University of Victoria
Notes for Math 413:
Applied Algebra

Peter Dukes

January 1, 2020

Contents

Part I

Finite Fields

Chapter 1

Introduction

Our notes begin by rougly following Chapter 3 of *Applied Abstract Algebra*, by Lidl and Pilz, [2]. The goal is to quickly obtain structural information on finite fields, assuming an introductory course in abstract algebra. Hungerford's text [1] is an excellent reference on the necessary background in algebra, as well as for additional depth and extra topics.

1.1 Integers modulo n

The ring of integers modulo n, written $\mathbb{Z}/n\mathbb{Z}$ and sometimes abbreviated $\mathbb{Z}_n$, is the set $\{0, 1, \ldots, n-1\}$ together with operations of addition and multiplication mod n. The additive group of $\mathbb{Z}/n\mathbb{Z}$ is a cyclic group with n elements, and we denote it by $(\mathbb{Z}/n\mathbb{Z})^+$.

Concerning the multiplicative structure, an element $a \in \mathbb{Z}/n\mathbb{Z}$ is a *unit* or is *invertible* if there exists $b \in \mathbb{Z}/n\mathbb{Z}$ such that $ab \equiv 1 \pmod{n}$.

Proposition 1.1. *An element $a \in \mathbb{Z}/n\mathbb{Z}$ is invertible if and only if* $\gcd(a, n) = 1$.

Proof. Suppose there exists $b \in \mathbb{Z}/n\mathbb{Z}$ such that $ab \equiv 1 \pmod{n}$. Then, for some integer t, we have $ab + nt = 1$. It follows that if $d \mid a$ and $d \mid n$, we have $d \mid 1$, So $\gcd(a, n) = 1$.

Conversely, suppose $\gcd(a, n) = 1$. Using Bézout's lemma, find integers s, t such that $as + nt = 1$. Letting b be the least residue of s modulo n, we have found $b \in \mathbb{Z}/n\mathbb{Z}$ such that $ab \equiv as \equiv as + nt = 1 \pmod{n}$. □

Note that a product of invertible elements is invertible; its inverse is just the product of inverses of its factors. So the set of invertible elements of $\mathbb{Z}/n\mathbb{Z}$ forms an abelian group under multiplication, sometimes called the *group of units mod* n, and denoted $(\mathbb{Z}/n\mathbb{Z})^\times$, or U_n for short.

Define $\phi(n)$ to be the number of positive integers $a \leq n$ satisfying $\gcd(a, n) = 1$. Then $|(\mathbb{Z}/n\mathbb{Z})^\times| = \phi(n)$. The function $\phi(n)$ is called *Euler's totient* or the *Euler-phi function*. It is clear that $\phi(p) = p - 1$ for primes p. More generally, $\phi(p^k) = p^k - p^{k-1}$ because there are exactly p^{k-1} multiples of p in the relevant range. Also, an integer a is relatively prime to n if and only if it is relatively prime to every prime divisor of n. Putting these together, we have the following identity.

Proposition 1.2. *For a positive integer n, we have $\phi(n) = n \prod_{p|n} (1 - \frac{1}{p})$.*

Example 1.3. We compute $\phi(12) = 12(\frac{1}{2})(\frac{2}{3}) = 4$. The group of units mod 12 is $U_{12} = \{1, 5, 7, 11\}$, and is isomorphic to the Klein 4-group $\mathbb{Z}_2 \oplus \mathbb{Z}_2$.

Recall that the order of every element of a group divides the group order. Applying this to the group of units mod n, we obtain a useful fact.

Theorem 1.4 (Euler's theorem)**.** *If $\gcd(x, n) = 1$ then $x^{\phi(n)} \equiv 1 \pmod{n}$.*

The RSA cipher

Let $n = pq$, where p and q are large primes. Each user gets a key pair of integers $k = (a, b)$ with $a, b > 1$ and $ab \equiv 1 \pmod{\phi(n)}$. Note that $\phi(n) = (p-1)(q-1) = n - p - q + 1$. Alice sends an encrypted message $x \in \mathbb{Z}/n\mathbb{Z}$ to Bob by exponentiation using Bob's public key: $E_k(x) = x^b \pmod{n}$. Bob decrypts messages by doing the same with his private key: $D_k(y) = y^a \pmod{n}$.

Proposition 1.5. *With functions defined as above, $D_k \circ E_k$ is the identity map on $\mathbb{Z}/n\mathbb{Z}$.*

Proof. It is clear from the definition that $D_k(E_k(0)) = D_k(0) = 0$. Let $x \in \mathbb{Z}/n\mathbb{Z}$, $x \neq 0$. If $\gcd(n, x) = 1$, then $D_k(E_k(x)) = (x^b)^a = x^{ab} \equiv x \pmod{n}$ by Euler's theorem. Assume now that $\gcd(n, x) = p$. Then $D_k(E_k(x)) \equiv 0 \pmod{p}$. And we have $\gcd(x, q) = 1$, so $D_k(E_k(x)) = x^{ab} = x \cdot x^{q-1} \equiv x \cdot 1 \pmod{q}$. It follows that $D_k(E_k(x)) \equiv x \pmod{pq}$ by the Chinese remainder theorem. The case with p and q interchanged is similar. □

1.2 Fields

A *field* is a triple $(\mathbb{F}, +, \times)$, where:

- $\mathbb{F}$ is an abelian group under $+$ (addition) with identity 0;
- $\mathbb{F} \setminus \{0\}$ is an abelian group under $\times$ (multiplication) with identity 1; and
- the distributive law(s) for rings hold: $a(b+c) = ab + ac$ for all $a, b, c \in \mathbb{F}$.

Various familiar laws can be proved from these properties, such as $0 \times a = a \times 0 = 0$ for all $a \in \mathbb{F}$.

The additive and multiplicative groups of $\mathbb{F}$ are here denoted $\mathbb{F}^+$ and $\mathbb{F}^\times$, respectively. Note that multiplication was written with juxtaposition in the distributive laws. This is quite standard and will continue to be used in what follows.

Example 1.6. $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ (the sets of rational, real, and complex numbers) are fields with the usual operations.

From our remarks in the previous section, we obtain instances of (finite) fields whose cardinalities are prime.

Proposition 1.7. *Let p be a prime. The integers mod p, $\mathbb{Z}/p\mathbb{Z}$, form a field under the usual operations.*

The field $\mathbb{Z}/p\mathbb{Z}$ is commonly denoted by $\mathbb{F}_p$ or $\mathrm{GF}(p)$.

Example 1.8. The smallest possible field $\mathbb{F}_2$ has two elements $\{0, 1\}$. The additive group $(\mathbb{F}_2)^+$ is the usual binary cyclic group and the multiplicative group $(\mathbb{F}_2)^\times$ is the trivial group.

$+$	0	1
0	0	1
1	1	0

and

$\times$	1
1	1

Example 1.9. The field $\mathbb{F}_5$ has elements $\{0, 1, 2, 3, 4\}$ and operation tables as shown.

$+$	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

and

$\times$	1	2	3	4
1	1	2	3	4
2	2	4	1	3
3	3	1	3	2
4	4	3	2	1

The *characteristic* of a field $\mathbb{F}$ is the least positive integer n such that

$$n \cdot 1 = \overbrace{1 + \cdots + 1}^{n} = 0,$$

or 0 if no such n exists. Since fields have no zero-divisors, the only possible positive characteristics are prime integers. Of course, the characteristic of $\mathbb{F}_p$ is p, while the fields $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ each have characteristic zero.

Recall that the field $\mathbb{C}$ of complex numbers can be constructed from $\mathbb{R}$ by appending the new element i satisfying the relation $i^2 = -1$. The notation $\mathbb{C} = \mathbb{R}(i)$ is often used. Something similar can be done to construct finite fields of non-prime order.

Example 1.10. Consider $\mathbb{F}_3(i) = \{a + bi : a, b \in \mathbb{F}_3\}$, where $i^2 = -1 \equiv 2 \pmod 3$. We have $|\mathbb{F}_3(i)| = 9$ and, to illustrate the arithmetic,

$$(2 + i) + (2 + 2i) = (2 + 2) + (1 + 2)i \equiv 1 + 0i = 1,$$

$$(1 + 2i)^2 = 1 + 4i + 4i^2 = 1 + i - 4 = i,$$

and

$$(1 + i)^{-1} = 2^{-1}(1 - i) = 2 + i.$$

Remark. The above construction works for $\mathbb{F}_p(i)$ for a general prime p. However, for existence of inverses in $\mathbb{F}_p(i)$, it is necessary that -1 be a non-square in $\mathbb{F}_p$. This occurs if and only if $p \equiv 3 \pmod 4$.

