CSE 533: Error-Correcting Codes **(Autumn 2006)**

Lecture 7: Explicit Linear Code Constructions

October 20, 2006

Lecturer: Venkatesan Guruswami *Scribe: Patrick Tague*

1 Introduction

To this point, we have focused on the derivation of asymptotic bounds for the existence/non-existence of $(n, k, d)_q$ codes with rate $R = \frac{k}{n}$ and relative distance $\delta = \frac{d}{n}$. We are now interested in explicit constructions of linear codes in an attempt to achieve or approach the previously derived bounds.

2 Reed-Solomon Codes

A Reed-Solomon (RS) code is an $[n, k, d]_q$ linear code with $\Sigma = \mathbb{F}_q$ for a prime power $q \geq n$ described in terms of the following encoding function. The encoding function

$$Enc : \Sigma^k \to \Sigma^n$$

maps a k-symbol message $(m_0, \dots, m_{k-1})$ to an n-symbol codeword $(M(\alpha_0), \dots, M(\alpha_{n-1}))$ where $M(x)$ is the polynomial

$$M(x) = \sum_{i=0}^{k-1} m_i x^i, \tag{1}$$

and $\alpha_0, \dots, \alpha_{n-1}$ are distinct elements in $\mathbb{F}_q$. Typically, $q = n$ and the α_i'a are all thje elements of $\mathbb{F}_q$, or $n = q - 1$ and the α_i's are all the nonzero elements of of $\mathbb{F}_q$.

The linearity of an RS code $\mathcal{C}$ can be easily verified by checking the conditions for closure under addition and scalar multiplication. Let $c, c' \in \mathcal{C}$ be codewords corresponding to the messages $m = (m_0, \dots, m_{k-1})$ and $m' = (m'_0, \dots, m'_{k-1})$, respectively. Then $c + c'$ is the encoding of the message $(m_0 + m'_0, \dots, m_{k-1} + m'_{k-1})$ since

$$\begin{aligned} M(\alpha_i) + M'(\alpha_i) &= \sum_{j=0}^{k-1} m_j \alpha_i^j + \sum_{j=0}^{k-1} m'_j \alpha_i^j \\ &= \sum_{j=0}^{k-1} (m_j + m'_j) \alpha_i^j \ . \end{aligned} \tag{2}$$

Also for $\gamma \in \mathbb{F}_q$, γc is the encoding of the polynomial with coefficients $(\gamma m_0, \dots, \gamma m_{k-1})$.

An RS code $\mathcal{C}$ can thus be described using the $n \times k$ generator matrix G. From the encoding function Enc defined using (1), it is clear that G is the Vandermonde matrix

$$G = \begin{pmatrix} 1 & \alpha_0 & \dots & \alpha_0^{k-1} \\ 1 & \alpha_1 & \dots & \alpha_1^{k-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_{n-1} & \dots & \alpha_{n-1}^{k-1} \end{pmatrix}. \tag{3}$$

The minimum distance d of an RS code $\mathcal{C}$ can be computed algebraicly using Lemma 2.1.

Lemma 2.1. *A polynomial of degree D over a field $\mathbb{F}$ has at most D roots (counting multiplicity).*

Proof. The theorem is proved by induction on the degree D. The case $D = 0$ is obvious. Let $f(X)$ be a nonzero polynomial of degree D over $\mathbb{F}$. let $\alpha \in \mathbb{F}$ be a root of $f(X)$. By the division theorem for polynomials over a field, we can write $f(X) = Q(X)(X - \alpha) + R(X)$, where $R(X)$ is the remainder polynomial with degree less than 1, and therefore a constant polynomial. Since $f(\alpha) = R(\alpha) = 0$, we must have $R(X) = 0$. Therefore $f(X) = (X_\alpha)Q(X)$. By induction hypothesis, $Q(X)$, which has degree $D - 1$, has at most $D - 1$ roots. These roots together with α can make up at most D roots for $f(X)$. $\square$

Since the degree of the encoded polynomial in (1) is $k - 1$, a codeword c can have at most $k - 1$ elements $M(\alpha_i)$ equal to zero. The minimum distance d, equal to the minimum weight of any codeword in $\mathcal{C}$, is therefore at least as $d \geq n - k + 1$. The Singleton bound (proven in Lecture 5) provides a bound of $d \leq n - k + 1$ for any code. Hence, the minimum distance of the RS code $\mathcal{C}$ is $d = n - k + 1$. The upper bound can also be demonstrated by constructing a codeword with exactly $d = n-k+1$ non-zero entries. Let $M(x) = (x-\alpha_0)(x-\alpha_1)\dots(x-\alpha_{k-2})$ be the encoding polynomial as in (1). Since the degree of $M(x)$ is $k - 1$, there exists a message $m = [m_0, \dots, m_{k-1}]$ which corresponds to the polynomial $M(x)$, simply by mathing coefficients in (1). Hence, evaluating $M(x)$ for all $\alpha_i, i = 0, \dots, q - 1$ yields a codeword with $k - 1$ zeros followed by $n - k + 1$ non-zero entries. We record the distance property of RS codes as:

Lemma 2.2. *Reed-Solomon codes meet the Singleton bound, i.e., a code of block length n and dimension k has distance $n - k + 1$.*

RS codes can thus be used to achieve a relative distance of $\delta = \frac{d}{n} = \frac{n-k+1}{n} = 1 - R + o(1)$ for any rate $R = \frac{k}{n}$. However, the alphabet size q scales as $q = \Omega(n)$. By the Plotkin bound, for codes over an alphabet of size q, we have $R \leq 1 - \frac{q}{q-1}\delta$, so to meet the Singleton bound q has to grow with the block legnth n. We now use similar algebraic ideas to construct codes over smaller alphabet size, at the expense of worse rate vs distance trade-offs.

3 Reed-Muller Codes

In what follows, a generalization is provided for the RS codes described in Section 2 by expanding the polynomial encoding in (1) to multivariate polynomials. The resulting codes are hereafter referred to as Reed-Muller (RM) codes.[1]

3.1 Bivariate RM Codes

We begin with the simplest extension, from univariate to bivariate polynomials. Let m be the matrix $[m_{ij}]$ for $0 \le i \le \ell - 1$ and $0 \le j \le \ell - 1$ denoting a message of $k = \ell^2$ symbols in $\mathbb{F}_q$. The encoding function

$$Enc : \mathbb{F}_q^{\ell \times \ell} \to \mathbb{F}_q^{q \times q}$$

is given by mapping a message m to a codeword c given by the matrix $[M(\alpha_x, \alpha_y)]$ for $\alpha_x \in \mathbb{F}_q$ and $\alpha_y \in \mathbb{F}_q$, where $M(x, y)$ is given by

$$M(x,y) = \sum_{i=0}^{\ell-1} \sum_{j=0}^{\ell-1} m_{ij} x^i y^j. \tag{4}$$

The resulting RM code is a $[q^2, \ell^2, d]_q$ linear code. Linearity can be verified as in Section 2. The minimum distance d of the RM code can be computed using the following result.

Lemma 3.1. *The tensor product of two $[q, \ell, d]_q$ RS codes $\mathcal{C}_1$ and $\mathcal{C}_2$ is the $[q^2, \ell^2, d^2]_q$ (bivariate) RM code $\mathcal{C}$.*

Proof. The tensor product of two codes $\mathcal{C}_1$ and $\mathcal{C}_2$ is defined as the code $\mathcal{C} = \mathcal{C}_1 \otimes \mathcal{C}_2$ given by

$$\mathcal{C}_1 \otimes \mathcal{C}_2 = \left\{ G_1 m G_2^T \mid m \in \{0,1\}^{\ell \times \ell} \right\},$$

where G_1 and G_2 are the generator matrices for $\mathcal{C}_1$ and $\mathcal{C}_2$, respectively. Since both $\mathcal{C}_1$ and $\mathcal{C}_2$ are RS codes, the matrices G_1 and G_2 are both equal to the RS generator matrix G given in (3). Hence, a message m is mapped to the codeword $M = GmG^T \in \mathcal{C}$. The entry $M(\alpha_x, \alpha_y)$ in row x and column y of the codeword M is given by the product $g_x m g_y^T$, where g_x denotes the row $[1, \alpha_x, \ldots, \alpha_x^{\ell-1}]$ of G, for $0 \le x \le q - 1$. Hence, the product code is such that

$$M(\alpha_x, \alpha_y) = \sum_{i=0}^{\ell-1} \sum_{j=0}^{\ell-1} m_{ij} \alpha_x^i \alpha_y^j,$$

which is consistent with the definition of the bivariate Reed-Muller code $\mathcal{C}$ in (4) with x and y replaced with α_x and α_y. $\square$

The use of tensor product codes and the result of Lemma 3.1 implies that the $[q^2, \ell^2, d]_q$ Reed-Muller code has distance $d = (q - \ell + 1)^2 = q^2 - 2q(\ell - 1) + (\ell - 1)^2$ and rate $R = \frac{\ell^2}{q^2}$. Note that the distance $d = (q - \ell + 1)^2$ no longer achieves equality in the Singleton bound $d \le q^2 - \ell^2 + 1$. However, the alphabet size q in this case scales as $q = \mathcal{O}(\sqrt{n})$. This demonstrates the trade-off between optimal distance and smaller alphabet size that is characteristic of RM codes over RS codes.

