CSE522/Math581EB Coding Theory **Fall 2001**

Lecture 3: October 9

Lecturer: Elchanan Mossel *Scribe: Elchanan Mossel*

3.1 Error correcting codes

Shannon theorem is highly non-constructive, as both the encoding and decoding functions are exponentially large look-up tables with no structure. We want to find more compact description of the encoding and decoding functions and to ensure that encoding and decoding is performed efficiently.

We will consider codes over an alphabet Σ. A code C is a subset of Σ^n. The minimum distance of C, is defined by $d = \Delta(C) = \min_{x,y \in C} \Delta(x, y)$, where Δ is Hamming-distance. We are interested in the following parameters:

- *block length*: n, where $C \subset \Sigma^n$.
- *message length*: $k = \log_q |C|$, the amount of information which is encoded in each codeword.
- *minimum distance*: $d = \Delta(C)$.
- *alphabet size*: $q = |\Sigma|$.

We will refer to a code with such parameters as a $(n, k, d)_q$ code. Usually, if all the other parameters are fixed, we want to

- minimize n - clear.
- maximize k - clear.
- maximize d - clear
- minimize q - It turns out that it is easier to construct codes for large q than for small q. Sometimes the codes for large q can be latter used for small values of q.

The distance d is related to the error correcting properties of the code. Suppose that code C has distance d, then it is possible to detect an error provided that the total number of errors is less than d. It is possible to correct errors provided the total number of errors is at most $\lfloor \frac{d-1}{2} \rfloor$. If a family of codes satisfies $\lim d(i)/n(i) > p$, then it can be used in order to transmit information over the binary symmetric channel with parameter $p/2$, or over the binary erasure channel with parameter p, in such a way that the error probability is $o(1)$.

3.2 Linear codes

$C \subset \mathcal{F}_q^n$ is a *linear code*, if C is a linear subspace of $\mathcal{F}_q^n$ (i.e. if $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in C$, and $a \in \mathcal{F}_q$, then $ax = (ax_1, \ldots, ax_n) \in C$ and $x + y = (x_1 + y_1, \ldots, x_n + y_n) \in C$). When $C \subset F_q^n$ s an $(n, k, d)_q$ linear code we will denote it by using square brackets, and say that C is an $[n, k, d]_q$ code.

If $C \subset \mathcal{F}_q^n$ is an $[n,k,d]_q$ code, then we may choose a basis $x_1,\ldots,x_k \in \mathcal{F}_q^n$ s.t.

$$C = \{\sum_{i=1}^{k} a_i x_i : a_i \in \mathcal{F}_q\} = \{aG : a \in \mathcal{F}_q^k\}$$

where the rows of G are $x_1,\ldots,x_k$, so that G is $k \times n$ matrix over $\mathcal{F}$. Matrix G is called a *generator* matrix of the code C.

The scalar product of $x \in \mathcal{F}^n$ and $y \in \mathcal{F}^n$ is defined as $< x,y >= \sum_{i=1}^{n} x_i y_i$. For a code C, let the *dual code* of C, be the dual space of C:

$$C^{\perp} = \{y : \mathcal{F}_q^n \text{ s.t. } \forall x \in C, < x,y >= 0\}.$$

Let $H^T \in M_{n-k,n}(\mathcal{F})$ be the generator matrix of $C^{\perp}$. Matrix H is the *parity-check* matrix of the code C. Note that C may be expressed as $C = \{y \in \mathcal{F}_q^n : yH = 0\}$.

Letting the *weight* of x, be $\mathrm{wt}(x) = \Delta(x,0)$, note that $\Delta(x,y) = \Delta(x-y,0) = wt(x-y)$ and therefore, when C is linear,

$$d = \min_{x \in C, y \in C, x \neq y} \Delta(x,y) = \min_{x \in C, y \in C, x \neq y} \Delta(x-y,0) = \min_{0 \neq x \in C} \mathrm{wt}(x) = \min\{\mathrm{wt}(x) : x \neq 0, xH = 0\}.$$

3.3 Simple linear codes

- $d = 1$: Taking the identity encoding and decoding functions we obtain an $[n,n,1]_q$ code.
- $d = 2$: Looking at the code

  $$C = \{(x_1,\ldots,x_{n-1},\sum_{i=1}^{n} x_i) : (x_1,\ldots,x_{n-1}) \in \mathcal{F}_q^{n-1}\},$$

  we obtain an $[n,n-1,2]_q$ code. Indeed if $x \in C$ and $x \neq 0$, then it cannot be that $x_i = 0$ for $1 \leq i \leq n-1$, as this would imply $x_n = 0$. Therefore there exists $1 \leq i \leq n-1$, s.t. $x_i \neq 0$. If there are two such i's, then $wt(x) \geq 2$. Otherwise $x_n = \sum_{i=1}^{n} x_i \neq 0$, so $wt(x) = 2$. We thus obtained that $d = 2$.

Both construction can be generalized to Σ^n for arbitrary Σ, by identifying Σ with the group $\mathbf{Z}_q$.

