Information and Coding Theory

Mahdi Cheraghchi

Monday 21 Nov 2015

m.cheraghchi@imperial.ac.uk

Algebraic codes

Reed-Solomon codes

Reed-Solomon (RS) code

- A Reed-Solomon code over GF(q) is determined by a set of ***distinct*** "evaluation points" $\alpha_1, \alpha_2, \ldots, \alpha_n \in GF(q)$.
- It's possible to define polynomials over GF(q) just in the same way as they are defined over real/complex numbers:
  - $f(x) = c_0 + c_1 x + c_2 x^2 + \cdots + c_{k-1} x^{k-1}$.
- Given *f*, "***evaluation vector***" of *f*(x) is defined as
  - $(f(\alpha_1), f(\alpha_2), \ldots, f(\alpha_n))$.
- The codewords of a ***Reed-Solomon code*** of dimension *k* is the set of evaluation vectors of all polynomials of degree < *k*.
- Immediate restriction: we have to have $q \geq n$.

Reed-Solomon (RS) code

- RS code is linear:
  - Adding to polynomials of degree < k gives a polynomial of degree < k.
  - Multiplying a polynomial of degree < k by a scalar in GF(q) gives a polynomial of degree < k.
  - => The code is a vector space and thus linear.

Reed-Solomon (RS) code

- Natural encoder for the code:
  - Enc$(b_0, b_1, \ldots, b_{k-1}) \coloneqq (f(\alpha_1), f(\alpha_2), \ldots, f(\alpha_n))$,
    where $f(x) \coloneqq b_0 + b_1 x + b_2 x^2 + \cdots + b_{k-1} x^{k-1}$.
- Generator matrix for this encoder is "***Vandermonde matrix***":
- $G = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1^{k-1} & \alpha_2^{k-1} & \cdots & \alpha_n^{k-1} \end{pmatrix}$.
- Rank of this matrix is exactly *k* if $\alpha_1, \alpha_2, \ldots, \alpha_n$ are distinct.

Reed-Solomon (RS) code

- What is the minimum distance?
- Recall: Same as "what is the minimum non-zero weight of codewords?"
- Weight of $(f(\alpha_1), f(\alpha_2), \ldots, f(\alpha_n))$ is at least ***n*** **- #(roots of *f* over GF(*q*))**
- Basic algebra: Same as real polynomials, a nonzero polynomial $f(x)$ of degree ***k*-1** over GF(q) has less than ***k*** roots.
- => $d(C) \geq n - (k-1) = n - k + 1.$
- But because of Singleton bound, $d(C) \leq n - k + 1.$
- So $d(C) = n - k + 1$ and the code is MDS ☺
- Familiar interpretation: A polynomial of degree at most *d* can be uniquely interpolated from any set of *d* known evaluations.

Reed-Solomon (RS) code

- Because of being MDS, a Reed-Solomon of rate ***R*** can correct up to **(1-*R*)/2** *fraction* of errors (half the minimum distance).
- Because the code is MDS, its dual is also MDS.
- In fact the dual code is "essentially" a Reed-Solomon code (those are called "***generalized RS codes***".) [exercise]
- A generalized RS code is defined by $\alpha_1, \alpha_2, \dots, \alpha_n \in GF(q)$ and also non-zero scalars $\gamma_1, \gamma_2, \dots, \gamma_n \in GF(q)$ (not necessarily distinct).
- Codewords are of the form:
$$(\gamma_1 f(\alpha_1), \gamma_2 f(\alpha_2), \dots, \gamma_n f(\alpha_n))$$
- This gives the same properties as ordinary RS codes (where we have $\gamma_1 = \gamma_2 = \cdots = \gamma_n = 1$).

Reed-Muller (RM) codes

- Problem with RS codes: The alphabet size has to be large, at least as large as the desired length.
- One solution is going from univariate polynomials to multivariate polynomials. This gives (q-ary) Reed-Muller (RM) codes.
- RM(r, m): ***m-variate*** Reed-Muller code of ***order r***.
- Codewords are evaluation vectors of polynomials of ***m*** variables of degree at most ***r*** at *all points* of $(GF(q))^m$.
- What's the degree of a multivariate polynomial?

Reed-Muller (RM) codes

- Degree of a polynomial is the maximum degree of its monomials (individual terms).
- Degree of $\alpha \cdot x_1^{d_1} \cdot x_2^{d_2} \cdots x_m^{d_m}$ is $d_1 + d_2 + \cdots + d_m$.
- Example:
  - Degree of $f(x, y) = 1 + x + y + xy^2$ is 3, number of variables = 2.
  - Degree of $f(x, y, z) = 1 + 2x + 3y - z$ is 1, number of variables = 3.

Reed-Muller (RM) codes

- The standard encoder of RM(d, m) interprets the message as the coefficient vector of a polynomial in ***m variables*** and ***degree at most r***, and outputs the vector consisting of evaluations of this polynomials at ***all*** points of $(GF(q))^m$.
- Example for RM(2, 3):
- $Enc(b_0, \dots, b_9) = (evaluations\ of\ f(x, y, z)\ everywhere)$ where $f(x, y, z) = b_0 + b_1 x + b_2 y + b_3 z + b_4 x^2 + b_5 y^2 + b_6 z^2 + b_7 xy + b_8 xz + b_9 yz$.
- Dimension: k = 10, length: $n = q^3$.
- The alphabet size for an RM code can be as small as 2 ☺
- In general, length of RM(r, m) is $n = q^m$ (because we evaluate *f* everywhere).
- What is the dimension of RM(r, m)?

