Introduction to Reed-Solomon Codes[1]

Yunghsiang S. Han

Graduate Institute of Communication Engineering,
National Taipei University
Taiwan
E-mail: yshan@mail.ntpu.edu.tw

Reed-Solomon Codes Construction (1)

- The first construction of Reed-solomon (RS) codes is simply to evaluate the information polynomials at all the non-zero elements of finite field $GF(q^m)$.
- Let α be a primitive element in $GF(q^m)$ and let $n = q^m - 1$.
- Let $u(x) = u_0 + u_1 x + \cdots + u_{k-1}x^{k-1}$ be the information polynomial, where $u_i \in GF(q^m)$ for all $0 \leq i \leq k-1$.
- The encoding is defined by the mapping $\rho : u(x) \longrightarrow \boldsymbol{v}$ by

$$(v_0, v_1, \ldots, v_{n-1}) = (u(1), u(\alpha), u(\alpha^2), \ldots, u(\alpha^{n-1})).$$

- The RS code of length n and dimensional k over $GF(q^m)$ is the image under all polynomials in $GF(q^m)[x]$ of degree less than or equal to $k-1$.

- The minimum distance of an (n, k) RS code is $d_{min} = n - k + 1$. It can be proved by follows.
- Since $u(x)$ has at most $k-1$ roots, there are at most $k-1$ zero positions in each nonzero codeword. Hence, $d_{min} \geq n - k + 1$. By the Singleton bound, $d_{min} \leq n - k + 1$. So $d_{min} = n - k + 1$.

Reed-Solomon Codes Construction (2)

- The RS codes can be constructed by finding their generator polynomials.
- In $GF(q^m)$, the minimum polynomial for any element α^i is simply $(x - \alpha^i)$.
- Let $g(x) = (x - \alpha^b)(x - \alpha^{b+1}) \cdots (x - \alpha^{b+2t-1})$ be the generator polynomial for the RS code. Since the degree of $g(x)$ is exactly equal to $2t$, by the BCH bound, $n = q^m - 1$, $n - k = 2t$, and $d_{min} \geq n - k + 1$.
- Again, by the Singleton bound, $d_{min} = n - k + 1$.
- Considering $GF(8)$ with the primitive polynomial

$x^3 + x + 1$. Let α be a root of $x^3 + x + 1$. Then

$$g(x) = (x-\alpha)(x-\alpha^2)(x-\alpha^3)(x-\alpha^4) = x^4+\alpha^3 x^3+x^2+\alpha x+\alpha^3$$

will generate a $(7, 3)$ RS code with $d_{min} = 2 \times 2 + 1 = 5$. The number of codewords of this code is $8^3 = 512$.

Encoding Reed-Solomon Codes

- RS codes can be encoded just as any other cyclic code.
- The systematic encoding process is

$$v(x) = u(x)x^{n-k} - \left[u(x)x^{n-k} \bmod g(x)\right].$$

- Typically, the code is over $GF(2^m)$ for some m. The information symbols u_i can be formed by grabbing m bits of data, then interpreting these as the vector representation of the $GF(2^m)$ elements.

Weight Distributions for RS Codes

- A code is called *maximum distance separable* (MDS) code when its d_{min} is equal to $n - k + 1$. A family of well-known MDS nonbinary codes is Reed-Solomon codes.
- The dual code of any (n, k) MDS code $\boldsymbol{C}$ is also an $(n, n-k)$ MDS code with $d_{min} = k + 1$.
- It can be proved as follows: We need to prove that the $(n, n-k)$ dual code $\boldsymbol{C}^{\perp}$, which is generated by the parity-check matrix $\boldsymbol{H}$ of $\boldsymbol{C}$, has $d_{min} = k + 1$. Let $\boldsymbol{c} \in \boldsymbol{C}^{\perp}$ have weight w, $0 < w \leq k$. Since $w \leq k$, there are at least $n - k$ coordinates of $\boldsymbol{c}$ are zero. Let $\boldsymbol{H}_s$ be an $(n-k) \times (n-k)$ submatrix formed by any collection of $n - k$ columns of $\boldsymbol{H}$ in the above coordinates. Since the

row rank of $\boldsymbol{H}_s$ is less than $n-k$ and consequently the column rank is also less than $n-k$. Therefore, we have found $n-k$ columns of $\boldsymbol{H}$ are linear dependent which contradicts to the facts that d_{min} of $\boldsymbol{C}$ is $n-k+1$ and then any combination of $n-k$ columns of $\boldsymbol{H}$ is linear independent.

- Any combination of k symbols of codewords in an MDS code may be used as information symbols in a systematic representation.
- It can be proved as follows: Let $\boldsymbol{G}$ be the $n \times k$ generator matrix of an MDS code $\boldsymbol{C}$. Then $\boldsymbol{G}$ is the parity check matrix for $\boldsymbol{C}^{\perp}$. Since $\boldsymbol{C}^{\perp}$ has minimum distance $k+1$, any combination of k columns of $\boldsymbol{G}$ must be linearly independent . Thus any $k \times k$ submatrix of $\boldsymbol{G}$ must be

nonsingular. So, by row reduction on $\boldsymbol{G}$, any $k \times k$ submatrix can be reduced to the $k \times k$ identity matrix.

- The number of codewords in a q-ary (n, k) MDS code $\boldsymbol{C}$ of weight $d_{min} = n - k + 1$ is

$$\boxed{A_{n-k+1} = (q-1)\binom{n}{n-k+1}.}$$

- It can be proved as follows: Select an arbitrary set of k coordinates as information positions for an information $\boldsymbol{u}$ of weight 1. The systemic encoding for these coordinates thus has $k-1$ zeros in it. Since the minimum distance of the code is $n-k+1$, all the $n-k$ parity check symbols must be nonzero. Since there are $\binom{n}{k-1} = \binom{n}{n-k+1}$

different ways of selecting the $k-1$ zero coordinates and $q-1$ ways of selecting the nonzero information symbols,

$$A_{n-k+1} = (q-1)\binom{n}{n-k+1}.$$

- The number of codewords of weight j in a q-qry (n, k) MDS code is

$$\boxed{A_j = \binom{n}{j}(q-1)\sum_{i=0}^{j-d_{min}}(-1)^i\binom{j-1}{i}q^{j-d_{min}-i}.}$$

References

[1] T.K. Moon, *Error Correction Coding: Mathematical Methods and Algorithms*, Hoboken, NJ: John Wiley & Sons, Inc., 2005.