Electronic Colloquium on Computational Complexity, Report No. 25 (2021)

Improved Maximally Recoverable LRCs using Skew Polynomials

Sivakanth Gopi[*] Venkatesan Guruswami[†]

Abstract

An (n, r, h, a, q)-Local Reconstruction Code is a linear code over $\mathbb{F}_q$ of length n, whose codeword symbols are partitioned into n/r local groups each of size r. Each local group satisfies 'a' local parity checks to recover from 'a' erasures in that local group and there are further h global parity checks to provide fault tolerance from more global erasure patterns. Such an LRC is Maximally Recoverable (MR), if it offers the best blend of locality and global erasure resilience—namely it can correct all erasure patterns whose recovery is information-theoretically feasible given the locality structure (these are precisely patterns with up to 'a' erasures in each local group and an additional h erasures anywhere in the codeword).

Random constructions can easily show the existence of MR LRCs over very large fields, but a major algebraic challenge is to construct MR LRCs, or even show their existence, over smaller fields, as well as understand inherent lower bounds on their field size. We give an explicit construction of (n, r, h, a, q)-MR LRCs with field size q bounded by $\left(O\left(\max\{r, n/r\}\right)\right)^{\min\{h, r-a\}}$. This improves upon known constructions in many relevant parameter ranges.

Moreover, it matches the lower bound from [GGY20] in an interesting range of parameters where $r = \Theta(\sqrt{n})$, $r - a = \Theta(\sqrt{n})$ and h is a fixed constant with $h \leqslant a + 2$, achieving the optimal field size of $\Theta_h(n^{h/2})$.

Our construction is based on the theory of skew polynomials. We believe skew polynomials should have further applications in coding and complexity theory; as a small illustration we show how to capture algebraic results underlying list decoding folded Reed-Solomon and multiplicity codes in a unified way within this theory.

[*] Microsoft Research. Email: sigopi@microsoft.com.

[†] Carnegie Mellon University. Email: venkatg@cs.cmu.edu. Research supported in part by NSF grants CCF-1563742 and CCF-1814603.

ISSN 1433-8092

Contents

1 Introduction

In this work, we present a construction of Maximally Recoverable Local Reconstruction Codes (MR LRCs) based on the theory of skew polynomials. Our construction matches or improves the field size of MR LRCs for most parameter regimes. We now describe the motivation of MR LRCs in the context of coding for distributed storage, and then formally define them and describe our results.

In distributed storage such as in data centers, data is partitioned and stored in individual servers; each with a small storage capacity of a few terabytes. A server can crash any time losing all the data it contains. More often than a crash, a server might become temporarily unavailable either due to system updates, network bottlenecks or it might be busy serving requests of some other user. Thus there are two design objectives for a distributed storage system. The first one is to never lose user data in the event of crashes (or at least make it highly improbable). The second is to service user requests with low latency despite some servers becoming temporarily unavailable. Instead of just replicating data which is wasteful, distributed storage systems use erasure codes. Using a Reed-Solomon code, if we add $n-k$ parity check servers to k data servers, we can recover user data from any k available servers. But as k gets larger, this doesn't satisfy our second objective of servicing user requests with low latency. Local Reconstruction Codes (LRCs) were invented precisely for achieving both the objectives while still maintaining storage efficiency and have been implemented in several large scale systems such as Microsoft Azure [HSX$^+$12] and Hadoop [SAP$^+$13]. These codes have *locality* which means that they can recover quickly from a small number of erasures by reading only a small number of available servers. But at the same time, they can also recover from the unlikely event of a large number of erasures (but can do so less efficiently). Locality in distributed storage was first introduced in [HCL07, CHL07], but LRCs were first formally defined and studied in [GHSY12] and [PD14]. We will now define them formally.

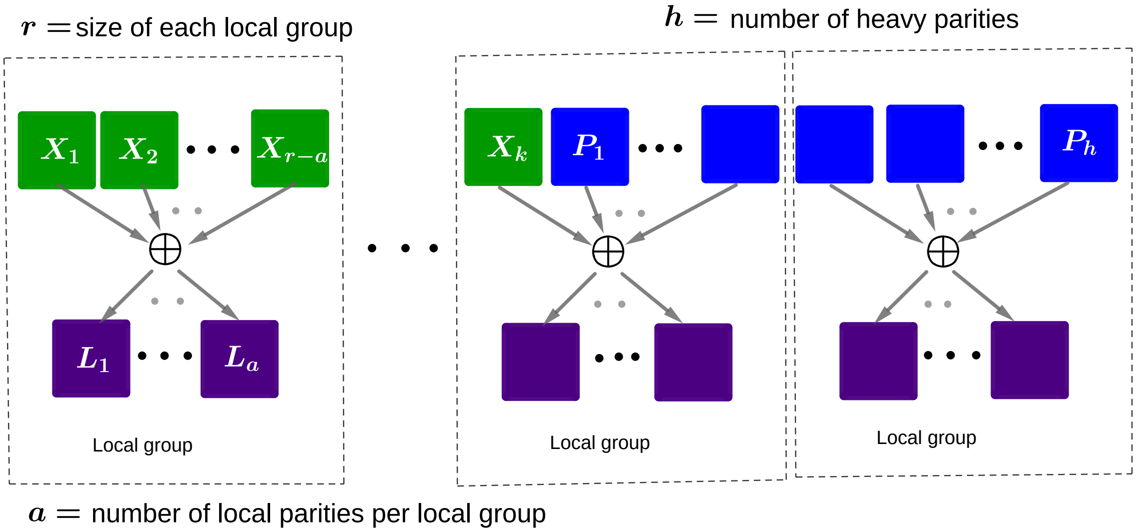
An (n, r, h, a, q)-LRC is a linear code over $\mathbb{F}_q$ of length n, whose codeword symbols are partitioned into n/r local groups each of size r. The coordinates in each local group satisfy 'a' local parity checks and there are further h global parity checks that all the n coordinates satisfy. The local parity checks are used to recover from up to 'a' erasures in a local group by reading at most $r-a$ symbols in that local group. The h global parities are used to correct more global erasure patterns which involve more than a erasures in each local group. The parity check matrix H of an (n, r, h, a, q)-LRC has the structure shown in Equation 1.

$$H = \left[\begin{array}{c|c|c|c} A_1 & 0 & \cdots & 0 \\ \hline 0 & A_2 & \cdots & 0 \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \cdots & A_g \\ \hline B_1 & B_2 & \cdots & B_g \end{array}\right]. \tag{1}$$

Here $g = n/r$ is the number of local groups. $A_1, A_2, \ldots, A_g$ are $a \times r$ matrices over $\mathbb{F}_q$ which correspond to the local parity checks that each local group satisfies. $B_1, B_2, \ldots, B_g$ are $h \times r$ matrices over $\mathbb{F}_q$ and together they represent the h global parity checks that the codewords should satisfy.

Equivalently, from an encoding point of view, an (n, r, h, a, q)-LRC is obtained by adding h global parity checks to k data symbols, partitioning these $k+h$ symbols into local groups of size r, and then adding 'a' local parity checks for each local group. As a result we have $n = k + h + a \cdot \frac{k+h}{r}$ codeword symbols. This is shown in Figure 1.

Information-theoretically, one can show that we can at best hope to correct an additional h erasures distributed across global groups on top of the 'a' erasures in each local group. LRCs



r = size of each local group
h = number of heavy parities
X_1
X_2
X_{r-a}
X_k
P_1
P_h
L_1
L_a
Local group
Local group
Local group
a = number of local parities per local group

Figure 1: An LRC with k data symbols, h heavy parities and 'a' local parities per local group. The length of the code $n = k + h + a \cdot \frac{k+h}{r}$.

which can correct all such erasure patterns which are information-theoretically possible to correct are called *Maximally Recoverable (MR) LRCs*. The notion of maximal recoverability was first introduced by [CHL07, HCL07] and extended to more general settings in [GHJY14]. But MR LRCs were specifically studied first by [BHH13, Bla13] where they are called *Partial-MDS (Maximum Distance Separable) codes.*

Definition 1.1. *Let C be an arbitrary (n, r, h, a, q)-local reconstruction code. We say that C is maximally recoverable if:*

1. *Any set of 'a' erasures in a local group can be corrected by reading the rest of the $r - a$ symbols in that local group.*

2. *Any erasure pattern $E \subseteq [n]$, $|E| = ga + h$, where E is obtained by selecting a symbols from each of g local groups and h additional symbols arbitrarily, is correctable by the code C.*

For a code C with parity check matrix H, an erasure pattern E is correctable iff the submatrix of H formed by columns corresponding the coordinates in E has full column rank. Therefore, we have the following characterization of an MR LRC in terms of its parity check matrix.

Proposition 1.2. *An (n, r, h, a, q)-LRC with parity check matrix given by H from Equation 1 is maximally recoverable iff:*

1. *Each of the local parity check matrices A_i are the parity check matrices of an MDS code, i.e., any a columns of A_i are linearly independent.*

2. *Any submatrix of H which can be formed by selecting a columns in each local group and an additional h columns has full column rank.*

It is known that MR-LRCs exist over exponentially large fields [GHSY12]. This can be seen by instantiating the parity check matrix H from Equation 1 randomly from an exponentially large field and verifying that the condition in Proposition 1.2 is satisfied with high probability by Schwartz-Zippel lemma. But codes deployed in practice require small fields for computational efficiency, typically fields such as $\mathbb{F}_{2^8}$ or $\mathbb{F}_{2^{16}}$ are preferred. Therefore a lot of prior work focused on explicit constructions of MR LRCs over small fields.

1.1 Prior Work

Upper Bounds. There are several known constructions of MR LRCs which are incomparable to each other in terms of the field size [GHJY14, GYBS17, GJX20, MK19, GGY20, Bla13, TPD16, HY16, GHK$^+$17, CK17, BPSY16]. Some constructions are better than others based on the range of parameters. A few of the important ones are shown in Table 1.1. The table is divided into two parts. The first part shows constructions which work for all ranges of parameters and the second part shows constructions which work for some special cases. The first bound by [GYBS17] is good when r is close to n. The second bound by [GJX20] is better when $h \ll r \ll n$. The bound by [MK19] is better when $r - a \leqslant h$. The construction in [MK19] is also significantly different from the previous constructions and our construction is inspired by the construction in [MK19]. Finally, the bound in [GHJY14] is best when $a = 1$ and $h \leqslant r = O(1)$ are constants (we note that the implicit constant hidden in $O_r(\cdot)$ has an exponential dependence in r). In the special case when $h = 2$, a construction over linear sized fields for all ranges of other parameters is given in [GGY20].

Field size q	
$O\big(r \cdot n^{(a+1)h-1}\big)$	[GYBS17]
$\max\big(O(n/r), O(r)^{\min\{r,h+a\}}\big)^{\min\{h,g\}}$	[GJX20]
$\big(O(\max\{n/r, r\})\big)^{r-a}$	[MK19]
$O_r\left(n^{\lceil (h-1)(1-1/2^r)\rceil}\right)$ when $a = 1$ and $r = O(1)$	[GHJY14]
$O(r)$ when $h = 0$ or $h = 1$	[BHH13]
$O(n)$ when $h = 2$	[GGY20]
$\widetilde{O}(n)$ when $h = 3, a = 1, r = 3$	[GGY20]

Table 1: Table showing the best known upper bounds on the field size of (n, r, h, a, q)-MR LRCs.

Lower Bounds. The best known lower bounds on the field size required for (n, r, h, a, q)-MR LRCs (with $g = n/r$ local groups) is from [GGY20] who show that for $h \geqslant 2$,

$$q \geqslant \Omega_{h,a}\left(n \cdot r^{\alpha}\right) \text{ where } \alpha = \frac{\min\{a, h - 2\lceil h/g\rceil\}}{\lceil h/g\rceil}. \tag{2}$$

The lower bound (2) simplifies to

$$q \geqslant \Omega_{h,a}\left(nr^{\min\{a,h-2\}}\right) \tag{3}$$

when $g = n/r \geqslant h$. When $2 \leqslant h \leqslant \min\{a + 2, g\}$, we have:

$$q \geqslant \Omega_h\left(\frac{n(r-a)^{h-1}}{r}\right). \tag{4}$$

Note that the hidden constant in (4) only depends on h.

1.2 Our Results

We are now ready to present our main result.

Theorem 1.3 (Main). *Let $q_0 \geqslant \max\{g+1, r-1\}$ be any prime power where $g = n/r$ is the number of local groups. Then there exists an explicit (n, r, h, a, q)-MR LRC with $q = q_0^{\min\{h, r-a\}}$. Asymptotically, the field size satisfies*

$$q \leqslant \left(O(\max\{r, n/r\})\right)^{\min\{h, r-a\}}. \tag{5}$$

Our construction is better than (or matches) the first three bounds in Table 1.1 for *all* parameter ranges. Moreover when h is a fixed constant with $h \leqslant a+2$ and $r = \Theta(\sqrt{n})$ and $r - a = \Theta(\sqrt{n})$, our construction matches the lower bound (4), achieving the optimal field size of $\Theta_h(n^{h/2})$. This is first non-trivial case (other than when $h = 2$ [GGY20]) where we know the optimal field size for MR LRCs.

