Coding Theory

(Linear codes)

Lector: Nikolai L. Manev

Institute of Mathematics and Informatics, Sofia, Bulgaria

Modes of error control

Error-control codes are used for two main goals, either separately, or together:

- *error detecting*: In this mode the decoder either detects the presence of error in the received word, or with previously given probability (≈ 1) asserts that the received word coincides with the sent codeword.
- *error correcting*: In this mode the decoder attempts to correct the errors in the received word. It is also referred to as *forward error correction (FEC)*.

Modes of error control

In error detecting mode, after detecting an error the decoder can proceed in two ways:

- Request a retransmission of the erroneous received codeword. This reaction is used in a very reliable set of error control strategies collectively referred to as *automatic repeat request (ARQ)* protocols.
- Tag the word as being incorrect and pass it further. It is referred to as *muting*. It is typical of applications (e.g., voice communications and digital audio) in which delay constraints do not allow for a possible retransmission. Usually a predetermined muted value is assigned to the received word.

Modes of error control

In error correcting mode, once an error has been detected, there are two ways to proceed:

- *complete decoding*. Every received word is decoded into a codeword.
- *incomplete decoding*. The decoder corrects only the received words for which the probability of decoding error is less than a previously given value (≈ 0). Otherwise, the decoder proceeds as in the error detecting mode.

Modes of error control

Let $\mathcal{C}$ be a code over $\mathbb{F}$. In order it to be used in error correcting mode a *decoding rule, or scheme (algorithm) of decoding* is required. Any such rule is a mapping

$$f : \left| \begin{array}{lcl} \mathbb{F}^n & \rightarrow & \mathcal{C} \\ v & \rightarrow & c = f(v) \end{array} \right.$$

- In complete decoding mode f is defined for every $v \in \mathbb{F}^n$.
- In incomplete decoding mode the domain of accepted values of F is the union of $\mathcal{C}$ and a special tag. For some $v \in \mathbb{F}^n$ this tag is the value of $f(v)$.

Probability of error decoding

Let $\{p(c) \mid c \in \mathcal{C}\}$ be the probability distribution (pmf) of the codewords of $\mathcal{C}$. Let $p_{ud}(c)$ be the probability of undetected error when the codeword c is sent, that is

$$p_{ud}(c) = \Pr(c' \neq c \text{ received } \mid c \text{ sent})$$

Therefore the *probability P_{ud} of undetected error* is given by

$$P_{ud} = \sum_{c \in \mathcal{C}} p(c) p_{ud}(c).$$

Let $p(err \mid c) = \Pr(\text{error decoding} \mid c \text{ sent})$ for a chosen decoding rule. The *probability of error decoding per word, or word error rate* is

$$P_e = \sum_{c \in C} p(c) p(err \mid c).$$

Decoders

Let the vectors of $\mathbb{F}^n$ appear at the output of the channel according to the probability mass function $\{p_r(v) \mid v \in \mathbb{F}^n\}$.
Let $p(c \mid v)$ *be the probability that codeword* c *is sent conditioned on the receipt of the vector* $v \in \mathbb{F}^n$. Let $p(v \mid c)$ *be the probability of receiving* v *provided that the codeword* c *is transmitted*.
These probabilities are related each other by Bayes' rule:

$$p(c \mid v) = \frac{p(c)\, p(v \mid c)}{p_r(v)}.$$

A *maximum posteriori decoder* looks for a codeword c_0 that maximizes $p(c \mid v)$.
A *maximum likelihood decoder* identifies the codeword that maximizes $p(v \mid c)$.

Hamming distance

Definition. The *Hamming distance* $d(x, y)$ between two vectors x and y of $\mathbb{F}^n$ is the number of places in which they differ:

$$d(x, y) \stackrel{\text{def}}{=} |\{i | x_i \neq y_i\}|.$$

(i) $d(x, y) \geq 0$ with equality iff $x = y$,

(ii) $d(x, y) = d(y, x)$

(iii) $d(x, y) \leq d(x, z) + d(y, z)$ (triangle inequality).

