CSCI-B609: A Theorist's Toolkit, Fall 2016 Oct 11

Reed-Solomon Codes and Existence of Good Codes

Lecturer: Yuan Zhou *Scribe: Ruiyu Zhu*

1 Reed-Solomon Codes

Reed-Solomon Codes are a scheme of linear codes.

Definition 1.

For alphabeta $\mathbb{F}_q$*, and* $1 \leq k < n \leq q$*, pick* $S = \{a_1, a_2, \cdots, a_n\} \subseteq \mathbb{F}_q$.

For message $(m_0, m_1, \cdots, m_{k-1}) \in \mathbb{F}_q^k$*, define* $P_m(x) = \sum_{i=0}^{l-1} m_i x^i$.

Define:

$$Enc : \mathbb{F}_q^k \to \mathbb{F}_q^n$$

$$Enc(m) = (P_m(a_1), P_m(a_2), \cdots, P_m(a_n))$$

1.1 Generator Matrix of Reed-Solomon Codes

$$G = \begin{bmatrix} a_1^0 & a_1^1 & a_1^2 & \dots & a_1^{k-1} \\ a_2^0 & a_2^1 & a_2^2 & \dots & a_2^{k-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_n^0 & a_n^1 & a_n^2 & \dots & a_n^{k-1} \end{bmatrix}$$

$$Enc(m) = G \cdot m^T$$

G is a Vandermonde matrix

1.2 Minimum distance

Proposition 2. *Reed-Solomon code is a* $[n, k, n-k+1]_q$ *code.*

Proof. Suppose for contradiction that $\exists m \neq \vec{0}$, s.t.

$$\begin{aligned} & Wt(Enc(m)) < n-k+1 \\ \Leftrightarrow\ & Wt(Enc(m)) \leq n-k \\ \Leftrightarrow\ & \exists \text{at most } n-k \text{ different } a_i, s.t. P_m(a_i) \neq 0 \\ \Leftrightarrow\ & \exists \text{at least } k \text{ different } a_i, s.t. P_m(a_i) = 0 \\ \Rightarrow\ & P_m \text{ has at least } k \text{ roots} \end{aligned}$$

which contradicts the fact that $P_m \neq 0$ and degree$(P_m) \leq k-1$. □

2 Decoding Reed-Solomon Codes

Berlekamp - Welch algorithm

Goal: efficient algorithm to correct up to $e = \frac{n-k+1}{2}$ errors in RS-code.

Let $y = (y_1, y_2, \cdots y_n)$ denote the "codeword" to decode, $P_m(x)$ to denote the polynomial corresponding to the original message.

Define

$$E_m(x) = x^t \prod_{y_i \neq P_m(a_i)} (x - a_i)$$

where t is a non-negative integer such that $E_m(x)$ is a e-degree polynomial.

Thus we have $\forall i, y_i E_m(a_i) = P_m(a_i) E_m(a_i)$.

Note that, E_m is a e-degree monic polynomial, P_m is a $k-1$-degree polynomial.

Let's assume $E(x)$ is a e-degree monic polynomial, $P(x)$ is a $k-1$-degree polynomial, such that $\forall i, y_i E(a_i) = P(a_i) E(a_i)$.

We will have n different equations and $e+k-1+1 < \frac{n-k+1}{2} + k < n$ unknowns. This system is solvable, one solution is $E(x) = E_m(x)$ and $P(x) = P_m(x)$.

However this is not a linear system, thus it is not efficiently solvable and there might exists more than one solution.

In order to fix this, we introduce $N(x) = E(x) \cdot P(x)$, thus $N(x)$ is a $e+k-1$-degree polynomial. And we have $\forall i, y_i E_m(a_i) = N(a_i)$.

There are n different equations and $e+e+k-1+1 < n-k+1+k = n+1$ unknowns. Note that this is a linear system with n constraints and up to n unknowns, which is solvable within $O(n^3)$ time.

Now we are going to show that solving this system really decodes the codeword.

