6.897 Algorithmic Introduction to Coding Theory September 24, 2001

Lecture 5

Lecturer: Madhu Sudan *Scribe: Mohammad Mahdian*

Today we will talk about:

- Asides on Reed-Solomon codes.
- Asymptotics of codes.
- Random codes.

1 Reed-Solomon codes

There are two equivalent ways to look at Reed-Solomon codes:

- Evaluation of polynomials: This is the definition that we have seen in the previous lecture. It's usually more convenient to prove theorems using this definition.
- Coefficients of polynomials: Some special cases of Reed-Solomon codes $\mathrm{RS}_{q,n,k}$ (q is the size of the alphabet (field), n is the block length, and k is the message length) can be described as follows:
  - **Generator polynomial:** A polynomial $g(x) \in \mathbb{F}_q[x]$ of degree $n-k$.
  - **Message:** coefficients of a polynomial $M(x)$ of degree less than k.
  - **Encoding:** coefficients of $g(x)M(x)$.

This definition gives us various codes depending on which g we pick. If g is a polynomial that divides $x^n - 1$, then we get a general family of codes called "cyclic codes". For some further restrictions, we get BCH codes, which in turn contain the special cases of RS codes. To define the generator polynomial for these RS codes, we first need to introduce the notion of a primitive element of $\mathbb{F}_q$.

Definition 1 *An element $\alpha \in \mathbb{F}_q$ is a* primitive element *if $\alpha, \alpha^2, \ldots, \alpha^{q-1}$ are all the nonzero elements of the field.*

It is well-known that every field has many primitive elements. Given a primitive element α, we can define the generator polynomial g for $RS_{q,n,k}$ as follows:

$$g(x) := \prod_{i=1}^{n-k} (x - \alpha^i).$$

Using this polynomial with $n = q - i$ and $\alpha_i = \alpha^i$, we get an RS code $\mathrm{RS}_{q,n,k}$. (Recall that $\alpha_1, \ldots, \alpha_n$ are the set of points on which the polynomial in the first definition is evaluated.) A proof of this fact can be found at the end of this lecture.

2 Alternant Codes

Given n distinct elements $\alpha_1, \ldots, \alpha_n$ and n nonzero elements $\beta_1, \ldots, \beta_n$ of $\mathbb{F}_q$, the *Alternant Code* is defined as follows:

- **Message:** Polynomial $M(x)$ of degree less than k.
- **Encoding:** $\langle \beta_1 M(\alpha_1), \dots, \beta_n M(\alpha_n) \rangle$.

In terms of the minimum distance, it is clear that alternant codes are equivalent to the RS codes. In particular a given coordinate of an encoding a given message in the alternant code is non-zero if and only if the same coordinate of the RS encoding of the same message is non-zero. However, the alternant code might have different properties in terms of its sub-codes. A sub-code of a code is defined as follows:

Let $q = 2^k$. We know that $\mathbb{F}_2$ is a subfield of $\mathbb{F}_q$. Consider an $[n, k, d]_q$ code C_1. The $\mathbb{F}_2$ sub-code of C_1, denoted C_2, is an $[n, k', d']_2$ code that consists of all codewords of C_1 that are also in $\mathbb{F}_2^n$ ($C_2 = C_1 \cap \mathbb{F}_2^n$). Such an operation can be carried out in general for any pair of field $\mathbb{F}^{(1)} \subseteq \mathbb{F}^{(2)}$, and the resulting codes are called sub-field sub-codes of the original code.

The resulting codes have minimum distance at least as much as that of the original code. However their message length may be much smaller. In fact, a priori it is unclear as to why the sub-field sub-code should contain any non-zero vector. However their performance is not as bad as it looks! Many interesting families of codes can be obtained as sub-field sub-codes of Alternant codes. BCH codes form one such example, for some clever choice of $\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n$.

Readers more interested in this material can find it in [2, Chapter 12].

3 Asymptotics of Codes

So far we have seen a variety of codes:

Hamming code These codes have a good relationship between k and n, but d is small.

RS codes They meet the Singleton bound, but need large alphabet size.

Hadamard code $d = n/2$, but $k = \log n$, i.e., encoding increases the length of the message exponentially.