Corollary 1.4. *Suppose $r = \Theta(\sqrt{n})$, $r - a = \Theta(\sqrt{n})$ and h is a fixed constant independent of n such that $h \leqslant a + 2$. Then the optimal field size of an (n, r, h, a, q)-LRC is $q = \Theta_h(n^{h/2})$.*

We note that our construction is worse compared to the constructions in the second part of Table 1.1 which work for some special setting of parameters.

MR LRCs used in practice typically have only 2 or 3 local groups i.e. $g = n/r$ is typically a constant [HSX+12]. We can further improve the construction from Theorem 1.3 in this regime, in the special case when the number of local parities $a = 1$.

Theorem 1.5. *Suppose the number of local groups $g = n/r$ is some fixed constant and the number of local parities $a = 1$. Let $q_0 \geqslant g + 1$ be any prime power and let s be such that $q_0^s \geqslant r$. Then there exists an explicit $(n, r, h, a = 1, q)$-LRC with field size $q = q_0^{s\lceil \min\{h, r-1\}(1-1/q_0)\rceil}$. Asymptotically, the field size satisfies*

$$q \leqslant (O(n))^{\lceil \min\{h, r-1\}(1-1/q_0)\rceil}.$$

Our Techniques. Our constructions are based on the theory of skew polynomials and is inspired by the construction from [MK19]. Skew polynomials are a non-commutative generalization of polynomials, but they retain many of the familiar and important properties of polynomials. Just as Reed-Solomon codes are constructed using the fact that a degree d polynomial can have at most d roots, our codes will use an analogous theorem that a degree d skew polynomial can have at most d roots *when counted appropriately* (see Theorem 2.21). Unlike the roots of the usual degree d polynomials which do not have any structure, the roots of degree d skew polynomials have an interesting linear-algebraic structure which we exploit in our constructions. The construction from [MK19] is also implicitly based on skew polynomials. In this paper, we make this connection explicit in the hope that the theory of skew polynomials will lead to further developments in the constructions of MR LRCs and coding theory more broadly. As an illustration, in Appendix C we show how skew polynomials can give an explanation of algebraic results concerning (generalizations of) Wronskian and Moore matrices that have recently been used in the context of list decoding algorithms [GW13], rank condensers [FS12, FSS14, FG15], and subspace designs [GK16, GXY18]. We also reproduce a construction of maximum sum-rank distance (MSRD) codes due to [MP18] using the framework of skew polynomials in Appendix D. Readers familiar with the theory of skew polynomials or who directly want to get to the construction can skip most of the preliminaries in Section 2 except for Section 2.4.

Related Work. Shortly before we published our results, we learned that [CMST20] have independently obtained a result analogous to Theorem 1.3 with a very similar construction. They construct (n, r, h, a, q)-MR LRCs with a field size of

$$q = \left(O(\max\{r, n/r\})\right)^h . \tag{6}$$

Compared to this, we have a $\min\{h, r-a\}$ in the exponent in our field size bound (5).

Soon after [CMST20], two more constructions of MR LRCs were published by [Mar20] with the following field sizes:

$$q \leqslant \left(\max\left\{(2r)^{r-a}, \frac{g}{r}\right\}\right)^{\min\{h, \lfloor g/r \rfloor\}} , \tag{7}$$

$$q \leqslant (2r)^{r-a} \left(\left\lfloor \frac{g}{r} \right\rfloor + 1\right)^{h-1} . \tag{8}$$

The constructions in (7) and (8) are incomparable to our construction in (5). For example when $r = O(1)$, the construction (8) achieves $O(n)^{h-1}$ field size, whereas our construction achieves $O(n)^{\min\{h,r-a\}}$ field size. In the regime when $r = \Theta(\sqrt{n})$ and $r - a = \Theta(\sqrt{n})$ and $h \leqslant a + 2$ is a fixed constant, our construction achieves the optimal field size of $\Theta_h(n^{h/2})$, whereas the constructions from [Mar20] require fields of size $n^{\Theta(\sqrt{n})}$.

Despite all these constructions, a particularly interesting setting of parameters, which remains challenging is the case when $h = O(1)$ and $r - a = n^{o(1)}$. The lower bound (4) only shows that $q = \Omega_h(n^{1+o(1)})$ whereas all the existing constructions need $q \gtrsim_h n^{h-1-o(1)}$.

Open Question 1.6. *When $h = O(1)$ and $r = n^{o(1)}$, do there exist MR LRCs with field size $q \leqslant n^{1+o(1)}$?*

Skew polynomials have been used directly and indirectly in coding theory before. As discussed in Appendix C, folded Reed-Solomon codes and multiplicity codes can be thought of as special cases of skew polynomial based codes. [BU14] used skew polynomials explicitly to define skew Reed-Solomon codes. [MP18] used skew Reed-Solomon codes to construct maximum sum-rank codes, see Appendix D for the construction.

2 Preliminaries

2.1 Skew polynomial ring

Skew polynomials generalize polynomials while inheriting many of the nice properties of polynomials. Skew polynomials can be defined over division rings* and most of the results about skew polynomials are true in this more general setting. It is known that every finite division ring is a field. Since we will only work with skew polynomial rings defined over fields, we will only define them over fields for simplicity. Most of the theory of skew polynomials presented here is from [LL88, Lam85], but we reprove the main results in a more accessible way. Skew polynomials were first defined by Ore [Ore33] in 1933 where it was shown that they are the unique non-commutative generalization of polynomials which satisfy (1) associativity (2) distributivity on both sides and (3) the fact that the degree of product of two polynomials is the sum of their degrees.

Let $\mathbb{K}$ be a field. We will first define the key concepts of 'endomorphism' and 'derivation'.

Definition 2.1 (Endomorphism)**.** *A map $\sigma : \mathbb{K} \to \mathbb{K}$ is called an endomorphism if:*

*Rings where every non-zero element has a multiplicative inverse, but multiplication may not be commutative.

1. *σ is a linear map i.e. $\sigma(a+b)=\sigma(a)+\sigma(b)$ for all $a,b\in K$ and*

2. *$\sigma(ab)=\sigma(a)\sigma(b)$ for all $a,b\in K$.*

Example 2.2. 1. *If $\mathbb{K}=\mathbb{F}_{q^m}$, then $\sigma(x)=x^q$ is an endomorphism.*

2. *If $\mathbb{K}=\mathbb{F}(x)$ is the field of rational functions and $\gamma\in\mathbb{F}^*$, then $\sigma(f(x))=f(\gamma x)$ is an endomorphism.*

Definition 2.3 (Derivation). *A map $\delta:\mathbb{K}\to\mathbb{K}$ is called a σ-derivation if:*

1. *δ is a linear map i.e. $\delta(a+b)=\delta(a)+\delta(b)$ for all $a,b\in K$ and*

2. *$\delta(ab)=\sigma(a)\delta(b)+\delta(a)b$ for all $a,b\in K$.*

We will now define the skew polynomial ring.

Definition 2.4 (Skew polynomial ring). *Let σ be an endomorphism of $\mathbb{K}$ and δ be a σ-derivation. The skew polynomial ring in variable t, denoted by $\mathbb{K}[t;\sigma,\delta]$, is a non-commutative ring of skew polynomials in t of the form $\{\sum_{i=0}^{d}a_it^i : d\geqslant 0, a_i\in\mathbb{K}\}$ (where we always write the coefficients to the left). Degree of a polynomial $f(t)=\sum_i a_it^i$, denoted by $\deg(f)$, is the largest d such that $a_d\neq 0$.*[*] *Addition in $\mathbb{K}[t;\sigma,\delta]$ is component wise. But multiplication is distributive and done according to the following rule:*

$$\text{For } a\in\mathbb{K},\ t\cdot a=\sigma(a)t+\delta(a). \tag{9}$$

To multiply $f(t)g(t)$, we can first use distributivity to get $f(t)g(t)=\sum_{ij}f_it^i\cdot g_jt^j$ where $f_i,g_j\in\mathbb{K}$ are coefficients of f,g respectively. Then we use the rule (9) for i times to move the coefficient g_j to the left of t^i. This multiplication turns out to be associative, but may not be commutative. Also $\deg(f\cdot g)=\deg(f)+\deg(g)$. Therefore the skew polynomial ring has no zero divisors. We will now give some examples of skew-polynomials.

Example 2.5 (Skew Polynomial Rings). 1. *The simplest example of a skew polynomial ring is when σ is the identity map and δ is the zero map. In this case, skew polynomials coincide with the usual notion of polynomials.*

2. *The simplest derivation is the zero map i.e. $\delta(a)=0$ for all $a\in\mathbb{K}$. In this case, the skew ring is denoted by $\mathbb{K}[t;\sigma]$ and is said to be of endomorphism type. Skew polynomials are interesting even in this case, and in fact the constructions in this paper only use skew polynomials with $\delta\equiv 0$. So the reader can imagine that the derivation is the zero map on a first reading. We include the general case in the hope that skew polynomial rings with non-zero derivations will find applications in future.*

3. *Let $\mathbb{K}$ be any field and let $\sigma:\mathbb{K}\to\mathbb{K}$ be an endomorphism. Then for any $\lambda\in\mathbb{K}$, $\delta(a)=\lambda(\sigma(a)-a)$ is a σ-derivation.*[†] *These are called inner-derivations and the skew polynomial ring defined using such a derivation is isomorphic to the skew polynomial ring over $\mathbb{K}$ with the same σ and $\delta=0$.*[‡] *The concept of q-derivatives [BSC+12] is a special case of this for $\mathbb{K}=\mathbb{F}(x)$. For some fixed $q\in\mathbb{F}$, the q-derivative $f\in\mathbb{F}(x)$ is defined as $(f(qx)-f(x))/(qx-x)$. This is a derivation w.r.t to the endomorphism $\sigma:f(x)\to f(qx)$.*

[*] We will define the degree of the zero polynomial to be ∞.

[†] If $\mathbb{K}$ is a division ring, then $\delta(a)=\sigma(a)\lambda-\lambda a$ is a σ-derivation.

[‡] The isomorphism is $\phi:\mathbb{K}[t;\sigma,\delta]\to\mathbb{K}[\tilde{t};\sigma]$ defined as $\phi(t)=\tilde{t}-\lambda$ and $\phi|_{\mathbb{K}}\equiv\mathrm{Id}$.

4. *Let $\mathbb{K} = \mathbb{F}(x)$ and σ be the identity map. Then $\delta(f(x))$ defined as the formal derivate of $f(x)$ is a σ-derivation. This can be extended to rational functions in a consistent way using power series. When σ is the identity map, the skew ring is denoted by $\mathbb{K}[t; \delta]$ and is said to be of derivation type.*

We remark that when $\mathbb{K}$ is a field (as opposed to being a division ring), up to isomorphisms, the only possible skew polynomial rings are either of endomorphism type (i.e., $\delta \equiv 0$) or derivation type (i.e., $\sigma \equiv \mathrm{Id}$). This is because if $\sigma \neq \mathrm{Id}$, then there exists some element $a_0 \in \mathbb{K}$ such that $\sigma(a_0) \neq a_0$. Now using commutativity of $\mathbb{K}$, we have $\delta(aa_0) = \delta(a_0a)$ for any $a \in \mathbb{K}$. Expanding both sides, we get that for any $a \in \mathbb{K}$, $\delta(a) = \lambda(\sigma(a) - a)$ where $\lambda = \delta(a_0)/(\sigma(a_0) - a_0)$ is a fixed constant, i.e., δ is an inner-derivation. As we discussed above, this skew polynomial ring is isomorphic to the the skew polynomial ring with $\delta \equiv 0$ and the same endomorphism σ.

We will now collect some simple facts about skew polynomials rings. Let $\mathbb{K}[t; \sigma, \delta]$ be a skew polynomial ring.

Lemma 2.6 ([LL88])**.** *$t^n a = \sum_{i=0}^{n} f_i^n(a) t^i$ where $f_0^n = \delta^n$, $f_1^n = \delta^{n-1}\sigma + \delta^{n-2}\sigma\delta + \cdots + \sigma\delta^{n-1}, \ldots, f_n^n = \sigma^n$ are linear maps.*

It turns out that the skew polynomial ring has Euclidean algorithm for right division.

Lemma 2.7 (Euclidean algorithm for right division [LL88])**.** *For every two polynomial $f, g \in \mathbb{K}[t; \sigma, \delta]$, there exist unique polynomials $q(t), r(t)$ such that $f = q \cdot g + r$ where $\deg(r) < \deg(g)$ or $r = 0$.*

This brings us to the most important definition about skew polynomial rings. In the usual polynomial world, we can define the evaluation of a polynomial $f(t) = \sum_i f_i t^i$ at $t = a$ as $\sum_i f_i a^i$. With this definition, it is true that $f(t) = q(t)(t - a) + f(a)$. But for skew polynomials, these two notions of evaluation differ with each other. And the right definition is the second one.

Definition 2.8 (Evaluation)**.** *The evaluation of a polynomial $f \in \mathbb{K}[t; \sigma, \delta]$ at a point $a \in \mathbb{K}$, denoted by $f(a)$, is defined as the remainder obtained when we divide f by $t - a$ on the right i.e. $f(t) = q(t)(t - a) + f(a)$.*

Note that evaluation is a linear map i.e. $(f + g)(a) = f(a) + g(a)$. But it is not always true that $(fg)(a) = f(a)g(a)$. We will see shortly how to compute $(fg)(a)$. The evaluation map can be expressed using "power functions", which are the evaluations of monomials of the form t^i.

