A repair scheme for a distributed storage system based on multivariate polynomials

Hiram H. López*, Gretchen L. Matthews, and Daniel Valvo

Abstract A distributed storage system stores data across multiple nodes, with the primary objective of enabling efficient data recovery even in the event of node failures. The main goal of an exact repair scheme is to recover the data from a failed node by accessing and downloading information from the rest of the nodes. In a groundbreaking paper, Guruswami and Wootters (2017) developed an exact repair scheme for a distributed storage system that is based on Reed-Solomon codes, which depend on single-variable polynomials. In these notes, we extend the repair scheme to the family of distributed storage systems based on Reed-Muller codes, which are linear codes based on multivariate polynomials. The repair scheme we propose repairs any single node failure and multiple node failures, provided the positions satisfy certain conditions.

Keywords

Reed-Solomon code, Reed-Muller code, repair scheme, distributed storage system

Hiram H. López

Department of Mathematics, Virginia Tech, Blacksburg, VA USA e-mail: `hhlopez@vt.edu`, Gretchen L. Matthews

Department of Mathematics, Virginia Tech, Blacksburg, VA USA e-mail: `gmatthews@vt.edu`, and Daniel Valvo

Department of Mathematics, Virginia Tech, Blacksburg, VA USA e-mail: `vdaniel1@vt.edu`.

Introduction

The goal of a distributed storage system is to store data over multiple storage nodes. A linear code, which is a vector space over a finite field, may be used in a distributed storage system setting to allow the information stored on a failed node to be recovered using the information stored on the remaining nodes. The general idea is the following.

- The information to be stored is encoded into codewords using a linear code.
- Every codeword is distributed across nodes so that each node stores a symbol.
- Recovering a failed node exactly is equivalent to fixing an erasure in the codeword Dimakis et al (2010), Dimakis et al (2011).

A repair scheme is an algorithm that recovers the value at any node using limited information from the values at the other nodes. Under certain conditions, some repair schemes require less information than standard approaches to repair. Thus, the mathematical goal is the following.

- Design a set of vectors in such a way that every entry of every vector can be recovered from the rest of the entries. In these notes, we use evaluation codes, meaning that the set of vectors is a vector space over a finite field, and every vector depends on the evaluation of a certain polynomial.
- Give an explicit description of an exact repair scheme. In other words, give the algorithm describing how every vector entry can be recovered from the rest of the entries. In these notes, we use the trace function from finite fields to describe the repair scheme.

An evaluation code is a linear code defined by evaluating a collection of polynomials on a set of points. Reed-Solomon codes, the most well-known family of evaluation codes, are defined by the evaluation of single-variable polynomials up to a certain degree on a set of points of the finite field $\mathbb{F}_q$ of size q. The design of linear exact repair schemes for distributed storage systems using evaluation codes began with the foundational work of Guruswami and Wootters in which they developed a repair scheme (GW scheme) to efficiently repair an erasure in a Reed-Solomon (RS) code; see Guruswami and Wootters (2017). The GW scheme highly depends on the dual of a Reed-Solomon code, which is a generalized Reed-Solomon code. For a general framework for evaluation codes, see Jaramillo et al (2021). For the dual of an evaluation code, see López et al (2021).

The GW scheme inspired the linear exact repair schemes for algebraic geometry codes Jin et al (2018) and Reed-Muller codes Chen and Zhang (2019). Similarly to Reed-Solomon codes, the Reed-Muller codes are defined by evaluating polynomials up to a certain degree in m variables on the points $\mathbb{F}_{q^t}^m$.

In these notes, we develop a repair scheme for several failures on a distributed storage system that is based on Reed-Muller codes, provided the positions satisfy a certain restriction. The approach we develop in these notes, which relies on the dual of an evaluation code López et al (2021), is different from the one used in Chen

and Zhang (2019), and it gives the basis to extend it to other families of codes, for instance, the family of Cartesian codes López et al (2014).

Preliminaries

Let q be a power of a prime p and $\mathbb{F}_{q^t}$ be a field extension of degree $t = [\mathbb{F}_{q^t} : \mathbb{F}_q]$ of $\mathbb{F}_q$. Let C be an $[n,k]$-linear code over $\mathbb{F}_{q^t}$, meaning a k-dimensional $\mathbb{F}_{q^t}$-subspace of $\mathbb{F}_{q^t}^n$. The elements of $\mathbb{F}_{q^t}$ are called *symbols* and the elements of $\mathbb{F}_q$ are called *subsymbols*. As $\mathbb{F}_{q^t}$ is a vector space of dimension t over $\mathbb{F}_q$, any codeword $c \in C$ (or more generally any element $w \in \mathbb{F}_{q^t}^n$) can be represented using n symbols or tn subsymbols.

Field trace

The *field trace* can be defined as the polynomial $\mathrm{Tr}_{\mathbb{F}_{q^t}/\mathbb{F}_q}(x) \in \mathbb{F}_{q^t}[x]$ given by

$$\mathrm{Tr}_{\mathbb{F}_{q^t}/\mathbb{F}_q}(x) := x^{q^{t-1}} + \cdots + x^q + x.$$

For the sake of convenience, we will often refer to $\mathrm{Tr}_{\mathbb{F}_{q^t}/\mathbb{F}_q}(x)$ as simply $\mathrm{Tr}(x)$ when the extension being used is obvious from context.

The following property of the trace function is crucial to developing the repair scheme.

Remark 1. Let $B = \{z_1, \ldots, z_t\}$ be a basis for $\mathbb{F}_{q^t}$ over $\mathbb{F}_q$. Then there exists a dual basis $B' = \{z'_1, \ldots, z'_t\}$ of $\mathbb{F}_{q^t}$ over $\mathbb{F}_q$, such that

$$\mathrm{Tr}(z_i z'_j) = \delta_{ij} := \begin{cases} 1 & i = j \\ 0 & \text{otherwise.} \end{cases}$$

In this case, B and B' are called *dual bases*. For $\alpha \in \mathbb{F}_{q^t}$,

$$\alpha = \sum_{i=1}^{t} \mathrm{Tr}(\alpha z_i) z'_i.$$

Thus, given dual bases B and B', determining α is equivalent to finding $\mathrm{Tr}(\alpha z_i)$ for all $i \in [t] := \{1, \ldots, t\}$; see Lidl and Niederreiter (1994).

