



Coding Theory

CO 331



Alfred Menezes

Preface

Disclaimer Much of the information on this set of notes is transcribed directly/indirectly from the lectures of CO 331 during Winter 2021 as well as other related resources. I do not make any warranties about the completeness, reliability and accuracy of this set of notes. Use at your own risk.

For any questions, send me an email via <https://notes.sibeliusp.com/contact/>.

You can find my notes for other courses on <https://notes.sibeliusp.com/>.

Sibelius Peng

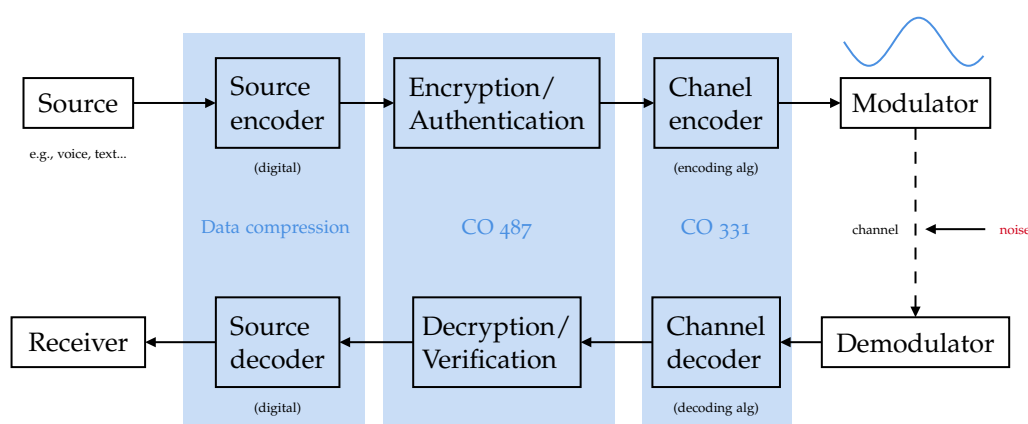
Contents

Preface	1
Introduction	3
1 Fundamentals	5
1.1 Basic Definitions and Concepts	5
1.2 Decoding Strategy	7
1.3 Error Correcting & Detecting Capabilities of a Code	9
2 Introduction to Finite Fields	11
2.1 Definitions	11
2.2 Finite fields: Non-existence	13
2.3 Existence of finite fields	14
2.4 Properties of finite fields	17
3 Linear codes	20

Introduction

Coding theory is about clever ways of adding redundancy to messages to allow (efficient) error detection and error correction.

Here is our communication model:



Example: Parity Code

Encoding algorithm Add a 0 bit to the (binary) msg m if the number of 1's in m is even; else add a 1 bit.

Decoding algorithm If the number of 1's in a received msg r is even, then accept r ; else declare that an error has occurred.

Example: Replication Code

Source msgs	Codeword	# err/codeword (always) detected	# err/codeword (always) corrected *	Information rate
0	0	0	0	1
1	1	0	0	1
0	00	1	0	$\frac{1}{2}$
1	11	1	0	$\frac{1}{2}$
0	000	2	1	$\frac{1}{3}$
1	111	2	1	$\frac{1}{3}$
0	0000	3	1	$\frac{1}{4}$
1	1111	3	1	$\frac{1}{4}$
0	00000	4	2	$\frac{1}{5}$
1	11111	4	2	$\frac{1}{5}$

encoding algorithm
→

*: using "nearest neighbour decoding"

Goal of Coding Theory

Design codes so that:

1. High information rate
2. High error-correcting capability
3. Efficient encoding & decoding algorithms

Course Overview

This course deals with *algebraic methods* for designing good (block) codes. The focus is on error correction (not on error detection). These codes are used in wireless communications, space probes, CD/DVD players, storage, QR codes, etc.

Some modern stuff are not covered: Turbo codes, LDPC codes, Raptor codes, ... Their math theories are not so elegant as algebraic codes.

The big picture

Coding theory in its broadest sense deals with techniques for the *efficient, secure* and *reliable* transmission of data over communication channels that may be subject to *non-malicious errors* (noise) and *adversarial intrusion*. The latter includes passive intrusion (eavesdropping) and active intrusion (injection/deletion/modification).

Fundamentals

1.1 Basic Definitions and Concepts

alphabet

An **alphabet** A is a finite set of $q \geq 2$ symbols.

word

A **word** is a finite sequence of symbols from A (also: vector, tuple).

length

The **length** of a word is the number of symbols it has.

code

A **code** C over A is a set of words (of size ≥ 2).

codeword

A **codeword** is a word in the code C .

block code

A **block code** is a code in which all codewords have the same length.

A **block code of length n containing M codewords over A** is a subset $C \subseteq A^n$ with $|C| = M$. C is called an $[n, M]$ -code over A .

Example:

$A = \{0, 1\}$. $C = \{00000, 11100, 00111, 10101\}$ is a $[5, 4]$ -code over $\{0, 1\}$.