Hamming distance

Definition. The *Hamming distance* $d(x,y)$ between two vectors x and y of $\mathbb{F}^n$ is the number of places in which they differ:

$$d(x,y) \stackrel{\text{def}}{=} |\{i|x_i \neq y_i\}|.$$

(i) $d(x,y) \geq 0$ with equality iff $x = y$,
(ii) $d(x,y) = d(y,x)$
(iii) $d(x,y) \leq d(x,z) + d(y,z)$ (triangle inequality).

Definition. Let $C \subset \mathbb{F}^n$ be a code of length n. Then the *minimum distance* of C is the smallest possible Hamming distance between two different codewords. That is,

$$d(C) \stackrel{\text{def}}{=} \min_{x \neq y \in C} d(x,y).$$

Hamming weight

Definition. The *Hamming weight* $\mathrm{wt}(c)$ of a codeword $c \in C$ is the number of nonzero entries of the codeword. The *minimum weight* $\mathrm{wt}(C)$ of the code C is the smallest nonzero value of $\mathrm{wt}(c)$. Obviously $\mathrm{wt}(c) = d(c, o)$.

If $\mathcal{C}$ is a linear code, then $d(\mathcal{C}) = \mathrm{wt}(C)$. Indeed

$$d(C) = \min_{i \neq j} d(c_i, c_j) = \min_{i \neq j} d(o, c_i - c_j) = \min_{i \neq j} \mathrm{wt}(c_i - c_j) = \mathrm{wt}(\mathcal{C}).$$

Hamming weight

Definition. The *Hamming weight* $\mathrm{wt}(c)$ of a codeword $c \in C$ is the number of nonzero entries of the codeword. The *minimum weight* $\mathrm{wt}(C)$ of the code C is the smallest nonzero value of $\mathrm{wt}(c)$. Obviously $\mathrm{wt}(c) = d(c, o)$.

If $\mathcal{C}$ is a linear code, then $d(\mathcal{C}) = \mathrm{wt}(C)$. Indeed

$$d(C) = \min_{i \neq j} d(c_i, c_j) = \min_{i \neq j} d(o, c_i - c_j) = \min_{i \neq j} \mathrm{wt}(c_i - c_j) = \mathrm{wt}(\mathcal{C}).$$

Example. Consider the $[4, 2]$ code $\mathcal{C}$ over $\mathbb{Z}_3 = \{0, 1, 2\}$:

$$C = \{0000, 1110, 2220, 2102, 1201, 0212, 0121, 2011, 1022\}$$

$$A_0 = 1,\ A_1 = A_2 = 0,\ A_3 = 8,\ A_4 = 0; \quad d(C) = \mathrm{wt}(C) = 3.$$

Error detecting and correcting capability

$(n, M, d)_q$ – q-ary code of length n, cardinality M, and minimum distance d.

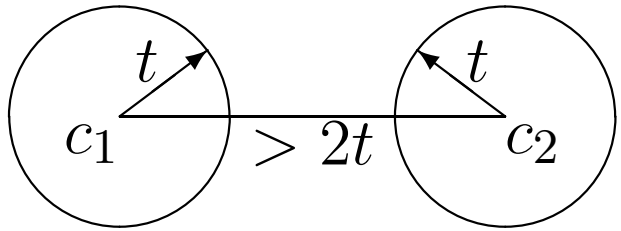
$[n, k, d]_q$ – q-ary linear code of length n, dimension k, and minimum distance d, $M = q^k$.

Definition. We say that a code $\mathcal{C}$ can *detect up to t errors*, or it is *t-error detecting*, if a valid codeword can not be obtained by changing from one to t positions of any codeword. In other words, any error pattern of Hamming weight at most t does not cause the received vector to be a valid codeword.

Definition. We say that a code $\mathcal{C}$ *correct up to t errors*, or it is *t-error correcting*, if for each vector $v \in \mathbb{F}^n$, there exists at most one codeword $c \in \mathcal{C}$, such that $d(v, c) \leq t$.