3.4 Hamming codes

We already constructed a $[n,n,1]$ code - the identity, and $[n,n-1,2]$ code - the parity code. We now follow [HR50] in constructing codes with $d = 3$. For simplicity we discuss first the $q = 2$ binary case. Recall that

$$d = \min\{\mathrm{wt}(x) : x \neq 0, xH = 0\}.$$

If we want to construct a code with $d = 3$, then the parity check matrix $H \in M_{n,n-k}(\mathcal{F})$ should satisfy $xH \neq 0$ for x with weight 1 or 2. A vector of weight 1 is of the form e_i, and $e_i H$ is the i'th row of the matrix. A vector of weight 2 is of the form $e_i + e_j$ so xH is the sum of the i'th row and j'th row of the matrix. We therefore conclude that all n rows of H should be non-zero, and there should be no identical rows. Fixing $\ell = n - k$, we obtain the maximal number of rows n when we enumerate all $2^{\ell} - 1$ non-zero ℓ-bit vectors.

Proposition 3.1. *For every positive ℓ there exists an $[2^\ell - 1, 2^\ell - \ell - 1, 3]_2$ code.*

Suppose that $y = x + e_i$, where y is a received word and x is a code word. Is there an efficient way of find e_i and recovering x? One way to do it is enumerate over all possible i and apply the parity check matrix, this takes $O(n^3)$ steps.

Another way is to note that $yH = e_iH = i^{\text{th}}$ row of H. So if we arrange the rows of H such that row i represent the number i in binary, then yH gives the location of the error. This procedure has complexity $O(n^2)$.

Note that we add logarithmic number of bits in order to correct one error! Is this necessary? Hamming bound, [HR50], proves that it is.

Theorem 3.2. *If an $(n, k, d)_q$ code exists, then*

$$q^k Vol_q(\lfloor \frac{d-1}{2} \rfloor, n) \leq q^n.$$

Note in particular that for $q = 2, d = 3$ this bound gives $2^k(n+1) \leq 2^n$, which is an inequality for Hamming codes:

$$2^k(n+1) = 2^{2^\ell - \ell - 1}(2^\ell - 1 + 1) = 2^{2^\ell - 1} = 2^n.$$

Proof of Theorem 2.2: Let C be an $(n, k, d)_q$ code. Let $x \neq y$ be two elements of the code. Since $d(x, y) \geq d$ it follows that

$$B(x, \lfloor \frac{d-1}{2} \rfloor) \cap B(y, \lfloor \frac{d-1}{2} \rfloor) = \emptyset.$$

Since

$$\cup_{x \in C} B(x, \lfloor \frac{d-1}{2} \rfloor) \subset \Sigma^n,$$

the proof follows, as the cardinality of the left hand side is $q^k Vol_q(\lfloor \frac{d-1}{2} \rfloor, n)$, and the cardinality of the right hand size is q^n. ■

A code is called *perfect*, if it achieves the Hamming bound exactly. It turns out the except the Hamming codes, there exist only two other perfect codes.

3.5 Singleton bound

Theorem 3.3. *If C is an $(n, k, d)_q$ code, then $d \leq n - k + 1$.*

Proof. Let $\pi : \Sigma^n \to \Sigma^{k-1}$ be the projection into the first $k - 1$ coordinates. Since there are q^k codewords, there exist $x, y \in C$ such that $\pi(x) = \pi(y)$. It follows that

$$d \leq \Delta(x, y) \leq n - k + 1.$$

■

Codes that meet the singleton bound are referred to as *Maximum distance separable (MDS)* codes. Singleton [SR64] proved this bound and gave several examples of MDS codes.

3.6 Reed Solomon codes

$RS_{q,n,k}$ for $k \leq n \leq q$ is defined as follows.

- We fix n distinct elements in $\mathcal{F}_q$: $a_1, \ldots, a_n$.
- Given a *message* $m = (m_0, \ldots, m_{k-1})$, construct the polynomial

$$m(x) = \sum_{i=0}^{k-1} m_i x^i \in \mathcal{F}_q^k,$$

- we let the *encoding* of m be

$$E(m) = (m(a_1), \ldots, m(a_n)).$$

Theorem 3.4. *For q prime power, and $k \leq n \leq q$, the code $RS_{q,n,k}$ is an $[n, k, n-k+1]_q$ code.*

Proof. First we have to justify the use of the square brackets, i.e. show that the code is linear. This follows from the fact that for any two polynomials m and n and all $a, b, c \in \mathcal{F}_q$

$$(bm + cn)(a) = bm(a) + cn(a).$$

In order to prove that $d = n - k + 1$, we first note that by the Singleton bound $d \leq n - k + 1$. Since a polynomial of degree $k-1$ is *determined* by its values on a set of k points, it follows that if m and n are two different polynomials of degree $\leq k$, then $m(x)$ and $n(x)$ agree at most on $k-1$ points in the set $\{a_1, \ldots, a_n\}$, i.e., $d(E(m), E(n)) \geq n - k + 1$. So $d \geq n - k + 1$, as needed. ■

Reed Solomon codes have a quite different description in the original paper [RS60] where they were constructed.

References

[1] Hamming, R. W. (1950) Error detecting and error correcting codes. *Bell System Tech. J.* **29**, 147–160.

[2] Reed, I. S.; Solomon, G. (1960) Polynomial codes over certain finite fields. *J. Soc. Indust. Appl. Math.* **8**, 300–304.

[3] Singleton, Richard C. (1964) Maximum distance q-nary codes. *IEEE Trans. Information Theory* **IT-10**, 116–118.