Reed-Muller codes

By Reed-Muller codes, we mean the evaluation codes obtained when polynomials in m variables up to a certain total degree $d \in \mathbb{Z}_{\geq 0}$ are evaluated at all the points of $\mathbb{F}_{q^t}^m$. The precise definition is as follows.

Let $\mathbb{F}_{q^t}[x_1, \ldots, x_m]$ be the polynomial ring of m variables over $\mathbb{F}_{q^t}$. Denote by $\mathbb{F}_{q^t}[x_1, \ldots, x_m]_{\leq d}$ the set of those polynomials up to a certain total degree d.

Definition 1. Assume $\{P_1, \ldots, P_n\} = \mathbb{F}_{q^t}^m$, fixing an order on the $n := q^{t^m}$ elements of $\mathbb{F}_{q^t}^m$. The *Reed-Muller code* of degree d is given by

$$\mathrm{RM}(\mathbb{F}_{q^t}^m, d) := \left\{ f(\mathbb{F}_{q^t}^m) \ : \ f \in \mathbb{F}_{q^t}[x_1, \ldots, x_m]_{\leq d} \right\} \subseteq \mathbb{F}_{q^t}^n,$$

where $f(\mathbb{F}_{q^t}^m) := (f(P_1), \ldots, f(P_n))$.

Definition 2. Note that a Reed-Solomon code of dimension k of length q^t is defined as

$$\mathrm{RS}(\mathbb{F}_q^m, k) := \mathrm{RM}(\mathbb{F}_{q^t}, k-1).$$

The dual of a Reed-Muller code, denoted by $\mathrm{RM}(\mathbb{F}_{q^t}^m, d)^{\perp}$, is the set of all $\alpha \in \mathbb{F}_{q^t}^n$ such that $\alpha \cdot \beta = 0$ for all $\beta \in \mathrm{RM}(\mathbb{F}_{q^t}^m, d)$, where $\alpha \cdot \beta$ is the ordinary inner product in $\mathbb{F}_{q^t}^n$. The dual code $\mathrm{RM}(\mathbb{F}_{q^t}^m, d)^{\perp}$ has been extensively studied in the literature. See, for instance, Delsarte et al (1970) and Huffman and Pless (2003), where it is shown that the dual of a RM-code is another RM-code:

$$\mathrm{RM}(\mathbb{F}_{q^t}^m, d)^{\perp} = \mathrm{RM}(\mathbb{F}_{q^t}^m, d^{\perp}),$$

where $d^{\perp} := m(|\mathbb{F}_{q^t}| - 1) - d - 1 = m(q^t - 1) - d - 1$.

Exact repair scheme

In terms of distributed storage systems, each entry c_i of a codeword $c \in C$ represents the information stored on one of n different storage nodes. Informally, when one of the storage nodes fails, meaning that it is unavailable to serve a data request, an exact repair scheme is an algorithm designed to recover the information of the erased node in terms of data held by the other storage nodes. Formally, we have the following definition.

Definition 3. Let C be a linear code over $\mathbb{F}_{q^t}$ of length n and dimension k, given by a collection of functions $\mathscr{F}$ and a set of evaluation points A. A linear exact repair scheme for C over $\mathbb{F}_q$ consists of the following.

- For each $\alpha^* \in A$, and for each $\alpha \in A \setminus \{\alpha^*\}$, a set of queries $Q_\alpha(\alpha^*) \subseteq \mathbb{F}_{q^t}$.

- For each $\alpha^* \in A$, a linear reconstruction algorithm that computes

$$f(\alpha^*) = \sum_i \lambda_i z_i$$

for coefficients $\lambda_i \in B$ and a basis $\{z_1, \ldots, z_t\}$ for $\mathbb{F}_{q^t}$ over $\mathbb{F}_q$ so that the coefficients λ_i are $\mathbb{F}_q$-linear combinations of the queries

$$\bigcup_{\alpha \in A \setminus \{\alpha^*\}} \left\{ \mathrm{Tr}_{\mathbb{F}_{q^t}/\mathbb{F}_q}(\gamma f(\alpha)) : \gamma \in Q_\alpha(\alpha^*) \right\}.$$

We often omit the word linear because we consider only linear exact repair schemes.

The *repair bandwidth b* is the number of subsymbols the scheme downloads to recover the erased entry. As any element $c \in \mathbb{F}_{q^t}^n$ depends on nt subsymbols, the number $\frac{b}{nt}$ can be seen as the fraction of the codeword that is needed by the exact repair scheme to recover the erased symbol. It is important to note that to recover a symbol $c_i \in \mathbb{F}_{q^t}$ of a codeword c, these b elements of $\mathbb{F}_q$ rely on the entries c_j, $j \neq i$, of c but are not necessarily b of them.

A one-erasure repair scheme of a Reed-Muller code

In this section, we adapt the GW scheme of one erasure from Reed-Solomon codes to Reed-Muller codes.

Remark 2. In López et al (2022), the authors developed a repair scheme of one erasure for certain decreasing monomial-Cartesian and augmented Reed-Muller codes. We decided to focus these notes on Reed-Muller codes to determine, in Theorem 1, to what degree a Reed-Muller code can be repaired using ideas similar to Guruswami and Wootters (2017).

Theorem 1. *Let $RM(\mathbb{F}_{q^t}^m, d)$ be a Reed-Muller code such that $d \leq m(q^t - 1) - q^{t-1}$. There exists a repair scheme for one erasure with bandwidth at most*

$$b = q^{mt} - 1 + (t-1)\left(q^{(m-1)t} - 1\right).$$

Proof. Let $c := (f(P_1), \ldots, f(P_n)) \in RM(\mathbb{F}_{q^t}^m, d)$. Assume that the entry of c corresponding to $P^* = (p_1, \ldots, p_m) \in \mathbb{F}_{q^t}^m$, meaning $f(P^*)$, has been erased.

