mathbf{e}_1 = \mathbf{y}_2 + \mathbf{e}_2$, then $\mathbf{x}_1 - \mathbf{x}_2 = \mathbf{e}_2 - \mathbf{e}_1$. Therefore $d(x_1, x_2) = \mathfrak{L}(\mathbf{e}_2 - \mathbf{e}_1)$. By Theorem 3, $d(x_1, x_2) \leq \mathfrak{L}(\mathbf{e}_2) + \mathfrak{L}(\mathbf{e}_1) < d$. This is in contradiction with the minimum distance of $\mathcal{S}$.

Let $\mathcal{S}$, of dimension k, be the OSS in Example 1. Suppose that we know $\mathbf{y} = \mathbf{x} + \mathbf{e}$ with $\mathbf{x} \in \mathcal{S}$ and $\mathfrak{L}(\mathbf{e}) < \frac{n-k+1}{2}$. By Proposition 4, we know that $\mathbf{e}$ is unique. Since the $n - k$ first entries of $\mathbf{x}$ are equal to zero. Then we know the first $n - k$ entries of $\mathbf{e}$. Now, since $\mathfrak{L}(\mathbf{e}) < \frac{n-k+1}{2}$, we can uniquely recover the LFSR generating $\mathbf{e}$ by using Berlekamp-Massey. on the first $n - k$ entries of e. We are therefore able to produce the whole $\mathbf{e}$ and then we compute $\mathbf{x} = \mathbf{y} - \mathbf{e}$. By Proposition 4, the resulting $\mathbf{x}$ is the only correct original codeword.

C Application for Decoding Reed-Solomon Codes

We can use linear complexity to get a decoding algorithm for Reed-Solomon coded (see Sect. 1). Let $\mathbb{F}_q^* = \{\alpha_1, \ldots, \alpha_n\}$, where $n = q - 1$. The Reed Solomon code $\mathcal{C}$ is defined as

$$\mathcal{C} = \{(f(\alpha_1), \ldots, f(\alpha_n)) \colon f(x) \in \mathbb{F}_q[x],\ \deg f(x) \leq k - 1\}.$$

Assume that the received codeword is $\mathbf{c} + \mathbf{e}$ and $w_H(\mathbf{e}) \leq \frac{n-k+1}{2}$. By Theorem 2, $\mathbf{c}$ corresponds to a polynomial $f_\mathbf{c}$ of degree at most $k - 1$, and $\mathbf{e}$ corresponds to a polynomial $f_\mathbf{e}$ of degree at most $q - 2$. The first step of decoding is to interpolate $\mathbf{c} + \mathbf{e}$ to get $f_\mathbf{c} + f_\mathbf{e}$. Now, since $f_\mathbf{c}$ has degree at most $k - 1$, the last $n - k + 1$ coefficients of $f_\mathbf{e}$ are the same as the last $n - k + 1$ coefficients of $f_\mathbf{c} + f_\mathbf{e}$. Since $\mathbf{e}$ has Hamming weight smaller or equal to $\frac{n-k+1}{2}$, the coefficients of $f_\mathbf{e}$ has linear complexity $t \leq \frac{n-k+1}{2}$. In particular the last $n - k + 1$ coefficients of $f_\mathbf{e}$ is generated by an LFSR of length t at most. Now, given that $t \leq \frac{n-k+1}{2}$ and since we know $n - k + 1$ coefficients, the Berlekamp-Massey algorithm gives the shortest LFSR generating these coefficients. The same LFSR also generates the whole array of coefficients of $f_\mathbf{e}$ periodically, and so we can recover the whole of $f_\mathbf{e}$ using simple linear algebra. Finally, evaluating f_e at $(\alpha_1, \ldots, \alpha_n)$ gives us $\mathbf{e}$.

References

1. Albrecht, M.R., et al: Classic McEliece: conservative code-based cryptography. https://classic.mceliece.org/
2. Aguilar Melchor, C., et al.: RQC - Rank Quasi-Cyclic. http://pqc-rqc.org/
3. Aragon, N., et al.: BIKE: Bit Flipping Key Encapsulation. https://bikesuite.org/
4. Aragon, N., Blazy, O., Gaborit, P., Hauteville, A., Zémor, G.: Durandal: a rank metric based signature scheme. In: Ishai, Y., Rijmen, V. (eds.) EUROCRYPT 2019. LNCS, vol. 11478, pp. 728–758. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-17659-4_25