Whether we append it to $\mathbb{R}$ or some $\mathbb{F}_p$, one can think of the imaginary unit i as a zero of the polynomial $x^2 + 1$. But note that we are not limited to this particular polynomial.

Example 1.11. Consider $\mathbb{F}_2(\alpha)$, where α is a new symbol satisfying $\alpha^2 + \alpha + 1 = 0$. We obtain a field on four elements whose operation tables are below.

$+$	0	1	α	$\alpha+1$
0	0	1	α	$\alpha+1$
1	1	0	$\alpha+1$	α
α	α	$\alpha+1$	0	1
$\alpha+1$	$\alpha+1$	α	1	0

$\times$	1	α	$\alpha+1$
1	1	α	$\alpha+1$
α	α	$\alpha+1$	1
$\alpha+1$	$\alpha+1$	1	α

We can similarly obtain a field on 8 elements as $\mathbb{F}_2(\alpha)$, where $\alpha^3 + \alpha + 1 = 0$. As we shall see, whenever we have an irreducible polynomial of degree k in $\mathbb{F}_p[x]$, we can append a zero of it to $\mathbb{F}_p$ to construct a field whose cardinality (also called its *order*) is of the form p^k, an integer power of a prime. In what follows, we show that these are the only possible finite fields, and we examine the structure of such fields in more detail.

1.3 Prime fields and extensions

We say that $\mathbb{F}$ is a *subfield* of $\mathbb{K}$ if $\mathbb{F} \subseteq \mathbb{K}$ and $\mathbb{F}$ is closed under the two operations of $\mathbb{K}$. Alternatively, $\mathbb{K}$ is an *extension* of $\mathbb{F}$.

Example 1.12. $\mathbb{Q}$ is a subfield of $\mathbb{R}$, which is in turn a subfield of $\mathbb{C}$.

Example 1.13. The field $\mathbb{F}_3(i)$ is an extension of $\mathbb{F}_3$.

Be careful: even though $\{0, 1\} \subseteq \mathbb{F}$ for every field $\mathbb{F}$, this does not mean $\mathbb{F}_2$ is a subfield of every field. Indeed, $\mathbb{F}_2$ is a subfield of precisely the fields of characteristic 2.

A *prime field* is a field with no proper subfields. If one takes the intersection of all subfields of a field $\mathbb{F}$, the result is the unique prime subfield of $\mathbb{F}$. The following result classifies the prime fields.

Theorem 1.14. *Up to isomorphism, the only prime fields are $\mathbb{Q}$ and $\mathbb{F}_p$, p prime.*

Proof. Let P be a prime field with identity 1. Define $D = \{n \cdot 1 : n \in \mathbb{Z}\} \subset P$. The mapping $\psi : \mathbb{Z} \to D$ defined by $\psi : n \mapsto n \cdot 1$ is a ring epimorphism (onto homomorphism) of $\mathbb{Z}$ onto D.

Case 1: If $\ker \psi = \{0\}$, then ψ is an isomorphism. Therefore, P is isomorphic to the smallest field containing $\mathbb{Z}$, namely its field of fractions $\mathbb{Q}$.

Case 2: If $\ker \psi \neq \{0\}$, then $\ker \psi$ is a nontrivial ideal of $\mathbb{Z}$. The integers are a PID, so $\ker \psi = \langle n \rangle$ for some positive $n \in \mathbb{Z}$. (This n is the characteristic of P.) So by the first ring isomorphism theorem, D is isomorphic to $\mathbb{Z}/\langle n \rangle = \mathbb{Z}/n\mathbb{Z}$. Since P is a field, it follows that n is prime, say p and $P = D = \mathbb{Z}/p\mathbb{Z}$ since P is prime. □

To summarize: if $\mathbb{F}$ has characteristic zero, then its prime field is $\mathbb{Q}$; otherwise, $\mathbb{F}$ has prime characteristic p, and its prime subfield is $\mathbb{Z}/p\mathbb{Z}$.

If $\mathbb{K}$ is an extension of $\mathbb{F}$, then it is a vector space over $\mathbb{F}$. The 'vector addition' takes place in $\mathbb{K}$, and 'scalar multiplication' by elements of $\mathbb{F}$ acts on $\mathbb{K}$. (The fact that elements of $\mathbb{K}$ admit a multiplication is ignored here.)

The dimension of $\mathbb{K}$ over $\mathbb{F}$ is $[\mathbb{K} : \mathbb{F}]$, called the *degree* of the field extension. An extension of degree n is also sometimes written

$$\begin{array}{c} \mathbb{K} \\ \Big| \, n \\ \mathbb{F} \end{array}$$

Example 1.15. We have $[\mathbb{C} : \mathbb{R}] = 2$ with one $\mathbb{R}$-basis for $\mathbb{C}$ being $\{1, i\}$. On the other hand, $\mathbb{R}$ is an infinite-degree extension of $\mathbb{Q}$.

Theorem 1.16 (Moore, 1896)**.** *If $\mathbb{F}$ is a finite field of characteristic p, then $|\mathbb{F}| = p^n$, where $n = [\mathbb{F} : \mathbb{F}_p]$.*

Proof. When considered as a vector space over its prime field $\mathbb{F}_p$, $\mathbb{F}$ contains a basis of n elements. Each element of $\mathbb{F}$ can be expressed as a unique linear combination of the n basis elements with coefficients in $\mathbb{F}_p$. In counting these linear combinations, there are independently p choices for each coefficient. Therefore, $|\mathbb{F}| = p^n$. □

A finite field of order $q = p^n$ is written $\mathbb{F}_q$, or sometimes $GF(q)$.

Corollary 1.17. *The additive group $\mathbb{F}_{p^n}^+$ is isomorphic to the 'elementary' abelian group*

$$\overbrace{\mathbb{Z}_p \oplus \cdots \oplus \mathbb{Z}_p}^{n}.$$

Remark. This also follows from the fact that $x \mapsto \frac{b}{a}x$ is an additive automorphism of $\mathbb{F}^+$ for any nonzero $a, b \in \mathbb{F}$. So all nonzero elements have the same order, p, and the only possibility is that $\mathbb{F}^+$ is an elementary abelian p-group.

Example 1.18. For $p \equiv 3 \pmod 4$, the field $\mathbb{F}_p(i)$ has order p^2 and a basis over its prime field $\mathbb{F}_p$ is $\{1, i\}$.

Example 1.19. A basis for $\mathbb{F}_8 = \mathbb{F}_2(\alpha)$, $\alpha^3 + \alpha + 1 = 0$, over $\mathbb{F}_2$ is $\{1, \alpha, \alpha^2\}$.

1.4 The multiplicative group of a finite field

Having now essentially fully covered the additive structure of finite fields, we turn to the multiplicative structure. First, we will need to recall that the polynomial ring $\mathbb{F}[x]$ is Euclidean; it follows that a polynomial of degree d over a field has at most d zeros.

Theorem 1.20. *The multiplicative group $\mathbb{F}_q^\times$ is cyclic (of order $q - 1$).*

Proof. Suppose $q > 2$ to avoid triviality. Write $q-1 = p_1^{r_1} p_2^{r_2} \cdots p_m^{r_m}$ is the unique prime factorization of $q-1$. Then the polynomial $f_i = x^{(q-1)/p_i} - 1$ has at most $(q-1)/p_i$ roots in $\mathbb{F}_q$. Pick a nonzero element a_i of $\mathbb{F}_q$ which is not a root of f_i. Put $b_i = a_i^{(q-1)/p_i^{r_i}}$. Then $b_i^{p_i^{r_i}} = 1$, and so $|b_i| \mid p_i^{r_i}$. From this it follows that $|b_i| = p_i^{s_i}$ for some integer $0 \leq s_i \leq r_i$. But by definition of a_i, we have $b_i^{p_i^{r_i-1}} = a_i^{(q-1)/p_i} \neq 1$. So $|b_i| = p_i^{r_i}$.