Recall that $\{z_1, \ldots, z_t\}$ is a basis for $\mathbb{F}_{q^t}$ over $\mathbb{F}_q$. For $i \in [t]$, define the following t polynomials, which we refer to as recovery polynomials:

$$\begin{aligned}
r_i(x) &= \frac{\mathrm{Tr}(z_i(x_1 - p_1))}{(x_1 - p_1)} \\
&= z_i + z_i^q (x_1 - p_1)^{q-1} + \cdots + z_i^{q^{t-1}} (x_1 - p_1)^{q^{t-1}-1}.
\end{aligned}$$

As $d \leq m(q^t - 1) - q^{t-1}$, then

$$
\begin{aligned}
d^{\perp} &= m(q^t - 1) - d - 1 \\
&\geq m(q^t - 1) - (m(q^t - 1) - q^{t-1}) - 1 \\
&= q^{t-1} - 1 \geq \deg r_i(x).
\end{aligned}
$$

It follows that every polynomial $r_i(x)$ defines an element in $\mathrm{RM}(\mathbb{F}_{q^t}^m, d)^{\perp}$, meaning $r_i(\mathbb{F}_{q^t}^m) \in \mathrm{RM}(\mathbb{F}_{q^t}^m, d)^{\perp}$.

As a consequence, we obtain the t equations

$$
r_i(P^*)f(P^*) = - \sum_{\mathbb{F}_{q^t}^m \setminus \{P^*\}} r_i(P)f(P) \quad \forall i \in [t]. \tag{1}
$$

As $r_i(P^*) = z_i$, applying the trace function to both sides of the previous equations and employing the linearity of the trace function, we obtain

$$
\mathrm{Tr}\left(z_i f(P^*)\right) = - \sum_{\mathbb{F}_{q^t}^m \setminus \{P^*\}} \mathrm{Tr}\left(r_i(P)f(P)\right) \quad \forall i \in [t].
$$

Define the set $\Gamma := \{p_1\} \times \mathbb{F}_{q^t} \times \cdots \times \mathbb{F}_{q^t} \subset \mathbb{F}_{q^t}^m$. Then,

$$
r_i(P) =
\begin{cases}
z_i & \text{if } P \in \Gamma \\
\\
\frac{\mathrm{Tr}(z_i(s_P - p_1))}{(s_P - p_1)} & \text{if } P \notin \Gamma,
\end{cases}
$$

where s_P is the first entry of the point $P \in \mathbb{F}_{q^t}^m$. Note $P^* \in \Gamma$. Therefore, we obtain that for $i \in [t]$,

$$
\begin{aligned}
\mathrm{Tr}\left(z_i f(P^*)\right) &= \sum_{\mathbb{F}_{q^t}^m \setminus \{P^*\}} \mathrm{Tr}\left(r_i(P)f(P)\right) \\
\\
&= \sum_{\Gamma \setminus \{P^*\}} \mathrm{Tr}\left(r_i(P)f(P)\right) + \sum_{\mathbb{F}_{q^t}^m \setminus \Gamma} \mathrm{Tr}\left(r_i(P)f(P)\right) \\
\\
&= \sum_{\Gamma \setminus \{P^*\}} \mathrm{Tr}\left(z_i f(P)\right) + \sum_{\mathbb{F}_{q^t}^m \setminus \Gamma} \mathrm{Tr}\left(\frac{\mathrm{Tr}(z_i(s_P - p_1))}{s_P - p_1} f(P)\right) \\
\\
&= \sum_{\Gamma \setminus \{P^*\}} \mathrm{Tr}\left(z_i f(P)\right) + \sum_{\mathbb{F}_{q^t}^m \setminus \Gamma} \mathrm{Tr}(z_i(s_P - p_1)) \mathrm{Tr}\left(\frac{f(P)}{s_P - p_1}\right).
\end{aligned}
$$

The properties of the trace function imply that the entry $f(P^*)$ can be recovered from its t independent traces $\mathrm{Tr}(z_i f(P^*))$, which can be obtained by downloading the following information from every element $P \neq P^*$:

- if $P \in \Gamma \setminus \{P^*\}$, download $f(P)$.
- if $P \notin \Gamma$, download $\mathrm{Tr}\left(\dfrac{f(P)}{s_P - p_1}\right)$.

Hence, the bandwidth is

$$
\begin{aligned}
b &= t(|\Gamma| - 1) + |\mathbb{F}_{q^t}^m \setminus \Gamma| = t(q^{(m-1)t} - 1) + q^{mt} - q^{(m-1)t} \\
&= tq^{(m-1)t} - t + q^{mt} - q^{(m-1)t} \\
&= (t-1)q^{(m-1)t} - t + q^{mt} \\
&= q^{mt} - 1 + (t-1)\left(q^{(m-1)t} - 1\right),
\end{aligned}
$$

which concludes the proof.

As a consequence of Theorem 1, when $m = 1$, we obtain the GW repair scheme for Reed-Solomon codes.

Corollary 1. *Let $RM(\mathbb{F}_{q^t}, d)$ be a Reed-Muller code such that $d \leq q^t - q^{t-1} - 1$, meaning a Reed-Solomon code. Then, there exists a repair scheme for one erasure with bandwidth at most*

$$
b = q^t - 1.
$$

Proof. By Definition 2, $\mathrm{RS}(\mathbb{F}_q^m, k) = \mathrm{RM}(\mathbb{F}_{q^t}, k-1)$. So, we obtain the result as a consequence of Theorem 1. ∎

A two-erasures repair scheme of a Reed-Muller code

In this section, we adapt the GW scheme of one erasure from Reed-Solomon codes to two erasures on a Reed-Muller code.

We keep the same notation as in previous sections and develop a repair scheme that repairs two simultaneous erasures $f(\boldsymbol{s}')$ and $f(\boldsymbol{s}^*)$ on a distributed storage system based on a Reed-Muller code $\mathrm{RM}(\mathbb{F}_{q^t}, d)$ provided the erasure positions satisfy a certain condition.

Remark 3. In López et al (2022), the authors developed a repair scheme of two erasures for certain decreasing monomial-Cartesian and augmented Reed-Muller codes. We focused these notes on Reed-Muller codes to determine in Theorem 1 to which degree a Reed-Muller code can be repaired using ideas similar to Guruswami and Wootters (2017).