We sense that none of these codes is "good enough", but we have not defined a concept of "good" codes. To do so, we need to study the *asymptotics* of codes. To do so, we will consider infinite families of codes $\mathcal{C} = \{(n_i, k_i, d_i)_{q_i}\}_{i=1}^{\infty}$, with $\lim_{i\to\infty}\{n_i\} = \infty$.

Definition 2 *The* (message) rate *of a family of codes $\mathcal{C} = \{(n_i, k_i, d_i)_{q_i}\}_{i=1}^{\infty}$, denoted $R(\mathcal{C})$, is defined to be* $\liminf_{i\to\infty} \left\{\frac{k_i}{n_i}\right\}$. *The* relative distance *of $\mathcal{C}$, denoted $\delta(\mathcal{C})$, is defined to be* $\liminf_{i\to\infty} \left\{\frac{d_i}{n_i}\right\}$.

Definition 3 *A family of codes $\mathcal{C}$ is* asymptotically good *if $R(\mathcal{C}), \delta(\mathcal{C}) > 0$.*

When the family $\mathcal{C}$ is clear from context, we will skip the argument and just refer to R and δ.

One of the early "holy grails" of coding theory was to construct asymptotically good codes. This was achieved early on. We will see in this lecture that such codes do exist, and in the next lecture we will show how to construct a family of asymptotically good codes.

Every result in coding theory tends to have an asymptotic interpretation, and often the asymptotic version is much more succinct. For example, the Singleton bound ($n - k + 1 \geq d$) implies

$$\delta \leq 1 - R.$$

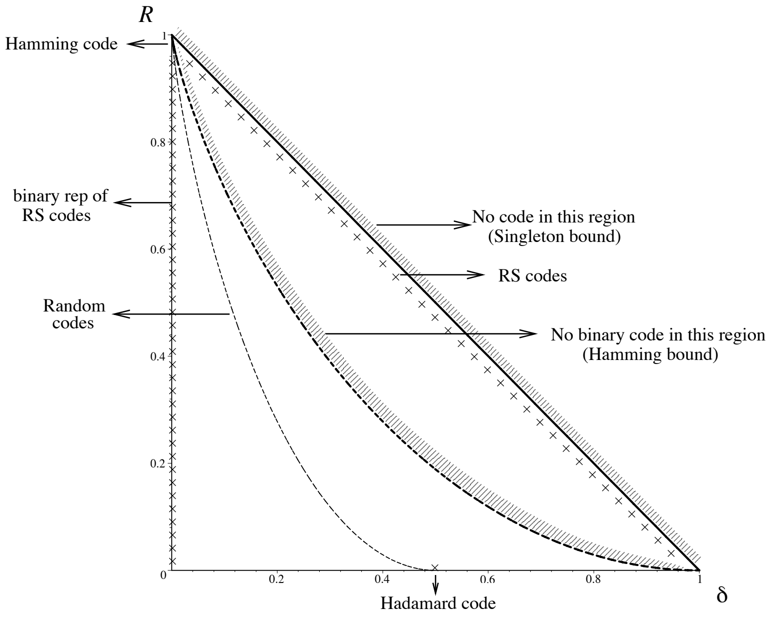

Figure 1:

Similarly, the Hamming bound has an asymptotic interpretation. Recall that this bound says that for binary codes, $2^k \mathrm{Vol}(\frac{d-1}{2}, n) \leq 2^n$. Using the approximations $\frac{d-1}{2} \approx \frac{\delta n}{2}$ and $\mathrm{Vol}(pn, n) \approx 2^{H(p)n}$, we have

$$\text{For } q = 2, \qquad R + H\left(\frac{\delta}{2}\right) \leq 1.$$

The above bounds are shown in Figure 1. In the rest of this lecture, we will show that random binary codes satisfy $R \geq 1 - H(\delta)$. We don't know of an explicit construction for a code satisfying this bound.

4 The Gilbert-Varshamov bound

The Gilbert-Varshamov bound says that there is an infinite family of codes $\mathcal{C}$ satisfying $R(\mathcal{C}) \geq 1 - H(\delta(\mathcal{C}))$. We will present three proofs for this fact. These proofs are due to:

- Gilbert [1], who showed essentially that a random code has this property
- Varshamov [4] who showed that random linear codes have this property.
- Wozencraft [5] who constructed a small space of linear codes most of whose members meet the Gilbert-Varshamov bound.