Now we claim that $b = b_1 b_2 \cdots b_m$ has order $q-1$. If not, b has order (wlog) dividing $(q-1)/p_1$. So

$$1 = b^{(q-1)/p_1} = b_1^{(q-1)/p_1} b_2^{(q-1)/p_1} \cdots b_m^{(q-1)/p_1} = b_1^{(q-1)/p_1},$$

since for $2 \leq j \leq m$ we have $b_j^{(q-1)/p_1} = 1$. This is a contradiction. □

By Lagrange's theorem, every element of $\mathbb{F}_q$ is a root of the polynomial equation $x^q - x = 0$ in $\mathbb{F}_p[x]$. Conversely, we see later that these q roots are distinct and form a field. This guarantees existence of a field of order q for every prime power q.

Multiplicative subgroups

Suppose $\mathbb{F}_q$ is a field of order q. For any divisor d of $q-1$, there is a unique subgroup of $\mathbb{F}_q^\times$ of *index* d, that is, of order $(q-1)/d$. Letting g denote a generator of $\mathbb{F}_q^\times$, we have $\langle g^d \rangle$ of index d in $\mathbb{F}_q^\times$. In particular, for odd q, the set of squares in $\mathbb{F}_q^\times$ is a subgroup of index two.

Example 1.21. Consider $\mathbb{F}_7$. It can be checked that 3 generates $\mathbb{F}_7^\times$. Since $3^2 = 2$ in $\mathbb{F}_7$, the set of nonzero squares is $\langle 2 \rangle = \{1, 2, 4\}$. Now consider $\mathbb{F}_{49} = \mathbb{F}_7(i)$. To find an element of order 24, it suffices to find an element of order 8 and an element of order 3, and take their product. As before, $2 \in \mathbb{F}_7(i)$ has order 3. To find an element of order 8, it is enough to find an element z satisfying $z^2 = i$. Letting $z = a + bi$ for $a, b \in \mathbb{F}_7$, we obtain the equations $a^2 - b^2 \equiv 0 \pmod 7$ and $2ab \equiv 1 \pmod 7$. By inspection, $a = b = 2$ is a solution. So the nonzero squares are generated by $2(2 + 2i) = 4 + 4i$.

The element -1 is a square in $\mathbb{F}_q$ if and only if $-1 = g^{(q-1)/2}$ is a power of g^2; that is, -1 is a square if and only if $q \equiv 1 \pmod 4$. When -1 is not a square, the set of squares has an interesting arithmetic structure. A *difference set* in an additive group G is a subset $D \subseteq G$ with the property that every element of $G \setminus \{0\}$ occurs equally often a difference of two distinct element of D. When $|G| = n$, $|D| = k$, and every element nonzero element of

G is a difference exactly λ times, the difference set is labelled by the triple of parameters (n, k, λ). The difference set condition implies $k(k-1) = \lambda(n-1)$.

Proposition 1.22. *Let $q \equiv 3 \pmod 4$ be a prime power. The set of squares in $\mathbb{F}_q^\times$ is a $(q, (q-1)/2, (q-3)/4)$-difference set in $\mathbb{F}_q^+$.*

Proof. Let g be a generator of $\mathbb{F}_q$, and put $D = \langle g^2 \rangle$, the set of squares in $\mathbb{F}_q^\times$. Since $|D| = (q-1)/2$, we need only check that every element of $\mathbb{F}_q \setminus \{0\}$ occurs equally often as a difference of squares. We can write $\mathbb{F}_q = \pm D$ since -1 is a non-square, and hence $-D$ is the set of all non-squares. Then, for each fixed $t = 1, 2, \ldots, (q-3)/4$, we have

$$\pm\{g^{2s+2t} - g^{2s} : s = 0, 1, \ldots, (q-3)/4\} = \pm(g^{2t} - 1)D = \mathbb{F}^q \setminus \{0\}.$$

It follows that every element of $\mathbb{F}_q \setminus \{0\}$ occurs exactly $(q-3)/4$ times as a difference of squares. □

Example 1.23. The set $D = \{1, 3, 4, 5, 9\}$ of squares in $\mathbb{F}_{11}$ is an $(11, 5, 2)$-difference set.

1.5 Existence of finite fields

Our main goal in this section is to prove existence and uniqueness of finite fields $\mathbb{F}_q$ of each prime power order q. It turns out that $\mathbb{F}_q$ gets delivered as the splitting field of a special polynomial: $x^q - x \in \mathbb{F}_p[x]$. First, though, we need some background on polynomials.

Quotient by poiynomial ideals

Let $m(x) \in \mathbb{F}[x]$. Declare two polynomials $u(x), v(x)$ to be *congruent modulo* $m(x)$ if

$$m(x) \mid (u(x) - v(x)).$$

This is usually written $u(x) \equiv v(x) \pmod{m(x)}$. When $\mathbb{F} = \mathbb{F}_p$, coefficients are also subject to reduction modulo p. For extra clarity when we wish to indicate the characteristic, it is also meaningful to write $u(x) \equiv v(x) \pmod{p, m(x)}$. The usual rules of modular arithmetic can be extended to polynomials in this way.

Example 1.24. Consider $\mathbb{F}_5$ with $m(x) = x^2 + 3x + 2$. Then

$$\begin{aligned} x^4 &\equiv (x^2)^2 = (-3x-2)^2 = (2x+3)^2 \\ &\equiv 4x^2 + 12x + 9 = -x^2 + 2x + 9 \\ &\equiv (3x+2) + (2x+9) \\ &\equiv 1 \pmod{5, x^2+3x+2}. \end{aligned}$$

As you can see, there are really just $5^2 = 25$ distinct equivalence classes of polynomials. Beware, though: these 25 classes do not form a field since $m(x)$ splits (more on this later).

Recall the ideal generated by polynomial $m \in \mathbb{F}[x]$ is

$$\langle m(x) \rangle = \{m(x)f(x) : f \in \mathbb{F}[x]\}.$$

Note that we said $u \equiv v \pmod{m}$ above if and only if $u(x) - v(x)$ is an element of $\langle m(x) \rangle$. So, just as arithmetic modulo n induces the quotient ring $\mathbb{Z}/\langle n \rangle$, in the same way polynomial arithmetic modulo $m(x)$ is really happening in

$$\mathbb{F}[x]/\langle m(x) \rangle.$$

Theorem 1.25. *If $f(x)$ is irreducible in $\mathbb{F}[x]$, then the quotient $\mathbb{F}[x]/\langle f(x) \rangle$ is a field.*

Proof. Since commutative rings mod maximal ideals are fields, it suffices to show $\langle f(x) \rangle$ is a maximal ideal in $\mathbb{F}[x]$. Recall $\mathbb{F}[x]$ is Euclidean, hence a PID. So, to verify maximality, we may consider the inclusion $\langle f(x) \rangle \subseteq \langle a(x) \rangle \subseteq \langle 1 \rangle = \mathbb{F}[x]$. In this case, $a(x) \mid f(x)$; thus by irreducibility of $f(x)$ we have either $\langle a(x) \rangle = \mathbb{F}[x]$ or $\langle f(x) \rangle$, as required. □

Conversely, $m(x)$ not irreducible yields zero divisors and renders the quotient not a field.

Corollary 1.26. *If $f(x)$ is an irreducible polynomial of degree n in $\mathbb{F}_p[x]$, then $\mathbb{F}_p[x]/\langle f(x) \rangle$ is a field of order p^n.*

This gives a concrete presentation of finite fields in which both operations (addition by collecting powers of x and multiplication by reducing mod $f(x)$) are natural. It remains, though, to find such irreducibles $f(x)$. They do exist for each p, n (though we have not seen a proof yet) and they can be found in a table or on computer.

Splitting fields

Definition 1.27. A polynomial $f \in \mathbb{F}[x]$ is said to *split* in an extension $\mathbb{K}$ of $\mathbb{F}$ if f can be expressed as a product of linear factors in $\mathbb{K}[x]$; that is, if

$$f(x) = c(x - a_1)(x - a_2) \cdots (x - a_n)$$

holds in $\mathbb{K}[x]$ for some $c \in \mathbb{F}$ and (a possibly empty) list of $a_i \in \mathbb{K}$.

Example 1.28. In $\mathbb{F}_{13}[x]$, $f(x) = x^2 + 1$ splits as $(x + 5)(x + 8)$. But in $\mathbb{F}_7[x]$ (as with $\mathbb{R}[x]$), $f(x) = x^2 + 1$ does not split in the ground field; so, being quadratic, it is an irreducible polynomial.

Definition 1.29. Say $\mathbb{K}$ is a *splitting field* of $f(x) \in \mathbb{F}[x]$ over $\mathbb{F}$ if f splits in $\mathbb{K}$ but does not split in any proper subfield of $\mathbb{K}$ containing $\mathbb{F}$.