Messages		Codewords
00	→	00000
10	→	11100
01	→	00111
11	→	10101

↑
Encoding of messages (1-1 map)

Assumptions about the communications channel

- (1) The channel only transmits symbols from A (“hard decision decoding”).
- (2) No symbols are deleted, added, interchanged or transposed during transmission.
- (3) The channel is a q -symmetric channel:

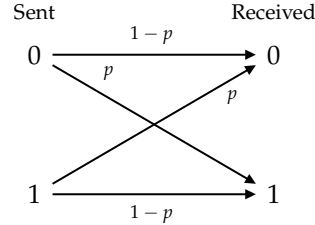
Let $A = \{a_1, \dots, a_q\}$. Let X_i = the i^{th} symbol sent. Let Y_i = the i^{th} symbol received. Then for all $i \geq 1$, and all $i \leq j, k \leq q$,

$$\Pr(Y_i = a_j | X_i = a_k) = \begin{cases} 1 - p, & \text{if } j = k \\ \frac{p}{q-1}, & \text{if } j \neq k. \end{cases}$$

p is called the **symbol error probability** of the channel ($0 \leq p \leq 1$).

Binary Symmetric Channel (BSC)

A 2-symmetric channel is called a binary symmetric channel.



For a BSC:

1. If $p = 0$, the channel is *perfect*.
2. If $p = 1/2$, the channel is *useless*.
3. If $1/2 < p \leq 1$, then flipping all received bits converts the channel to a BSC with $0 \leq p < 1/2$.
4. Henceforth, we will assume that $0 < p < 1/2$ for a BSC.

Exercise:

For a q -symmetric channel, show that one can take $0 < p < \frac{q-1}{q}$ WLOG.

One can first consider the case $q = 3$.

information rate

The **information rate** (or rate) R of an $[n, M]$ -code C over A is $R = \frac{\log_q M}{n}$.

If C encodes messages that are k -tuples over A (so $M = |A^k| = q^k$), then $R = \frac{k}{n}$.

Note:

$0 \leq R \leq 1$. Ideally, R should be close to 1.

Example:

The rate of the binary code $C = \{00000, 11100, 00111, 10101\}$ is $R = \frac{2}{5}$.

Hamming distance

The **Hamming distance** (or distance) between two n -tuples over A is the number of coordinate positions in which they differ.

The Hamming distance (or distance) of an $[n, M]$ -code C is $d(C) = \min\{d(x, y) : x, y \in C, x \neq y\}$.

Example:

The distance of $C = \{00000, 11100, 00111, 10101\}$ is $d(C) = 2$.

Theorem 1.1: properties of Hamming distance

For all $x, y, z \in A^n$,

1. $d(x, y) \geq 0$, with $d(x, y) = 0$ iff $x = y$.
2. $d(x, y) = d(y, x)$.
3. $d(x, y) + d(y, z) \geq d(x, z)$ (\triangle inequality).

1.2 Decoding Strategy

Example:

Let $C = \{00000, 11100, 00111, 10101\}$. C is a $[5, 4]$ -code over $\{0, 1\}$ (a binary code).

Error Detection If C is used for error detection only, the strategy is the following: A received word $r \in A^n$ is accepted if and only if $r \in C$.

Error Correction Let C be an $[n, M]$ -code over A with distance d . Suppose $c \in C$ is transmitted, and $r \in A^n$ is received. The (channel) decoder must decide one of the following:

- (i) No errors have occurred; *accept* r .
- (ii) Errors have occurred; *correct*¹ (decode) r to a codeword $c \in C$?
- (iii) Errors have occurred; *no correction is possible*.

Nearest Neighbour Decoding

- (i) Incomplete Maximum Likelihood Decoding (IMLD):

If there is a unique codeword $c \in C$ such that $d(r, c)$ is minimum, then correct r to c . If no such c exists, then report that errors have occurred, but correction is not possible (ask for retransmission, or disregard information).

- (ii) Complete Maximum Likelihood Decoding (CMLD):

Same as IMLD, except that if there are two or more $c \in C$ for which $d(r, c)$ is minimum, correct r to an arbitrary one of these.

¹Error correction does not guarantee that the channel decoder always makes the correct decision. For example, 00000 $\xrightarrow{\text{transmit}}$ 11100 which is accepted.

Is IMLD a reasonable strategy?

Theorem 1.2

IMLD chooses the codeword c for which the conditional probability

$$P(r|c) = P(r \text{ is received} | c \text{ is sent})$$

is largest.

Proof:

Suppose $c_1, c_2 \in C$ with $d(c_1, r) = d_1$ and $d(c_2, r) = d_2$. Suppose $d_1 > d_2$.

Now

$$P(r|c_1) = (1-p)^{n-d_1} \left(\frac{p}{q-1} \right)^{d_1}$$

and

$$P(r|c_2) = (1-p)^{n-d_2} \left(\frac{p}{q-1} \right)^{d_2}$$

So,

$$\frac{P(r|c_1)}{P(r|c_2)} = (1-p)^{d_2-d_1} \left(\frac{p}{q-1} \right)^{d_1-d_2} = \left(\frac{p}{(1-p)(q-1)} \right)^{d_1-d_2}$$

Recall

$$\begin{aligned} p < \frac{q-1}{q} &\implies pq < q-1 \implies 0 < q-pq-1 \\ \implies p < p+q-pq-1 &\implies p < (1-p)(q-1) \implies \frac{p}{(1-p)(q-1)} < 1 \end{aligned}$$

Hence

$$\frac{P(r|c_1)}{P(r|c_2)} < 1$$

and so

$$P(r|c_1) < P(r|c_2)$$

and the result follows. \square

Minimum Error Probability Decoding (MED)

An *ideal strategy* would be to correct r to a codeword $c \in C$ for which $P(c|r) = P(r \text{ is received} | c \text{ is sent})$ is largest. This is MED.

