

MATH 433
Applied Algebra

Lecture 20:
Binary codes.
Linear codes.

Error-detecting/correcting codes

Messages sent over electronic and other channels are subject to distortions of various sorts. Therefore it is important to encode a message so that a possible error can be detected. Then the receiver may ask that the message be repeated. Such codes are called **error-detecting**.

To achieve this, the message should carry a certain degree of redundancy. One way to do this is a **checksum**. Namely, the sender adds to a message one or several check symbols, which are functions of the message. Then the receiver reevaluates these additional symbols.

In some cases, requesting that the message be repeated is too expensive. For such cases, we need a code that not only can detect an error, but also allows to correct it. Such codes are called **error-correcting**.

ISBN

International Standard Book Number (ISBN) is assigned to all published books. It is an example of an error-detecting code.

- ISBN-10 (old standard) consists of 9 decimal digits that constitute the number followed by a check symbol, which is a digit in base 11 (0–9 or X, the Roman notation for 10).

If $a_1a_2 \dots a_9a_{10}$ is the number, then

$$10a_1 + 9a_2 + 8a_3 + \dots + 3a_8 + 2a_9 + a_{10}$$

is to be divisible by 11. This happens for a unique choice of a_{10} .

The code allows to detect one wrong digit or exchange of two digits.

Example. 0 521 54050 X (ISBN-10 of the textbook).

ISBN

- ISBN-13 (new standard) consists of 13 decimal digits, the last one being a checksum. If $b_1b_2 \dots b_{12}b_{13}$ is the number, then $b_1 + 3b_2 + b_3 + 3b_4 + \dots + 3b_{12} + b_{13}$ is to be divisible by 10. This happens for a unique choice of b_{13} .

The code allows to detect one wrong digit or exchange of two neighboring digits.

Old numbers are converted into new ones by adding 978 at the beginning and recalculating the checksum.

Example. ISBN-10 of the textbook is 052154050X.

Therefore ISBN-13 of the textbook is 978-052154050d, where

$$\begin{aligned} &9 + 3 \cdot 7 + 8 + 3 \cdot 0 + 5 + 3 \cdot 2 + 1 \\ &+ 3 \cdot 5 + 4 + 3 \cdot 0 + 5 + 3 \cdot 0 + d \equiv 0 \pmod{10}. \end{aligned}$$

We obtain that $d = 6$.

Problem 1. Find the missing digit in an ISBN-10:
04*5011614.

Let d be the missing digit. Then

$$\begin{aligned} 10 \cdot 0 + 9 \cdot 4 + 8d + 7 \cdot 5 + 6 \cdot 0 + 5 \cdot 1 \\ + 4 \cdot 1 + 3 \cdot 6 + 2 \cdot 1 + 4 \equiv 0 \pmod{11}, \end{aligned}$$

which simplifies to $8d + 5 \equiv 0 \pmod{11}$. The inverse of 8 modulo 11 is 7 (as $7 \cdot 8 = 56 \equiv 1 \pmod{11}$). It follows that $d \equiv 7 \cdot (-5) \equiv 9 \pmod{11}$. Thus $d = 9$.

Problem 2. Could this be a valid ISBN-13:
978-0495022613 ?

$$\begin{aligned} 9 + 3 \cdot 7 + 8 + 3 \cdot 0 + 4 + 3 \cdot 9 + 5 \\ + 3 \cdot 0 + 2 + 3 \cdot 2 + 6 + 3 \cdot 1 + 3 \equiv 4 \not\equiv 0 \pmod{10}, \end{aligned}$$

therefore this could not be a valid ISBN-13.

Binary codes

Let us assume that a message to be transmitted is in binary form. That is, it is a word in the alphabet $\mathbf{B} = \{0, 1\}$. For any integer $k \geq 1$, the set of all words of length k is identified with \mathbf{B}^k .

A **binary code** (or a **binary coding function**) is an injective function $f : \mathbf{B}^m \rightarrow \mathbf{B}^n$.

For any $w \in \mathbf{B}^m$, the word $f(w)$ is called the **codeword** associated to w .

The code f is **systematic** if $f(w) = wu$ for any $w \in \mathbf{B}^m$ (that is, w is the beginning of the associated codeword). This condition clearly implies injectivity of the function f .

Encoding / decoding

The code $f : \mathbf{B}^m \rightarrow \mathbf{B}^n$ is used as follows.