Theorem 2. *Let $c = (f(\boldsymbol{s}_1), \ldots, f(\boldsymbol{s}_n)) \in RM(\mathbb{F}_{q^t}^m, d)$, where $RM(\mathbb{F}_{q^t}^m, d)$ is a Reed-Muller code with $d \leq m(q^t - 1) - q^{t-1}$ and $n = q^{tm}$. Assume that c has the two erasures $f(\boldsymbol{s}')$ and $f(\boldsymbol{s}^*)$, where $\boldsymbol{s}' = (s_1', \ldots, s_n')$ and $\boldsymbol{s}^* = (s_1^*, \ldots, s_n^*)$. If there is $j \in n$ such that $s_j' - s_j^* \in \mathbb{F}_q^*$, then there exists a repair scheme for the two erasures with bandwidth at most*

$$
b = 2\left[n - 2 + (t-1)\left(q^{(m-1)t} - 2\right)\right].
$$

Proof. Note that the kernel of the trace $\ker \mathrm{Tr} = \left\{ \alpha \in \mathbb{F}_{q^t} : \mathrm{Tr}(\alpha) = 0 \right\}$ has dimension $t-1$ as an $\mathbb{F}_q$-vector space. Let $\{z_1, \ldots, z_{t-1}\}$ be an $\mathbb{F}_q$-basis for $\ker \mathrm{Tr}$ and z_t an element in $\mathbb{F}_{q^t}$ such that $\{z_1, \ldots, z_{t-1}, z_t\}$ is an $\mathbb{F}_q$-basis for $\mathbb{F}_{q^t}$. Finally, let $\tau \in \ker \mathrm{Tr}$. We are ready to define the repair polynomials. For $i \in [t]$, take

$$p_i(\boldsymbol{x}) = \tau \frac{\mathrm{Tr}\left(z_i(x_j - s_j^*)\right)}{\left(x_j - s_j^*\right)} \qquad \text{and} \qquad q_i(\boldsymbol{x}) = \frac{\mathrm{Tr}\left(z_i(x_j - s_j')\right)}{(x_j - s_j')}.$$

As $d \leq m(q^t - 1) - q^{t-1}$, the polynomials $p_i(\boldsymbol{x})$ and $q_i(\boldsymbol{x})$ define elements in the dual code $\mathrm{RM}(\mathbb{F}_{q^t}^m, d)^\perp$. Therefore, in a similar way to the proof of Theorem 1, we obtain the following $2t$ equations:

$$p_i(\boldsymbol{s}^*)f(\boldsymbol{s}^*) + p_i(\boldsymbol{s}')f(\boldsymbol{s}') = -\sum_{\boldsymbol{s} \in \mathbb{F}_{q^t}^m \setminus \{\boldsymbol{s}^*, \boldsymbol{s}'\}} p_i(\boldsymbol{s})f(\boldsymbol{s}), \qquad i \in [t] \qquad (2)$$

and

$$q_i(\boldsymbol{s}^*)f(\boldsymbol{s}^*) + q_i(\boldsymbol{s}')f(\boldsymbol{s}') = -\sum_{\boldsymbol{s} \in \mathbb{F}_{q^t}^m \setminus \{\boldsymbol{s}^*, \boldsymbol{s}'\}} q_i(\boldsymbol{s})f(\boldsymbol{s}), \qquad i \in [t]. \qquad (3)$$

By definition of the p_i's, we have

$$p_i(\boldsymbol{s}^*) = \tau z_i \qquad \text{and} \qquad p_i(\boldsymbol{s}') = \tau \mathrm{Tr}(z_i), \qquad i \in [t].$$

By definition of the q_i's, we have

$$q_i(\boldsymbol{s}^*) = \mathrm{Tr}(z_i) \qquad \text{and} \qquad q_i(\boldsymbol{s}') = z_i, \qquad i \in [t].$$

Equations 2 and 3 become

$$\tau z_i f(\boldsymbol{s}^*) + \tau \mathrm{Tr}(z_i) f(\boldsymbol{s}') = -\sum_{\boldsymbol{s} \in \mathbb{F}_{q^t}^m \setminus \{\boldsymbol{s}^*, \boldsymbol{s}'\}} p_i(\boldsymbol{s})f(\boldsymbol{s}), \qquad i \in [t]$$

and

$$\mathrm{Tr}(z_i) f(\boldsymbol{s}^*) + z_i f(\boldsymbol{s}') = -\sum_{\boldsymbol{s} \in \mathbb{F}_{q^t}^m \setminus \{\boldsymbol{s}^*, \boldsymbol{s}'\}} q_i(\boldsymbol{s})f(\boldsymbol{s}), \qquad i \in [t].$$

As $\{z_1, \ldots, z_{t-1}\}$ is an $\mathbb{F}_q$-basis for $\ker \mathrm{Tr}$, the last two equations imply

$$\mathrm{Tr}\left(\tau z_i f(\boldsymbol{s}^*)\right) = -\sum_{\boldsymbol{s} \in \mathbb{F}_{q^t}^m \setminus \{\boldsymbol{s}^*, \boldsymbol{s}'\}} \mathrm{Tr}\left(p_i(\boldsymbol{s})f(\boldsymbol{s})\right), \quad i \in [t-1],$$

(4)

$$\mathrm{Tr}\left(\tau z_t f(\boldsymbol{s}^*)\right) + \mathrm{Tr}\left(z_t\right)\mathrm{Tr}\left(\tau f(\boldsymbol{s}')\right) = -\sum_{\boldsymbol{s} \in \mathbb{F}_{q^t}^m \setminus \{\boldsymbol{s}^*, \boldsymbol{s}'\}} \mathrm{Tr}\left(p_t(\boldsymbol{s})f(\boldsymbol{s})\right), \quad (5)$$

$$\mathrm{Tr}\left(z_i f(\boldsymbol{s}')\right) = -\sum_{\boldsymbol{s} \in \mathbb{F}_{q^t}^m \setminus \{\boldsymbol{s}^*, \boldsymbol{s}'\}} \mathrm{Tr}\left(q_i(\boldsymbol{s})f(\boldsymbol{s})\right), \quad i \in [t-1], \quad (6)$$

$$\mathrm{Tr}\left(z_t\right)\mathrm{Tr}\left(f(\boldsymbol{s}^*)\right) + \mathrm{Tr}\left(z_t f(\boldsymbol{s}')\right) = -\sum_{\boldsymbol{s} \in \mathbb{F}_{q^t}^m \setminus \{\boldsymbol{s}^*, \boldsymbol{s}'\}} \mathrm{Tr}\left(q_t(\boldsymbol{s})f(\boldsymbol{s})\right). \quad (7)$$

