SUPPLEMENT TO SECT. 5.4
ERROR-DETECTING AND ERROR-CORRECTING CODES

1. Binary codes

1.1. **Vector spaces over $\mathbb{Z}_2$.** Let $\mathbf{B}^n$ be the set of all binary words $x_1 \dots x_n$ of length n, where $x_i = 0$ or 1. Note that $\mathbf{B}^n$ is isomorphic to

$$\mathbb{Z}_2^n = \underbrace{\mathbb{Z}_2 \times \cdots \times \mathbb{Z}_2}_{n \text{ times}}$$

and is a vector space over the finite field $\mathbb{Z}_2$ of two elements 0 and 1. The addition of two vectors $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ is component-wise mod 2 and multiplication of a vector x by elements of $\mathbb{Z}_2$ is also component-wise, $1 \cdot x = x$ and $0 \cdot x = \mathbf{0}$, where $\mathbf{0} = (0, \dots, 0)$.

1.2. **Coding function.** Suppose the original message is composed of binary words of length m. The *coding function* is a function

$$f : \mathbf{B}^m \to \mathbf{B}^n,$$

which replaces the original word $u \in \mathbf{B}^m$ by the coded word $f(u) \in \mathbf{B}^n$, where $n > m$. The words in the image $f(\mathbf{B}^m)$ of the coding function f in $\mathbf{B}^n$ are called *codewords*. The goal is to detect and correct as many errors as possible; check Examples 1 and 2 on pp. 232-233.

1.3. **The Hamming distance.** For every $v \in \mathbf{B}^n$ define it *weight* $\mathrm{wt}(v)$ to be the number of 1s in its binary expression.

Lemma 1. *For all $u, v \in \mathbf{B}^n$,*

$$\mathrm{wt}(u + v) \leq \mathrm{wt}(u) + \mathrm{wt}(v).$$

Proof. Since the addition is mod 2, it is clear that number of 1s in $u + v$ is not greater than number of 1s in u plus number of 1s in v. □

Define the *distance* $d(u, v)$ between binary words $u, v \in \mathbf{B}^n$ by

$$d(u, v) = \mathrm{wt}(u - v).$$

Note that the distance between u and v is the number of places in which these two words differ. Also, since we are working over $\mathbb{Z}_2$, $u - v = u + v$ and $d(u, v) = \mathrm{wt}(u + v)$.

The following result is fundamental.

Theorem 2. *The distance on $\mathbf{B}^n$ is symmetric, $d(u, v) = d(v, u)$ and satisfies the triangle inequality*

$$d(u, v) \leq d(u, w) + d(w, v)$$

for all $u, v, w \in \mathbf{B}^n$.

Proof. Using that $w + w = \mathbf{0}$ in $\mathbf{B}^n$, we get from Lemma 1

$$\begin{aligned} d(u, v) &= \mathrm{wt}(u + v) = \mathrm{wt}((u + w) + (w + v)) \leq \mathrm{wt}(u + w) + \mathrm{wt}(w + v) \\ &= d(u, w) + d(w, v). \end{aligned}$$ □

2. Error-detection and error-correction

Let $f : \mathbf{B}^m \to \mathbf{B}^n$ be a coding function. We say that $u \in \mathbf{B}^n$ *contains l errors*, if it is obtained from a codeword v by the alteration of l digits. In other words, $u \in \mathbf{B}^n$ contains l errors if there is $v \in f(\mathbf{B}^m)$ such that $d(u, v) = l$. We say that a coding function f *detects k or fewer errors*, if every $u \in \mathbf{B}^n$, which is obtained from some codeword v by l errors, $1 \leq l \leq k$, is not a codeword, i.e., $u \notin f(\mathbf{B}^m)$. We say that a coding function f *corrects k or fewer errors*, if it detects such errors and for any $u \in \mathbf{B}^n$ with l errors, $1 \leq l \leq k$, there is a *unique* $v \in f(\mathbf{B}^m)$ such that $d(u, v) = l$.