Claim:

$$P_m(x) = \frac{N(x)}{E(x)}$$

Proof. Consider $R(x) = E(x)E_m(x)P_m(x) - E_m(x)N(x)$.

We can verify that

$$R(a_i) = \begin{cases} 0, & E_m(a_i) = 0 \text{ if } P_m(a_i) \neq y_i, \\ (y_i E(a_i) - N(a_i))E_m(a_i) = 0 & \text{if } P_m(a_i) = a_i \end{cases}$$

Thus $R(x)$ has at least n roots.

However, degree$(R) \leq (2e + k - 1, 2e + k - 1) = 2e + k - 1 < n$. Thus it is clear that R is a constant zero, which suggests that $E(x)E_m(x)P_m(x) = E_m(x)N(x)$, a.k.a. $P_m(x) = \frac{N(x)}{E(x)}$ □

3 Bounds

3.1 Relationship between n, k, d

Theorem 3. *The Singleton Bound*

For any block code, $k \leq n - d + 1$

Proof. Suppose not, thus we have $|C| > |\Sigma|^{n-d+1}$

By pigeonhole principle, $\exists$ 2 different codeword are the same in the first $n - d + 1$ symbols, a.k.a. $\exists c \neq c' \in C$ s.t. $c_i = c'_i, \forall i \leq n-d+1$. Thus their distance is at most $n-(n-d+1) = d - 1$. □

3.2 Asymptotically good

Let $C = \{C_1, C_2, \cdots C_n, \cdots\}$ be a family of code with C_n being a $[n, k(n), d(n)]_q$ code for a given q.

Define code rate $R(C) = \underline{\lim}_{n\to\infty} \frac{k(n)}{n}$.

Define relative distance $\delta(C) = \underline{\lim}_{n\to\infty} \frac{d(n)}{n}$

C is said to be Asymptotically good if $R(C) > 0, \delta(C) > 0$.

Note that $R(\text{Hamming Code}) = 1$, $\delta(\text{Hamming Code}) = 0$, $R(\text{Hadamard Code}) = 0$,

$\delta(\text{Hadamard Code}) = \frac{1}{2}$

Claim:

Good code exists.

Theorem 4. *Gilbert-Varshamov bound*

Let $q \geq 2$, $\forall 0 \leq \delta \leq 1 - \frac{1}{q}$ *and* $0 < \epsilon \leq 1 - H_q(\delta)$,$\exists C$ *s.t.* $\delta(C) = \delta > 0, R(C) \geq 1 - H_q(\delta) - \epsilon \geq 0$

where $H_q(x) = -x \log_q(x) - (1-x) \log_q(1-x)$. *The equation holds only if* $\epsilon = H_q(\delta)$

Proof. We focus on the case that the field is $\mathbb{F}_2$ (Formal proof could be found in the link to wiki).

Start with $C_n = \emptyset$, add codeword greedily.

While $\exists Z \in \mathbb{F}_2^n$ s.t. $\forall Z' \in C_n, \Delta(Z, Z') \geq \delta_n$, add Z to C_n.

When the process stops, we must have

$$\cup_{Z \in C_n} B_{n\delta}(Z) = \mathbb{F}_2^n$$

Thus we have $|C_n| \cdot |B_{n\delta}(Z)| \geq 2^n$.

However $|B_{n\delta}(Z)| = \sum_{i=0}^{n\delta} \binom{n}{i} \leq 2^{nH_2(\delta)}$[1], we must have $|C_N| = \frac{2^n}{|B_\delta(Z)|} \geq 2^{n(1-H_2(\delta))}$

□

Theorem 5. *For* $q = 2$, $\forall 0 \leq k \leq 1$, $\exists$ *polynomial time constructible encodable decodable family of codes* C *s.t.* $\delta(C) \approx \frac{1-k}{2}$ *where decode up to* $\lfloor \frac{\delta(C)n}{2} \rfloor$ *errors.*

[1] the upper bound of the Volume of Hamming ball