Next comes an important first result in the direction of explicitly obtaining splitting fields.

Fundamental Theorem of Fields (Kronecker, 1887). Let $\mathbb{F}$ be a field and $f \in \mathbb{F}[x]$ be non-constant. There exists an extension field $\mathbb{K}$ of $\mathbb{F}$ such that f has a zero in $\mathbb{K}$.

Proof. Take $\mathbb{K} = \mathbb{F}[x]/\langle g(x)\rangle$, where g is an irreducible factor of f in $\mathbb{F}[x]$. By the above discussion, $\mathbb{K}$ is a field. A copy of $\mathbb{F}$ exists as a subfield inside $\mathbb{K}$ as $a + \langle g(x)\rangle$. Finally, it is easily checked that a zero of f in $\mathbb{K}$ is furnished by $x + \langle g(x)\rangle$. □

Theorem 1.30. *For* $\deg f > 0$*, there exists a splitting field for* $f \in \mathbb{F}[x]$*.*

Proof idea. Use strong induction on $\deg f$, together with the Fundamental Theorem of Fields. □

Example 1.31. Let's return to $f(x) = x^2 + 1$. As a polynomial in $\mathbb{R}[x]$, a splitting field for f over $\mathbb{R}$ is $\mathbb{C}$, the complex numbers. As we all know, $x^2 + 1 = (x + i)(x - i)$ in $\mathbb{C}[x]$.

On the other hand, as a polynomial in $\mathbb{F}_7[x]$, a splitting field for f over $\mathbb{F}_7$ is obtained as $\mathbb{F}_7[x]/\langle x^2 + 1\rangle$. This is a finite field of order 49, with elements $a + bx + \langle x^2 + 1\rangle$, $a, b \in \mathbb{F}_7$.

For an element $\alpha \in \mathbb{K} \setminus \mathbb{F}$, where $\mathbb{K}$ is an extension of $\mathbb{F}$, let $\mathbb{F}(\alpha)$ denote the smallest subfield of $\mathbb{K}$ which contains both $\mathbb{F}$ and α. Alternatively, $\mathbb{F}(\alpha)$ is the field of fractions of $\mathbb{F}[\alpha]$.

Theorem 1.32. *Let $\mathbb{F}$ be a field and $f(x)$ irreducible in $\mathbb{F}[x]$. Suppose α is a zero of $f(x)$ in some extension. Then*

$$\mathbb{F}(\alpha) \cong \mathbb{F}[x]/\langle f(x)\rangle.$$

Proof. Consider the 'evaluation' homomorphism $\phi : \mathbb{F}[x] \to \mathbb{F}(\alpha)$ defined by $f(x) \mapsto f(\alpha)$. Its kernel is an ideal containing $f(x)$; hence $\ker(\phi) = \langle f(x)\rangle$ by maximality of this ideal. The first isomorphism theorem completes the proof. □

Example 1.33. We can identify $\mathbb{F}_7[x]/\langle x^2+1\rangle$ with $\mathbb{F}_7(i)$, where as usual i satisfies the relation $i^2 = -1$.

Corollary 1.34. *If α, β are two zeros of $m(x)$, then $\mathbb{F}(\alpha) \cong \mathbb{F}(\beta)$, since each is isomorphic to $\mathbb{F}[x]/\langle m(x)\rangle$.*

Example 1.35. Let $f(x) = x^2 + x + 1 \in \mathbb{F}_2[x]$. It is clear by checking possible roots that f is irreducible. Let α be a symbol representing a root of f, so that α satisfies $\alpha^2 = \alpha + 1$. Now put $\mathbb{K} = \mathbb{F}_2[x]/(x^2+x+1) \cong \mathbb{F}_2(\alpha) = \{0, 1, \alpha, \alpha+1\}$. Then $\mathbb{K}$ is a splitting field for f since $f(x) = (x-\alpha)(x-\alpha-1)$ in $\mathbb{K}[x]$, and yet is irreducible over the only proper subfield. Note that in this case $\mathbb{F}_2(\alpha) \cong \mathbb{F}_2(\alpha+1)$.

Applying induction to Corollary 1.34 gives uniqueness of splitting fields. Details are omitted.

Theorem 1.36. *Any two splitting fields for $f \in \mathbb{F}[x]$ over $\mathbb{F}$ are isomorphic (via an isomorphism which fixes the ground field $\mathbb{F}$).*

Multiplicity of zeros

If $\alpha \in \mathbb{K}$ is a zero of $f(x)$, then $(x-\alpha) \mid f(x)$ in $\mathbb{K}[x]$. There exists a largest integer k such that $(x-\alpha)^k \mid f(x)$. This k is the *multiplicity* of α, and α is *simple* if $k = 1$.

Definition 1.37. A polynomial $f \in \mathbb{F}[x]$ is *separable* if and only if, in the splitting field $\mathbb{K}$ of f, $f(x)$ has no multiple zeros.

To characterize the separable polynomials, we introduce the notion of *formal derivative* $D : \mathbb{F}[x] \to \mathbb{F}[x]$. The operator D acts just as ordinary differentiation of polynomials, where

$$D : a_0 + a_1 x + \cdots + a_n x^n \mapsto a_1 + 2a_2 x + \cdots + n a_n x^{n-1}.$$

It is easy to see that D is a linear transformation on $\mathbb{F}[x]$ and, after an induction argument, that the product rule holds. One usually writes f' instead of $D(f)$.

An important property is that the formal derivative tests for separability of polynomials without the need to explicity produce a splitting field.

Theorem 1.38. *$f \in \mathbb{F}[x]$ is separable if and only if* $\gcd(f, f') = 1$.

Proof. Suppose $f \in \mathbb{F}[x]$ is separable and let $\mathbb{K}$ be its splitting field. Then every zero of f, say $\alpha \in \mathbb{K}$ is simple. We have $f(x) = (x - \alpha)g(x)$ where $g(\alpha) \neq 0$. By the product rule, $f'(x) = g(x) + (x - \alpha)g'(x)$ and so $f'(\alpha) = g(\alpha) \neq 0$. It follows that f has no common linear factors with f' and $\gcd(f, f') = 1$.

On the other hand, suppose $f(x) = (x - \alpha)^2 g(x)$ for some $g(x) \in \mathbb{K}[x]$. Then $f'(x) = (x - \alpha)[2g(x) + (x - \alpha)g'(x)]$, so that $x - \alpha \mid \gcd(f, f')$. □

Whether a polynomial is separable depends on the characteristic of the ground field.

Example 1.39. In $\mathbb{R}[x]$, $f(x) = x^2 + 1$, since, over its splitting field $\mathbb{C}$, f factors as $(x + i)(x - i)$, and these are distinct linear factors.

On the other hand, in $\mathbb{F}_2[x]$, $f(x) = x^2 + 1$ is not separable, since $f(x) = (x + 1)^2$. In this case, 1 is a zero of multiplicity two. Note $f'(x) = 2x = 0$, giving $\gcd(f, f') = f \neq 1$.

Freshman's dream

We now investigate an important property of polynomials in characteristic p.

Definition 1.40. For a nonnegative integer t and real number x,

$$\binom{x}{t} = \frac{x(x-1)\cdots(x-t+1)}{t!}.$$

(Note: if $t = 0$, the product in the numerator is empty and $\binom{x}{0} = 1$.)

Theorem 1.41. *If p is prime and $0 < t < p$, then* $\binom{p}{t} \equiv 0 \pmod{p}$.

Proof. $t! \not\equiv 0 \pmod{p}$, so $t!$ has a multiplicative inverse in $\mathbb{F}_p$. Therefore,

$$\binom{p}{t} \equiv (t!)^{-1} p(p-1)\cdots(p-t+1) \equiv 0 \pmod{p}.$$

□

Corollary 1.42. *In characteristic p, $(x+y)^p = x^p + y^p$. And, more generally, $(x+y)^{p^r} = x^{p^r} + y^{p^r}$.*

We now have the background required to construct finite fields of any prime power order. Incidentally, although Galois is credited with this (after his death) it took about another 50 years before the notion of 'field' was made precise! But the proof idea is essentially the same as Galois used.

Main result

Theorem 1.43 (Galois, 1846)**.** *For any prime p and positive integer n, there exists a unique finite field of order $q = p^n$.*

Proof. Let $\mathbb{K}$ be the splitting field of $f(x) = x^q - x$ over $\mathbb{F}_p$. Since $f' = p^n x^{q-1} - 1 = -1$, it follows by Theorem 1.38 that f is separable. So there are q distinct zeros of f in $\mathbb{K}$.