5. Bardet, M., et al.: An algebraic attack on rank metric code-based cryptosystems. In: Canteaut, A., Ishai, Y. (eds.) EUROCRYPT 2020. LNCS, vol. 12107, pp. 64–93. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-45727-3_3
6. Bardet, M., et al.: Improvements of algebraic attacks for solving the rank decoding and MinRank problems. In: Moriai, S., Wang, H. (eds.) ASIACRYPT 2020. LNCS, vol. 12491, pp. 507–536. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-64837-4_17
7. Barenghi, A., Biasse, J.-F., Persichetti, E., Santini, P.: LESS-FM: fine-tuning signatures from the code equivalence problem. In: Cheon, J.H., Tillich, J.-P. (eds.) PQCrypto 2021 2021. LNCS, vol. 12841, pp. 23–43. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-81293-5_2
8. Barg, S.: Some new NP-complete coding problems. Problemy Peredachi Informatsii **30**(3), 23–28 (1994). ISSN 0555-2923
9. Barg, A.: Complexity issues in coding theory. In: Pless, V., Brualdi, R., Huffman, W. (eds.) Handbook of Coding Theory, chap. 7, pp. 649–754. Elsevier, New York (1998). ISBN 978-0-444-50088-5
10. Berger, T.P., Cayrel, P.-L., Gaborit, P., Otmani, A.: Reducing key length of the McEliece cryptosystem. In: Preneel, B. (ed.) AFRICACRYPT 2009. LNCS, vol. 5580, pp. 77–97. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02384-2_6
11. Biasse, J.-F., Micheli, G., Persichetti, E., Santini, P.: LESS is more: code-based signatures without syndromes. In: Nitaj, A., Youssef, A. (eds.) AFRICACRYPT 2020. LNCS, vol. 12174, pp. 45–65. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-51938-4_3
12. Blahut, R.E.: Transform techniques for error control codes. IBM J. Res. Dev. **23**(3), 299–315 (1979). ISSN 0018-8646
13. Berlekamp, E., McEliece, R., van Tilborg, H.: On the inherent intractability of certain coding problems (corresp.). IEEE Trans. Inf. Theory **24**(3), 384–386 (1978). ISSN 0018-9448
14. Feneuil, T., Joux, A., Rivain, M.: Shared Permutation for Syndrome Decoding: New Zero-Knowledge Protocol and Code-Based Signature. Cryptology ePrint Archive, Report 2022/188 (2022). https://ia.cr/2021/1576
15. Feneuil, T., Joux, A., Rivain, M.: Syndrome decoding in the head: shorter signatures from zero-knowledge proofs. Cryptology ePrint Archive, Report 2022/188 (2022). https://ia.cr/2022/188
16. Gueron, S., Persichetti, E., Santini, P.: Designing a practical code-based signature scheme from zero-knowledge proofs with trusted setup. Cryptography **6**(1), 5 (2022)
17. Gustavson, F.G.: Analysis of the berlekamp-massey linear feedback shift-register synthesis algorithm. IBM J. Res. Dev. **20**(3), 204–212 (1976). https://doi.org/10.1147/rd.203.0204. ISSN 0018-8646
18. Lidl, R., Niederreiter, H.: Finite Fields, 2nd edn. Cambridge University Press, Cambridge (1996). ISBN 0-521-39231-4/hbk
19. Misoczki, R., Barreto, P.S.L.M.: Compact McEliece keys from Goppa codes. In: Jacobson, M.J., Rijmen, V., Safavi-Naini, R. (eds.) SAC 2009. LNCS, vol. 5867, pp. 376–392. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-05445-7_24
20. McEliece, R.: A public-key cryptosystem based on algebraic coding theory. Deep Space Network Progress Report 44:114116 (1978)

21. Misoczki, R., Tillich, J., Sendrier, N., Barreto, P.: MDPC-McEliece: new McEliece variants from moderate density parity-check codes. In: 2013 IEEE International Symposium on Information Theory, pp. 2069–2073, July 2013. ISSN 2157-8095
22. Niederreiter, H.: Knapsack type cryptosystems and algebraic coding theory. Prob. Control Inf. Theory. Problemy Upravlenija i Teorii Informacii **15**, 19–34 (1986)
23. NIST. https://csrc.nist.gov/projects/post-quantum-cryptography. Accessed 9 June 2022
24. Persichetti, E.: Compact McEliece keys based on quasi-dyadic Srivastava codes. J. Math. Cryptol. **6**(2), 149–169 (2012)
25. Samardjiska, S., Santini, P., Persichetti, E., Banegas, G.: A reaction attack against cryptosystems based on LRPC codes. In: Schwabe, P., Thériault, N. (eds.) LATINCRYPT 2019. LNCS, vol. 11774, pp. 197–216. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-30530-7_10