Definition 2.9 (Power functions)**.** *The power functions are defined inductively as follows. For every $a \in \mathbb{K}$*

1. *$N_0(a) = 1$ and*

2. *$N_{i+1}(a) = \sigma(N_i(a))a + \delta(N_i(a))$.*

When $\delta \equiv 0$, we have $\mathbb{N}_i(a) = \sigma^{i-1}(a)\sigma^{i-2}(a) \cdots \sigma(a)a$. Additionally if $\sigma \equiv \mathrm{Id}$, then $N_i(a) = a^i$ which explains the terms "power functions".

Lemma 2.10. *Let $f = \sum_i f_i t^i$. Then $f(a) = \sum_i f_i N_i(a)$.*

Proof. It is easy to prove by induction that evaluation of t^i at a is $N_i(a)$. The general claims follows by linearity of evaluation. □

We now come to the problem of evaluating $(fg)(a)$. For this, it is useful to define the notion of *conjugates*, which play a big role in this theory.

2.2 Conjugation and Product Rule

Definition 2.11 (Conjugation)**.** *Let $a \in \mathbb{K}$ and $c \in \mathbb{K}^*$. We define the c-conjugate of a, denoted by ${}^{c}a$, as*

$${}^{c}a = \sigma(c)ac^{-1} + \delta(c)c^{-1}.$$

We say that b is a conjugate of a if there exists some $c \in \mathbb{K}^*$ such that $b = {}^{c}a$.

We have the following lemma which shows that conjugacy is an equivalence relation, the proof of which is in Appendix A.

Lemma 2.12. *1. ${}^{d}({}^{c}a) = {}^{dc}a$*

2. Conjugacy is an equivalence relation, i.e., we can partition $\mathbb{K}$ into conjugacy classes where elements in each part are conjugates of each other, but elements in different parts are not conjugates.

So $\mathbb{K}$ will get partitioned into conjugacy classes. To understand the structure of each conjugacy class, we need the notion of *centralizer*.

Definition 2.13 (Centralizer)**.** *The centralizer of $a \in \mathbb{K}$ is defined as:*

$$\mathbb{K}_a = \{c \in \mathbb{K}^* : {}^{c}a = a\} \cup \{0\}.$$

The following lemma shows that centralizers are subfields, the proof of which appears in Appendix A.

Lemma 2.14. *1. $\mathbb{K}_a$ is a subfield of $\mathbb{K}$.*[*]

2. If $a, b \in \mathbb{K}$ are conjugates, then $\mathbb{K}_a = \mathbb{K}_b$. [†]

Because of the above lemma, we can associate a centralizer subfield to each conjugacy class.

Example 2.15. *Let $\mathbb{K} = \mathbb{F}_{q^m}$, $\sigma(a) = a^q$ and $\delta \equiv 0$. Then ${}^{c}a = c^{q-1}a$. Suppose γ is a generator for $\mathbb{F}_{q^m}^*$. There are q equivalence classes, $E_{-1}, E_0, E_1, \ldots, E_{q-2}$, where $E_\ell = \{\gamma^i : i \equiv \ell \mod (q-1).\}$ and $E_{-1} = \{0\}$. The centralizer of an element $a \in \mathbb{K}^*$ is*

$$\mathbb{K}_a = \{c : c^{q-1}a = a\} \cup \{0\} = \{c : c^{q-1} = 1\} \cup \{0\} = \mathbb{F}_q.$$

Therefore the centralizer of every non-zero element is $\mathbb{F}_q$ and the the centralizer of 0 is $\mathbb{K}_0 = \mathbb{K}$.

We will now show how to evaluate $(fg)(a)$. And conjugates play a key role. The proof of this really important lemma is given in Appendix A.

Lemma 2.16 (Product evaluation rule)**.** *If $g(a) = 0$, then $(fg)(a) = 0$. If $g(a) \neq 0$ then*

$$(fg)(a) = f\left({}^{g(a)}a\right) g(a).$$

Using the product rule, one can prove an interpolation theorem for skew polynomials just like ordinary polynomials. For any $A \subset \mathbb{K}$ be of size n, there exists a non-zero degree $\leqslant n$ skew polynomial $f \in \mathbb{K}[t; \sigma, \delta]$ which vanishes on A [LL88]. We will later need the following lemma.

Lemma 2.17. *Let f be any skew polynomial. Fix some $a \in \mathbb{K}$. Then $D_{f,a}(y) = f({}^{y}a)y$ is an $\mathbb{K}_a$-linear map from $\mathbb{K} \to \mathbb{K}$.*

Proof. Linearity follows since $f({}^{y}a)y$ is equal to the evaluation of the polynomial $f(t)y$ at a by Lemma 2.16. And clearly the evaluation is linear in y. $\mathbb{K}_a$-linearity follows since $\forall c \in \mathbb{K}_a$,

$$D_{f,a}(yc) = f({}^{yc}a)yc = f({}^{y}({}^{c}a))yc = f({}^{y}a)yc = D_{f,a}(y)c. \quad \square$$

[*] When $\mathbb{K}$ is a division ring, $\mathbb{K}_a$ will be a sub-division ring of $\mathbb{K}$.

[†] When $\mathbb{K}$ is a division ring and not a field, we have $\mathbb{K}_{({}^{x}a)} = x\mathbb{K}_a x^{-1}$.

2.3 Roots of skew polynomials

The most important and useful fact about usual polynomials is that a degree d non-zero polynomial can have at most d roots. It turns out that this statement is false for skew polynomials! A skew polynomial can have many more roots than its degree. But when counted in the right way, we can recover an analogous statement for skew polynomials. In this section, we will prove the "fundamental theorem" about roots of skew polynomials which shows that a degree d skew polynomial cannot have more than d roots when counted the right way. We will begin with showing that any non-zero degree d skew polynomial can have at most d roots in distinct conjugacy classes.

Lemma 2.18. *Let $f \in \mathbb{K}[t;\sigma,\delta]$ be a degree d non-zero polynomial. Then f can have at most d roots in distinct conjugacy classes.*

Proof. We will prove it using induction on the degree. For the base case, it is clear that a degree 0 polynomial which is a non-zero constant cannot have any roots. Suppose $a_0, a_1, \ldots, a_d \in \mathbb{K}$ be roots of f in distinct conjugacy classes. Since $f(a_0) = 0$, we can write $f(t) = h(t)(t - a_0)$ where $\deg(h) = d-1$. By Lemma 2.16, $f(a_i) = h({}^{a_i-a_0}a_i)(a_i-a_0)$. Therefore $b_i = {}^{a_i-a_0}a_i$ for $i \in \{1,\ldots,d\}$ are d roots of h and they lie in distinct conjugacy classes because a_i lie in distinct conjugacy classes. Thus by induction $h = 0$ and therefore $f = 0$ which is a contradiction. □

Now let us try to understand, the roots of a skew polynomial in the same conjugacy class. The following lemma shows that they form a vector space over a subfield of $\mathbb{K}$.

Lemma 2.19. *Let $f \in \mathbb{K}[t;\sigma,\delta]$ be a non-zero polynomial and fix some $a \in \mathbb{K}$ and let $\mathbb{F} = \mathbb{K}_a$ be the centralizer of a (which is a subfield of $\mathbb{K}$). Define $V_f(a) = \{y \in \mathbb{K}^* : f({}^{y}a) = 0\} \cup \{0\}$. Then $V_f(a)$ is a vector space over $\mathbb{F}$.*

Proof. For any $\lambda \in \mathbb{F}$ and $y \in V_f(a)$, $f({}^{\lambda y}a) = f({}^{y}({}^{\lambda}a)) = f({}^{y}a) = 0$. Therefore $\lambda y \in V_f(a)$. If $y_1, y_2 \in V_f(a)$ where $y_1+y_2 \neq 0$, then by Lemma 2.17, $f({}^{y_1+y_2}a) = 0$. Therefore $y_1+y_2 \in V_f(a)$. □

The next lemma shows that the dimension of $V_f(a)$ can be at most $\deg(f)$.

Lemma 2.20. *Let $f \in \mathbb{K}[t;\sigma,\delta]$ be a degree d non-zero polynomial and fix some $a \in \mathbb{K}$ and let $\mathbb{F} = \mathbb{K}_a$ be the centralizer subfield of a. Define $V_f(a) = \{y \in \mathbb{K}^* : f({}^{y}a) = 0\} \cup \{0\}$. Then $V_f(a)$ is a vector space over $\mathbb{F}$ of dimension at most d.*

Proof. We will use induction on the degree. For the base case, it is clear that for a degree 0 polynomial, which is a non-zero constant, $\dim_{\mathbb{F}}(V_f(a)) = 0$. Suppose for contradiction that there exists $y_0, y_1, \ldots, y_d \in V_f(a)$ which are linearly independent over $\mathbb{F}$. WLOG, we can assume that $y_0 = 1$ (by redefining a to be equal to ${}^{y_0}a$). Since $f(a) = 0$, we can write $f(t) = h(t)(t-a)$ where $\deg(h) = d-1$. By Lemma 2.16, $f({}^{y_i}a) = h({}^{y_i({}^{y_i}a-a)}a)({}^{y_i}a-a)$. Since $y_0 = 1$ and y_i is linearly independent from y_0 over $\mathbb{F}$, $y_i \notin \mathbb{F}$. Therefore ${}^{y_i}a - a \neq 0$, and so $b_i = {}^{y_i({}^{y_i}a-a)}a$ for $i \in \{1,\ldots,d\}$ are d roots of h. If we show that $y_i({}^{y_i}a-a)$ for $i \in \{1,\ldots,d\}$ are linearly independent over $\mathbb{F}$, then we are done by induction.

Suppose they are not independent. Then there exists $c_1, \ldots, c_d \in \mathbb{F}$ s.t. $\sum_{i=1}^{d} c_i y_i({}^{y_i}a-a) = 0$.

Therefore,

$$\begin{aligned} a\sum_{i=1}^{d} c_i y_i &= \sum_{i=1}^{d} c_i y_i \cdot {}^{y_i}a \\ &= \sum_{i=1}^{d} c_i y_i \cdot {}^{c_i y_i}a && (c_i \in \mathbb{F} = \mathbb{K}_a) \\ &= \left(\sum_{i=1}^{d} c_i y_i\right)^{\left(\sum_{i=1}^{d} c_i y_i\right)} a && ({}^{x+y}a(x+y) = {}^{x}ax + {}^{y}ay \text{ for all } x, y \in \mathbb{K}^*) \end{aligned}$$

Since $y_1, \ldots, y_d$ are independent over $\mathbb{F}$, $\sum_{i=1}^{d} c_i y_i \neq 0$. Therefore ${}^{\left(\sum_{i=1}^{d} c_i y_i\right)}a = a$ i.e. $\sum_{i=1}^{d} c_i y_i \in \mathbb{K}_a = \mathbb{F}$. But this contradicts the fact that $\{y_0 = 1, y_1, \ldots, y_d\}$ are linearly independent over $\mathbb{F}$. □

The following theorem is the "fundamental theorem" about roots of skew polynomials. It immediately implies Lemma 2.18 and Lemma 2.20 as corollaries. But we have proved them before, just to convey some intuition.

Theorem 2.21. *Let $f \in \mathbb{K}[t; \sigma, \delta]$ be a degree d non-zero polynomial. Let A be the set of roots of f in $\mathbb{K}$ and let $A = \cup_i A_i$ be a partition of A into conjugacy classes. Fix some representatives $a_i \in A_i$. Let $V_i = \{y : {}^{y}a_i \in A_i\} \cup \{0\}$ which is a linear subspace over $\mathbb{F}_i = \mathbb{K}_{a_i}$ by Lemma 2.19. Then*

$$\sum_i \dim_{\mathbb{F}_i}(V_i) \leqslant d.$$

The proof of Theorem 2.21 is given in Appendix B.

2.4 Vandermonde matrix

Definition 2.22 (Vandermonde matrix)**.** *Let $A = \{a_1, \ldots, a_n\} \subset \mathbb{K}$. The Vandermonde matrix formed by A, denoted by $V(a_1, \ldots, a_n)$, is defined as:*

$$V_d(a_1, \ldots, a_n) = \begin{bmatrix} N_0(a_1) & N_0(a_2) & \cdots & N_0(a_n) \\ N_1(a_1) & N_1(a_2) & \cdots & N_1(a_n) \\ \vdots & \vdots & & \vdots \\ N_{d-1}(a_1) & N_{d-1}(a_2) & \cdots & N_{d-1}(a_n) \end{bmatrix}$$

If $f(t) = \sum_{i=0}^{d-1} f_i t^i$ is a skew polynomial of degree at most $d-1$, then by Lemma 2.10,

$$[f_0, f_1, \ldots, f_{d-1}] \cdot V_d(a_1, a_2, \ldots, a_n) = [f(a_1), f(a_2), \ldots, f(a_n)]. \tag{10}$$

Lemma 2.23. *Let $a_1, \ldots, a_d \in \mathbb{K}$ be in distinct conjugacy classes. Then $V_d(a_1, \ldots, a_d)$ is full-rank.*

Proof. If not, then there exists a non-zero vector $(f_0, f_1, \ldots, f_{d-1}) \in \mathbb{K}^d$ such that $[f_0, f_1, \ldots, f_{d-1}] \cdot V_d(a_1, \ldots, a_d) = 0$. By Equation (10), this implies that the skew polynomial $f(t) = \sum_{i=0}^{d-1} f_i t^i$ has d roots in distinct conjugacy classes. This is a contradiction by Lemma 2.18. □

Corollary 2.24. *Let $\gamma \in \mathbb{F}_{q^m}^*$ be a generator of the multiplicative group. Let $d \leqslant q-1$ and $\ell_1, \ldots, \ell_d \in \{0, 1, 2, \ldots, q-2\}$ be distinct. Then the following matrix M is full rank.*

$$M = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \gamma^{\ell_1} & \gamma^{\ell_2} & \cdots & \gamma^{\ell_d} \\ \gamma^{\ell_1(1+q)} & \gamma^{\ell_2(1+q)} & \cdots & \gamma^{\ell_d(1+q)} \\ \vdots & \vdots & & \vdots \\ \gamma^{\ell_1(1+q+\cdots+q^{d-2})} & \gamma^{\ell_2(1+q+\cdots+q^{d-2})} & \cdots & \gamma^{\ell_d(1+q+\cdots+q^{d-2})} \end{bmatrix}$$

Proof. Let $\mathbb{K} = \mathbb{F}_{q^m}$, $\sigma(a) = a^q$ and $\delta \equiv 0$. Then $N_i(a) = a^{1+q+q^2+\cdots+q^{i-1}}$. By Lemma 2.23, it is enough to show that $\ell_1, \ldots, \ell_d$ fall in distinct conjugacy classes. This is shown in Example 2.15. □

Note that when $m = 1$, the matrix in the above corollary reduces to the usual Vandermonde matrix one is familiar with.