We have the following main results, which are, respectively, Theorems 5.4.1 and 5.4.2 in the textbook.

Theorem 3. *Let $f : \mathbf{B}^m \to \mathbf{B}^n$ be a coding function. The f allows the detection of k or fewer errors if and only if the minimal distance between distinct codewords is at least $k + 1$.*

Proof. If the minimal distance between distinct codewords is at least $k + 1$ and $u \in \mathbf{B}^n$ is obtained from a codeword v by l errors, then $d(u, v) = l$ and $u \notin f(\mathbf{B}^m)$ for $1 \leq l \leq k$. Conversely, if f allows the detection of k or fewer errors, then no two codewords could be at a distance k. Indeed, then by k errors one codeword would be converted to another codeword, a contradiction. □

Theorem 4. *Let $f : \mathbf{B}^m \to \mathbf{B}^n$ be a coding function. The f allows the correction of k or fewer errors if and only if the minimal distance between distinct codewords is at least $2k + 1$.*

Proof. Suppose that the minimal distance between distinct codewords is at least $2k + 1$. Then if for $u \in \mathbf{B}^n$ there exist two distinct codewords v and w such that $d(u, v) = l$ and $d(u, w) = l$, where $1 \leq l \leq k$, then by the triangle inequality

$$2k + 1 \leq d(v, w) \leq d(u, v) + d(u, w) = 2l \leq 2k$$

— a contradiction! Conversely, suppose that f allows the correction of k or fewer errors and there are two codewords u and v such that $d(u, v) = 2k$. The codewords u and v differ in $2k$ places. Changing k of them in v, we obtain $w \in \mathbf{B}^n$ such that $d(w, v) = d(w, u) = k$ — a contradiction to the assumption that f allows the correction of k errors. □

Using these theorems, check that the code in Example 3 on p. 236 detects up to two errors and corrects any single error.

3. Linear codes

A coding function $f : \mathbf{B}^m \to \mathbf{B}^n$ gives a *linear code* if its image $f(\mathbf{B}^m)$ is a subgroup of $\mathbf{B}^n$. This is equivalent to the statement that if u and v are the codewords, than $u+v$ is also a codeword. (Recall that $u+v = u-v$ in $\mathbf{B}^n$). In particular, $\mathbf{0} = u+u$ is a codeword.

Theorem 5.4.3 asserts that for a linear code the minimal distance between distinct codewords is the lowest weight of the non-zero codeword. Indeed, let $d(u,v) = d$ be the minimal distance and $x = \mathrm{wt}(w)$ be the minimal weight of a non-zero codeword. Than

$$d \le d(w, \mathbf{0}) = \mathrm{wt}(w) = x \quad \text{and} \quad d = d(u,v) = \mathrm{wt}(u+v) \ge \mathrm{wt}(w) = x,$$

which shows that $d = x$.

Every *linear transformation* $f : \mathbf{B}^m \to \mathbf{B}^n$ of $\mathbb{Z}_2$-vector spaces $\mathbf{B}^m$ and $\mathbf{B}^n$ produces a linear code. Indeed, this follows from the defining property of a linear transformation

$$f(u+v) = f(u) + f(v) \quad \text{for all} \quad u, v \in \mathbf{B}^m,$$

so that the range W of f is a subgroup of $\mathbf{B}^n$. Every such linear transformation f is given by a $m \times n$ matrix G by the formula

$$f(u) = u \cdot G,$$

where $u \in \mathbf{B}^m$ is considered as $1 \times m$ matrix (a row vector) and $\cdot$ stands for the matrix multiplication of the $1 \times m$ and $m \times n$ matrices, so that $f(u) \in \mathbf{B}^n$. The matrix G of the form $G = (I_m, A)$, where I_m is the $m \times m$ identity matrix and A is $m \times (n-m)$ matrix, is called a *generator matrix.* We have $w = f(u) = uv$, where $v = u \cdot A$. Thus the codeword $w = uv$ consists of the original word u and the "check digits" $v = u \cdot A$. Consider Examples 1-4 on pp. 239-240.