We claim that the elements $f(\boldsymbol{s}^*)$ and $f(\boldsymbol{s}')$ can be recovered from the set

$$R := \left\{ \mathrm{Tr}\left(p_i(\boldsymbol{s})f(\boldsymbol{s})\right), \mathrm{Tr}\left(q_i(\boldsymbol{s})f(\boldsymbol{s})\right) : i \in [t], \boldsymbol{s} \in \mathbb{F}_{q^t}^m \setminus \{\boldsymbol{s}^*, \boldsymbol{s}'\} \right\}.$$

To prove this claim, we take the following steps.

Step 1: For $i \in [t-1]$, $\mathrm{Tr}\left(z_i f(\boldsymbol{s}')\right)$ can be recovered from set R and Equation 6.

Step 2: We have $\tau \in \ker \mathrm{Tr}$, whose $\mathbb{F}_q$-basis is $\{z_1, \ldots, z_{t-1}\}$. Hence, there exist $\alpha_1, \ldots, \alpha_{t-1}$ in $\mathbb{F}_q$ such that $\tau = \alpha_1 z_1 + \ldots + \alpha_{t-1} z_{t-1}$ and

$$\mathrm{Tr}\left(\tau f(\boldsymbol{s}')\right) = \sum_{i=1}^{t-1} \alpha_i \mathrm{Tr}\left(z_i f(\boldsymbol{s}')\right).$$

Thus, $\mathrm{Tr}\left(\tau f(\boldsymbol{s}')\right)$ can be recovered from Step 1.

Step 3: From Step 2 and Equations 4 and 5, $\left(\tau z_i f(\boldsymbol{s}^*)\right)$ can be recovered for $i \in [t]$. The element $\tau f(\boldsymbol{s}^*)$ can then be recovered from the t traces $\mathrm{Tr}(\tau z_i f(\boldsymbol{s}^*))$ by

$$\tau f(\boldsymbol{s}^*) = \mathrm{Tr}(\tau z_1 f(\boldsymbol{s}^*))z_1' + \ldots + \mathrm{Tr}(\tau z_t f(\boldsymbol{s}^*))z_t',$$

where $\{z_1', \ldots, z_t'\}$ is the dual basis of $\{z_1, \ldots, z_t\}$; see Remark 1. Thus, $f(\boldsymbol{s}^*)$ can be recovered by

$$f(\boldsymbol{s}^*) = \tau^{-1}\mathrm{Tr}(\tau z_1 f(\boldsymbol{s}^*))z_1' + \ldots + \tau^{-1}\mathrm{Tr}(\tau z_t f(\boldsymbol{s}^*))z_t'.$$

Step 4: From Step 3, and Equations 6 and 7, $\mathrm{Tr}\left(z_i f(\boldsymbol{s}')\right)$ can be recovered for $i \in [t]$ by the set R. Then, similarly to Step 3, the element $f(\boldsymbol{s}')$ can be recovered from the traces $\mathrm{Tr}\left(z_i f(\boldsymbol{s}')\right)$. This completes the proof of the claim.

Recall that for $i \in [t]$, we have the following expressions:

$$p_i(\boldsymbol{x}) = \tau \frac{\mathrm{Tr}\left(z_i(x_j - s_j^*)\right)}{\left(x_j - s_j^*\right)} \qquad \text{and} \qquad q_i(\boldsymbol{x}) = \frac{\mathrm{Tr}\left(z_i(x_j - s_j')\right)}{(x_j - s_j')}.$$

Define the sets

$$\Gamma^* := \mathbb{F}_{q^t} \times \cdots \times \{s_j^*\} \times \cdots \times \mathbb{F}_{q^t} = \{(\gamma_1, \ldots, \gamma_n) \in \mathbb{F}_{q^t}^m : \gamma_j = s_j^*\}$$

and

$$\Gamma' := \mathbb{F}_{q^t} \times \cdots \times \{s_j'\} \times \cdots \times \mathbb{F}_{q^t} = \{(\gamma_1, \ldots, \gamma_n) \in \mathbb{F}_{q^t}^m : \gamma_j = s_j'\}.$$

As a consequence of the claim, both erasures $f(\boldsymbol{s}')$ and $f(\boldsymbol{s}^*)$ can be recovered from the set

$$
\begin{aligned}
R &= \left\{ \mathrm{Tr}\left(p_i(\boldsymbol{s})f(\boldsymbol{s})\right), \mathrm{Tr}\left(q_i(\boldsymbol{s})f(\boldsymbol{s})\right) : i \in [t], \boldsymbol{s} \in \mathbb{F}_{q^t}^m \setminus \{\boldsymbol{s}^*, \boldsymbol{s}'\} \right\} \\
&= \left\{ \mathrm{Tr}\left(p_i(\boldsymbol{s})f(\boldsymbol{s})\right) : i \in [t], \boldsymbol{s} \in \mathbb{F}_{q^t}^m \setminus \{\boldsymbol{s}^*, \boldsymbol{s}'\} \right\} \\
&\quad \cup \left\{ \mathrm{Tr}\left(q_i(\boldsymbol{s})f(\boldsymbol{s})\right) : i \in [t], \boldsymbol{s} \in \mathbb{F}_{q^t}^m \setminus \{\boldsymbol{s}^*, \boldsymbol{s}'\} \right\} \\
&= \left\{ \mathrm{Tr}\left(p_i(\boldsymbol{s})f(\boldsymbol{s})\right) : i \in [t], \boldsymbol{s} \in \Gamma^* \setminus \{\boldsymbol{s}^*, \boldsymbol{s}'\} \right\} \\
&\quad \cup \left\{ \mathrm{Tr}\left(p_i(\boldsymbol{s})f(\boldsymbol{s})\right) : i \in [t], \boldsymbol{s} \notin \Gamma^* \right\} \\
&\quad \cup \left\{ \mathrm{Tr}\left(q_i(\boldsymbol{s})f(\boldsymbol{s})\right) : i \in [t], \boldsymbol{s} \in \Gamma' \setminus \{\boldsymbol{s}^*, \boldsymbol{s}'\} \right\} \\
&\quad \cup \left\{ \mathrm{Tr}\left(q_i(\boldsymbol{s})f(\boldsymbol{s})\right) : i \in [t], \boldsymbol{s} \notin \Gamma' \right\}.
\end{aligned}
$$