Example: (IMLD/CMLD) is not the same as MED

Consider $C = \{000, 111\}$. Suppose $P(c_1) = 0.1$ and $P(c_2) = 0.9$. Suppose $p = \frac{1}{4}$ (for a BSC).

$\begin{matrix} \uparrow & \uparrow \\ c_1 & c_2 \end{matrix}$

Suppose $r = 100$ is the received word. Then

$$P(c_1|r) = \frac{P(r|c_1) \cdot P(c_1)}{P(r)} = \frac{p(1-p)^2 \times 0.1}{P(r)} = \frac{9}{640} \cdot \frac{1}{P(r)}$$

$$P(c_2|r) = \frac{P(r|c_2) \cdot P(c_2)}{P(r)} = \frac{(1-p)p^2 \times 0.9}{P(r)} = \frac{27}{640} \cdot \frac{1}{P(r)}$$

So, MED decodes r to c_2 . But IMLD decodes r to c_1 .

IMLD vs. MED

- IMLD maximizes $P(r|c)$. MED maximizes $P(c|r)$.

- (i) MED has the drawback that the decoding algorithm depends on the probability distribution of source messages.
- (ii) If all source messages are equally likely, then CMLD and MED are equivalent:

$$P(r|c_i) = P(c_i|r) \cdot P(c_i)/P(r) = P(c_i|r) \cdot \underbrace{\left[\frac{1}{M \cdot P(r)} \right]}_{\text{does not depend on } c_i}$$

(iii) In practice IMLD (or CMLD) is used.

- In this course, we will use IMLD/CMLD.

1.3 Error Correcting & Detecting Capabilities of a Code

Detection Only

Strategy: If r is received, then accept r if and only if $r \in C$.

e-error detecting code

A code C is an **e-error detecting code** if the decoder always makes the correct decision if e or fewer errors per codeword are introduced by the channel.

Example:

Consider $C = \{000, 111\}$.

C is a 2-error detecting code.

C is not a 3-error detecting code.

Theorem 1.3

A code C of distance d is a $(d-1)$ -error detecting code (but is not a d -error detecting code).

Proof:

Suppose $c \in C$ is sent.

If no errors occur, then c is received (and is accepted).

Suppose that # of errors is ≥ 1 and $\leq d-1$; let r be the received word. Then $1 \leq d(r, c) \leq d-1$, so $r \notin C$. Thus r is rejected. This proves that it is $(d-1)$ -error detecting code.

Since $d(C) = d$, there exist $c_1, c_2 \in C$ with $d(c_1, c_2) = d$. If c_1 is sent and c_2 is received, then c_2 is accepted; the d errors go undetected. \square

Correction

Strategy: IMLD/CMLD

e-error correcting code

A code C is an **e-error correcting code** if the decoder always makes the correct decision if e or fewer errors per codeword are introduced by the channel.

Example:

Consider $C = \{000, 111\}$.

C is a 1-error correcting code.

C is not a 2-error correcting code.

Theorem 1.4

A code C of distance d is an e -error correcting code, where $e = \lfloor \frac{d-1}{2} \rfloor$.

Proof:

Suppose that $c \in C$ is sent, at most $\frac{d-1}{2}$ errors are introduced, and r is received. Then $d(r, c) \leq \frac{d-1}{2}$.

On the other hand, if c_1 is any other codeword, then

$$\begin{aligned} d(r, c_1) &\geq d(c, c_1) - d(r, c) && \triangle \text{ ineq} \\ &\geq d - \frac{d-1}{2} && \text{since } d(C) = d \\ &= \frac{d+1}{2} \\ &> \frac{d-1}{2} \geq d(r, c) \end{aligned}$$

Hence c is the unique codeword at minimum distance from r , so the decoder correctly concludes that c was sent. \square

Exercise:

Suppose $d(C) = d$, and let $e = \lfloor \frac{d-1}{2} \rfloor$. Show that C is *not* an $(e+1)$ -error correcting code.

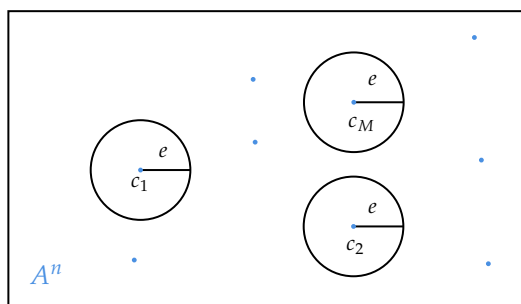
A natural question to ask is: given A, n, M, d , does there exist an $[n, M]$ -code C over A of distance $\geq d$. This can be phrased as an equivalent sphere packing problem:

Sphere packing

Can we place M spheres of radius $e = \lfloor \frac{d-1}{2} \rfloor$ in A^n so that no two spheres overlap?