If α and β are zeros of f, then from Corollary 1.42,

$$(\alpha + \beta)^q - (\alpha + \beta) = \alpha^q - \alpha + \beta^q - \beta = 0,$$

$$(\alpha\beta)^q - (\alpha\beta) = (\alpha^q - \alpha)(\beta^q - \beta) + (\alpha^q - \alpha)\beta + \alpha(\beta^q - \beta) = 0,$$

and

$$(\alpha^{-1})^q - \alpha^{-1} = \alpha^{-q-1}(\alpha - \alpha^q) = 0.$$

So the zeros of f are closed under sum, product, and inverse. It follows that they form a field with q elements, and so this is $\mathbb{K}$ itself.

To see the uniqueness, any $\mathbb{F}_q$ contains such a splitting field $\mathbb{K}$, since $x^q - x$ is satisfied by all elements of $\mathbb{F}_q$. Being of size q, we have $\mathbb{F}_q \cong \mathbb{K}$, which is unique by Theorem 1.36. □

Refer again to Example 1.35 and the operation tables therein.

Subfield structure and algebraic closure

To understand the subfield structure of $\mathbb{F}_q$, we use a simple fact about polynomial divisibility.

Lemma 1.44. *$x^m - 1 \mid x^n - 1$ in $\mathbb{Z}[x]$ if and only if $m \mid n$.*

Proof. If $n = km$, then $x^n - 1 = (x^m - 1)(1 + x^m + x^{2m} + \cdots + x^{(k-1)m})$. Conversely, take the derivative of $x^n - 1 = (x^m - 1)f(x)$ and use the product rule to obtain $nx^{n-1} = mx^{m-1}f(x) + (x^m - 1)f'(x)$. Substituting $x = 1$, we have $n = mf(1)$. Since $f(x)$ has integral coefficients, it follows that $m \mid n$. □

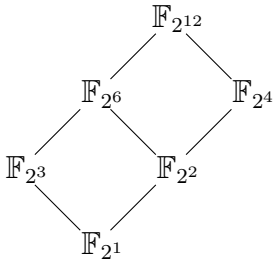
Theorem 1.45. *$\mathbb{F}_{p^m}$ is a subfield of $\mathbb{F}_{p^n}$ if and only if $m \mid n$.*

Proof. We have $[\mathbb{F}_{p^n} : \mathbb{F}_p] = [\mathbb{F}_{p^n} : \mathbb{F}_{p^m}] \cdot [\mathbb{F}_{p^m} : \mathbb{F}_p]$, and so $[\mathbb{F}_{p^n} : \mathbb{F}_{p^m}] = n/m$. Conversely, if $m \mid n$, then $p^m - 1 \mid p^n - 1$ and so

$$x^{p^m-1} - 1 \mid x^{p^n-1} - 1.$$

Therefore, the splitting field of $x^{p^m} - x$ is contained in the splitting field of $x^{p^n} - x$. Given uniqueness of splitting fields, the proof is complete. □

Example 1.46. We have the following lattice of subfields of $\mathbb{F}_{2^{12}}$.



The algebraic closure of $\mathbb{F}_p$ can be explicitly built as a direct limit of the chain of field extensions

$$\mathbb{F}_p \subset \mathbb{F}_{p^2} \subset \mathbb{F}_{p^6} \subset \cdots \subset \mathbb{F}_{p^{n!}} \subset \cdots \to \overline{\mathbb{F}}_p.$$

(This is simply the infinite union, where multiplication and addition of two elements take place in a common field containing them.)

Exercises

1. (a) Explain why the unique prime subfield of $\mathbb{F}$ equals the intersection of all subfields of $\mathbb{F}$.

(b) Find, with proof, two subfields of $\mathbb{R}$ such that neither of them is $\mathbb{Q}$ but their intersection is $\mathbb{Q}$.

2. (a) Prove that if $\{a_1, \ldots, a_m\}$ is a basis of $\mathbb{L}$ over $\mathbb{K}$ and $\{b_1, \ldots, b_n\}$ is a basis of $\mathbb{K}$ over $\mathbb{F}$, then $\{a_i b_j : 1 \le i \le m, 1 \le j \le n\}$ is a basis of $\mathbb{L}$ over $\mathbb{F}$.

   (b) Find a basis of $\mathbb{Q}(\sqrt[3]{2}, i)$ over $\mathbb{Q}$.

3. Let $\mathbb{F}_q$ be a field of order q. Compute (a) $\displaystyle\sum_{a \in \mathbb{F}_q^+} a$ and (b) $\displaystyle\prod_{a \in \mathbb{F}_q^\times} a$.

4. Prove that an irreducible polynomial f is separable over $\mathbb{K}$ if and only if its derivative f' is nonzero. But show how non-constant polynomials, say in $\mathbb{F}_p[x]$, can have zero derivative; characterize all such polynomials.

5. Find a prime p so that the smallest generator (i.e. primitive root) of $\mathbb{F}_p^\times$ is 6. Justify that 6 is smallest possible. You are encouraged to use a computer for both the research and calculation.

6. Let a and b be elements of $\mathbb{F}_{2^n}$, n odd. Prove that $a^2 + ab + b^2 = 0$ implies $a = b = 0$.

7. Show that in a finite field, every element is a sum of two squares. (*Hint*: More than half of the elements in $\mathbb{F}_q$ are squares!)

8. (a) Prove that -1 is a square in $\mathbb{F}_q$ if and only if $q \not\equiv 3 \pmod 4$.

   (b) When $q \equiv 1 \pmod 4$, give an explicit decomposition of every element $a \in \mathbb{F}_q$ as a sum of two squares.

9. Suppose $f \in \mathbb{F}_p[x]$ and $\alpha \in \mathbb{F}_{p^n}$. If $f(\alpha) = 0$, prove that $m_\alpha(x) \mid f(x)$.

10. The *trace* function Tr maps $\mathbb{F}_{q^n}$ to $\mathbb{F}_q$ and is defined by $\mathrm{Tr}(\beta) = \beta + \beta^q + \cdots + \beta^{q^{n-1}}$.

    (a) Prove that $\mathrm{Tr}(\beta)$ actually does belong to $\mathbb{F}_q$ for any $\beta \in \mathbb{F}_{q^n}$.

    (b) Prove that $\mathrm{Tr}(\beta) = 0$ if and only if $\beta = \alpha^q - \alpha$ for some α. (*Hint:* For the "only if" direction, begin by computing $a^p - a$, where

$$a = \beta + (\beta + \beta^p) + \cdots + (\beta + \beta^p + \cdots + \beta^{p-2}).)$$

11. Consider the finite field $\mathbb{F}_{27}$.

    (a) Express it in the form $\mathbb{F}_p[x]/\langle f(x) \rangle$ for some polynomial f.

    (b) Express its additive and multiplicative groups using various $\mathbb{Z}/n\mathbb{Z}$ and $\oplus$ only.

(c) Find a polynomial $g(x)$ with the property that, for $\alpha \in \mathbb{F}_{27}$, $g(\alpha) = 0$ if and only if α is not a generator.

12. (a) Show that $x^5 + x^3 + 1 \in \mathbb{F}_2[x]$ is irreducible.

    (b) Is there an irreducible polynomial in $\mathbb{F}_2[x]$ with exactly four nonzero terms?

13. Let p be an odd prime, and let $q = p^2$. Give the subfields of the field of order q^q, and describe the structure of inclusions.

14. Consider the irreducible polynomial $f(x) = x^4 + x^3 + x^2 + x + 1$ over $\mathbb{F}_2$. Let α be a root of f, and let $\mathbb{F} = \mathbb{F}_2[x]/\langle f(x)\rangle$.

    (a) Regard the field element $x + \langle f\rangle \in \mathbb{F}$ as α. Multiply all elements of $\mathbb{F}$ by α. What is α^{-1}?

    (b) Show that α is not a generator of $\mathbb{F}^\times$.

    (c) Find a generator β of $\mathbb{F}$, and a degree 4 polynomial $g(x)$ with $g(\beta) = 0$.

Chapter 2

Polynomials

2.1 Minimal and primitive polynomials

The splitting field construction for finite fields is theoretically convenient, but lacks some concreteness. We do know that $\mathbb{F}_p[x]/\langle f(x)\rangle$ is a field of order p^n when $f(x)$ is irreducible of degree n. From this next section, we are able to conclude that irreducibles of every possible degree over $\mathbb{F}_p$ do exist (and see how to find nice ones).