Observe that

$$
p_i(\boldsymbol{s}) = \begin{cases} \tau z_i & \text{if } \boldsymbol{s} \in \Gamma^* \\[2mm] \frac{\tau \mathrm{Tr}(z_i(s_P - s_j^*))}{(s_P - s_j^*)} & \text{if } \boldsymbol{s} \notin \Gamma^* \end{cases}
\quad \text{and} \quad
q_i(\boldsymbol{s}) = \begin{cases} z_i & \text{if } \boldsymbol{s} \in \Gamma' \\[2mm] \frac{\mathrm{Tr}(z_i(s_P - s_j'))}{(s_P - s_j')} & \text{if } \boldsymbol{s} \notin \Gamma' \end{cases}
$$

where s_P is the first entry of the point $\boldsymbol{s} \in \mathbb{F}_{q^t}^m$. Thus, the set R can be written as

$$
\begin{aligned}
R &= \left\{ \mathrm{Tr}\left(\tau z_i f(\boldsymbol{s})\right) : i \in [t], \boldsymbol{s} \in \Gamma^* \setminus \{\boldsymbol{s}^*, \boldsymbol{s}'\} \right\} \\
&\quad \cup \left\{ \mathrm{Tr}\left(\frac{\tau \mathrm{Tr}(z_i(s_P - s_j^*))}{(s_P - s_j^*)} f(\boldsymbol{s}) \right) : i \in [t], \boldsymbol{s} \notin \Gamma^* \right\} \\
&\quad \cup \left\{ \mathrm{Tr}\left(z_i f(\boldsymbol{s})\right) : i \in [t], \boldsymbol{s} \in \Gamma' \setminus \{\boldsymbol{s}^*, \boldsymbol{s}'\} \right\} \\
&\quad \cup \left\{ \mathrm{Tr}\left(\frac{\mathrm{Tr}(z_i(s_P - s_j'))}{(s_P - s_j')} f(\boldsymbol{s}) \right) : i \in [t], \boldsymbol{s} \notin \Gamma' \right\} \\
&= \left\{ \mathrm{Tr}\left(\tau z_i f(\boldsymbol{s})\right) : i \in [t], \boldsymbol{s} \in \Gamma^* \setminus \{\boldsymbol{s}^*, \boldsymbol{s}'\} \right\} \\
&\quad \cup \left\{ \mathrm{Tr}(z_i(s_P - s_j^*)) \mathrm{Tr}\left(\frac{\tau f(\boldsymbol{s})}{(s_P - s_j^*)} \right) : i \in [t], \boldsymbol{s} \notin \Gamma^* \right\} \\
&\quad \cup \left\{ \mathrm{Tr}\left(z_i f(\boldsymbol{s})\right) : i \in [t], \boldsymbol{s} \in \Gamma' \setminus \{\boldsymbol{s}^*, \boldsymbol{s}'\} \right\} \\
&\quad \cup \left\{ \mathrm{Tr}(z_i(s_P - s_j')) \mathrm{Tr}\left(\frac{f(\boldsymbol{s})}{(s_P - s_j')} \right) : i \in [t], \boldsymbol{s} \notin \Gamma' \right\}.
\end{aligned}
$$

Thus, both erasures $f(\mathbf{s}')$ and $f(\mathbf{s}^*)$ can be recovered by downloading the following information from every element $\mathbf{s} \in \mathbb{F}_{q^t}^m \setminus \{\mathbf{s}^*, \mathbf{s}'\}$:

- if $\mathbf{s} \in \Gamma^* \setminus \{\mathbf{s}^*, \mathbf{s}'\}$, download $f(\mathbf{s})$.
- if $\mathbf{s} \notin \Gamma^*$, download $\mathrm{Tr}\left(\dfrac{\tau f(\mathbf{s})}{s_P - s_j^*}\right)$.
- if $\mathbf{s} \in \Gamma' \setminus \{\mathbf{s}^*, \mathbf{s}'\}$, download $f(\mathbf{s})$.
- if $\mathbf{s} \notin \Gamma'$, download $\mathrm{Tr}\left(\dfrac{\tau f(\mathbf{s})}{s_P - s_j'}\right)$.

Note $|\Gamma^*| = |\Gamma'|$. Hence, the bandwidth is at most

$$
\begin{aligned}
b &= 2\left(t(|\Gamma| - 2) + |\mathbb{F}_{q^t}^m \setminus \Gamma|\right) = 2\left(t(q^{(m-1)t} - 2) + q^{mt} - q^{(m-1)t}\right) \\
&= 2\left(tq^{(m-1)t} - 2t + q^{mt} - q^{(m-1)t}\right) \\
&= 2\left((t-1)q^{(m-1)t} - 2t + q^{mt}\right) \\
&= 2\left(q^{mt} - 2 + (t-1)\left(q^{(m-1)t} - 2\right)\right) \\
&= 2\left(n - 2 + (t-1)\left(q^{(m-1)t} - 2\right)\right),
\end{aligned}
$$

which concludes the proof.

An ℓ-erasures repair scheme of a Reed-Muller code

In this section, we adapt the GW scheme of one erasure from Reed-Solomon codes to several erasures on a Reed-Muller code. We give the sketch to prove the case of three erasures. Such a sketch gives the key steps for the general case.