$C = \{c_1, \dots, c_M\}$, $e = \lfloor \frac{d-1}{2} \rfloor$, S_c = sphere of radius e centered at c = all words within distance e of c .

We proved: if $c_1, c_2 \in C$, $c_1 \neq c_2$, then $S_{c_1} \cap S_{c_2} = \emptyset$.



Let $n = 128, q = 2, M = 2^{64}$. Does there exist a binary $[n, M]$ -code with $d \geq 22$? If so, can encoding and decoding be done efficiently?

We'll view $\{0, 1\}^{128}$ as a vector space of dimension 128 over \mathbb{Z}_2 . We'll choose C to be a 64-dimensional subspace of this vector space. We will construct such a code at the end of the course. The main tools used will be linear algebra (over finite fields) and abstract algebra (rings and fields).

Introduction to Finite Fields

2.1 Definitions

ring

A **(commutative) ring** $(R, +, \cdot)$ consists of a set R and two operations $+: R \times R \rightarrow R$ and $\cdot: R \times R \rightarrow R$, such that

1. $a + (b + c) = (a + b) + c \quad \forall a, b, c \in R.$
2. $a + b = b + a, \quad \forall a, b \in R.$
3. $\exists 0 \in R$ such that $a + 0 = a, \forall a \in R.$
4. $\forall a \in R, \exists -a \in R$ such that $a + (-a) = 0.$
5. $a \cdot (b \cdot c) = (a \cdot b) \cdot c, \quad \forall a, b, c \in R.$
6. $a \cdot b = b \cdot a, \quad \forall a, b \in R.$
7. $\exists 1 \in R, 1 \neq 0$, such that $a \cdot 1 = a, \quad \forall a \in R.$
8. $a \cdot (b + c) = a \cdot b + a \cdot c, \quad \forall a, b, c \in R.$

Notation We will denote $(R, +, \cdot)$ by R .

Example:

$\mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{Z}$ are commutative rings.

field

A **field** $(F, +, \cdot)$ is a commutative ring with the additional property:

9. $\forall a \in F, a \neq 0, \exists a^{-1} \in F$ such that $a \cdot a^{-1} = 1.$

Example:

$\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are fields. \mathbb{Z} is *not* a field.

infinite/finite field

A field $(F, +, \cdot)$ is a **finite field** if F is a finite set; otherwise it is an **infinite field**. If F is a finite field, its **order** is $|F|$.

Example:

$\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are infinite fields.

For which integers $n \geq 2$ does there *exist* a finite field of order n ? How does one *construct* such a field, i.e., what are the field elements, and how are the field operations performed?

The Integers Modulo n

Let $n \geq 2$. Recall that \mathbb{Z}_n consists of the set of equivalence classes of integers modulo n , $\mathbb{Z}_n = \{[0], [1], [2], \dots, [n-1]\}$, with addition and multiplication: $[a] + [b] = [a + b]$, $[a] \cdot [b] = [a \cdot b]$.

More simply, we write $\mathbb{Z}_n = \{0, 1, 2, \dots, n-1\}$, and perform addition and multiplication modulo n .

Example:

$\mathbb{Z}_9 = \{0, 1, 2, \dots, 8\}$. In \mathbb{Z}_9 , $3 + 7 = 1$ and $3 \cdot 7 = 3$.

More precisely, $3 + 7 \equiv 1 \pmod{9}$ and $3 \cdot 7 \equiv 3 \pmod{9}$.

\mathbb{Z}_n is a commutative ring (i.e., axioms 1-8 in the definition are satisfied).

When is \mathbb{Z}_n a field?

Theorem 2.1

\mathbb{Z}_n is a field if and only if n is prime.

Proof:

\Leftarrow Suppose n is prime. Let $a \in \mathbb{Z}_n$, $a \neq 0$ (so $1 \leq a \leq n-1$). Since n is prime, $\gcd(a, n) = 1$. Hence $\exists s, t \in \mathbb{Z}$ such that $as + nt = 1$. Reducing both sides modulo n gives $as \equiv 1 \pmod{n}$. Hence $a^{-1} = s$. Thus \mathbb{Z}_n is a field.

\Rightarrow Suppose n is composite, say $n = ab$ where $2 \leq a, b \leq n-1$.

Now, if a^{-1} exists, say $ac \equiv 1 \pmod{n}$, then $abc \equiv b \pmod{n}$, so $nc \equiv b \pmod{n}$. Thus $b \equiv 0 \pmod{n}$, so $n \mid b$ which is absurd since $2 \leq b \leq n-1$. Thus \mathbb{Z}_n is not a field.

□

We have established the existence of finite fields of order n , for each prime n . What about finite fields of order n , where n is composite? In particular, is there a field of order 4? Order 6?

characteristic

Let F be a field. The **characteristic** of F , denoted $\text{char}(F)$, is the smallest positive integer m such that $\underbrace{1 + \dots + 1}_m = 0$. If no such m exists, then $\text{char}(F) = 0$.

Example:

$\mathbb{Q}, \mathbb{R}, \mathbb{C}$ have characteristic 0. \mathbb{Z}_p (p prime) has characteristic p .