4.1 Gilbert's code (Greedy code)

Gilbert showed the family of codes $\mathcal{C}$ with its nth element picked greedily according to the following procedure satisfies the bound $R(\mathcal{C}) \geq 1 - H(\delta(\mathcal{C}))$. Later we will view the result as showing that a randomized procedure leads to good codes with high probability.

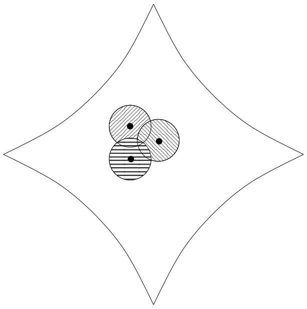

Figure 2: Hypercube and the greedy algorithm

- GREEDY(n, d)
- Intitialize: $S \leftarrow \{0,1\}^n, C \leftarrow \emptyset$.
- Iterate until $S = \emptyset$:
  1. Pick $\mathbf{x} \in S$ and add $\mathbf{x}$ to C.
  2. Delete $B(\mathbf{x}, d)$ (ball of radius d around $\mathbf{x}$) from S. (See Figure 2.)

Lemma 4 *Fix $0 < \delta < \frac{1}{2}$ and $\epsilon > 0$ and let $R \geq 1 - H(\delta) - \epsilon$. Then for all sufficiently large n, the procedure* GREEDY$(n, \lceil \delta n \rceil)$ *produces a code with at least 2^{Rn} codewords.*

Proof Let n be large enough so that $\mathrm{Vol}(d, n) \leq 2^{(H(\delta)+\epsilon)n}$.

Assume the algorithm picks K codewords. At every step, the greedy algorithm deletes at most $\mathrm{Vol}(d, n)$ elements from S. Therefore, since S started with 2^n elements, we have

$$K \geq \frac{2^n}{\mathrm{Vol}(d, n)} \geq 2^{(1-H(\delta)-\epsilon)n} = 2^{Rn}.$$

■

The following theorem follows:

Theorem 5 *There exists a family of codes $\mathcal{C}$ with $R(\mathcal{C}) \geq 1 - H(\delta(\mathcal{C}))$.*

An alternate way to get a similar result is to follow the following probabilistic procedure.

1. Pick a subset $C' \subset \{0,1\}^n$ of size 2^k at random.
2. Let $B \subseteq C'$ be the set $\{\mathbf{x} \in C' | \exists \mathbf{y} \in C' - \{\mathbf{x}\} \text{ s.t. } \Delta(\mathbf{x}, \mathbf{y}) \leq d\}$.
3. Let $C = C' - B$. Output C.

We argue (informally) that for an appropriate setting of parameters, C' still has $\Theta(2^k)$ codewords and that its distance is at least d. To do so, note that the expected number of neighbours from $C' - \mathbf{x}$ that a vector $\mathbf{x}$ has of distance at most d is approximately $2^{k-n}\mathrm{Vol}(d, n)$. For appropriate setting of parameters (still maintaining $R \approx 1 - H(\delta)$) this expected number of neighbours can be made arbitrarily small, say γ. Thus the probability that $\mathbf{x}$ belongs to B is at most γ. By Markov's inequaltiy, we have that the event that half of the elements of C' are in B occurs with probability at most 2γ. Deleting this set still leaves us with a C of size 2^{k-1} and this happens with probability at least $1 - 2\gamma$.

4.2 Varshamov's linear code

Varshamov's linear codes are constructed using the following probabilistic procedure:

> Random-Linear(n, k)
>
> Pick the entries of the $k \times n$ generator matrix $\mathbf{G}$ uniformly and independently at random from $\mathbb{F}_2^{k\times n}$. Let $C = \{\mathbf{yG} : \mathbf{y} \in \mathbb{F}_2^k\}$.