Definition 2.1. The *Frobenius automorphism* $\theta : \mathbb{F}_q \to \mathbb{F}_q$ is defined by $\theta(x) = x^p$, where $q = p^n$.

It's clear that $\theta(xy) = \theta(x)\theta(y)$. Additivity $\theta(x + y) = \theta(x) + \theta(y)$ is another form of Corollary 1.42. Also, observe that the fixed points of θ are precisely the elements of the prime field $\mathbb{F}_p$; this follows from $x^p - x$ having exactly p zeros in $\mathbb{F}_q$.

We have $\theta^n(x) = x^{p^n} = x$, for all $x \in \mathbb{F}_q$; hence, θ^n is the identity on $\mathbb{F}_q$. In fact, θ generates the (cyclic) group of all automorphisms of $\mathbb{F}_q$ fixing $\mathbb{F}_p$.

Definition 2.2. Let $\beta \in \mathbb{F}_{p^n}^{\times}$. The least degree nonzero monic polynomial $f \in \mathbb{F}_p[x]$ with $f(\beta) = 0$ is called the *minimal polynomial* of β, usually denoted $m_\beta(x)$.

It is clear (but deserves mentioning) that m_β is unique, since the difference of two minimal polynomials of β can only be zero. Furthermore, an easy consequence of the (polynomial) division algorithm is that m_β divides *any* polynomial which annihilates (evaluates to zero

at) β. The minimal polynomial of an element represented as a polynomial in $\mathbb{F}_p[x]/\langle f(x)\rangle$ can be found by solving a system of linear equations.

Since β is a zero of $m_\beta(x) \in \mathbb{F}_p[x]$, it follows that its Frobenius iterates $\beta, \beta^p, \beta^{p^2}, \ldots$ are also zeros; these are also called the *conjugates* of β. In fact, the zeros of m_β are precisely the conjugates of β.

Theorem 2.3. *Let $\beta \in \mathbb{F}_{p^n}^\times$ such that $\beta^{p^r} = \beta$ for the least positive r. Then its minimal polynomial is*

$$m_\beta(x) = (x-\beta)(x-\beta^p)\cdots(x-\beta^{p^{r-1}}).$$

Proof sketch. Put $q = p^n$. Extend the Frobenius automorphism θ to an (algebra) automorphism $\widehat{\theta} : \mathbb{F}_q[x] \to \mathbb{F}_q[x]$ by

$$\widehat{\theta} : a_0 + a_1 x + \cdots + a_k x^k \mapsto \theta(a_0) + \theta(a_1)x + \cdots + \theta(a_k)x^k.$$

Since this mapping simply cycles the factors of the asserted $m_\beta(x)$, that polynomial is fixed by $\widehat{\theta}$. It follows that it has coefficients in $\mathbb{F}_p$. Moreover, no polynomial in $\mathbb{F}_p[x]$ of smaller degree can annihilate β, due to the choice of r. □

Corollary 2.4. *Suppose α is a generator of $\mathbb{F}_{p^n}^\times$. Then m_α is an irreducible polynomial of degree n in $\mathbb{F}_p[x]$.*

Proof. That m_α has degree n follows from α being a generator. Consider the irreducibility claim. If m_α were to split in $\mathbb{F}_p[x]$, each factor would have to be invariant under the Frobenius automorphism θ. But the zeros of those factors would partition $\{\alpha, \alpha^p, \ldots, \alpha^{p^{n-1}}\}$, and no such partition is invariant under θ. □

When α is a generator of $\mathbb{F}_{p^n}^\times$, the finite field presentation $\mathbb{F}_{p^n} \cong \mathbb{F}_p[x]/\langle m_\alpha(x)\rangle$ has the extra feature that it is (multiplicatively) generated by x. This leads to another associated definition.

Definition 2.5. For a polynomial $f(x) \in \mathbb{F}_p[x]$ with $f(0) \neq 0$, the *order* of f is the least positive integer e such that $f \mid x^e - 1$. If f is monic, irreducible of degree n, and order $p^n - 1$, then it is said to be *primitive*.

Note that the order of any degree n polynomial f is well defined: since the quotient $\mathbb{F}_p[x]/\langle f(x)\rangle$ has $p^n - 1$ nonzero elements, it follows that the powers of x must eventually repeat. In fact, 'order' of irreducible polynomials is related to 'order' of $\mathbb{F}_q^\times$ group elements.

Lemma 2.6. *The order of $\alpha \in \mathbb{F}_q^\times$ equals the order of $m_\alpha(x)$ as a polynomial in $\mathbb{F}_p[x]$.*

Proof. Suppose $m_\alpha(x) \mid x^e - 1$. Then, since $m_\alpha(\alpha) = 0$, it follows that $\alpha^e = 1$. On the other hand, suppose $\alpha^e = 1$. Then, apply Frobenius to get $\theta(\alpha)^e = 1$. It follows that every zero of $m_\alpha(x)$ is also a zero of $x^e - 1$, and we obtain that $m_\alpha(x) \mid x^e - 1$. □

We summarize our most important findings in the next result.

Theorem 2.7. *Given a prime p and positive integer n, there exists a primitive polynomial g of degree n in $\mathbb{F}_p[x]$. Moreover, $\mathbb{F}_{p^n} \cong \mathbb{F}_p[x]/\langle g(x)\rangle$ has x as a (multiplicative) generator.*

Example 2.8. Consider $g(x) = x^3 + x + 1$ in $\mathbb{F}_2[x]$. Note g is irreducible over $\mathbb{F}_2$ since it is of degree 3 and has no zeros in $\mathbb{F}_2$. Consider powers of x in $\mathbb{F}_2[x]/\langle g(x)\rangle$:

$$x^0 = 1,\ x^1 = x,\ x^2 = x^2,\ x^3 = x + 1,\ x^4 = x^2 + x,\ x^5 = x^2 + x + 1,\ x^6 = x^2 + 1,$$
$$x^7 = 1, \ldots$$

This gives a representation of the finite field $\mathbb{F}_8$. We have

$$\mathbb{F}_8 = \{0, 1, \alpha, \alpha + 1, \alpha^2, \alpha^2 + 1, \alpha^2 + \alpha, \alpha^2 + \alpha + 1\},$$

where polynomial addition and polynomial multiplication (mod 2,g) are the operations.

Now consider $\alpha + 1$. Since $\alpha + 1 = \alpha^3$, its conjugates under θ are $\alpha^6 = \alpha^2 + 1$ and $\alpha^{12} = \alpha^5 = \alpha^2 + \alpha + 1$. we have

$$m_{\alpha+1}(x) = (x - \alpha^3)(x - \alpha^6)(x - \alpha^5) = x^3 + x^2 + 1.$$

This is the other primitive polynomial of degree 3 in $\mathbb{F}_2[x]$, and its quotient furnishes an isomorphic copy of $\mathbb{F}_8$.

Generating elements in a group are sometimes also known as *primitive elements*. What we have seen is that primitive elements in $\mathbb{F}_{p^n}^\times$ (together with their conjugates) are in correspondence with primitive polynomials of degree n in $\mathbb{F}_p[x]$.

2.2 Cyclotomic polynomials

Given the importance of primitive polynomials over $\mathbb{F}_p[x]$ for presenting/constructing finite fields, the problem of finding such polynomials deserves a closer look. For this, we venture briefly back to characteristic zero.

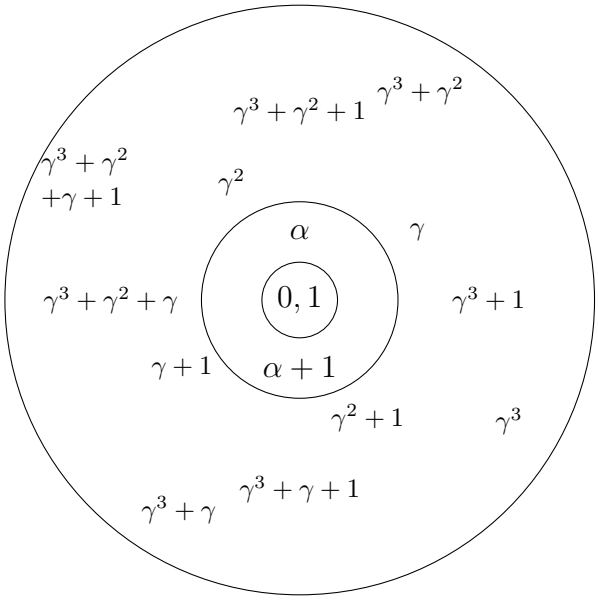


Figure 2.1: The finite field $\mathbb{F}_{16}$, its subfields, and Frobenius automorphisms

Recall that in $\mathbb{C}$, the Nth *roots of unity* are the N powers of $\zeta = e^{2\pi i/N}$. When $\gcd(k, N) = 1$, one says that ζ^k is *primitive*; these are generators for the cyclic group of all Nth roots of unity. It follows that there exist $\phi(N)$ primitive Nth roots of unity, where ϕ is the Euler-phi function.