Theorem 2.2

If $\text{char}(F) = 0$, then F is an infinite field.

Proof:

The elements $1, 1+1, 1+1+1, \dots$ are distinct, because if $\underbrace{1+\dots+1}_a = \underbrace{1+\dots+1}_b$ where $a < b$, then $\underbrace{(1+\dots+1)}_b - \underbrace{(1+\dots+1)}_a = 1+\dots+1_{b-a} = 0$, contradicting $\text{char}(F) = 0$. \square

Theorem 2.3

Let F be a field with $\text{char}(F) = m \neq 0$. Then m is prime.

Proof:

Suppose m is composite, say $m = ab$ where $2 \leq a, b \leq m-1$. Let $s = \underbrace{1+\dots+1}_a$ and $t = \underbrace{1+\dots+1}_b$; note that $s, t \neq 0$. Then $s \cdot t = \underbrace{(1+\dots+1)}_a \cdot \underbrace{(1+\dots+1)}_b = \underbrace{1+\dots+1}_{ab=m} = 0$. Thus

$$s \cdot t \cdot t^{-1} = s \cdot 1 = s = 0,$$

a contradiction. We then conclude that m is prime. \square

Let F be a finite field of characteristic p . Consider the subset of elements of F :

$$E = \{0, 1, 1+1, 1+1+1, \dots, \underbrace{1+\dots+1}_{p-1}\}.$$

The elements of E are distinct. One can verify that E is a field, using the same operations as F . E is a **subfield** of F . If we identify the elements of E with the elements of \mathbb{Z}_p in the natural way, then E is essentially the same field as \mathbb{Z}_p . We have proven:

Theorem 2.4

Let F be a finite field of char p . Then \mathbb{Z}_p is a subfield of F .

Finite fields as vector spaces

Let F be a finite field of characteristic p . Identify:

vectors	\leftrightarrow	elements of F
scalars	\leftrightarrow	elements of \mathbb{Z}_p
vector addition	\leftrightarrow	addition of F
scalar multiplication	\leftrightarrow	multiplication of F

Then F is a vector space over \mathbb{Z}_p (i.e., the axioms of what it means to be a vector space are satisfied).

2.2 Finite fields: Non-existence**Theorem 2.5**

Let F be a finite field of characteristic p . Then the order of F is p^n , for some positive integer n .

Proof:

Let the dimension of F as a vector space over \mathbb{Z}_p be n . Let $\alpha_1, \dots, \alpha_n$ be a basis for F over \mathbb{Z}_p . Then each element $\beta \in F$ can be written uniquely in the form $\beta = c_1\alpha_1 + \dots + c_n\alpha_n$, where $c_i \in \mathbb{Z}_p$. Thus

$$F = \left\{ \sum_{i=1}^n c_i \alpha_i : c_i \in \mathbb{Z}_p \right\}, \text{ so } |F| = p^n. \quad \square$$

For example, there do not exist finite fields of order 6, 10, 12, 14, 15, ...

Do finite fields of orders 4, 8, 9, 16, 25, 27, ... exist?

2.3 Existence of finite fields

Polynomial rings Let F be a field. $F[x]$ denotes the set of all polynomials in x with coefficients from F . Addition and multiplication of polynomials in $F[x]$ is done in the usual way, with coefficient arithmetic done in F .

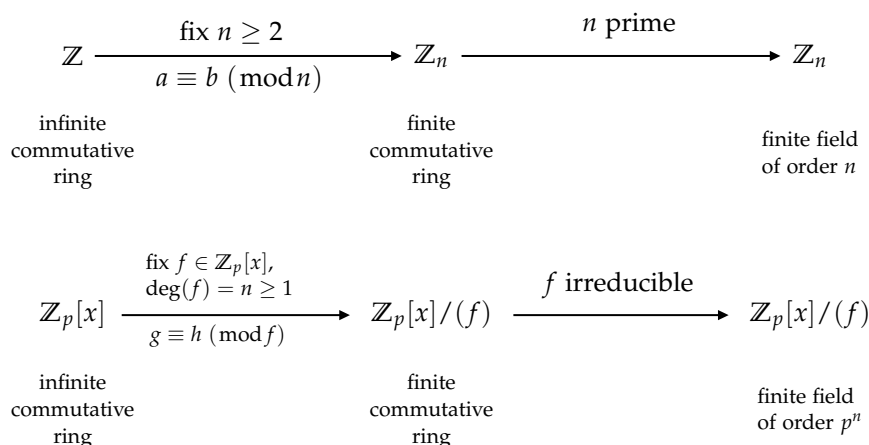
Example:

In $\mathbb{Z}_5[x]$,

$$\begin{aligned}(3x^4 + 2x^3 + x + 4) + (x^5 + 2x^4 + x^2 + 2x + 3) &= x^5 + 2x^3 + x^2 + 3x + 2 \\ (3x^2 + 4x + 1) \cdot (2x^2 + x + 2) &= (x^4 + x^3 + 2x^2 + 4x + 2).\end{aligned}$$

Note that $F[x]$ is an infinite commutative ring.