Error detecting and correcting capability

Theorem. A code $\mathcal{C}$ is t-error correcting, if and only if $d(\mathcal{C}) \geq 2t+1$. It detects s errors if and only if $d(\mathcal{C}) \geq s+1$.



Theorem. A code $\mathcal{C}$ with minimum distance d can detect all error patterns of weight less than or equal to $d-1$.

Theorem. A code $\mathcal{C}$ with minimum distance d can correct all error patterns of weight less than or equal to $\lfloor (d-1)/2 \rfloor$.

Linear codes

Let $\mathcal{C}$ be a q-ary $[n, k]$ code.

Definition. *Generator matrix* of $\mathcal{C}$ is any $k \times n$ matrix whose rows form a basis of $\mathcal{C}$ as a linear space.

If $\mathbf{G}$ is a generating matrix of $\mathcal{C}$, then

$$\mathcal{C} \equiv \{a\mathbf{G} \mid a \in \mathbb{F}^k\}.$$

Any generating matrix gives an *encoding rule*. Varying the generator matrix we obtain different encoding rules.

Definition. We say that $\mathbf{G}$ is in a *systematic form*, if $\mathbf{G} = (\mathbf{I}_k \mid \mathbf{P})$ where $\mathbf{I}_k$ is the $k \times k$ identity matrix and $\mathbf{P}$ is a $k \times (n-k)$ matrix.

$$\mathbf{G} = (1 \underbrace{1 \ldots 1}_{n-1}) \qquad \mathbf{G} = \begin{pmatrix} 1 & & & 1 & 1 & 0 \\ & 1 & & 1 & 0 & 1 \\ & & 1 & 0 & 1 & 1 \end{pmatrix}.$$

An encoding rule based on a generator matrix in systematic form is called *systematic encoding*. Systematic encoding simplifies encoding and recovering the data block from a codeword. The data block $\mathbf{i} = (i_1, \ldots, i_k) \in \mathbb{F}^k$ is embedded without modification in the first k coordinates of the corresponding codeword:

$$\mathbf{c} = \mathbf{iG} = (i_1, \ldots, i_k \mid \mathbf{iP}).$$

Definition. The q-ary code $\mathcal{C}$ is called *systematic on given k positions* (and the symbols in these positions are called information symbols) if $|\mathcal{C}| = q^k$ and exactly one codeword exists for every choice of values in these k positions, i.e. every k-tuple $\mathbf{i} \in \mathbb{F}^k$ occurs in exactly one codeword at these k-positions. For example, $\mathbf{G} = (\mathbf{P} \mid \mathbf{I}_k)$ for cyclic codes.

Inner product of two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{F}^n$ is

$$\mathbf{a} \circ \mathbf{b} \stackrel{\text{def}}{=} a_1 b_1 + a_2 b_2 + \cdots + a_n b_n.$$

If $\mathbf{a} \circ \mathbf{b} = 0$ we say that $\mathbf{a}$ and $\mathbf{b}$ are *orthogonal*.

Definition. The *dual code (written $\mathcal{C}^\perp$)* of a code $\mathcal{C}$ is called the set of those vectors of $\mathbb{F}^n$, which are orthogonal to each codeword of $\mathcal{C}$, that is

$$\mathcal{C}^\perp \stackrel{\text{def}}{=} \{v \in \mathbb{F}^n \mid v \circ c, \, \forall c \in \mathcal{C}\}.$$

If $\mathcal{C}^\perp = \mathcal{C}$, the code $\mathcal{C}$ is called *selfdual*.

Definition. Any generator matrix $\mathbf{H}$ of $\mathcal{C}^\perp$ is called *parity-check matrix* of $\mathcal{C}$.

Theorem. For any $c \in \mathcal{C}$ and $x \in \mathcal{C}^{\perp}$,

(1) $c\mathbf{H}^{\tau} = o$, (2) $x\mathbf{G}^{\tau} = o$, and (3) $\mathbf{H}\mathbf{G}^{\tau} = \mathbf{O}$

hold.