Encoding: The sender splits the message into words of length m : w_1, w_2, \dots, w_s . Then he applies f to each of these words and produces a sequence of codewords $f(w_1), f(w_2), \dots, f(w_s)$, which is to be transmitted.

Decoding: The receiver obtains a sequence of words of length n : w'_1, w'_2, \dots, w'_s , where w'_i is supposed to be $f(w_i)$ but it may be different due to errors during transmission. Each w'_i is checked for being a codeword. If it is, $w'_i = f(w)$, then w'_i is decoded to w . Otherwise an error (or errors) is detected. In the case of an error-correcting code, the receiver attempts to correct w'_i by applying a correction function $c : \mathbf{B}^n \rightarrow \mathbf{B}^n$, then decodes the word $c(w'_i)$.

Examples. • Parity bit.

$f : \mathbf{B}^m \rightarrow \mathbf{B}^{m+1}$, $f(w) = wx$, where x is the parity bit of w , which means that $x = 0$ if there is an even number of 1s in w and $x = 1$ otherwise.

The number of 1s in any codeword is even. This code detects any single error (or an odd number of errors), but correction is not possible.

- Tell three times.

$f : \mathbf{B}^m \rightarrow \mathbf{B}^{3m}$, $f(w) = www$.

This code detects two errors (for sure) and also can correct one error (using “split decision”).

We say that a binary code **detects** k **errors** if a wrong word is detected whenever there are k or fewer errors. We say that the code **corrects** k **errors** if the correction is successful whenever there are k or fewer errors.

The distance $d(w_1, w_2)$ between binary words w_1, w_2 of the same length is the number of positions in which they differ. The **weight** of a word w is the number of nonzero digits, which is the distance to the zero word.

Theorem Let $f : \mathbf{B}^m \rightarrow \mathbf{B}^n$ be a coding function. Then
(i) f allows detection of k or fewer errors if and only if the minimum distance between distinct codewords is at least $k + 1$; **(ii)** f allows correction of k or fewer errors if and only if the minimum distance between distinct codewords is at least $2k + 1$.

The correction function c is usually chosen so that $c(w)$ is the codeword closest to w .

Linear codes

The binary alphabet $\mathbf{B} = \{0, 1\}$ is naturally identified with \mathbb{Z}_2 , the field of 2 elements. Then \mathbf{B}^n can be regarded as the n -dimensional vector space over the field \mathbb{Z}_2 .

A binary code $f : \mathbf{B}^m \rightarrow \mathbf{B}^n$ is **linear** if f is a linear transformation of vector spaces. Any linear code is given by a **generator matrix** G , which is an $m \times n$ matrix with entries from \mathbb{Z}_2 such that $f(w) = wG$ (here w is regarded as a row vector). For a systematic code, G is of the form $(I_m | A)$.

Theorem If $f : \mathbf{B}^m \rightarrow \mathbf{B}^n$ is a linear code, then

- the set W of all codewords forms a subspace (and a subgroup) of \mathbf{B}^n ;
- the zero word is a codeword;
- the minimum distance between distinct codewords is equal to the minimum weight of nonzero codewords.

Coset leaders

A received word w' is represented as $w_0 + w_1$, where w_0 is the codeword that was sent and w_1 is an **error pattern** (namely, w_1 has 1s in those positions where transmission errors occurred).

The set of words received with a particular error pattern w_1 is a coset $w_1 + W$ of the subgroup W of codewords in the group \mathbf{B}^n of all words of length n .

Error-correction is based on the following assumption. For every coset C of W , we assume that all words from C are received with the same error pattern w_C . Note that $w_C \in C$ so that $C = w_C + W$. Given a word $w \in C$, the corrected codeword is $w - w_C$ ($= w + w_C$).

The word w_C is a **coset leader**. It is chosen as the most likely error pattern, namely, the word of smallest weight in the coset C (the choice may not be unique).

Coset decoding table

Example. Generator matrix: $G = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \end{pmatrix}$

Coding function: $\left\{ \begin{array}{l} 00 \rightarrow 00000 \\ 01 \rightarrow 01011 \\ 10 \rightarrow 10110 \\ 11 \rightarrow 11101 \end{array} \right.$

Coset decoding table (each row is a coset, the 1st row consists of codewords, the 1st column consists of the coset leaders):

00000	01011	10110	11101
00001	01010	10111	11100
00010	01001	10100	11111
00100	01111	10010	11001
01000	00011	11110	10101
10000	11011	00110	01101
10001	11010	00111	01100
00101	01110	10011	11000