Construction of finite fields: main idea



Polynomial division

Let $f, g \in F[x]$, with $g \neq 0$. Then there exist unique polynomials $r, s \in F[x]$ such that $f = \overset{\substack{\uparrow \\ \text{quotient}}}{s}g + \overset{\substack{\uparrow \\ \text{remainder}}}{r}$, $\deg(r) < \deg(g)$. By convention, $\deg(0) = -\infty$.

Example:

Consider $f = 3x^4 + 2x^2 + x + 1, g = 2x^2 + 3x + 4 \in \mathbb{Z}_5[x]$.

$$\begin{array}{r} \overline{4x^2 } \\ 2x^2 + 3x + 4 \overline{) 3x^4 + 2x^3 + 2x^2 + x + 1} \\ \underline{3x^4 + 2x^3 + x^2} \\ x^2 + x + 1 \\ \underline{x^2 + 4x + 2} \\ 2x + 4 \end{array}$$

So, $f = (4x^2 + 3)g + (2x + 4)$.

The ring $F[x]/(f)$

$$g \equiv h \pmod{f}$$

Let $f \in F[x]$ with $\deg(f) \geq 1$. Let $g, h \in F[x]$. Then g is congruent to h modulo f , written $g \equiv h \pmod{f}$, if $g - h = \ell f$ for some $\ell \in F[x]$ (equivalently, $f \mid (g - h)$, or g, h leave the same remainder upon division by f).

The relation $\equiv \pmod{f}$ is an equivalence relation, and partitions $F[x]$ into equivalence classes:

$$[g] = \{h \in F[x] : g \equiv h \pmod{f}\}.$$

Addition & multiplication: $[g] + [h] = [g + h]$, $[g] \cdot [h] = [g \cdot h]$.

$$F[x]/(f)$$

The set of equivalence classes is denoted $F[x]/(f)$.

Theorem 2.6

$F[x]/(f)$ is a commutative ring.

Suppose now that $\deg(f) = n$. Let $g \in F[x]$. Then we can write $g = sf + r$, where $s, r \in F[x]$, and $\deg(r) < n$. Thus $g \equiv r \pmod{f}$, so $[g] = [r]$.

If $r_1, r_2 \in F[x]$, $r_1 \neq r_2$, $\deg(r_1), \deg(r_2) < n$, then $f \nmid (r_1 - r_2)$, so $r_1 \not\equiv r_2 \pmod{f}$. Thus $[r_1] \neq [r_2]$.

Thus the polynomials in $F[x]$ of degree $< n$ are a complete set of representatives of the equivalence classes of $F[x]/(f)$.

Now, let $F = \mathbb{Z}_p$. Then $\mathbb{Z}_p[x]/(f) = \{[r] : r \in \mathbb{Z}_p[x], \deg(r) < n\}$. Thus $|\mathbb{Z}_p[x]/(f)| = p^n$, so $\mathbb{Z}_p[x]/(f)$ is a commutative ring of order p^n .

When is $F[x]/(f)$ a field?

irreducible over F

Let $f \in F[x]$, with $\deg(f) \geq 1$. Then **irreducible over F** if f cannot be written as $f = g \cdot h$, $g, h \in F[x]$, $\deg(g) \geq 1, \deg(h) \geq 1$.

Example:

$x^2 + 1$ is irreducible over \mathbb{R} , since it has no roots in \mathbb{R} .

$x^2 + 1$ is reducible over \mathbb{C} , since $x^2 + 1 = (x + i)(x - i)$.

$x^2 + 1$ is reducible over \mathbb{Z}_2 , since $x^2 + 1 = (x + 1)(x + 1)$.

$x^2 + 1$ is irreducible over \mathbb{Z}_3 , since it has no roots in \mathbb{Z}_3 .

Theorem 2.7

$F[x]/(f)$ is a field if and only if f is irreducible over F .

Proof:

Analogous to the proof of the theorem: \mathbb{Z}_n is a field if and only if n is prime. □

Now let's construct finite fields.

Theorem 2.8

Let $f \in \mathbb{Z}_p[x]$ be an irreducible polynomial of degree $n \geq 1$. Then $\mathbb{Z}_p[x]/(f)$ is a finite field of order p^n and characteristic p . The elements are the polynomials in $\mathbb{Z}_p[x]$ of degree $< n$.

Example: finite field of order $4 = 2^2$

Here $p = 2$ and $n = 2$. Let $f(x) = x^2 + x + 1 \in \mathbb{Z}_2[x]$. Then $f(0) = 1$, $f(1) = 1$, so f has no roots in \mathbb{Z}_2 . Thus f is irreducible over \mathbb{Z}_2 .

So, $F = \mathbb{Z}_2[x]/(x^2 + x + 1)$ is a finite field of order $2^2 = 4$. The elements are $\{0, 1, x, x + 1\}$ where $[\]$ is omitted.

Example of addition: $x + (x + 1) = 1$

Example of multiplication: $x \cdot (x + 1) = x^2 + x = 1$

Example: finite field of order $2^3 = 8$

Here, $p = 2$ and $n = 3$. We need an irreducible polynomial. We need an irreducible polynomial of degree 3 over \mathbb{Z}_2 .