In general we would want to compute the rank of $V_n(a_1, \ldots, a_n)$ for any given $a_1, \ldots, a_n$. The following lemma generalizes Lemma 2.23.

Lemma 2.25. *Let $A = \{a_1, \ldots, a_n\} \subset \mathbb{K}$. Let $A = A_1 \cup A_2 \cup \cdots \cup A_r$ be the partition of A into conjugacy classes. Then $\operatorname{rank}(V_n(A)) = \sum_i \operatorname{rank}(V_n(A_i))$.*

Proof. Follows from Theorem 2.21. □

By the above lemmas, we reduced the problem to computing $\operatorname{rank}(V_n(A))$ when all elements of A belong to the same conjugacy class. The following lemma shows how to compute this.

Lemma 2.26. *Let $a \in \mathbb{K}$ and $\mathbb{F} = \mathbb{K}_a$ which is a subfield of $\mathbb{K}$. Then for any $\{c_1, \ldots, c_n\} \subset \mathbb{K}^*$, we have*

$$\operatorname{rank}(V_n({}^{c_1}a, \ldots, {}^{c_n}a)) = \dim_{\mathbb{F}} \operatorname{span}_{\mathbb{F}}\{c_1, \ldots, c_n\}.$$

In particular, $V_n({}^{c_1}a, \ldots, {}^{c_n}a)$ is full-rank iff $\{c_1, \ldots, c_n\}$ are linearly independent over $\mathbb{F}$.

Proof. Follows from Lemma 2.20. □

Corollary 2.27. *Let $\gamma \in \mathbb{F}_{q^m}^*$ be a generator of the multiplicative group and let $\ell \in \{0, 1, \ldots, q-2\}$. Let $\beta_1, \ldots, \beta_m \in \mathbb{F}_{q^m}$ be linearly independent over $\mathbb{F}_q$. Then the following matrix M is full rank.*

$$M = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \gamma^{\ell}\beta_1^{q-1} & \gamma^{\ell}\beta_2^{q-1} & \cdots & \gamma^{\ell}\beta_m^{q-1} \\ \gamma^{\ell(1+q)}\beta_1^{q^2-1} & \gamma^{\ell(1+q)}\beta_2^{q^2-1} & \cdots & \gamma^{\ell(1+q)}\beta_m^{q^2-1} \\ \vdots & \vdots & & \vdots \\ \gamma^{\ell(1+q+\cdots+q^{m-2})}\beta_1^{q^{m-1}-1} & \gamma^{\ell(1+q+\cdots+q^{m-2})}\beta_2^{q^{m-1}-1} & \cdots & \gamma^{\ell(1+q+\cdots+q^{m-2})}\beta_m^{q^{m-1}-1} \end{bmatrix}$$

Proof. Let $\mathbb{K} = \mathbb{F}_{q^m}$, $\sigma(a) = a^q$ and $\delta \equiv 0$. Then $N_i(a) = a^{1+q+q^2+\cdots+q^{i-1}}$. Let $a = \gamma^{\ell}$ then $M = V_m({}^{\beta_1}a, \ldots, {}^{\beta_m}a)$. Therefore M is full rank by Lemma 2.26. □

3 Skew polynomials based MR LRC constructions

Let us recall that an (n, r, h, a, q)-LRC admits a parity check matrix H of the following form

$$H = \left[\begin{array}{c|c|c|c} A_1 & 0 & \cdots & 0 \\ \hline 0 & A_2 & \cdots & 0 \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \cdots & A_g \\ \hline B_1 & B_2 & \cdots & B_g \end{array}\right]. \tag{11}$$

Here $A_1, A_2, \cdots, A_g$ are $a \times r$ matrices over $\mathbb{F}_q$ which represent the local parity checks, $B_1, B_2, \cdots, B_g$ are $h \times r$ matrices over $\mathbb{F}_q$ which together represent the h global parity checks. The rest of the matrix is filled with zeros. By Proposition 1.2, C is an MR LRC iff (1) any 'a' columns of each matrix A_i are linearly independent and (2) any submatrix of H formed by selecting a columns in each local group and any h additional columns is full rank.

3.1 Construction: Proof of Theorem 1.3

In this section, we will prove Theorem 1.3 by presenting a construction of MR LRCs over fields of size $q = O\left(\max(g,r)\right)^{\min\{h,r-a\}}$. The construction presented here is inspired from [MK19], where they achieve a field size of $O\left(\max(g,r)\right)^{r-a}$.*

Let $q_0 \geqslant \max\{g+1, r\}$ be a prime power. Choose $\alpha_1, \alpha_2, \dots, \alpha_r \in \mathbb{F}_{q_0}$ to be distinct. Define

$$A_\ell = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \alpha_1 & \alpha_2 & \dots & \alpha_r \\ \alpha_1^2 & \alpha_2^2 & \dots & \alpha_r^2 \\ \vdots & \vdots & & \vdots \\ \alpha_1^{a-1} & \alpha_2^{a-1} & \dots & \alpha_r^{a-1} \end{bmatrix}.$$

Note that $A_1 = A_2 = \dots = A_g$. Let $m = \min\{r-a, h\}$ and let γ be a generator for $\mathbb{F}_{q_0^m}^*$. Our codes will be defined over the field $\mathbb{F}_q = \mathbb{F}_{q_0^m}$. Define $\beta_1, \beta_2, \dots, \beta_r \in \mathbb{F}_{q_0^m}$ as

$$\beta_i = \begin{bmatrix} \alpha_i^a \\ \alpha_i^{a+1} \\ \vdots \\ \alpha_i^{a+m-1} \end{bmatrix},$$

where we are expressing β_i in some basis for $\mathbb{F}_{q_0^m}$ (which is a $\mathbb{F}_{q_0}$-vector space of dimension m). Define

$$B_\ell = \begin{bmatrix} \beta_1 & \beta_2 & \dots & \beta_r \\ \gamma^\ell \beta_1^{q_0} & \gamma^\ell \beta_2^{q_0} & \dots & \gamma^\ell \beta_r^{q_0} \\ \gamma^{\ell(1+q_0)} \beta_1^{q_0^2} & \gamma^{\ell(1+q_0)} \beta_2^{q_0^2} & \dots & \gamma^{\ell(1+q_0)} \beta_r^{q_0^2} \\ \vdots & \vdots & & \vdots \\ \gamma^{\ell(1+q_0+\dots+q_0^{h-2})} \beta_1^{q_0^{h-1}} & \gamma^{\ell(1+q_0+\dots+q_0^{h-2})} \beta_2^{q_0^{h-1}} & \dots & \gamma^{\ell(1+q_0+\dots+q_0^{h-2})} \beta_r^{q_0^{h-1}} \end{bmatrix}.$$

To prove that the above construction is an MR LRC, we will use properties of the skew field $\mathbb{F}_{q_0^m}[x;\sigma]$ where $\sigma(a) = a^{q_0}$. We know that $\mathbb{F}_{q_0^m}$ will get partitioned into $q_0 - 1$ conjugacy classes as shown in Example 2.15. If $\gamma \in \mathbb{F}_{q_0^m}^*$ is a generator of $\mathbb{F}_{q_0^m}^*$, then $\{1, \gamma, \gamma^2, \dots, \gamma^{q_0-2}\}$ fall in distinct conjugacy classes. Intuitively, in the construction each local group corresponds to one conjugacy class. This is possible since we chose $q_0 \geqslant g+1$. The stabilizer subfield of each conjugacy class is $\mathbb{F}_{q_0}$ as shown in Example 2.15. Therefore we choose the matrices B_i for local group i as a (skew) Vandermonde matrix where the evaluation points $\beta_1, \cdots, \beta_r$ are from the conjugacy class of γ^i, but are linearly independent over the stabilizer subfield $\mathbb{F}_{q_0}$.

Claim 3.1. *The above construction is an MR LRC over fields of size $q = q_0^{\min\{h,r-a\}}$.*

Proof. Given an erasure pattern E of size $|E| = ag + h$, composed of a erasures in each local group and h additional erasures, we want to argue that the submatrix $H(E)$, H restricted columns in E, is full rank. WLOG, assume that the h additional erasures happen in local groups $1, 2, \dots, t \in [g]$ for $t \leqslant h$. Let E_i be the set of erasures that happen in the i^{th} local group. Let $S_i \subset E_i$ be an arbitrary subset of size $|S_i| = a$ and let $T_i = E_i \setminus S_i$. Note that $|T_i| \leqslant m$ for all i. For a matrix M

[*] The improvement comes from choosing $\beta_1, \dots, \beta_r$ carefully in our construction. Moreover [MK19] constructs a generator matrix for the code, whereas we construct a parity check matrix.

and a subset X of its columns, we will use $M(X)$ to denote the submatrix of M formed by columns in X. We need to show that $H(E)$ (which is an $(ag+h) \times (ag+h)$ matrix) is full rank where

$$H(E) = \left[\begin{array}{c|c|c|c} A_1(S_1 \cup T_1) & 0 & \cdots & 0 \\ \hline 0 & A_2(S_2 \cup T_2) & \cdots & 0 \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \cdots & A_g(S_g \cup T_g) \\ \hline B_1(S_1 \cup T_1) & B_2(S_2 \cup T_2) & \cdots & B_g(S_g \cup T_g) \end{array}\right].$$

Note that $A_1(S_1), A_2(S_2), \cdots, A_g(S_g)$ are $a \times a$ matrices of full rank. By doing column operations on $H(E)$, in each local group we can use the columns of $A_i(S_i)$ to remove the columns of $A_i(T_i)$. This results in the lower block $B_i(T_i)$ to change into a Schur complement as follows:

$$\left[\begin{array}{c|c} A_i(S_i) & A_i(T_i) \\ \hline B_i(S_i) & B_i(T_i) \end{array}\right] \rightarrow \left[\begin{array}{c|c} A_i(S_i) & 0 \\ \hline B_i(S_i) & B_i(T_i) - B_i(S_i)A_i(S_i)^{-1}A_i(T_i) \end{array}\right].$$

Note that $T_i = \phi$ for $i > t$. So by doing row and column operations on $H(E)$, we can set it in a block diagonal form, where the diagonal blocks are given by $A_1(S_1), A_2(S_2), \ldots, A_g(S_g)$ and one additional $h \times h$ block given by

$$C = \left[\begin{array}{c|c|c} B_1(T_1) - B_1(S_1)A_1(S_1)^{-1}A_1(T_1) & \cdots & B_t(T_t) - B_t(S_t)A_t(S_t)^{-1}A_t(T_t) \end{array}\right].$$

Note that all the entries in $A(S_i)^{-1}A_i(T_i)$ are in the base field $\mathbb{F}_{q_0}$. Also column operations on B_i with $\mathbb{F}_{q_0}$ coefficients retain its structure with β's replaced by their corresponding $\mathbb{F}_{q_0}$-linear combinations. Therefore by Lemma 2.25 and Lemma 2.26, it is enough to show that the following t matrices $D_1, D_2, \ldots, D_t$ are full rank:

$$D_i = \left[\beta(T_i) - \beta(S_i)A_i(S_i)^{-1}A_i(T_i)\right]$$

where $\beta = [\beta_1, \ldots, \beta_r]$ is a $m \times r$ matrix over $\mathbb{F}_{q_0}$. Note that $[D_1|D_2|\ldots|D_t]$ is just the first row of C (with entries in $\mathbb{F}_{q_0^m}$) expressed as a matrix over $\mathbb{F}_{q_0}$. Consider following matrices given by

$$F_i = \left[\begin{array}{c|c} A_i(S_i) & A_i(T_i) \\ \hline \beta(S_i) & \beta(T_i) \end{array}\right]$$

where each F_i is of size $(a+m) \times (a + |T_i|)$. Each F_i is a Vandermonde matrix by construction. Since $|T_i| \leqslant m$, each F_i is full rank. Now if we do column operations to get F_i into block diagonal form we get:

$$\left[\begin{array}{c|c} A_i(S_i) & 0 \\ \hline \beta(S_i) & \beta(T_i) - \beta(S_i)A_i(S_i)^{-1}A(T_i) \end{array}\right] = \left[\begin{array}{c|c} A_i(S_i) & 0 \\ \hline \beta(S_i) & D_i \end{array}\right].$$