Lemma 6 *Let $0 < \delta < \frac{1}{2}$ and $\epsilon > 0$. Let $R = 1 - H(\delta) - \epsilon$. For sufficiently large n and $k = \lceil Rn \rceil$, the procedure* Random-Linear(n, k) *produces a code with 2^k codewords and distance at least δn, with high probability.*

Proof We really need to prove two separate assertions here: First, that the matrix $\mathbf{G}$ has full column rank k so that the code does have 2^k codewords. Next, we need to show that no pair of distinct codewords in the code generated by $\mathbf{G}$ are within distance δn of each other. We will combine the two steps into one and simply argue that for every non-zero vector $\mathbf{y}$, it is the case that $\mathbf{yG}$ does not lie in $B(\mathbf{0}, \delta n)$. The first part (rank of $\mathbf{G}$) is implied by the fact that $\mathbf{yG} \neq \mathbf{0}$ for any non-zero $\mathbf{y}$. The second part follows, since we would have proved that no codeword has Hamming weight less than δn and we know that the the minimum distance of a linear code equals the minimum weight of a non-zero codeword.

Suppose n is large enough so that $\mathrm{Vol}(\delta n, n) \leq 2^{(H(\delta)+\epsilon/2)n}$. Let $d = \delta n$. For every fixed $\mathbf{y} \neq \mathbf{0}$ in $\mathbb{F}_2^k$, it is easy to see that $\mathbf{yG}$ is a random vector in $\{0,1\}^n$, and therefore,

$$\begin{aligned} \Pr[\mathrm{wt}(\mathbf{yG}) \leq d] &= \Pr[\mathbf{yG} \in B(\mathbf{0}, d)] \\ &= \frac{\mathrm{Vol}(d, n)}{2^n} \\ &\leq 2^{(H(\delta)+\epsilon/2-1)n}. \end{aligned}$$

Therefore, by the union bound, the probability that there is a $\mathbf{y}$ with $\mathrm{wt}(\mathbf{yG}) \leq d$ is at most $2^k 2^{(H(\delta)+\epsilon/2-1)n}$. If $R = \frac{k}{n} \leq 1 - H(\delta) - \epsilon$, then this probability is at most $2^{-(\epsilon/2)n}$ which goes to zero as $n \to \infty$. Therefore with high probability, the random procedure outputs a linear code with minimum distance at least δn. ■

4.3 Wozencraft construction

Varmashov's construction gives an algorithm of running time 2^{kn} for constructing the code. Wozencraft uses a clever idea to reduce this running time to 2^n. The idea is to find a family of disjoint sets $S_1, S_2, \ldots, S_t \subseteq \{0,1\}^n - \bar{\mathbf{0}}$ such that for every i, $S_i \cup \{\bar{\mathbf{0}}\}$ is a linear subspace of $\{0,1\}^n$.

Claim 7 *If such a family exists and $t \geq \mathrm{Vol}(d, n)$, then there is an i such that $S_i \cup \{\bar{\mathbf{0}}\}$ is a linear code with distance at least d.*

Proof Every vector $\mathbf{x} \in B(\mathbf{0}, d) - \{\mathbf{0}\}$ lies in at most one of the S_i's (since the S_i's are disjoint). Since $t > \mathrm{Vol}(d, n) - 1$, it follows that at least one of the S_i does not contain any of the elements of $B(\mathbf{0}, d) - \{\mathbf{0}\}$. Such an S_i has minimum weight at least d, and since $S_i \cup \{\bar{\mathbf{0}}\}$ is linear, it has distance at least d. ■

Furthermore, if we can construct this partition with the additional property that all $|S_i|$'s are equal, we will get a linear code of size $2^n/t$. Such a construction will be presented in the next lecture.

5 Appendix: Equivalence of Reed Solomon code

Here we show that the two definitions of Reed-Solomon codes coincide for appropriate choice of parameters. To be explicit let us reintroduce the two definitions (with more parameters).

Definition 8 *For prime power q, integers $k \leq n \leq q$ and a vector $\alpha = \langle \alpha_1, \ldots, \alpha_n \rangle \in \mathbb{F}_q^n$ of distinct elements of $\mathbb{F}_q$, the Reed-Solomon code $\mathrm{RS}_{q,n,k,\alpha}$ is the collection of vectors $\{\langle M(\alpha_1), \ldots, M(\alpha_n)\rangle | M \in \mathbb{F}_q[x], \deg(M) < k\}$.*

The second definition below is via the coefficients of polynomials.