Let $c = (f(\mathbf{s}_1), \ldots, f(\mathbf{s}_n))$ be an element of a Reed-Muller code $\mathrm{RM}(\mathbb{F}_{q^t}^m, d)$, where $d \le m(q^t - 1) - q^{t-1}$ and $n = q^{tm}$. Assume that c has the three erasures $f(\mathbf{s}^1)$, $f(\mathbf{s}^2)$, and $f(\mathbf{s}^3)$, where $\mathbf{s}^i = (s_1^i, \ldots, s_n^i)$ for $i = 1, 2, 3$. If there is $j \in [n]$ such that $s_j^{i_1} - s_j^{i_2} \in \mathbb{F}_q^*$, for every $i_1 \ne i_2 \in [3]$, then there exists a repair scheme for the three erasures with bandwidth at most

$$
b = 3\left[n - 3 + (t-1)\left(q^{(m-1)t} - 3\right)\right].
$$

The sketch of the proof is as follows. Let $\{z_1, \ldots, z_{t-1}\}$ be an $\mathbb{F}_q$-basis for $\ker \mathrm{Tr}$ and z_t an element in $\mathbb{F}_{q^t}$ such that $\{z_1, \ldots, z_{t-1}, z_t\}$ is an $\mathbb{F}_q$-basis for $\mathbb{F}_{q^t}$. Let τ_1 and τ_2 be two elements of $\ker \mathrm{Tr}$ that are independent over $\mathbb{F}_q$. For $i \in [t]$, take

$$p_i(\boldsymbol{x}) = \frac{\mathrm{Tr}\left(z_i(x_j - s_j^1)\right)}{\left(x_j - s_j^1\right)}, \qquad q_i(\boldsymbol{x}) = \frac{\tau\,\mathrm{Tr}_1\left(z_i(x_j - s_j^2)\right)}{(x_j - s_j^2)},$$

$$\text{and} \qquad r_i(\boldsymbol{x}) = \frac{\tau_2\,\mathrm{Tr}\left(z_i(x_j - s_j^3)\right)}{(x_j - s_j^3)}.$$

As $d \leq m(q^t - 1) - q^{t-1}$, the polynomials $p_i(\boldsymbol{x})$, $q_i(\boldsymbol{x})$, and $r_i(\boldsymbol{x})$ define elements in the dual code $\mathrm{RM}(\mathbb{F}_{q^t}^m, d)^{\perp}$ López et al (2021). Define the set $S = \{\boldsymbol{s}^1, \boldsymbol{s}^2, \boldsymbol{s}^3\}$. Similarly to the proof of Theorem 2, we obtain the following $3t$ equations:

$$\sum_{j=1}^{3} p_i(\boldsymbol{s}^j) f(\boldsymbol{s}^j) = - \sum_{\boldsymbol{s} \in \mathbb{F}_{q^t}^m \setminus S} p_i(\boldsymbol{s}) f(\boldsymbol{s}), \qquad i \in [t],$$

$$\sum_{j=1}^{3} q_i(\boldsymbol{s}^j) f(\boldsymbol{s}^j) = - \sum_{\boldsymbol{s} \in \mathbb{F}_{q^t}^m \setminus S} q_i(\boldsymbol{s}) f(\boldsymbol{s}), \qquad i \in [t],$$

and

$$\sum_{j=1}^{3} r_i(\boldsymbol{s}^j) f(\boldsymbol{s}^j) = - \sum_{\boldsymbol{s} \in \mathbb{F}_{q^t}^m \setminus S} r_i(\boldsymbol{s}) f(\boldsymbol{s}), \qquad i \in [t].$$

The last equations give rise to

$$\mathrm{Tr}\left(z_i f(\boldsymbol{s}^1)\right) = - \sum_{\boldsymbol{s} \in \mathbb{F}_{q^t}^m \setminus S} \mathrm{Tr}\left(p_i(\boldsymbol{s}) f(\boldsymbol{s})\right),$$

$$\mathrm{Tr}\left(z_t f(\boldsymbol{s}^1)\right) + \sum_{j=2,3} \mathrm{Tr}(z_t)\,\mathrm{Tr}\left(f(\boldsymbol{s}^j)\right) = - \sum_{\boldsymbol{s} \in \mathbb{F}_{q^t}^m \setminus S} \mathrm{Tr}\left(p_t(\boldsymbol{s}) f(\boldsymbol{s})\right),$$

$$\mathrm{Tr}\left(\tau_1 z_i f(\boldsymbol{s}^2)\right) = - \sum_{\boldsymbol{s} \in \mathbb{F}_{q^t}^m \setminus S} \mathrm{Tr}\left(q_i(\boldsymbol{s}) f(\boldsymbol{s})\right),$$

$$\sum_{j=1,3} \mathrm{Tr}(z_t)\,\mathrm{Tr}\left(\tau_1 f(\boldsymbol{s}^j)\right) + \mathrm{Tr}\left(\tau_1 z_t f(\boldsymbol{s}^2)\right) = - \sum_{\boldsymbol{s} \in \mathbb{F}_{q^t}^m \setminus S} \mathrm{Tr}\left(q_t(\boldsymbol{s}) f(\boldsymbol{s})\right),$$

$$\mathrm{Tr}\left(\tau_2 z_i f(\boldsymbol{s}^3)\right) = - \sum_{\boldsymbol{s} \in \mathbb{F}_{q^t}^m \setminus S} \mathrm{Tr}\left(r_i(\boldsymbol{s}) f(\boldsymbol{s})\right),$$

$$\sum_{j=1,2} \mathrm{Tr}(z_t)\,\mathrm{Tr}\left(\tau_2 f(\boldsymbol{s}^j)\right) + \mathrm{Tr}\left(\tau_2 z_t f(\boldsymbol{s}^3)\right) = - \sum_{\boldsymbol{s} \in \mathbb{F}_{q^t}^m \setminus S} \mathrm{Tr}\left(r_t(\boldsymbol{s}) f(\boldsymbol{s})\right),$$

for $i \in [t-1]$.