This implies that $D_1, D_2, \ldots, D_t$ are full rank over $\mathbb{F}_{q_0}$ which completes the proof. □

A slightly better construction which only requires $q_0 \geqslant \max\{g+1, r-1\}$ can be obtained by choosing

$$A_\ell = \begin{bmatrix} 1 & \alpha_2^{m+a-1} & \alpha_3^{m+a-1} & \ldots & \alpha_r^{m+a-1} \\ 0 & \alpha_2^{m+a-2} & \alpha_3^{m+a-2} & \ldots & \alpha_r^{m+a-2} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & \alpha_2^{m+1} & \alpha_3^{m+1} & \ldots & \alpha_r^{m+1} \\ 0 & \alpha_2^{m} & \alpha_3^{m} & \ldots & \alpha_r^{m} \end{bmatrix}$$

and $\beta_1, \beta_2, \dots, \beta_r \in \mathbb{F}_{q_0}^m$ as:

$$\beta_1 = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \text{ and } \beta_i = \begin{bmatrix} \alpha_i^{m-1} \\ \vdots \\ \alpha_i \\ 1 \end{bmatrix} \text{ for } i \in \{2, 3, \dots, r\}.$$

3.2 Construction: Proof of Theorem 1.5

When $a = 1$ and g is a fixed constant, we can improve the construction from the previous section using ideas from BCH codes. Let $q_0 \geqslant g + 1$ be a prime power. Define

$$A_\ell = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}.$$

Note that $A_1 = A_2 = \dots = A_g$. We will construct $\beta_1, \beta_2, \dots, \beta_r$ similarly as in the previous construction, but since we will only need $\mathbb{F}_{q_0}$ linear independence of β's, we can improve the construction by using BCH codes. Let $q_1 = q_0^s$ where $s = \lceil \log_{q_0}(r) \rceil$, note that $r \leqslant q_1 \leqslant q_0 r = O(gr) = O(n)$. Let $\alpha_1, \alpha_2, \dots, \alpha_r \in \mathbb{F}_{q_1}$ be distinct. Let $m = \min\{r-1, h\}$ and let $m' = m - \lceil m/q_0 \rceil$ and define $\beta_1, \beta_2, \dots, \beta_r \in \mathbb{F}_{q_1^{m'}}$ as

$$\beta_i = \begin{bmatrix} \alpha_i \\ \alpha_i^2 \\ \vdots \\ \alpha_i^{q_0-1} \\ \alpha_i^{q_0+1} \\ \vdots \\ \alpha_i^m \end{bmatrix},$$

where we are expressing β_i in some basis for $\mathbb{F}_{q_1^{m'}}$ (which is a $\mathbb{F}_{q_1}$-vector space of dimension m'). Note that we are skipping powers of α_i which are divisible by q_0. Therefore the dimension of β_i with entries in $\mathbb{F}_{q_1}$ is $m' = m - \lfloor m/q_0 \rfloor$. Let γ be a generator of $\mathbb{F}_{q_1^{m'}}^* = \mathbb{F}_{q_0^{sm'}}^*$. Define

$$B_\ell = \begin{bmatrix} \beta_1 & \beta_2 & \cdots & \beta_r \\ \gamma^\ell \beta_1^{q_0} & \gamma^\ell \beta_2^{q_0} & \cdots & \gamma^\ell \beta_r^{q_0} \\ \gamma^{\ell(1+q_0)} \beta_1^{q_0^2} & \gamma^{\ell(1+q_0)} \beta_2^{q_0^2} & \cdots & \gamma^{\ell(1+q_0)} \beta_r^{q_0^2} \\ \vdots & \vdots & & \vdots \\ \gamma^{\ell(1+q_0+\dots+q_0^{h-2})} \beta_1^{q_0^{h-1}} & \gamma^{\ell(1+q_0+\dots+q_0^{h-2})} \beta_2^{q_0^{h-1}} & \cdots & \gamma^{\ell(1+q_0+\dots+q_0^{h-2})} \beta_r^{q_0^{h-1}} \end{bmatrix}.$$

Claim 3.2. *The above construction is an MR LRC over fields of size*

$$q = q_1^{m'} \leqslant (O(n))^{m - \lfloor m/q_0 \rfloor}$$

where $q_0 \geqslant g + 1$ is any prime power and $m = \min\{r - 1, h\}$.

Proof. The proof is analogous to the proof of Theorem 1.3. We only need $\mathbb{F}_{q_0}$-linear independence of any $m + 1$ columns of

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \beta_1 & \beta_2 & \cdots & \beta_r \end{bmatrix}.$$

This follows from the properties of the BCH code construction. Since we only care about $\mathbb{F}_{q_0}$-linear independence, it is enough to show linear independence of any $m+1$ columns of the $(m+1) \times r$ matrix over $\mathbb{F}_{q_1}$ given by

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \alpha_1 & \alpha_2 & \cdots & \alpha_r \\ \alpha_1^2 & \alpha_2^2 & \cdots & \alpha_r^2 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1^m & \alpha_2^m & \cdots & \alpha_r^m \end{bmatrix}$$

where we added back all the rows where the powers are multiples of q_0. This follows trivially, since this is a Vandermonde matrix. □

Acknowledgment

We thank Sergey Yekhanin for several illuminating discussions about MR-LRCs and Umberto Martínez-Peñas for helpful comments on an earlier version of this paper.

References

[Ber15] Elwyn R Berlekamp. *Algebraic coding theory (revised edition).* World Scientific, 2015.

[BHH13] Mario Blaum, James Lee Hafner, and Steven Hetzler. Partial-MDS codes and their application to RAID type of architectures. *IEEE Transactions on Information Theory*, 59(7):4510–4519, 2013.

[Bla13] Mario Blaum. Construction of PMDS and SD codes extending RAID 5. Arxiv 1305.0032, 2013.

[BPSY16] Mario Blaum, James Plank, Moshe Schwartz, and Eitan Yaakobi. Construction of partial MDS and sector-disk codes with two global parity symbols. *IEEE Transactions on Information Theory*, 62(5):2673–2681, 2016.

[BSC+12] Alin Bostan, Bruno Salvy, Muhammad FI Chowdhury, Éric Schost, and Romain Lebreton. Power series solutions of singular (q)-differential equations. In *Proceedings of the 37th International Symposium on Symbolic and Algebraic Computation*, pages 107–114, 2012.

[BU14] Delphine Boucher and Felix Ulmer. Linear codes using skew polynomials with automorphisms and derivations. *Designs, codes and cryptography*, 70(3):405–431, 2014.

[CHL07] Minghua Chen, Cheng Huang, and Jin Li. On maximally recoverable property for multi-protection group codes. In *IEEE International Symposium on Information Theory (ISIT)*, pages 486–490, 2007.

[CK17] Gokhan Calis and Ozan Koyluoglu. A general construction fo PMDS codes. *IEEE Communications Letters*, 21(3):452–455, 2017.

[CMST20] Han Cai, Ying Miao, Moshe Schwartz, and Xiaohu Tang. A construction of maximally recoverable codes with order-optimal field size. *arXiv preprint arXiv:2011.13606*, 2020.

[FG15] Michael A. Forbes and Venkatesan Guruswami. Dimension expanders via rank condensers. In *Proceedings of the 19th International Workshop on Randomization and Computation (RANDOM)*, pages 800–814, 2015.

[FS12] Michael A. Forbes and Amir Shpilka. On identity testing of tensors, low-rank recovery and compressed sensing. In *Proceedings of the 44th ACM Symposium on Theory of Computing*, pages 163–172. ACM, 2012.

[FSS14] Michael A. Forbes, Ramprasad Saptharishi, and Amir Shpilka. Hitting sets for multilinear read-once algebraic branching programs, in any order. In *Proceedings of the ACM Symposium on Theory of Computing*, pages 867–875, 2014.

[GGY20] Sivakanth Gopi, Venkatesan Guruswami, and Sergey Yekhanin. Maximally recoverable LRCs: A field size lower bound and constructions for few heavy parities. *IEEE Trans. Inf. Theory*, 66(10):6066–6083, 2020.

[GHJY14] Parikshit Gopalan, Cheng Huang, Bob Jenkins, and Sergey Yekhanin. Explicit maximally recoverable codes with locality. *IEEE Transactions on Information Theory*, 60(9):5245–5256, 2014.

[GHK$^+$17] Parikshit Gopalan, Guangda Hu, Swastik Kopparty, Shubhangi Saraf, Carol Wang, and Sergey Yekhanin. Maximally recoverable codes for grid-like topologies. In *28th Annual Symposium on Discrete Algorithms (SODA)*, pages 2092–2108, 2017.

[GHSY12] Parikshit Gopalan, Cheng Huang, Huseyin Simitci, and Sergey Yekhanin. On the locality of codeword symbols. *IEEE Transactions on Information Theory*, 58(11):6925 –6934, 2012.

[GJX20] Venkatesan Guruswami, Lingfei Jin, and Chaoping Xing. Constructions of maximally recoverable local reconstruction codes via function fields. *IEEE Trans. Inf. Theory*, 66(10):6133–6143, 2020.

[GK16] Venkatesan Guruswami and Swastik Kopparty. Explicit subspace designs. *Combinatorica*, 36(2):161–185, 2016.

[GRX18] Venkatesan Guruswami, Nicolas Resch, and Chaoping Xing. Lossless dimension expanders via linearized polynomials and subspace designs. In *33rd Computational Complexity Conference (CCC 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

[Gur11] Venkatesan Guruswami. Linear-algebraic list decoding of folded Reed-Solomon codes. In *Proceedings of the 26th IEEE Conference on Computational Complexity*, pages 77–85, 2011.

[GW11] Venkatesan Guruswami and Carol Wang. Optimal rate list decoding via derivative codes. In *Proceedings of APPROX/RANDOM 2011*, pages 593–604, August 2011.

[GW13] Venkatesan Guruswami and Carol Wang. Linear-algebraic list decoding for variants of Reed-Solomon codes. *IEEE Transactions on Information Theory*, 59(6):3257–3268, 2013.

[GXY18] Venkatesan Guruswami, Chaoping Xing, and Chen Yuan. Constructions of subspace designs via algebraic function fields. *Trans. Amer. Math. Soc.*, 370:8757–8775, 2018.

[GYBS17] Ryan Gabrys, Eitan Yaakobi, Mario Blaum, and Paul Siegel. Construction of partial MDS codes over small finite fields. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 1–5, 2017.

[HCL07] Cheng Huang, Minghua Chen, and Jin Li. Pyramid codes: flexible schemes to trade space for access efficiency in reliable data storage systems. In *6th IEEE International Symposium on Network Computing and Applications (NCA 2007)*, pages 79–86, 2007.

[HSX$^+$12] Cheng Huang, Huseyin Simitci, Yikang Xu, Aaron Ogus, Brad Calder, Parikshit Gopalan, Jin Li, and Sergey Yekhanin. Erasure coding in Windows Azure Storage. In *USENIX Annual Technical Conference (ATC)*, pages 15–26, 2012.

[HY16] Guangda Hu and Sergey Yekhanin. New constructions of SD and MR codes over small finite fields. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 1591–1595, 2016.

[Lam85] Tsit-Yuen Lam. *A general theory of Vandermonde matrices.* Center for Pure and Applied Mathematics, University of California, Berkeley, 1985.

[LL88] Tsit-Yuen Lam and André Leroy. Vandermonde and wronskian matrices over division rings. *Journal of Algebra*, 119(2):308–336, 1988.

[Mar20] Umberto Martínez-Peñas. A general family of MSRD codes and PMDS codes with smaller field sizes from extended Moore matrices. *CoRR*, abs/2011.14109, 2020.

[MK19] Umberto Martínez-Peñas and Frank R. Kschischang. Universal and dynamic locally repairable codes with maximal recoverability via sum-rank codes. *IEEE Trans. Inf. Theory*, 65(12):7790–7805, 2019.

[MP18] Umberto Martínez-Peñas. Skew and linearized reed–solomon codes and maximum sum rank distance codes over any division ring. *Journal of Algebra*, 504:587–612, 2018.

[MPK19] Umberto Martínez-Peñas and Frank R Kschischang. Reliable and secure multishot network coding using linearized reed-solomon codes. *IEEE Transactions on Information Theory*, 2019.

[MV13] Hessam Mahdavifar and Alexander Vardy. Algebraic list-decoding of subspace codes. *IEEE Transactions on Information Theory*, 59(12):7814–7828, 2013.

[NUF10] Roberto W Nóbrega and Bartolomeu F Uchôa-Filho. Multishot codes for network coding using rank-metric codes. In *2010 Third IEEE International Workshop on Wireless Network Coding*, pages 1–6. IEEE, 2010.

[Ore33] Oystein Ore. Theory of non-commutative polynomials. *Annals of mathematics*, pages 480–508, 1933.

[PD14] Dimitris Papailiopoulos and Alexandros Dimakis. Locally repairable codes. *IEEE Transactions on Information Theory*, 60(10):5843–5855, 2014.