Definition 2.9. The Nth *cyclotomic polynomial* is

$$\Phi_N(x) = (x - \zeta_1)(x - \zeta_2)\cdots(x - \zeta_{\phi(N)}),$$

where $\zeta_1, \ldots, \zeta_{\phi(N)}$ are the primitive Nth roots of unity.

Fact 2.10.

$$x^N - 1 = \prod_{d|N} \Phi_d(x).$$

Proof idea. Any Nth root of unity (primitive or not) is a primitive dth root of unity for some $d \mid N$. □

This immediately allows us to calculate certain cyclotomic polynomials. The case when $N = p$, a prime, is easiest; in that case, every pth root of unity is primitive, except 1 of course.

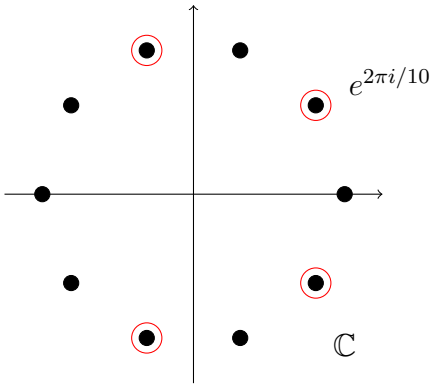


Figure 2.2: Complex tenth roots of unity, with primitive roots circled

Lemma 2.11.

$$\Phi_p(x) = \frac{x^p - 1}{x - 1} = 1 + x + x^2 + \cdots + x^{p-1}.$$

Extending to prime-power indices is not much harder.

Lemma 2.12.

$$\Phi_{p^m}(x) = \frac{x^{p^m} - 1}{x^{p^{m-1}} - 1} = 1 + x^{p^{m-1}} + x^{2p^{m-1}} + \cdots + x^{(p-1)p^{m-1}}.$$

We can also conclude that the coefficients are integers.

Theorem 2.13. *The cyclotomic polynomial $\Phi_N(x)$ has integer coefficients.*

Proof. We use induction on N and prove the stronger claim that the gcd of coefficients of $\Phi_N(x)$ equals 1. This is true for $N = 1$, since $\Phi_1(x) = x - 1$. Now suppose the claim is true for all indices less than some $N \geq 2$. We have

$$x^N - 1 = \Phi_N(x) \prod_{d|N, d \neq N} \Phi_d(x).$$

By a lemma of Gauss, the product of polynomials having relatively prime coefficients is another such polynomial. Take the smallest positive integer d such that $\Phi_N(x) \in \frac{1}{d}\mathbb{Z}[x]$. Then d times the left side is a polynomial with relatively prime coefficients, yet d times the right side has gcd of coefficients equal to d. It follows that $d = 1$. □

Remark. $\Phi_N(x)$ is irreducible in $\mathbb{Z}[x]$. We omit the proof.

Here are the first few cyclotomic polynomials.

N	$\phi(N)$	$\Phi_N(x)$
1	1	$x-1$
2	1	$x+1$
3	2	x^2+x+1
4	2	x^2+1
5	4	$x^4+x^3+x^2+x+1$
6	2	x^2-x+1

Applying Mobiüs inversion to Fact 2.10, we have

$$\Phi_N(x) = \prod_{d|N} (x^{N/d} - 1)^{\mu(d)},$$

where $\mu(1) = 1$, $\mu(p_1 p_2 \cdots p_t) = (-1)^t$ for distinct primes p_i, and $\mu(m) = 0$ if m has any square prime divisor.

Example 2.14. Let $N = 15$. Then

$$\Phi_{15}(x) = \frac{(x^{15}-1)(x-1)}{(x^5-1)(x^3-1)} = x^8 - x^7 + x^5 - x^4 + x^3 - x + 1$$

can be found using geometric series and long division.

Remark. It is tempting to guess that the coefficients of Φ_N are always in $\{-1, 0, 1\}$; however, Φ_{105} breaks the pattern with a coefficient of -2. In general, coefficients of Φ_N can get arbitrarily large in magnitude.

As in the characteristic zero case, a root $\alpha \in \overline{\mathbb{F}_p}$ of $x^N - 1 = 0$ is called an Nth root of unity with respect to $\mathbb{F}_p$. When viewed in $\mathbb{F}_p[x]$, we may just take Φ_N to have coefficients mod p.

Although Φ_N is irreducible in $\mathbb{Z}[x]$, it often reduces in $\mathbb{F}_p[x]$. Here is an important factorization in the case $N = p^n - 1$.

Proposition 2.15. *In* $\mathbb{F}_p[x]$,

$$\Phi_{p^n-1}(x) = \prod_{\substack{f \in \mathbb{F}_p[x] \text{ primitive,} \\ \deg(f) = n}} f(x).$$

Proof. Both sides are monic and contain precisely the same zeros (generators of $\mathbb{F}_{p^n}$), the right side being organized according to minimal polynomials of these elements. □

Corollary 2.16. *There are exactly* $\frac{1}{n}\phi(p^n-1)$ *primitive polynomials of degree* n *in* $\mathbb{F}_p[x]$.

Beware that in this context we actually care about $\Phi_N(x)$ for $N = p^n - 1$, not p^n. So Lemma 2.12 is usually not of much (direct) use. There are, however, coincidences such as Mersenne primes in which such $\Phi_N(x)$ are easy to compute.

Example 2.17. Recall the two primitive cubic polynomials in $\mathbb{F}_2[x]$ from Example 2.8. We have

$$(x^3+x+1)(x^3+x^2+1) = x^6+x^5+x^4+x^3+x^2+x+1 = \Phi_7(x).$$

To summarize, finding primitive polynomials of a given degree n over $\mathbb{F}_p$ is equivalent to factoring cyclotomic polynomials $\Phi_{p^n-1}(x)$. Equivalently, one can factor $x^{p^n-1}-1$ and discard those factors which are of order $< p^n - 1$. By Fact 2.10, $\Phi_{p^n-1}(x)$ (hence all primitive polynomials) will occur in this factorization.

2.3 Factoring via idempotents in characteristic two

We now turn our attention to the problem of factoring $x^n - 1$ over $\mathbb{F}_2$. (Some warnings: (1) this n takes the role of N in the previous section; (2) this method can be extended to other polynomials and odd characteristics, but the generalization is deeper and we will get to it after refreshing some linear algebra; (3) we will here prefer to write polynomials with ascending exponents, such as $1 + x^n$ for $x^n - 1$. Don't forget that subtraction is actually addition here too!)

Note that since the characteristic is two, we have $p(x)^2 = p(x^2)$ for any polynomial $p(x) \in \mathbb{F}_2[x]$. So if $n = 2^r s$ for r a nonnegative integer and s odd, then

$$1 + x^n = (1 + x^s)^{2^r}.$$

So it suffices to consider factoring $1+x^n$ for n odd. To this end, we introduce a definition.

Definition 2.18. Let n be a fixed positive integer. A polynomial $i(x) \in \mathbb{F}_2[x]$ is *idempotent* if $i(x)^2 \equiv i(x) \pmod{1+x^n}$.

Example 2.19. With $n = 7$, one idempotent is $i(x) = x + x^2 + x^4$. We have

$$i(x)^2 = i(x^2) = x^2 + x^4 + x^8 \equiv x^2 + x^4 + x \pmod{x^7 + 1}.$$

Check that the set of idempotent polynomials with respect to a given n is a vector space (actually an algebra) over $\mathbb{F}_2$. The idempotents include zero and are closed under addition.