The equations (1) are called *parity-check equations*. Their number is $n - k$ for an $[n, k]$ code.

Theorem. If $\mathcal{C}$ is a systematic code with $\mathbf{G} = (\mathbf{I}_k \mid \mathbf{P})$, then the $(n - k) \times n$ matrix $\mathbf{H} = (-\mathbf{P}^{\tau} \mid \mathbf{I}_{n-k})$ is a parity-check matrix of $\mathcal{C}$.

Example. Consider the $[7, 4]$ binary code (*Hamming code*) $\mathcal{H} = \{(i_1, i_2, i_3, i_4, p_1, p_2, p_3) \mid i_j \in \mathbb{Z}_2\}$ defined by parity-check equations

$$p_1 = i_1 + i_2 + i_3, \quad p_2 = i_2 + i_3 + i_4, \quad p_3 = i_1 + i_2 + i_4.$$

The corresponding parity-check matrix is

$$\mathbf{H} = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

Theorem. A code $\mathcal{C}$ has minimum distance d if and only if any $d-1$ or less columns are of a parity-check matrix of $\mathcal{C}$ are linearly independent.

Theorem. A linear code $\mathcal{C}$ corrects t errors, if and only if any $2t$ columns of its parity-check matrix are linearly independent.

Theorem. (*Singleton bound*) The minimum distance of an $[n, k]$ linear code $\mathcal{C}$ is bounded by

$$d(\mathcal{C}) \leq n - k + 1.$$

Equivalent codes

Let us note that elementary row operations on $\mathbf{G}$ and $\mathbf{H}$ do not affect on the code. They only give another generator, respectively parity-check matrix. In contrast, a permutation of the columns may change the code.

Definition. Two codes over $\mathbb{F} = GF(q)$ are *equivalent* if one can be obtained from the other by permuting the coordinate positions of the code and/or multiplying each coordinate by a non-zero scalar. The codes are referred to as *isomorphic* if a positional permutation suffices to take one to the other. (When $q = 2$ both the notions are obviously one and the same.) Any isomorphism of $\mathcal{C}$ onto itself is called an *automorphism*. The set of all automorphisms of C is called the *automorphism group* of $\mathcal{C}$.

Hamming codes 1

Definition. The *Hamming code* $\mathcal{H}_r$ is called the binary code defined by a $r \times (2^r - 1)$ parity-check matrix whose columns are binary representations of the numbers $1, 2, 3, \ldots, 2^r - 1$. Here is the parity-check matrix of $\mathcal{H}_4$:

$$\mathbf{H}_4 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

$\mathcal{H}_r$ is $[2^r - 1,\ 2^r - r - 1,\ 3]$ binary code.

$$\mathbf{H}_{r+1} = \begin{pmatrix} 0 & 0 & \ldots & 0 & 1 & 1 & 1 & \ldots & 1 \\ & & & & 0 & & & & \\ & \mathbf{H}_r & & & \vdots & & \mathbf{H}_r & & \\ & & & & 0 & & & & \end{pmatrix}.$$

Definition. The dual code of the Hamming code $\mathcal{H}_r$ is called *simplex code, written* $\mathcal{S}_r$. That is, $\mathbf{H}_r$ is a generator matrix of $\mathcal{S}_r$.

Theorem. The simplex code $\mathcal{S}_r$ is a $[2^r-1,\ r,\ 2^{r-1}]$ code and every nonzero codeword has weight 2^{r-1}.

Definition. The q*-ary Hamming code* $\mathcal{H}_r$ is the code over $\mathbb{F}=GF(q)$ defined by a $r\times\frac{q^r-1}{q-1}$ parity-check matrix $\mathbf{H}$ having as columns all distinct representatives of one-dimensional subspaces of $\mathbb{F}^r$. Usually r-tuples for which the uppermost nonzero element is 1 are selected.