Reed-Muller (RM) codes

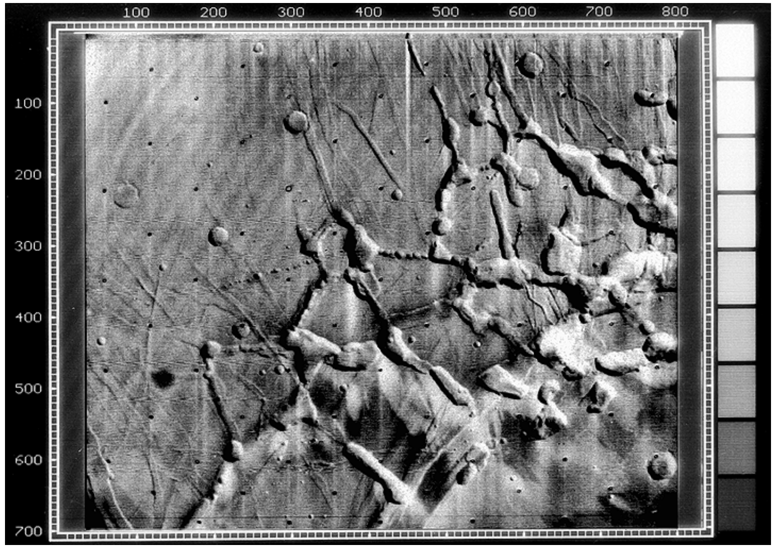
- Dimension is the number of monomials of a polynomial $f(x_1, \dots, x_m)$ of degree at most ***r***.
- A monomial of of the form $x_1^{d_1} \cdot x_2^{d_2} \cdots x_m^{d_m}$ where $d_1 + d_2 + \cdots + d_m \le r$.
- How many combinations of non-negative integers $d_1, d_2, \dots, d_m$ are there such that $d_1 + d_2 + \cdots + d_m \le d$?
  - Equivalently, How many combinations of *positive* integers $d_1, d_2, \dots, d_m$ are there such that $d_1 + d_2 + \cdots + d_m \le m + r$?
  - Answer: $k = \binom{m+r}{m} = \binom{m+r}{r}$.
  - For the example of RM(2, 3): $k = \binom{m+r}{m} = \binom{5}{2} = 10$.
- Note: this is only valid if $q \ge r$. Otherwise over GF(q) we have $x^q = x$ and we may count some monomials several times. For example over GF(q), $x^2 y^2 z^2 = xyz$.

Minimum distance of RM(r, m)

- Schwartz-Zippel Lemma: Any polynomial (possibly multivariate) of degree at most ***r*** is zero on at most ***r/q*** fraction of all possible points (can be proved by induction on the number of variables).
- => Minimum distance of RM(r, m) = min{Hamming weight of codewords} = min{n - #zeros} $\geq q^m - rq^{m-1}$.
- Note: This is nontrivial only when *r* < q.
- => Relative distance of RM(r, m) (=min-distance divided by length) is $\geq (1 - \frac{r}{q})$.
- Important special case: r = 1.

First order Reed-Muller codes

- Special case when q=2 and r=1.
- Also known (with a little difference) as the ***Hadamard code***.
- Has been used by NASA in Mariner 9 for image transmission from Mars.
- Length: $n = 2^m$ (m = # of variables)
- Dimension: $k = \binom{m+r}{r} = m + 1$.
- Distance $\geq n\left(1 - \frac{r}{q}\right) = 2^{m-1}$ (in fact, exactly 2^{m-1}).



First order Reed-Muller codes

- An encoder: $Enc(a_0, \dots, a_m) = (f(0,0,\dots,0), \dots, f(1,1,\dots,1))$ where $$f(x_1, \dots, x_m) := a_0 + a_1 x_1 + \cdots + a_m x_m.$$
- The corresponding generator matrix for m=3:
- $G = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$
- In general, first order RM code is the dual of extended Hamming code (=Hamming code with an extra parity bit added to the encoding; recall Tutorial sheet 16/11 exercise 1 and 09/11 exercise 3.1).
- Amazing error tolerance, but quickly becomes inefficient…

How good can a code be?

- **A(n, d)**: Maximum possible size of a binary code of length ***n*** and minimum distance at least ***d***.
- $\boldsymbol{A_q(n, d)}$: The same, over ***q***-ary alphabet.
- In general, we don't know much about the exact value of **A(n, d)** and $\boldsymbol{A_q(n, d)}$**.**
- Known values maintained at http://www.codetables.de/
- Easy: non-increasing in ***d***, non-decreasing in ***n*** and ***q***.
- To make the question easier, we fix ***q*** (say ***q*** = 2 etc), omit the parameter ***n*** and look at the asymptotic situation:
$$R_q(\delta) := \limsup_{n \to \infty} \frac{\log_q A_q(n, \lfloor \delta n \rfloor)}{n}.$$
- In other words: Maximum possible **rate** for a given "***relative distance***" δ.