Candidates: $x^3, x^3 + 1, x^3 + x, x^3 + x + 1, x^3 + x^2, x^3 + x^2 + 1, x^3 + x^2 + x, x^3 + x^2 + x + 1$

- Try $f(x) = x^3 + x + 1$.

Since $f(0) = f(1) = 1$, f has no roots in \mathbb{Z}_2 , and thus no linear factors in $\mathbb{Z}_2[x]$. Thus f is irreducible over \mathbb{Z}_2 , and $F_1 = \mathbb{Z}_2[x]/(x^3 + x + 1)$ is a finite field of order $2^3 = 8$.

The elements of F_1 are $\{0, 1, x, x + 1, x^2, x^2 + 1, x^2 + x, x^2 + x + 1\}$.

Example of addition: $(x^2 + x) + (x^2 + x + 1) = 1$.

Example of multiplication: $(x^2 + x) \cdot (x^2 + x + 1) = x^4 + x = x^2$.

Example of inversion: $x^{-1} = x^2 + 1$, since $x \cdot (x^2 + 1) = 1$.

- $x^3 + x^2 + 1$ is irreducible over \mathbb{Z}_2 , so $F_2 = \mathbb{Z}_2[x]/(x^3 + x^2 + 1)$ is a finite field of order 8. The elements of F_2 are $\{0, 1, x, x + 1, x^2, x^2 + 1, x^2 + x, x^2 + x + 1\}$.

Note that F_1 and F_2 are not the same field. For example, in F_1 , $x \cdot x^2 = x + 1$, whereas in F_2 , $x \cdot x^2 = x^2 + 1$. However, F_1 and F_2 are isomorphic (essentially the same). Formally, there is a bijection $\phi : F_1 \rightarrow F_2$ such that $\phi(a + b) = \phi(a) + \phi(b)$ and $\phi(a \cdot b) = \phi(a) \cdot \phi(b) \forall a, b \in F_1$.

Existence and uniqueness of finite fields

Let p be prime and $n \geq 1$. Then there exists an irreducible polynomial of degree n over \mathbb{Z}_p .

Theorem 2.9

There exists a finite field of order q if and only if $q = p^n$ for some prime p and $n \geq 1$.

Actually, any two finite fields of the same order are isomorphic.

We will denote *the* finite field of order q by $\text{GF}(q)$ “the Galois Field of order q ”.

In the previous example, we saw two ways of representing the finite field $\text{GF}(2^3)$.

2.4 Properties of finite fields

Theorem 2.10: Frosh's dream

Let F be a finite field of characteristic p , and let $\alpha, \beta \in F$. Then $(\alpha + \beta)^{p^m} = \alpha^{p^m} + \beta^{p^m} \forall m \geq 1$.

Proof ($m = 1$):

By the Binomial Theorem,

$$(\alpha + \beta)^p = \binom{p}{0}\alpha^p + \sum_{i=1}^{p-1} \binom{p}{i}\alpha^i\beta^{p-i} + \binom{p}{p}\beta^p.$$

Now for $1 \leq i \leq p-1$,

$$\binom{p}{i} = \frac{p(p-1)(p-2)\cdots(p-i+1)}{1 \cdot 2 \cdot 3 \cdots i} \equiv 0 \pmod{p},$$

since p divides the numerator but not the denominator, and since $\binom{p}{i}$ is an integer. Thus

$$\binom{p}{i}\alpha^i\beta^{p-i} = \underbrace{\alpha^i\beta^{p-i} + \cdots + \alpha^i\beta^{p-i}}_{\binom{p}{i}} = (1 + \cdots + 1)\alpha^i\beta^{p-i} = 0.$$

Hence $(\alpha + \beta)^p = \alpha^p + \beta^p$. The statement for $m \geq 1$ can be proven by induction. \square

The multiplicative group $\text{GF}(q)^*$

multiplicative group of $\text{GF}(q)$

The **multiplicative group** of $\text{GF}(q)$ is $\text{GF}(q)^* = \text{GF}(q) \setminus \{0\}$.

Theorem 2.11

Let $\alpha \in \text{GF}(q)^*$. Then $\alpha^{q-1} = 1$.

Note that if $\text{GF}(q) = \mathbb{Z}_p$, this is Fermat's Little Theorem.

Proof:

Let the (distinct) elements of $\text{GF}(q)^*$ be $\alpha_1, \alpha_2, \dots, \alpha_{q-1}$. Consider the (nonzero) elements $\alpha\alpha_1, \alpha\alpha_2, \dots, \alpha\alpha_{q-1}$. These elements are distinct because if $\alpha\alpha_i = \alpha\alpha_j$ for some $i \neq j$, then $\alpha^{-1}(\alpha\alpha_i) = \alpha^{-1}(\alpha\alpha_j)$, so $\alpha_i = \alpha_j$, a contradiction. Hence $\{\alpha\alpha_1, \alpha\alpha_2, \dots, \alpha\alpha_{q-1}\} = \{\alpha_1, \alpha_2, \dots, \alpha_{q-1}\}$, so

$$(\alpha\alpha_1)(\alpha\alpha_2) \cdots (\alpha\alpha_{q-1}) = \alpha_1\alpha_2 \cdots \alpha_{q-1}.$$

Cancelling gives $\alpha^{q-1} = 1$. \square

Corollary 2.12

Let $\alpha \in \text{GF}(q)$. Then $\alpha^q = \alpha$.