   In a similar way to the proof of Theorem 2, the elements $f(\boldsymbol{s}^1)$, $f(\boldsymbol{s}^2)$, and $f(\boldsymbol{s}^3)$ can be recovered from the set

$$R := \left\{ \mathrm{Tr}\left(p_i(\boldsymbol{s})f(\boldsymbol{s})\right), \mathrm{Tr}\left(q_i(\boldsymbol{s})f(\boldsymbol{s})\right), \mathrm{Tr}\left(r_i(\boldsymbol{s})f(\boldsymbol{s})\right) : i \in [t], \boldsymbol{s} \in \mathbb{F}_{q^t}^m \setminus S \right\}.$$

For $i \in [3]$, define the sets

$$\Gamma^i := \mathbb{F}_{q^t} \times \cdots \times \{s_i^*\} \times \cdots \times \mathbb{F}_{q^t} = \{(\gamma_1, \ldots, \gamma_n) \in \mathbb{F}_{q^t}^m : \gamma_j = s_j^i\}.$$

As a consequence, following the proof of Theorem 2, the three erasures $f(\boldsymbol{s}^1)$, $f(\boldsymbol{s}^2)$, and $f(\boldsymbol{s}^3)$ can be recovered by downloading the following information for every element $\boldsymbol{s} \in \mathbb{F}_{q^t}^m \setminus S$ and $i = 1, 2, 3$:

- if $\boldsymbol{s} \in \Gamma^i \setminus S$, download $f(\boldsymbol{s})$.
- if $\boldsymbol{s} \notin \Gamma^i$, download $\mathrm{Tr}\left(\dfrac{\tau f(\boldsymbol{s})}{s_P - s_j^i} \right)$.

As $|\Gamma^1| = |\Gamma^2| = |\Gamma^3|$, the bandwidth is at most

$$\begin{aligned}
b &= 3\left(t(|\Gamma| - 3) + |\mathbb{F}_{q^t}^m \setminus \Gamma| \right) \\
&= 3\left(t(q^{(m-1)t} - 3) + q^{mt} - q^{(m-1)t} \right) \\
&= 3\left(tq^{(m-1)t} - 3t + q^{mt} - q^{(m-1)t} \right) \\
&= 3\left((t-1)q^{(m-1)t} - 3t + q^{mt} \right) \\
&= 3\left(q^{mt} - 3 + (t-1)\left(q^{(m-1)t} - 3 \right) \right) \\
&= 3\left(n - 3 + (t-1)\left(q^{(m-1)t} - 3 \right) \right),
\end{aligned}$$

which concludes the sketch of the proof.

In general, the case of ℓ erasures for the Reed-Muller code can be stated in the following way.

Theorem 3. *Let $c = (f(\boldsymbol{s}_1), \ldots, f(\boldsymbol{s}_n)) \in RM(\mathbb{F}_{q^t}^m, d)$, where $RM(\mathbb{F}_{q^t}^m, d)$ is a Reed-Muller code with $d \leq m(q^t - 1) - q^{t-1}$ and $n = q^{tm}$. Assume that c has ℓ erasures $f(\boldsymbol{s}^1), \ldots, f(\boldsymbol{s}^\ell)$, where $\boldsymbol{s}^i = (s_1^i, \ldots, s_n^i)$ for $i = 1, \ldots, \ell$. If there is $j \in [n]$ such that $s_j^{i_1} - s_j^{i_2} \in \mathbb{F}_q^*$, for every $i_1 \neq i_2 \in [\ell]$, then there exists a repair scheme for the ℓ erasures with bandwidth at most*

$$b = \ell \left[n - \ell + (t-1)\left(q^{(m-1)t} - \ell \right) \right].$$

Summary/Conclusion

A distributed storage system stores data across multiple nodes, with the primary objective of enabling efficient data recovery even in the event of node failures. The

main goal of an exact repair scheme is to recover the data from a failed node by accessing and downloading information from the rest of the nodes. In these notes, we extended the exact repair scheme developed in Guruswami and Wootters (2017) from Reed-Solomon codes to Reed-Muller codes with several erasures that satisfy certain conditions.

References

Chen T, Zhang X (2019) Repairing generalized reed-muller codes. CoRR abs/1906.10310, URL `http://arxiv.org/abs/1906.10310`, 1906.10310

Delsarte P, Goethals J, Mac Williams F (1970) On generalized reedmuller codes and their relatives. Information and Control 16(5):403–442, DOI https://doi.org/10.1016/S0019-9958(70)90214-7, URL `https://www.sciencedirect.com/science/article/pii/S0019995870902147`

Dimakis AG, Godfrey PB, Wu Y, Wainwright MJ, Ramchandran K (2010) Network coding for distributed storage systems. IEEE Transactions on Information Theory 56(9):4539–4551, DOI 10.1109/TIT.2010.2054295

Dimakis AG, Ramchandran K, Wu Y, Suh C (2011) A survey on network codes for distributed storage. Proceedings of the IEEE 99(3):476–489, DOI 10.1109/JPROC.2010.2096170

Guruswami V, Wootters M (2017) Repairing Reed-Solomon codes. IEEE Transactions on Information Theory 63(9):5684–5698, DOI 10.1109/TIT.2017.2702660

Huffman WC, Pless V (2003) Fundamentals of Error-Correcting Codes. Cambridge University Press

Jaramillo D, Vaz Pinto M, Villarreal RH (2021) Evaluation codes and their basic parameters. Designs, Codes and Cryptography 89(2):269–300, DOI 10.1007/s10623-020-00818-8, URL `https://doi.org/10.1007/s10623-020-00818-8`

Jin L, Luo Y, Xing C (2018) Repairing algebraic geometry codes. IEEE Transactions on Information Theory 64(2):900–908, DOI 10.1109/TIT.2017.2773089

Lidl R, Niederreiter H (1994) Introduction to Finite Fields and their Applications, 2nd edn. Cambridge University Press

López HH, Rentería-Márquez C, Villarreal RH (2014) Affine cartesian codes. Designs, Codes and Cryptography 71(1):5–19, DOI 10.1007/s10623-012-9714-2, URL `https://doi.org/10.1007/s10623-012-9714-2`

López HH, Soprunov I, Villarreal RH (2021) The dual of an evaluation code. Designs, Codes and Cryptography 89(7):1367–1403, DOI 10.1007/s10623-021-00872-w, URL `https://doi.org/10.1007/s10623-021-00872-w`

López HH, Matthews GL, Valvo D (2022) Erasures repair for decreasing monomial-cartesian and augmented reed-muller codes of high rate. IEEE Transactions on Information Theory 68(3):1651–1662, DOI 10.1109/TIT.2021.3130096