The q-ary Hamming code has parameters

$$\left[n=\frac{q^r-1}{q-1},\quad k=\frac{q^r-1}{q-1}-r,\quad d=3\right].$$

Slepian standard array

Definition. Let $\mathcal{C}$ be an $[n,\ k]$ code with a parity-check matrix $\mathbf{H}$. For any $\mathbf{v} \in \mathbb{F}^n$, the matrix product $\mathbf{s} = \mathbf{v}\mathbf{H}^\tau$ is called *syndrome for* $\mathbf{v}$.

Note that the syndrome for $\mathbf{v}$ is a vector of $\mathbb{F}^{n-k}$ and it is zero vector if and only if $\mathbf{v}$ is a codeword.

Let $\mathcal{C} = \{\mathbf{c}_1{=}o, \mathbf{c}_2, \ldots, \mathbf{c}_M\},\ M = q^k$. Consider the set of all cosets of $\mathcal{C}$ in $\mathbb{F}^n$, that is

$$\mathbf{v} + \mathcal{C} = \{\mathbf{v} + \mathbf{c} \mid \mathbf{c} \in \mathcal{C}\}.$$

Leader of the coset $\mathbf{v} + \mathcal{C}$ is the vector of minimum weight in the coset. (If there are several such vectors we take and fix one at random.) The weight of the leader is referred to as the *weight of the coset*. Let α_i denote the number of cosets of weight i. *Coset weight distributions* is the set of numbers

$$\alpha_0 = 1,\ \alpha_1,\ \alpha_2,\ \ldots,\ \alpha_n.$$

Slepian standard array 2

Theorem. Two vectors $\mathbf{u}$ and $\mathbf{v} \in \mathbb{F}^n$ are in the same coset of $\mathcal{C}$ if and only if they have the same syndrome (That is, there is a one-to-one correspondence between cosets and syndromes.)

Definition. A *Standard Array* for an $[n, k]_q$ code $\mathcal{C}$ is a $q^{n-k} \times q^k$ array of all vectors of $\mathbb{F}^n$ in which the first row consists of the code $\mathcal{C}$, and the other cosets $v_i + \mathcal{C}$, each arranged in corresponding to $\mathcal{C}$ order with the coset leader on the left. Cosets are arranged in an increasing order of cosets' weights.

Slepian standard array

Cosets	Coset elements				Syndrome
$o + C$	o	c_2	$\dots$	c_M	o
$v_2 + C$	v_2	$v_2 + c_2$	$\dots$	$v_2 + c_M$	s_2
$v_3 + C$	v_3	$v_3 + c_2$	$\dots$	$v_3 + c_M$	s_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$v_N + C$	v_N	$v_N + c_2$	$\dots$	$v_N + c_M$	s_N
$v_{N+1} + C$	v_{N+1}	$v_{N+1} + c_2$	$\dots$	$v_{N+1} + c_M$	s_{N+1}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$v_{q^{n-k}} + C$	$v_{q^{n-k}}$	$v_{q^{n-k}} + c_2$	$\dots$	$v_{q^{n-k}} + c_M$	$s_{q^{n-k}}$

Slepian standard array

Theorem. All vectors with weight $\leq t$ in $\mathbb{F}^n$ belong to different cosets of $\mathcal{C}$ if and only if the minimum distance $d(\mathcal{C}) \geq 2t+1$.

Therefore, when $t = \lfloor (d-1)/2 \rfloor$, that is, $d = 2t+1$ or $2t+2$, we have

$$N = 1 + \binom{n}{1}(q-1) + \binom{n}{2}(q-1)^2 + \cdots + \binom{n}{t}(q-1)^t.$$

Theorem. For BSC with error rate $p < 1/2$, the maximum likelihood decoding is equivalent to the *nearest neighbour decoding* rule.

$$\Pr(v \text{ received} \mid c \text{ sent}) = p^i(1-p)^{n-i}$$

$$p(1-p)^{n-1} > p^2(1-p)^{n-2} > p^3(1-p)^{n-3} > \cdots > p^i(1-p)^{n-i} > \cdots$$

Decoding linear codes 1

The standard array can be simplified by storing only the first and last column, that is, coset leaders and their syndromes.