[SAP$^+$13] Maheswaran Sathiamoorthy, Megasthenis Asteris, Dimitris S. Papailiopoulos, Alexandros G. Dimakis, Ramkumar Vadali, Scott Chen, and Dhruba Borthakur. XORing elephants: novel erasure codes for big data. In *Proceedings of VLDB Endowment (PVLDB)*, pages 325–336, 2013.

[TPD16] Itzhak Tamo, Dimitris Papailiopoulos, and Alexandros G. Dimakis. Optimal locally repairable codes and connections to matroid theory. *IEEE Transactions on Information Theory*, 62:6661–6671, 2016.

A Missing Proofs

Lemma A.1 (Lemma 2.12)**.** *1.* ${}^{d}({}^{c}a) = {}^{dc}a$

2. Conjugacy is an equivalence relation, i.e., we can partition $\mathbb{K}$ into conjugacy classes where elements in each part are conjugates of each other, but elements in different parts are not conjugates.

Proof. (1) follows easily from the definition of conjugation and the using the fact that $\delta(cd) = \sigma(c)\delta(d) + \delta(c)d$.

$$\begin{aligned}{}^{d}({}^{c}a) &= \sigma(d) \cdot {}^{c}a \cdot d^{-1} + \delta(d)d^{-1} \\ &= \sigma(d)(\sigma(c)ac^{-1} + \delta(c)c^{-1})d^{-1} + \delta(d)d^{-1} \\ &= \sigma(dc)ac^{-1}d^{-1} + \sigma(d)\delta(c)c^{-1}d^{-1} + \delta(d)d^{-1} \\ &= \sigma(dc)a(dc)^{-1} + (\sigma(d)\delta(c) + \delta(d)c)c^{-1}d^{-1} \\ &= \sigma(dc)a(dc)^{-1} + \delta(dc)(dc)^{-1} \\ &= {}^{dc}a.\end{aligned}$$

We now prove (2). Suppose a is a conjugate of b, i.e., $a = {}^{x}b$ for some $x \in \mathbb{K}^*$. Then ${}^{x^{-1}}a = {}^{x^{-1}}({}^{x}b) = {}^{x^{-1}x}b = b$. Therefore b is a conjugate of a. Suppose a is a conjugate of b, with $a = {}^{x}b$, and c is a conjugate of b, with $b = {}^{y}c$. Then $a = {}^{x}b = {}^{x}({}^{y}c) = {}^{xy}c$. So a is a conjugate of c. □

Lemma A.2 (Lemma 2.14)**.** *1. $\mathbb{K}_a$ is a subfield of $\mathbb{K}$.*[*]

2. If $a, b \in \mathbb{K}$ are conjugates, then $\mathbb{K}_a = \mathbb{K}_b$. [†]

Proof. (1) Let $x, y \in \mathbb{K}_a \setminus \{0\}$ i.e. ${}^{x}a = {}^{y}a = a$. Then

$$\begin{aligned}{}^{x+y}a(x+y) &= \sigma(c+d)a + \delta(c+d) \\ &= \sigma(c)a + \sigma(d)a + \delta(c) + \delta(d) \\ &= {}^{c}ac + {}^{d}ad \\ &= ac + ad = a(c+d).\end{aligned}$$

Therefore ${}^{x+y}a = a$. Also ${}^{yx}a = {}^{y}({}^{x}a) = a$. And finally ${}^{x^{-1}}a = {}^{x^{-1}}({}^{x}a) = {}^{x^{-1}x}a = a$.

(2) Suppose $b = {}^{d}a$ and let $c \in \mathbb{K}_a$.Then ${}^{c}b = {}^{c}({}^{d}a) = {}^{cd}a = {}^{dc}a = {}^{d}({}^{c}a) = {}^{d}a = b$. Therefore $\mathbb{K}_a \subset \mathbb{K}_b$. By symmetry, $\mathbb{K}_b \subset \mathbb{K}_a$. □

Lemma A.3 (Product evaluation rule (Lemma 2.16))**.** *If $g(a) = 0$, then $(fg)(a) = 0$. If $g(a) \neq 0$ then*

$$(fg)(a) = f\left({}^{g(a)}a\right)g(a).$$

[*] When $\mathbb{K}$ is a division ring, $\mathbb{K}_a$ will be a sub-division ring of $\mathbb{K}$.

[†] When $\mathbb{K}$ is a division ring and not a field, we have $\mathbb{K}_{({}^{x}a)} = x\mathbb{K}_a x^{-1}$.

Proof. If $g(a) = 0$, then $g(t) = b(t)(t-a)$ for some $b(t) \in \mathbb{K}[t;\sigma,\delta]$. Therefore $f(t)g(t) = f(t)b(t)(t-a)$, and so $(fg)(a) = 0$. Suppose $g(a) \neq 0$. Let $g(t) = b(t)(t-a) + g(a)$ and $f(t) = a(t)\left(t - {}^{g(a)}a\right) + f\left({}^{g(a)}a\right)$. Then

$$\begin{aligned}
f(t)g(t) &= f(t)\cdot(b(t)(t-a)+g(a)) \\
&= f(t)b(t)(t-a) + f(t)g(a) \\
&= f(t)b(t)(t-a) + \left(a(t)\left(t - {}^{g(a)}a\right) + f\left({}^{g(a)}a\right)\right)g(a) \\
&= f(t)b(t)(t-a) + a(t)\left(tg(a) - {}^{g(a)}a \cdot g(a)\right) + f\left({}^{g(a)}a\right)g(a) \\
&= f(t)b(t)(t-a) + a(t)\left(\sigma(g(a))t + \delta(g(a)) - \sigma(g(a))a - \delta(g(a))\right) + f\left({}^{g(a)}a\right)g(a) \\
&= f(t)b(t)(t-a) + a(t)\sigma(g(a))(t-a) + f\left({}^{g(a)}a\right)g(a) \\
&= \left(f(t)b(t) + a(t)\sigma(g(a))\right)(t-a) + f\left({}^{g(a)}a\right)g(a).
\end{aligned}$$

Therefore $(fg)(a) = f\left({}^{g(a)}a\right)g(a)$. □

B Roots of Skew Polynomials: Proof of Theorem 2.21

We restate Theorem 2.21 for convenience.

Theorem B.1. *Let $f \in \mathbb{K}[t;\sigma,\delta]$ be a degree d non-zero polynomial. Let A be the set of roots of f in $\mathbb{K}$ and let $A = \cup_i A_i$ be a partition of A into conjugacy classes. Fix some representatives $a_i \in A_i$. Let $V_i = \{y : {}^{y}a_i \in A_i\} \cup \{0\}$ which is a linear subspace over $\mathbb{F}_i = \mathbb{K}_{a_i}$ by Lemma 2.19. Then*

$$\sum_i \dim_{\mathbb{F}_i}(V_i) \leqslant d.$$

Proof. We will use induction on the degree. For the base case, it is clear that for a degree 0 polynomial, which is a non-zero constant, $\dim_{\mathbb{F}_i}(V_i) = 0$ for every i. We will now show the induction step.

For each i, let $d_i = \dim_{\mathbb{F}_i}(V_i)$. Fix some basis $y(i,1), y(i,2), \ldots, y(i,d_i) \in \mathbb{K}^*$ which span V_i with coefficients in $\mathbb{F}_i = \mathbb{K}_{a_i}$. WLOG, we can assume that $y(i,1) = 1$ for every i, by reassigning $a_i = {}^{y(i,1)}a_i$.

Fix some conjugacy class i^* s.t. $d_{i*} \geqslant 1$. Since $f(a_{i^*}) = 0$, we can write $f(t) = h(t)(t - a_{i^*})$ where $\deg(h) = d-1$. Now let A'_i be the roots of h in conjugacy class i and $V'_i = \{y : {}^{y}a_i \in A'_i \cup \{0\}$. We claim that $\dim_{\mathbb{F}_i}(V'_i) \geqslant \dim_{\mathbb{F}_i}(V_i)$ for every $i \neq i^*$ and $\dim_{\mathbb{F}_{i^*}}(V'_{i^*}) \geqslant \dim_{\mathbb{F}_{i^*}}(V_{i^*}) - 1$. By induction $\sum_i \dim_{\mathbb{F}_i}(V'_i) \leqslant d-1$. Therefore we have $\sum_i \dim_{\mathbb{F}_i}(V_i) \leqslant d$. We will now prove the claim in two parts.

Claim B.2. $\dim_{\mathbb{F}_i}(V'_i) \geqslant \dim_{\mathbb{F}_i}(V_i)$ *for every $i \neq i^*$.*

Proof. Fix some conjugacy class $i \neq i^*$. By Lemma 2.16,

$$f\left({}^{y(i,j)}a_i\right) = h\left({}^{y(i,j)\left({}^{y(i,j)}a_i - a_{i^*}\right)}a_i\right)\left({}^{y(i,j)}a_i - a_{i^*}\right).$$

Since a_i, a_{i^*} are in different conjugacy classes, ${}^{y(i,j)}a_i - a_{i^*} \neq 0$. So $b_j = {}^{y(i,j)({}^{y(i,j)}a_i - a_{i^*})}a_i$ for $j \in \{1,\ldots,d_i\}$ are d_i roots of h in the i^{th} conjugacy class A'_i. If we show that $y(i,j)({}^{y(i,j)}a_i - a_{i^*})$ for $j \in \{1,\ldots,d_i\}$ are linearly independent over $\mathbb{F}_i$, then this proves the claim.

Suppose they are not independent. Then there exists $c_1, \ldots, c_{d_i} \in \mathbb{F}_i$ s.t. $\sum_{j=1}^{d_i} c_j y(i,j)({}^{y(i,j)}a_i - a_{i^*}) = 0$. Therefore,

$$\begin{aligned} a_{i^*} \sum_{j=1}^{d_i} c_j y(i,j) &= \sum_{j=1}^{d_i} c_j y(i,j) \cdot {}^{y(i,j)}a_i \\ &= \sum_{j=1}^{d_i} c_j y(i,j) \cdot {}^{c_j y(i,j)}a_i && (c_j \in \mathbb{F}_i = \mathbb{K}_{a_i}) \\ &= \left(\sum_{i=1}^{d_i} c_j y(i,j)\right) {}^{\left(\sum_{j=1}^{d_i} c_j y(i,j)\right)}a_i && ({}^{x+y}a(x+y) = {}^{x}ax + {}^{y}ay \text{ for all } x, y \in \mathbb{K}^*) \end{aligned}$$

Since $y(i,1), \ldots, y(i,d_i)$ are independent over $\mathbb{F}_i$, $\sum_{j=1}^{d_i} c_j y(i,j) \neq 0$. Therefore ${}^{\left(\sum_{j=1}^{d_i} c_j y(i,j)\right)}a_i = a_{i^*}$. This is a contradiction because a_i, a_{i^*} are in different conjugate classes. □

Claim B.3. $\dim_{\mathbb{F}_{i^*}}(V'_{i^*}) \geqslant \dim_{\mathbb{F}_{i^*}}(V_{i^*}) - 1$.

Proof. The proof is exactly similar to that of the previous claim, up until the last. Let $j \in \{2, 3, \ldots, d_{i^*}\}$. By Lemma 2.16,

$$f\left({}^{y(i^*,j)}a_{i^*}\right) = h\left({}^{y(i^*,j)\left({}^{y(i^*,j)}a_{i^*} - a_{i^*}\right)}a_{i^*}\right)\left({}^{y(i^*,j)}a_{i^*} - a_{i^*}\right) .$$

Since $y(i^*,1) = 1$ and $y(i^*,j)$ are linearly independent over $\mathbb{F}_{i^*}$, $y(i^*,j) \notin \mathbb{F}_{i^*}$. Therefore ${}^{y(i^*,j)}a_{i^*} - a_{i^*} \neq 0$. So $b_j = {}^{y(i^*,j)({}^{y(i^*,j)}a_{i^*} - a_{i^*})}a_{i^*}$ for $j \in \{2, \ldots, d_{i^*}\}$ are $d_{i^*} - 1$ roots of h in the i^{*th} conjugacy class A'_{i^*}. If we show that $y(i^*,j)({}^{y(i^*,j)}a_{i^*} - a_{i^*})$ for $j \in \{2, \ldots, d_{i^*}\}$ are linearly independent over $\mathbb{F}_{i^*}$, then this proves the claim.

Suppose they are not independent. Then there exists $c_2, \ldots, c_{d_{i^*}} \in \mathbb{F}_{i^*}$ s.t.

$$\sum_{j=2}^{d_{i^*}} c_j y(i^*,j)({}^{y(i^*,j)}a_{i^*} - a_{i^*}) = 0 .$$

Therefore,

$$\begin{aligned} a_{i^*} \sum_{j=2}^{d_{i^*}} c_j y(i^*,j) &= \sum_{j=2}^{d_{i^*}} c_j y(i^*,j) \cdot {}^{y(i^*,j)}a_{i^*} \\ &= \sum_{j=2}^{d_{i^*}} c_j y(i^*,j) \cdot {}^{c_j y(i^*,j)}a_{i^*} && (c_j \in \mathbb{F}_{i^*} = C(a_{i^*})) \\ &= \left(\sum_{j=2}^{d_{i^*}} c_j y(i^*,j)\right) {}^{\left(\sum_{j=2}^{d_{i^*}} c_j y(i^*,j)\right)}a_{i^*} && ({}^{x+y}a(x+y) = {}^{x}ax + {}^{y}ay \text{ for all } x, y \in \mathbb{K}^*) \end{aligned}$$

Since $y(i^*,1), \ldots, y(i^*,d_{i^*})$ are independent over $\mathbb{F}_{i^*}$, $\sum_{j=2}^{d_{i^*}} c_j y(i^*,j) \neq 0$. Therefore

$${}^{\left(\sum_{j=2}^{d_{i^*}} c_j y(i^*,j)\right)}a_{i^*} = a_{i^*} ,$$

and thus $\sum_{j=2}^{d_{i^*}} c_j y(i^*,j) \in C(a_{i^*}) = \mathbb{F}_{i^*}$. But this contradicts the fact that

$$\{y(i^*,1) = 1, y(i^*,2), \ldots, y(i^*,d_{i^*})\}$$

are linearly independent over $\mathbb{F}_{i^*}$. □

The above two claims finish the proof of Theorem 2.21. □

C Skew Polynomial Wronskian and Moore matrices

In this section, we will discuss generalizations of Wronskian and Moore matrices using skew polynomials. The non-singularity of special cases of these matrices has been instrumental in work on list decoding [GW11, GW13] and algebraic pseudorandomness such as constructions of rank condensers and subspace designs [FG15, GK16, GXY18]. We will need the following simple lemmas.