Definition 9 *For a prime power q, a primitive element $\alpha \in \mathbb{F}_q$ and integer k, the alternate Reed-Solomon code $\mathrm{RS}'_{q,k,\alpha}$ is the collection of vectors $\{\langle c_0, \ldots, c_{q-2}\rangle | g(x)$ divides $C(x) = \sum_{i=0}^{q-2} c_i x^i\}$, where $g(x) = \prod_{j=1}^{q-1-k}(x - \alpha^j)$.*

The following proposition gives the equivalence:

Proposition 10 *For every prime power q, primitive element α and $k \leq q - 1$ and for $n = q - 1$ and $\alpha = \langle \alpha^0, \ldots, \alpha^{n-1}\rangle$, $\mathrm{RS}_{q,n,k,\alpha} = \mathrm{RS}'_{q,k,\alpha}$.*

Proof Since both codes have the same number of codewords, it suffices to prove that every codeword according to the first definition is a codeword according to the second definition.

Consider a vector $\langle c_0, \ldots, c_{n-1}\rangle$ and the associated polynomial $C(x) = \sum_{i=0}^{n-1} c_i x^i$. To prove this vector is a codeword according to the second definition, it suffices to prove that $C(\alpha^j) = 0$ for every $j \in [n-k]$ (since this implies that $\prod_j (x - \alpha^j)$ divides $C(x)$).

Consider a message $\mathbf{m} = \langle m_0, \ldots, m_{k-1}\rangle$ and the associated polynomial $M(x) = \sum_{l=0}^{k-1} m_l x^l$. Let $\mathbf{c} = \langle M(\alpha^0), \ldots, M(\alpha^{n-1})\rangle$ be the encoding of $\mathbf{m}$ according to the first definition. Let $C_{\mathbf{m}}(x)$ be the polynomial with $\mathbf{c}$ as its coefficients, i.e., the coefficient of x^i is $M(\alpha^i)$. We show below that $C_{\mathbf{m}}(\alpha^j) = 0$ for every $j \in [n-k]$.

$$\begin{aligned}
C_{\mathbf{m}}(\alpha^j) &= \sum_{i=0}^{n-1} M(\alpha^i)(\alpha^j)^i \\
&= \sum_{i=0}^{n-1}\sum_{l=0}^{k-1} m_l (\alpha^i)^l (\alpha^j)^i \\
&= \sum_{l=0}^{k-1} m_l \sum_{i=0}^{n-1} (\alpha^l \alpha^j)^i \\
&= \sum_{l=0}^{k-1} m_l \sum_{i=0}^{q-2} \gamma_{j,l}^i
\end{aligned}$$

where $\gamma_{j,l} = \alpha^{j+l}$. Notice that for every j, l s.t. $j + l \neq q - 1$, $\gamma_{j,l} \neq 1$. Notice further that for every such $\gamma_{j,l}$ the summation $\sum_{i=0}^{q-2} \gamma_{j,l}^i = 0$[1]. Since $l \in \{0, \ldots, k-1\}$, we find that $\gamma_{j,l} \neq 1$ for every $j \in [n-k]$. Thus for every $j \in [n-k]$, we find that $C_{\mathbf{m}}(\alpha^j) = 0$. This concludes the proof. ■

[1] This identity is obtained as follows: Recall that Fermat's little theorem asserts that $\gamma^{q-1} - 1 = 0$ for every non-zero γ in $\mathbb{F}_q$. Factoring the left hand side, we find that either $\gamma - 1 = 0$ or $\sum_{i=0}^{q-2} \gamma^i = 0$. Since $\gamma \neq 1$, the latter must be the case.

References

[1] E. N. Gilbert. A comparison of signalling alphabets. *Bell System Technical Journal*, 31:504–522, May 1952.

[2] F. J. MacWilliams and Neil J. A. Sloane. *The Theory of Error-Correcting Codes.* Elsevier/North-Holland, Amsterdam, 1981.

[3] James L. Massey. *Threshold decoding.* MIT Press, Cambridge, Massachusetts, USA, 1963.

[4] R. R. Varshamov. Estimate of the number of signals in error correcting codes. *Doklady Akadamii Nauk*, 117:739–741, 1957.

[5] J. M. Wozencraft. Threshold decoding. Personal communication in [3, Section 2.5], 1963.