Order of finite elements

order of α

Let $\alpha \in \text{GF}(q)^*$. The **order of** α , denoted $\text{ord}(\alpha)$, is the smallest positive integer t such that $\alpha^t = 1$.

Theorem 2.13

Let $\alpha \in \text{GF}(q)^*$, $\text{ord}(\alpha) = t$. Then $\alpha^s = 1$ if and only if $t \mid s$.

Proof:

Let $s \in \mathbb{Z}$. Then long division of s by t yields

$$s = \ell t + r, \quad \text{where } 0 \leq r < t.$$

Now, $\alpha^s = \alpha^{\ell t + r} = (\alpha^t)^\ell \cdot \alpha^r = \alpha^r$. Hence $\alpha^s = 1 \iff \alpha^r = 1 \iff r = 0 \iff t \mid s$. \square

Corollary 2.14

Let $\alpha \in \text{GF}(q)^*$. Then $\text{ord}(\alpha) \mid (q - 1)$.

Example:

There is only one element in $\text{GF}(q)$ of order 1, namely the element 1.

Example:

Consider $\text{GF}(2^3) = \mathbb{Z}_2[x]/(x^3 + x + 1)$. The order of $\alpha = x^2 + 1$ is 7.

Example:

Consider $\text{GF}(2^4) = \mathbb{Z}_2[x]/(x^4 + x + 1)$.

$f(x) = x^4 + x + 1$ has no roots in \mathbb{Z}_2 , thus no linear factors. Also, $f(x)$ has no irreducible quadratic factors, since $(x^2 + x + 1) \nmid f(x)$. Thus f is irreducible over \mathbb{Z}_2 .

Find $\text{ord}(x)$ in $\text{GF}(2^4)$.

Solution: We have $x^1 = x, x^2 = x^2, x^4 = x + 1, x^5 = x^2 + x \neq 1$. Thus $\text{ord}(x) \neq 1, 3, 5$. Since $\text{ord}(x) \mid 15$, we must have $\text{ord}(x) = 15$.

Let $\alpha \in \text{GF}(q)^*$ with $\text{ord}(\alpha) = t$. Then the elements $\alpha^0, \alpha^1, \alpha^2, \dots, \alpha^{t-1}$ are distinct. In particular, if $\text{ord}(\alpha) = q - 1$, then $\text{GF}(q)^* = \{\alpha^0, \alpha^1, \alpha^2, \dots, \alpha^{q-2}\}$.

generator

A **generator** of $\text{GF}(q)^*$ is an element of order $q - 1$.

Example:

$\alpha = x$ is a generator of $\text{GF}(2^4) = \mathbb{Z}_2[x]/(x^4 + x + 1)$ since $\text{ord}(x) = 15$. Let's verify the above fact:

$$\begin{array}{llll} x^0 = 1, & x^1 = x, & x^2 = x^2, & x^3 = x^3, \\ x^4 = x + 1, & x^5 = x^2 + x, & x^6 = x^3 + x^2, & x^7 = x^3 + x + 1, \\ x^8 = x^2 + 1, & x^9 = x^3 + x, & x^{10} = x^2 + x + 1, & x^{11} = x^3 + x^2 + x, \\ x^{12} = x^3 + x^2 + x + 1, & x^{13} = x^3 + x^2 + x + 1, & x^{14} = x^3 + 1, & x^{15} = 1. \end{array}$$

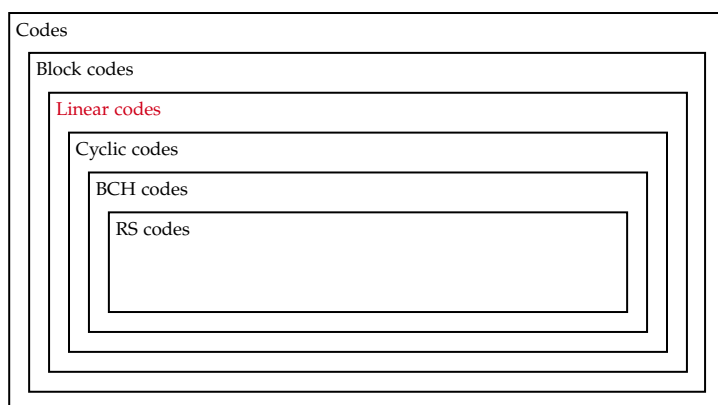
Theorem 2.15

Every finite field $\text{GF}(q)$ has a generator.

Proof:

Optional. □

Linear codes



Index

A

alphabet 5

B

block code 5

C

characteristic 12
code 5
codeword 5

E

e-error correcting code 9
e-error detecting code 9

F

field 11
finite field 12

G

generator 18

H

Hamming distance 7

I

infinite field 12
information rate 6
irreducible over F 15

L

length 5

M

multiplicative group of $\text{GF}(q)$ 17

O

order of α 18
order of a field 12

R

ring 11

S

subfield 13
symbol error probability 6

W

word 5