Lemma C.1. *Let $\mathbb{F}(x)$ be the field of rational functions in x and let $\mathbb{L} = \mathbb{F}(x^r)$ which is a subfield of $\mathbb{F}(x)$.*[*] *Let $g_1, g_2, \ldots, g_m \in \mathbb{F}[x]^{<r}$ be polynomials of degree strictly less than r. Then $g_1, g_2, \ldots, g_m$ are $\mathbb{L}$-linearly independent iff they are $\mathbb{F}$-linearly independent.*

Proof. One direction is obvious since $\mathbb{F}$ is a subfield of $\mathbb{L}$. To prove the other direction, suppose $g_1, g_2, \ldots, g_m$ are $\mathbb{L}$-linearly dependent, i.e., $\sum_i c_i(x^r)g_i(x) = 0$ for some $c_i \in \mathbb{F}(x)$. WLOG, by clearing denominators and common factors, we can assume that c_i are also polynomials (i.e., $c_i \in \mathbb{F}[x]$) with no common factor. By comparing the coefficients of powers of x between 0 and $r-1$, we immediately get that $\sum_i c_i(0)g_i(x) = 0$. Note that all $c_i(0)$ cannot be zero simultaneously since then x would be a common factor for all c_i. Therefore we get a non-trivial $\mathbb{F}$-linear dependency for $g_1, g_2, \cdots, g_m$. □

Lemma C.2. *Let $\mathbb{K}[x; \sigma, \delta]$ be a skew polynomial ring. For $a \in \mathbb{K}$, define $\phi_a : \mathbb{K} \to \mathbb{K}$ as $\phi_a(y) = \sigma(y)a + \delta(y)$. Then*

1. *$\phi_a^i(y) = N_i({}^y a)y$ where ϕ_a^i is ϕ_a composed with itself i times and*
2. *ϕ_a is a linear map over the subfield $\mathbb{K}_a$.*

Proof. (1) This can be proved by induction, it is true for $i = 1$.

$$\begin{aligned} N_{i+1}({}^y a)y &= \sigma(N_i({}^y a)){}^y a y + \delta(N_i({}^y a))y \\ &= \sigma(N_i({}^y a))(\sigma(y)a + \delta(y)) + \delta(N_i({}^y a))y \\ &= \sigma(N_i({}^y a)y)a + \sigma(N_i({}^y a))\delta(y) + \delta(N_i({}^y a))y \\ &= \sigma(N_i({}^y a)y)a + \delta(N_i({}^y a)y) \\ &= \phi_a(N_i({}^y a)y) = \phi_a(\phi_a^i(y)) = \phi_a^{i+1}(y). \end{aligned}$$

(2) $\mathbb{K}_a$-linearity follows since $\forall c \in \mathbb{K}_a$,

$$\phi_a(yc) = N_1({}^{yc}a)yc = N_1({}^y({}^c a))yc = N_1({}^y a)yc = \phi_a(y)c \ . \quad \square$$

Using Lemma C.2, one can linearize the evaluation of skew-polynomials on any conjugacy class. This gives a bijection between evaluation of skew-polynomials on a particular conjugacy class and *linearized polynomials* which found several applications in coding theory and linear-algebraic pseudorandomness [MV13, GRX18, Ber15]. In fact this is a ring isomorphism and the product operation denoted by $\otimes$ in [MV13] is equivalent to the product operation for skew polynomials in the appropriate skew polynomial ring.

[*] $\mathbb{F}(x^r)$ is the set of rational functions of the form $f(x^r)$ for $f \in \mathbb{F}(x)$ i.e. rational functions which only have terms whose powers are multiples of r.

C.1 Wronskian matrix

The theory of skew polynomials allows us to calculate rank of Wronskian matrices. Let $\mathbb{K}[x;\delta]$ be a skew-polynomial of derivation type i.e. $\sigma \equiv \mathrm{Id}$ is the identity map.

Definition C.3 (Wronskian)**.** *Let $c_1, \ldots, c_n \in \mathbb{K}^*$. Define the Wronskian*

$$W_n(c_1, \ldots, c_n) = \begin{bmatrix} c_1 & c_2 & \cdots & c_n \\ \delta(c_1) & \delta(c_2) & \cdots & \delta(c_n) \\ \delta^2(c_1) & \delta^2(c_2) & \cdots & \delta^2(c_n) \\ \vdots & \vdots & & \vdots \\ \delta^{n-1}(c_1) & \delta^{n-1}(c_2) & \cdots & \delta^{n-1}(c_n) \end{bmatrix}.$$

Corollary C.4. *$W_n(c_1, \ldots, c_n)$ is full-rank iff $c_1, \ldots, c_n$ are linearly independent over $\mathbb{F} = \mathbb{K}_0$, the centralizer of* 0.

Proof. By Lemma C.2, $\delta^i(c) = N_i({}^c0)c$. Thus the claim follows from Lemma 2.26. □

Note that when δ is the formal derivative of polynomials, the above is the usual Wronskian of polynomials. Applying the above corollary in this special case, we can relate the non-singularity of the Wronskian to the linear independence of the polynomials.

Proposition C.5. *Let $f_1, f_2, \ldots, f_s \in \mathbb{F}[x]$ be polynomials of degree at most d. Suppose $\delta^j(f_i)$ is the j^{th} derivative of f_i. Define*

$$M = \begin{bmatrix} f_1(x) & f_2(x) & \ldots & f_s(x) \\ \vdots & \vdots & & \vdots \\ \delta^j(f_1)(x) & \delta^j(f_2)(x) & \ldots & \delta^j(f_s)(x) \\ \vdots & \vdots & & \vdots \\ \delta^{s-1}(f_1)(x) & \delta^{s-1}(f_2)(x) & \ldots & \delta^{s-1}(f_s)(x) \end{bmatrix}.$$

Then the following are true:

1. *If* $\mathrm{char}(\mathbb{F}) = p$ *then*[*], $\det(M) \neq 0$ *iff* $f_1, f_2, \ldots, f_2$ *are linearly independent over* $\mathbb{F}(x^p)$.
2. *If* $\mathrm{char}(\mathbb{F}) > d$ *or* $\mathrm{char}(\mathbb{F}) = 0$ *then,* $\det(M) \neq 0$ *iff* $f_1, f_2, \ldots, f_s$ *are linearly independent over* $\mathbb{F}$.

Proof. It is clear that if $f_1, f_2, \ldots, f_d$ are linearly dependent over $\mathbb{F}$, then $\det M = 0$. Now we will prove the converse.

Consider the skew polynomial ring defined in Example 2.5 where $\mathbb{K} = \mathbb{F}(x), \sigma \equiv \mathrm{Id}$ and $\delta(p)$ is the derivative of p. By Corollary C.4, $\det(M)$ is zero iff $f_1, f_2, \ldots, f_s$ are linearly independent over $\mathbb{K}_0$, the centralizer of 0. We have

$$\mathbb{K}_0 = \{g : {}^g0 = 0\} \cup \{0\} = \{g : \delta(g) = 0\}.$$

If $\mathrm{char}(\mathbb{F}) = 0$, then $\mathbb{K}_0 = \mathbb{F}$ and we are done. If $\mathrm{char}(\mathbb{F}) = p$ for some prime p, then we claim below that $\mathbb{K}_0 = \mathbb{F}(x^p)$, which finishes the proof using Lemma C.1. □

Claim C.6. *If* $\mathrm{char}(\mathbb{F}) = p$*, then* $\mathbb{K}_0 = \mathbb{F}(x^p)$.

[*] $\mathrm{char}(\mathbb{F})$ is the characteristic of $\mathbb{F}$.

Proof. $\mathbb{K}_0 = \{g \in \mathbb{F}(x) : \delta(g) = 0\}$. If $g \in \mathbb{F}[x]$, then it is easy to see that $\delta(g) = 0$ iff $g \in \mathbb{F}[x^p]$. Now suppose g is a rational function of the form $g = a/b$ where $a, b \in \mathbb{F}[x]$ do not have any common factors. By product rule, $\delta(g) = 0 \iff \delta(a)b = a\delta(b)$. Since a, b do not have any common factors, this implies that a divides $\delta(a)$ and b divides $\delta(b)$. Since degree of $\delta(a)$ is smaller than a, this is not possible unless $\delta(a) = 0$ and similarly we can conclude that $\delta(b) = 0$. Therefore $a, b \in \mathbb{F}[x^p]$ and so $g \in \mathbb{F}(x^p)$. □

Using the above, we can now deduce the following result which is the basis of list-size bound for list decoding univariate multiplicity codes [GW11] and the analysis of the associated subspace design constructed in [GK16].

Proposition C.7. *Let* $\mathrm{char}(\mathbb{F}) = p$. *Let* δ *be the derivative operator on polynomials in* $\mathbb{F}[x]$ *and* $\delta^i(\cdot)$ *be the* i^{th} *derivative of a polynomial. Let* $Q(x, y_0, y_1, \ldots, y_{s-1}) = A(x) + \sum_{i=0}^{s-1} A_i(x)y_i$ *where* $A(x), A_i(x) \in \mathbb{F}[x]$ *and not all* A_i *are zero. The set of all* $f \in \mathbb{F}[x]$ *of degree less than p, such that*

$$Q(x, f(x), \delta(f)(x), \ldots, \delta^{s-1}(f)(x)) = 0, \tag{12}$$

form an $\mathbb{F}$*-affine subspace of* $\mathbb{F}[x]$ *of dimension at most* $s - 1$.

Proof. Equation (12) can be rewritten as $A + \sum_{i=0}^{s-1} A_i\delta^i(f) = 0$. Suppose that the set of solutions to this equation in $\mathbb{F}[x]^{<p}$ form an $\mathbb{F}$-affine subspace of $\mathbb{F}[x]$ of dimension at least s. Then there exist solutions $f_0, f_1, \ldots, f_s \in \mathbb{F}[x]^{<p}$ where $f_1 - f_0, \ldots, f_s - f_0$ are $\mathbb{F}$-linearly independent. Let $g_i = f_i - f_0$. Then for $j \in [s]$ we have, $\sum_{i=0}^{s-1} A_i\delta^i(g_j) = 0$. Therefore the determinant of the matrix $[\delta^i(g_j)]_{ij}$ is zero. Therefore by Proposition C.5, $g_1, g_2, \ldots, g_s$ should be $\mathbb{F}$-linearly dependent, which is a contradiction. □

We also remark that solving equation (12) when $A = 0$ is equivalent to finding roots of a skew polynomial of degree $s - 1$ in a conjugacy class. This also intuitively explains why the set of solutions is an affine subspace of dimension at most $s-1$. Consider the skew polynomial ring $\mathbb{K}[t; \delta]$ of derivation type where $\mathbb{K} = \mathbb{F}(x)$, $\sigma \equiv \mathrm{Id}$ and δ is the derivative operator. Then by Lemma C.2, $N_i(^f0)f = \delta^i(f)$. Therefore the Equation (12), when $A = 0$, can be rewritten as:

$$\sum_{i=0}^{s-1} A_i\delta^i(f) = 0 \iff \sum_{i=0}^{s-1} A_iN_i(^f0)f = 0.$$

Define $G(t) \in \mathbb{K}[t; \delta]$ as $G(t) = \sum_{i=0}^{s-1} A_it^i$ which is a skew polynomial of degree at most $s-1$. Then $G(0^f)f = \sum_{i=0}^{s-1} A_iN_i(^f0)f$. Therefore the solutions of (12) when $A = 0$ are precisely $\{0\} \cup \{f : G(0^f) = 0\}$.

C.2 Moore matrix

The theory of skew polynomials also allows us to calculate the rank of Moore matrices. Let $\mathbb{K}[t; \sigma]$ be a skew polynomial ring of endomorphism type i.e. $\delta \equiv 0$. This is completely analogous to Wronskian matrices (Section C.1) once we use the skew polynomial framework.

Definition C.8 (Moore matrix)**.** *Let* $c_1, \ldots, c_n \in \mathbb{K}^*$. *Define the Moore matrix*

$$M_n(c_1, \ldots, c_n) = \begin{bmatrix} c_1 & c_2 & \cdots & c_n \\ \sigma(c_1) & \sigma(c_2) & \cdots & \sigma(c_n) \\ \sigma^2(c_1) & \sigma^2(c_2) & \cdots & \sigma^2(c_n) \\ \vdots & \vdots & & \vdots \\ \sigma^{n-1}(c_1) & \sigma^{n-1}(c_2) & \cdots & \sigma^{n-1}(c_n) \end{bmatrix}.$$

Corollary C.9. *$M_n(c_1, \ldots, c_n)$ is full-rank iff $c_1, \ldots, c_n$ are linearly independent over $\mathbb{F} = \mathbb{K}_1$, the centralizer of* 1.