Algorithm.

1. At receiving $\mathbf{v}$ compute its syndrome $\mathbf{s} = \mathbf{v}\mathbf{H}^{\tau}$.
2. Locate the syndrome in the syndrome column.
3. Determine the corresponding coset leader. This is the error vector, $\mathbf{e}$.
4. Calculate the codeword which is sent by $\mathbf{c} = \mathbf{v} - \mathbf{e}$.

Decoding linear codes

Leader	Syndrome
o	o
v_2	s_2
$\vdots$	$\vdots$
v_N	s_N
v_{N+1}	s_{N+1}
$\vdots$	$\vdots$
$v_{q^{n-k}}$	$s_{q^{n-k}}$

Leader	Syndrome
0000000	000
1000000	001
0100000	010
0010000	011
0001000	100
0000100	101
0000010	110
0000001	111
$\emptyset$	$\emptyset$

Error detection performance of linear codes

Let $\mathcal{C}$ be an $[n,\, k]$ code. Consider BSC with channel error $0 < p < \frac{1}{2}$. Let a codeword c is sent across the channel. The received vector $c + e = c'$ is a codeword if and only if the error-vector $e \in \mathcal{C}$. Thus

$$\begin{aligned} p_{ud}(c) &= \Pr(c' \neq c \text{ received} \mid c \text{ sent}) = \Pr(e \in \mathcal{C}) \\ &= \Pr(u \in \mathcal{C} \text{ received} \mid o \text{ sent}) = P \end{aligned}$$

Let $\mathrm{wt}(e) = i$. Then $p^i(1-p)^{n-i}$ is the probability of transforming the zero codeword into e. Hence

$$P = \sum_{i=0}^{n} \Pr(e \in \mathcal{C}, \mathrm{wt}(e) = i \mid o \text{ sent}) = \sum_{i=0}^{n} A_i p^i (1-p)^{n-i},$$

where A_i is the number of codewords of weight i. Therefore

$$P_{ud} = \sum_{c \in \mathcal{C}} p(c) p_{ud}(c) = P \sum_{c \in \mathcal{C}} p(c) = \sum_{i=0}^{n} A_i p^i (1-p)^{n-i}.$$

Error correction performance of linear codes

Let $\mathcal{C}$ be an $[n,\, k]$ code with $d(\mathcal{C}) = 2t+1$ or $2t+2$ and syndrome (standard array) decoding rule. Consider BSC with channel error p. Let a codeword c is sent across the channel and $v = c + e$ is received. The vector v will be erroneously corrected if and only if $v = e_1 + c' \in e_1 + \mathcal{C}$, where e_1 is a leader of a coset and $e \neq e_1$. Thus, a decoding error occurs if and only if $e = e_1 + (c' - c) \in e_1 + \mathcal{C}$ is not a coset leader. Hence, $p(err \mid c)$ does not depend on c and

$$P_e = \sum_{c \in C} p(c) p(err \mid c) = p(err \mid o) \sum_{c \in C} p(c) = p(err \mid o) = 1 - P_{cor},$$

where P_{cor} is the probability of correct decoding.

$$P_{cor} = \Pr(e \text{ is a coset leader}) = \sum_{i=0}^{n} \alpha_i p^i (1-p)^{n-i},$$

where α_i is the number of coset leaders with weight i.

Error correction performance of linear codes

Therefore

$$P_e = 1 - \sum_{i=0}^{n} \alpha_i p^i (1-p)^{n-i},$$

for complete decoding. For incomplete decoding this probability is slightly less.

Since $\alpha_j = \binom{n}{j}$ for $j \leq t$, we have

$$P_e \leq 1 - \sum_{i=0}^{t} \binom{n}{i} p^i (1-p)^{n-i}.$$

The end of the part

Thank You for Attention!