[1] An alternate definition of Reed-Muller codes is common, but Prof. Guruswami claims the multivariate polynomial interpretation is more clear.

3.2 Multivariate RM Codes

The bivariate extension of Section 3.1 generalizes in the natural way to multivariate polynomials. A multivariate RM code $\mathcal{C}$ with v variables $x_1, \ldots, x_v$ can be interpreted as the tensor product code of v RS codes $\mathcal{C}_1, \ldots, \mathcal{C}_v$. The encoding function

$$Enc : \mathbb{F}_q^{\ell_1 \times \cdots \times \ell_v} \to \mathbb{F}_q^{q \times \cdots \times q}$$

maps a message $m = [m_{i_1 \ldots i_v}]$ to a codeword $M(x_1, \ldots, x_v)$ as

$$M(x_1, \ldots, x_v) = \sum_{i_1=0}^{\ell_1 - 1} \cdots \sum_{i_v=0}^{\ell_v - 1} m_{i_1 \ldots i_v} \prod_{j=1}^{v} x_j^{i_j}. \tag{5}$$

The resulting RM code is a $\left[q^v, \prod_{j=1}^v \ell_j, \prod_{j=1}^v d_j\right]_q$ linear code. Linearity can be verified using an identical method to that of Section 2. The minimum distance d of the multivariate RM code can be proven using the multivariate extension to Lemma 3.1 or using the following result.

Lemma 3.2. *A non-zero polynomial $P(x_1, \ldots, x_v)$ over a field $\mathbb{F}$ with maximum degree d_i for the variable x_i is non-zero in at least $\prod_{i=1}^v (q - d_i)$ points in $\mathbb{F}^v$.*

Proof. We use induction on v. The case $v = 1$ is the content of Lemma 2.1. Fix $x_1, \ldots, x_{v-1}$ and express $P(x_1, \ldots, x_v)$ as

$$P(x_1, \ldots, x_v) = R_{d_v}(x_1, \ldots, x_{v-1}) x_v^{d_v} + \ldots + R_0(x_1, \ldots, x_{v-1}),$$

which is a polynomial of degree d_v in the variable x_v. By Lemma 2.1, there are at least $q - d_v$ values of x_v for which $P(x_1, \ldots, x_v)$ is a non-zero olynomial in $x_1, \ldots, x_v$. For each of the (at least $q - d_v$) values of x_v which yield non-zero $P(x_1, \ldots, x_v)$, by induction there are at least $\prod_{i=1}^{v-1}(q - d_i)$ values to $x_1, \ldots, x_v$ that lead to a nonzero evaluation. $\square$

The following construction demonstrates how equality is achieved in the bound provided by Lemma 3.2. Since the bound results from fewer than $q - d_i$ roots for any given x_i, equality is achieved whenever there are exactly $q - d_i$ roots for each x_i. Hence, let $M_i(x_i)$ be the product $(x_i - \alpha_{i,1}) \ldots (x_i - \alpha_{i,\ell_i - 1})$, where the $\alpha_{i,j}$ are distinct, and let $M(x_1, \ldots, x_v) = \prod_{i=1}^v M_i(x_i)$.

3.3 Variant on Multivariate Reed-Muller Codes

We next relax the condition on multivariate RM codes independently bounding the maximum degree of each variable x_i and allow for codeword polynomials $M(x_1, \ldots, x_v)$ with total degree at most ℓ. The encoding function is similar to that in Section 3.2 with the encoding polynomial M given by

$$M(x_1, \ldots, x_v) = \sum_{\substack{i_1, \ldots, i_v \geq 0, \\ i_1 + \ldots + i_v \leq \ell}} m_{i_1 \ldots i_v} \prod_{j=1}^{v} x_j^{i_j}. \tag{6}$$

The resulting code $\mathcal{C}$ is a $[q^v, k, d]_q$ linear code, where k is the total number of tuples $(i_1, \ldots, i_v)$ of nonnegative integers satisfying $i_1 + \ldots + i_v \leq \ell$. The values of k and d are computed using the following results.