Proof. By Lemma C.2, $\sigma^i(c) = N_i({}^c1)c$. Thus the claim follows from Lemma 2.26. □

We now apply the above to the case when $\mathbb{K} = \mathbb{F}_q(x)$ and σ is the automorphism which maps $f(x) \in \mathbb{F}_q(x)$ to $f(\gamma x)$ for a generator γ of $\mathbb{F}_q^*$. In this case, the Moore matrix was called the folded Wronskian in [GK16]. Analogous Moore matrices for function fields were studied in [GXY18].

Proposition C.10. *Let $f_1, f_2, \ldots, f_s \in \mathbb{F}_q[x]$ be polynomials of degree at most d. Let γ be generator for $\mathbb{F}_q^*$. Define*

$$M = \begin{bmatrix} f_1(x) & f_2(x) & \ldots & f_s(x) \\ \vdots & \vdots & & \vdots \\ f_1(\gamma^j x) & f_2(\gamma^j x) & \ldots & f_s(\gamma^j x) \\ \vdots & \vdots & & \vdots \\ f_1(\gamma^{s-1}x) & f_2(\gamma^{s-1}x) & \ldots & f_s(\gamma^{s-1}x) \end{bmatrix}.$$

Then the following are true:

1. *$\det(M) \neq 0$ iff $f_1, f_2, \ldots, f_2$ are linearly independent over $\mathbb{F}_q(x^{q-1})$.*
2. *If $q - 1 > d$ then, $\det(M) \neq 0$ iff $f_1, f_2, \ldots, f_s$ are linearly independent over $\mathbb{F}_q$.*

Proof. It is clear that if $f_1, f_2, \ldots, f_d$ are linearly dependent over $\mathbb{F}$, then $\det M = 0$. Now we will prove the converse.

Consider the skew polynomial ring defined in Example 2.5 where $\mathbb{K} = \mathbb{F}(x), \sigma(g(x)) = g(\gamma x)$ and $\delta \equiv 0$. By Corollary C.9, $\det(M)$ is zero iff $f_1, f_2, \ldots, f_s$ are linearly independent over $\mathbb{K}_1$, the centralizer of 1. We have

$$\mathbb{K}_1 = \{g : {}^g1 = 1\} \cup \{0\} = \{g : g(\gamma x) = g(x)\}.$$

We now claim that $\mathbb{K}_1 = \mathbb{F}(x^{q-1})$ and the rest follows from Lemma C.1. □

Claim C.11. $\mathbb{K}_1 = \mathbb{F}_q(x^{q-1})$.

Proof. $\mathbb{K}_1 = \{g \in \mathbb{F}(x) : g(\gamma x) = g(x)\}$. If $g \in \mathbb{F}[x]$, then it is easy to see that $g(\gamma x) = g(x)$ iff $g \in \mathbb{F}[x^{q-1}]$. Now suppose g is a rational function of the form $g = a/b$ where $a, b \in \mathbb{F}[x]$ do not have any common factors and we can assume that the constant term of a or b is 1. $g(\gamma x) = g(x) \iff a(\gamma x)b(x) = a(x)b(\gamma x)$. Since a, b do not have any common factors, this implies that a divides $a(\gamma x)$ and b divides $b(\gamma x)$. Since degree of $a(\gamma x)$ is the same as that of $a(x)$ and the degree of $b(\gamma x)$ is the same as that of $b(x)$, this implies that $a(\gamma x) = \lambda a(x)$ and $b(\gamma x) = \lambda b(x)$ for some $\lambda \in \mathbb{F}_q$. Since we assumed that a or b has constant term 1, we can conclude that $\lambda = 1$. Therefore $a, b \in \mathbb{F}_q[x^{q-1}]$ and so $g \in \mathbb{F}_q(x^{q-1})$. □

Using the above, we can now deduce the following result which is the basis of list-size bound for list decoding folded Reed-Solomon codes [Gur11, GW13] and the analysis of the subspace design constructed using folded Reed-Solomon codes [GK16].

Lemma C.12. *Let γ be a generator for $\mathbb{F}_q^*$. Let $Q(x, y_0, y_1, \ldots, y_{s-1}) = A(x) + \sum_{i=0}^{s-1} A_i(x)y_i$ where $A(x), A_i(x) \in \mathbb{F}_q[x]$. The set of all $f \in \mathbb{F}_q[x]$ of degree less than $q - 1$, such that*

$$Q(x, f(x), f(\gamma x), \ldots, f(\gamma^{s-1}x)) = 0, \tag{13}$$

form an $\mathbb{F}_q$-affine subspace of $\mathbb{F}_q[x]$ of dimension at most $s - 1$.

Proof. Equation (13) can be rewritten as $A+\sum_{i=0}^{s-1} A_i f(\gamma^i x)=0$. Suppose that the set of solutions to this equation in $\mathbb{F}_q[x]^{<q-1}$ form an $\mathbb{F}_q$-affine subspace of $\mathbb{F}_q[x]$ of dimension at least s. Then there exist solutions $f_0, f_1, \ldots, f_s \in \mathbb{F}_q[x]^{<q-1}$ where $f_1-f_0, \ldots, f_s-f_0$ are $\mathbb{F}_q$-linearly independent. Let $g_i=f_i-f_0$. Then for $j \in [s]$ we have, $\sum_{i=0}^{s-1} A_i g_j(\gamma^i x)=0$. Therefore the determinant of the matrix $[g_j(\gamma^i x)]_{ij}$ is zero. Therefore by Proposition C.10, $g_1, g_2, \ldots, g_s$ should be $\mathbb{F}_q$-linearly dependent, which is a contradiction. □

Just as we did in Section C.1, we remark that solving Equation (13), when $A=0$, is equivalent to finding roots of the degree $s-1$ skew polynomial $G(t)=\sum_{i=0}^{s-1} A_i t^i$ in the conjugacy class of 1, where the underlying skew polynomial ring is $\mathbb{K}[t;\sigma]$ where $\mathbb{K}=\mathbb{F}(x)$ and $\sigma(f(x))=f(\gamma x)$.

D Maximum sum rank distance codes

In this section, we will present a construction of Maximum Sum-Rank Distance (MSRD) codes due to [MP18] using the skew polynomial framework. We will first define sum-rank distance codes.

Fix some basis $\mathcal{B}$ for $\mathbb{F}_{q^m}$ as vector space over $\mathbb{F}_q$. Given $z=(z_1, z_2, \ldots, z_r) \in \mathbb{F}_{q^m}^r$, we can think of z as an $m \times r$ matrix with entries in $\mathbb{F}_q$ by expressing each coordinate z_i as a $\mathbb{F}_q^m$ vector using basis $\mathcal{B}$; define $\operatorname{rank}_{\mathbb{F}_q}(z)$ to be the $\mathbb{F}_q$-rank of that matrix. Let $\mathcal{P}=A_1 \sqcup A_2 \sqcup \cdots \sqcup A_s$ be a partition of $[n]$ into s parts. Given $x \in \mathbb{F}_{q^m}^n$, let $x=(x_1, x_2, \ldots, x_s)$ be the partition of of x according to $\mathcal{P}$ where $x_i \in \mathbb{F}_{q^m}^{A_i}$. Define $\text{sum-rank}_{\mathcal{P}}(x)=\sum_{i=1}^{s} \operatorname{rank}_{\mathbb{F}_q}(x_i)$.

Definition D.1 (sum-rank distance)**.** *Fix some partition $\mathcal{P}=A_1 \sqcup A_2 \sqcup \cdots \sqcup A_s$ of $[n]$ into s parts. An $\mathbb{F}_{q^m}$-linear subspace C of $\mathbb{F}_{q^m}^n$ is said to have sum-rank distance d (w.r.t. partition $\mathcal{P}$) if every non-zero codeword $c \in C$,* $\text{sum-rank}_{\mathcal{P}}(c) \geqslant d$.

Note that the sum-rank distance generalizes both Hamming metric (by choosing $\mathcal{P}=\{1\} \sqcup \{2\} \sqcup \cdots \sqcup \{n\}$) and rank metric (by choosing $\mathcal{P}=[n]$). Moreover for any partition $\mathcal{P}$ and any $x \in \mathbb{F}_{q^m}^n$, $\text{sum-rank}_{\mathcal{P}}(x)$ is most the Hamming weight of x (as rank is upper bounded by the number of non-zero columns). Therefore by the Singleton bound, any k-dimensional code of $\mathbb{F}_{q^m}^n$, can have sum-rank distance at most $n-k+1$. A code achieving this bound is called an MSRD code. Therefore MSRD codes generalize both MDS codes and Gabidulin codes. Sum-rank distance was introduced by [NUF10] for applications in network coding. We will now present the construction of MSRD codes.

Theorem D.2 (Construction of maximum sum rank distance codes [MP18])**.** *Let γ be a generator for $\mathbb{F}_{q^m}$ and let $\beta_1, \ldots, \beta_m \in \mathbb{F}_{q^m}$ be linearly independent over $\mathbb{F}_q$. Let $n=(q-1)m$. For $k \leqslant n$, define a $k \times n$ matrix $M=[M_0|M_1|\ldots|M_{q-2}]$ where*

$$M_\ell = \begin{bmatrix} \beta_1 & \beta_2 & \cdots & \beta_m \\ \gamma^\ell \beta_1^q & \gamma^\ell \beta_2^q & \cdots & \gamma^\ell \beta_m^q \\ \gamma^{\ell(1+q)} \beta_1^{q^2} & \gamma^{\ell(1+q)} \beta_2^{q^2} & \cdots & \gamma^{\ell(1+q)} \beta_m^{q^2} \\ \vdots & \vdots & & \vdots \\ \gamma^{\ell(1+q+\cdots+q^{k-2})} \beta_1^{q^{k-1}} & \gamma^{\ell(1+q+\cdots+q^{k-2})} \beta_2^{q^{k-1}} & \cdots & \gamma^{\ell(1+q+\cdots+q^{k-2})} \beta_m^{q^{k-1}} \end{bmatrix}.$$

Then M is the generator matrix of a maximum sum rank distance code, i.e., for every non-zero vector $\lambda \in \mathbb{F}_{q^m}^k$, $\sum_{\ell=0}^{q-2} \operatorname{rank}_{\mathbb{F}_q}(\lambda^T M_\ell) \geqslant n-k+1$.[*]

[*] Here we are interpreting a row vector $c \in \mathbb{F}_{q^m}^r$ as an $m \times r$ matrix over $\mathbb{F}_q$. $\operatorname{rank}_{\mathbb{F}_q}(c)$ is the $\mathbb{F}_q$-rank of this matrix. We will also use $\ker_{\mathbb{F}_q}(c)$ in the proof to denote the kernel of the matrix.

Proof. Suppose $\lambda \in \mathbb{F}_{q^m}^k$ is a non-zero vector such that $\sum_{\ell=0}^{q-2} \operatorname{rank}_{\mathbb{F}_q}(\lambda^T M_\ell) \leqslant n-k$. This is equivalent to $\sum_{\ell=0}^{q-2} \dim_{\mathbb{F}_q}(\ker_{\mathbb{F}_q}(\lambda^T M_\ell)) \geqslant k$.

Let $\mathbb{K} = \mathbb{F}_{q^m}$, $\sigma(a) = a^q$ and $\delta \equiv 0$. See Example 2.15 for the conjugation relation and conjugacy classes in this case. Define $f(t) = \sum_{i=0}^{k-1} \lambda_i t^i$ which is a non-zero skew polynomial of degree at most $k-1$ in $\mathbb{F}_{q^m}[t;\sigma]$. We will find many roots for f which would violate Theorem 2.21 to get a contradiction.

Fix some $\ell \in \{0,1,\ldots,q-2\}$. Suppose $\dim_{\mathbb{F}_q}(\ker_{\mathbb{F}_q}(\lambda^T M_\ell)) = d_\ell$. Let $\mu_1, \ldots, \mu_{d_\ell} \in \mathbb{F}_q^m$ be a basis for the kernel. Let $\beta = (\beta_1, \beta_2, \ldots, \beta_m) \in \mathbb{F}_{q^m}^m$. Now $\lambda^T M_\ell \mu_i = 0$ implies that $\beta^T \mu_i$ is root of f. Moreover the d_ℓ roots $\beta^T \mu_1, \ldots, \beta^T \mu_{d_\ell} \in \mathbb{F}_{q^m}$ are linearly independent over $\mathbb{F}_q$ since $\operatorname{rank}_{\mathbb{F}_q}(\beta) = m$.

Thus we get $\sum_{\ell=0}^{q-2} d_\ell \geqslant k$ roots for f. And the roots in each conjugacy class are linearly independent over $\mathbb{F}_q$ (which is the centralizer). Therefore by Theorem 2.21, we get a contradiction. □

It is easy to see that the above construction can be easily modified to work for any partition $\mathcal{P}$ of $[n]$ into at most $(q-1)$ parts, where each part has size at most m. In [MPK19], an efficient decoding algorithm for these codes is given.