4. Detecting and correcting errors

Consider a linear code $f : \mathbf{B}^m \to \mathbf{B}^n$ with a subgroup $W \le \mathbf{B}^n$ of the codewords. The *coset decoding table* is obtained as follows. In the top row list, in any order, all elements of W starting with $\mathbf{0}$. Next, choose a word v of minimal weight in $\mathbf{B}^n$ which is not in W (if there are several, choose any one). In the second row list all elements of the coset $v + W$ exactly in the same order as in the first row, element $v + w$ beneath a codeword w for all $w \in W$. Then look for an element u of minimal weight which is not already listed above and list its coset $u + W$; $u + w$ should be beneath $v + w$. Continue until all elements of $\mathbf{B}^n$ are being listed. The table has $|W|$ columns, the order of the subgroup W, and $2^n/|W|$ rows, the order of the coset space $\mathbf{B}^n/W$. The elements of the first column of the coset decoding table are called *coset leaders.*

To decode a message, we correct each word which is not a codeword by finding it in the coset decoding table and replacing it by the codeword from the first row which is in the same column as the given word. This is called a *maximum likelihood decoding* since we correctly decode only words with few errors. Consider Example on pp. 242-243 with the generating matrix

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

The minimal distance between codewords is 3 so the code detects 2 errors and corrects one error. Now consider the received message on the bottom of p. 243. Decoding it by using the coset decoding table, you will see that it corrects all single errors but does not correct double or triple errors.

If the linear code $f : \mathbf{B}^m \to \mathbf{B}^n$ is given by a generator matrix, there is no need to construct and store in the memory the whole coset decoding table. Namely let $G = (I_m, A)$, where A is $m \times (n-m)$ matrix, be the corresponding generator matrix. Then $w \in W$ if and only if $w = u \cdot G = uv$, where $v = u \cdot A$. Equivalently, the last equation can be written as $w \cdot H = 0$, where $n \times (n-m)$ matrix H, called *parity-check matrix*, is given by

$$H = \begin{pmatrix} A \\ I_{n-m} \end{pmatrix},$$

and I_{n-m} is $(n-m) \times (n-m)$ identity matrix. Indeed, for $w \in W$ we have

$$w \cdot H = uv \cdot \begin{pmatrix} A \\ I_{n-m} \end{pmatrix} = u \cdot A + v = u \cdot A + u \cdot A = \mathbf{0},$$

since $w = uv \in W$ if and only if $v = u \cdot A$. (Note that we working over in the finite field $\mathbb{Z}_2$).

In the coding theory, for a word $w \in \mathbf{B}^n$ the $n - m$ word $w \cdot H$ is called a *syndrome* of w. Thus we have the following result (Theorem 5.4.5 and Corollary 5.4.6 in the textbook).

Theorem 5. *Let H be the parity-check matrix associated with a given generator matrix G. Then $w \in \mathbf{B}^n$ is a codeword if and only if its syndrome $w \cdot H = \mathbf{0}$ in $\mathbf{B}^{n-m}$ and two words are in the same row of the coset decoding table if and only if they have the same syndrome.*

Then from a coset decoding table one gets a two column decoding table, the second column being coset leaders with the first column being their syndromes. One can construct such table without constructing coset decoding table first.

The corresponding correction algorithm is the following.

(1) Compute the syndromes $w \cdot H$, where w ranges over all received words in a message and H is the parity-check matrix.
(2) If the syndrome $w \cdot H$ is not zero, add it to the coset leader in the same row as the syndrome. The result is a codeword.
(3) Take the initial m digits of all obtained codewords.

This will correct as many errors as the code allows, see Theorems 5.4.1 and 5.4.2. Some multiple errors will still remain. In case when a message contains few errors, the method works. Check Example on p. 250.