Observation 3.3. *The value k for the given code $\mathcal{C}$ is $\binom{v+l}{v}$ (stated without proof).*

Lemma 3.4. *A non-zero polynomial $P(x_1, \ldots, x_v)$ of total degree at most ℓ over $\mathbb{F}_q$ is zero on at most a fraction $\frac{\ell}{q}$ of points in $\mathbb{F}_q^v$.*

Proof. The statement is proved via induction. The case $v = 1$ states that a univariate polynomial of degree ℓ has at most ℓ roots and is proved using Lemma 2.1. We next note that such a polynomial can be written as

$$P(x_1, \ldots, x_v) = R_{\ell_1}(x_1, \ldots, x_{v-1})x_v^{\ell_1} + \ldots + R_0(x_1, \ldots, x_{v-1}).$$

The probability that $P(\alpha_1, \ldots, \alpha_v) = 0$ is computed using as

$$\begin{aligned}\Pr[P(\alpha_1, \ldots, \alpha_v) = 0] =& \Pr[P(\alpha_1, \ldots, \alpha_v) = 0 \mid R_{\ell_1}(\alpha_1, \ldots, \alpha_{v-1}) = 0] \\ &\times \Pr[R_{\ell_1}(\alpha_1, \ldots, \alpha_{v-1}) = 0] \\ =& \Pr[P(\alpha_1, \ldots, \alpha_v) = 0 \mid R_{\ell_1}(\alpha_1, \ldots, \alpha_{v-1}) \neq 0] \\ &\times \Pr[R_{\ell_1}(\alpha_1, \ldots, \alpha_{v-1}) \neq 0] \\ \leq& 1 \times \frac{\ell - \ell_1}{q} + 1 \times \frac{\ell_1}{q} = \frac{\ell}{q} \end{aligned} \tag{7}$$

where we used the induction step for R_ℓ which has degree $\ell - \ell_1$, and the fact that a univariate polynomial in x_v of degree ℓ_1 has at most ℓ_1 roots. $\square$

The result of Lemma 3.4 can then be used to yield the result that (assuming $\ell \leq q$) the distance of the code $\mathcal{C}$ can be bounded as $d \geq \left(1 - \frac{\ell}{q}\right) q^v$. This suggests that RM codes do not provide $R, \delta > 0$ for constant q, i.e. q increases with n.

4 Binary Reed-Muller Codes

We now shift our attention to the "original" Reed-Muller codes. These were binary codes defined by Muller (1954) and Reed (1954) gave a polynomial time majority logic decoder for these (which we will discuss later). The binary RM code $\mathcal{C}$ results from encoding a *multilinear* encoding polynomial M given by

$$M(x_1, \ldots, x_v) = \sum_{S \,:\, |S| \leq \ell} c_S \prod_{i \in S} x_i,$$

at all 2^v points in $\mathbb{F}_2^v$ (the coefficients c_S are the message bits). The binary RM code $\mathcal{C}$ is a $[2^v, \sum_{i=0}^{\ell} \binom{v}{i}, d]_2$ linear code, where the distance d is given by the following lemma.

Lemma 4.1. *The minimum distance d of the binary RM code described above is $d = 2^{v-\ell}$.*

Proof. Consider the encoding polynomial $M(x_1, \ldots, x_v) = \prod_{i=1}^{\ell} x_i$ resulting from the message leading to the coefficient $c_S = 1$ if and only if $S = \{1, \ldots, \ell\}$. There are exactly $2^{v-\ell}$ choices for $(x_1, \ldots, x_v)$ that make M non-zero, namely those with $x_1 = \ldots = x_\ell = 1$. The distance d is thus bounded as $d \leq 2^{v-\ell}$. Next, consider the non-zero polynomial $M(x_1, \ldots, x_v)$ and let $\prod_{i=1}^{r} x_i$ be the maximal monomial of M, i.e. reorder the indices $\{1, \ldots, v\}$ such that

$$M(x_1, \ldots, x_v) = \prod_{i=1}^{r} x_i + R(x_1, \ldots, x_v)$$

where there is no monomial term in $R(x_1, \ldots, x_v)$ with more than r variables. There are 2^{v-r} ways to choose the variables $x_{r+1}, \ldots, x_v$, but none of them can cause the maximal monomial to be cancelled. This leads to the bound $d \geq 2^{v-r}$, which implies $d \geq 2^{v-\ell}$ since $r \leq \ell$ by the definition of M. $\square$

5 Summary

Two families of linear codes, Reed-Solomon and Reed-Muller, were presented and analyzed using various algebraic properties. Though the Reed-Solomon codes can be used to achieve $R, \delta > 0$, and in fact achieve the optimal trade-off matching the Singleton bound, this can only be done if the alphabet size q increases linearly in the block length, i.e., $q \geq n$. Reed-Muller codes use multivariate polynomials to give codes over smaller alphabets, although they are unable to give codes with $R, \delta > 0$ over a bounded alphabet size.