Theorem 2.20. *Let n be odd. For every divisor $g(x)$ of $1 + x^n$ in $\mathbb{F}_2[x]$, there exists a unique idempotent polynomial $i(x) \pmod{1 + x^n}$ such that $\gcd(1 + x^n, i(x)) = g(x)$.*

Proof. Write $1 + x^n = g(x)h(x)$. Since n is odd, the derivative of $1 + x^n$ is x^{n-1}. It follows that $1 + x^n$ is separable over $\mathbb{F}_2$, and hence has no repeated factors. Therefore, $\gcd(g(x), h(x)) = 1$.

We first show the existence of $i(x)$. Take $a, b \in \mathbb{F}_2[x]$ such that $a(x)g(x) + b(x)h(x) = 1$. Define $i(x) = a(x)g(x) = 1 - b(x)h(x)$. We have $\gcd(1 + x^n, i(x)) = g(x)$, since this gcd is clearly relatively prime with $h(x)$. Moreover,

$$(i(x))^2 = i(x) - a(x)b(x)g(x)h(x) \equiv i(x) \pmod{1 + x^n}.$$

For uniqueness, suppose $i(x)$ and $j(x)$ both satisfy the given property. Then $g(x) \mid i(x)$ and $j(x)$, so $g(x) \mid i(x) + j(x)$. On the other hand, $h(x) \mid 1 + i(x)$ since $i(x)(1 + i(x)) \equiv 0 \pmod{2, 1 + x^n}$. Likewise, $h(x) \mid 1 + j(x)$ and, taken together, $h(x) \mid i(x) + j(x)$. Since g, h are relatively prime, $i(x) + j(x) \equiv 0 \pmod{2, gh = 1 + x^n}$. This means $i(x) \equiv j(x)$, as desired. □

In light of this result, we would now like to classify all idempotent polynomials; that is, all polynomials whose "exponent set" is invariant under doubling (mod n). For n odd, take the partition of $\mathbb{Z}/n\mathbb{Z}$ according to the orbits of the map $a \mapsto 2a \pmod{n}$. Let these orbits be $C_0 = \{0\}$, $C_1 = \{1, 2, \ldots, 2^{r-1}\}$, where r is the multiplicative order of 2 mod n, and more generally $C_i = \{i, 2i, \ldots\}$ is the orbit containing i.

Now define the polynomial

$$c_i(x) = \sum_{j \in C_i} x^j.$$

By definition $c_i(x)$ is idempotent. In fact, the span of the c_is is the set of *all* idempotents with respect to n. The idempotents are therefore in correspondence with the distinct factors of $1 + x^n$.

Corollary 2.21. *The number of irreducible factors of* $1 + x^n$ *in* $\mathbb{F}_2[x]$ *equals the number of orbits of the doubling map* $a \mapsto 2a$ *in* $\mathbb{Z}/n\mathbb{Z}$.

Example 2.22. Let $n = 9$. The partition of $\mathbb{Z}/9\mathbb{Z}$ into orbits under $a \mapsto 2a$ is $C_0 \cup C_1 \cup C_3$, where $C_0 = \{0\}$, $C_1 = \{1, 2, 4, 5, 7, 8\}$ and $C_3 = \{3, 6\}$. So we have $c_0(x) = 1$, $c_1(x) = x + x^2 + x^4 + x^5 + x^7 + x^8$ and $c_3(x) = x^3 + x^6$. By taking the gcd of various idempotents with $1 + x^9$, we get all irreducible factors of $1 + x^9$. For example, $c_0 + c_3 = 1 + x^3 + x^6$ divides $1 + x^9$ and is an irreducible factor.

It is worth mentioning that factoring any polynomial is a finite problem in $\mathbb{F}_q$. In particular, to see if a polynomial of degree 2 or 3 is irreducible, we only need to check at most q possible zeros. Although, this method fails for degrees greater than 4, there are always just a finite number of possible factors ($< q^{1+n}$), where $n = \deg f$.

Exhausting all these possible factors takes exponential time in n. On the other hand, the computationally intensive part of the factoring procedure we presented here involves just a handful of polynomial gcd calculations, and these take just polynomial time in n using the Euclidean algorithm.

2.4 Linear algebra over finite fields

Here, we very briefly consider vectors and matrices whose elements come from the finite field $\mathbb{F}_q$. Rather than a comprehensive reference, this material is meant to highlight a few important similarities and some key differences with the 'familiar' case of ground fields of characteristic zero. Since these entries form a field, the set $\mathbb{F}_q^d$ of vectors with d coordinates furnishes a vector space of dimension d. Vector sum and scalar products are computed componentwise, but with arithmetic in $\mathbb{F}_q$.

As usual, linear independence of $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ means that the only linear combination of the $\mathbf{v}_i$ which vanishes is the trivial combination. And the span of S is

$$\langle S \rangle = \{\alpha_1 \mathbf{v}_1 + \cdots + \alpha_k \mathbf{v}_k : \alpha_i \in \mathbb{F}_q\}.$$

Definition 2.23. A set B of vectors in a vector space V is a *basis* if it is linearly independent and spanning. Alternatively, B is a basis if $\langle B \rangle = V$ but $\langle S \rangle \neq V$ for any proper subset S of B.

Any element in a vector space with basis B has a unique representation as a linear combination in $\langle B \rangle$.

Theorem 2.24. *If a vector space V over $\mathbb{F}_q$ has a finite basis, say of size d, then every basis of V has size d and $V \cong \mathbb{F}_q^d$. In particular, $|V| = q^d$.*

In fact, $\mathbb{F}_q^d$ can be regarded as $\mathbb{F}_{q^d}$, but where multiplication in this field extension is ignored.

As in characteristic zero, matrix elimination steps can be used to solve linear systems over $\mathbb{F}_q$.

Example 2.25. Over $\mathbb{F}_3$, the matrix A shown below is singular, since each of its rowsums is zero. The reduced row-echelon form of A is R, showing that A has rank 2, and providing a basis for its row space (the two nonzero rows in R) and null space ($\{(-2,-2,1)\} = \{(1,1,1)\}$). Since A is symmetric, the column space and left null space are, respectively, the same.

$$A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 0 \\ 2 & 0 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix} = R.$$

Compare with the fact that A is invertible over $\mathbb{Q}$.

Vector dot product is also naturally extended to general fields, and gives rise to the familiar matrix multiplication. But some counter-intuitive phenomena can occur when arithmetic is done in $\mathbb{F}_p$.

Example 2.26. In characteristic two, a vector $\mathbf{v} \in \mathbb{F}_2^n$ having an even number of ones is 'orthogonal' to itself, in the sense that $\mathbf{v} \cdot \mathbf{v} = 0$.

More generally, it is possible that a subspace U of $\mathbb{F}_q^d$ might have nontrivial intersection with (or even equal) its own orthogonal complement $U^\perp$. So certain projection relations and dimension formulas that you are used to in characteristic zero may become invalid.

Calculating the determinant of a matrix is done in the usual way; this can lead to the characteristic polynomial $\chi_A(x)$ and shows that a matrix over $\mathbb{F}_q$ has an (algebraically) full set of eigenvalues in the closure $\overline{\mathbb{F}}_q$. The eigenspace of A at eigenvalue λ is, as usual, the null space of $A - \lambda I$.

Recall that the Cayley-Hamilton Theorem says that the characteristic polynomial $\chi_A(x)$ evaluated at A equals zero. Here is an interesting consequence.

Theorem 2.27. *Let A be a nonsingular $n \times n$ matrix over $\mathbb{F}_q$. Then $A^N = I$ for some positive integer N, and in particular, if the characteristic polynomial of A is irreducible over $\mathbb{F}_q$, then $A^{q^n-1} = I$.*

Proof. This follows from our work in Chapter 2, since $\chi_A(x) \mid x^N - 1$ for some N, provided x is not a factor. Likewise, every irreducible of degree n divides $x^{q^n-1} - 1$. □

Counting subspaces

Theorem 2.28. *The number of different subspaces of $\mathbb{F}_q^n$ with dimension k equals*

$$\begin{bmatrix} n \\ k \end{bmatrix}_q := \frac{(q^n-1)(q^{n-1}-1)\cdots(q^{n-k+1}-1)}{(q^k-1)(q^{k-1}-1)\cdots(q-1)}.$$

Proof. There are $(q^n-1)(q^n-q)\cdots(q^n-q^{k-1})$ ways to pick an ordered list of k linearly independent vectors in $\mathbb{F}_q^n$. Now, on the other hand, each k-dimensional subspace has $(q^k-1)(q^k-q)\cdots(q^k-q^{k-1})$ ordered bases. The count follows by dividing and some cancellation of powers of q. □

Remark. The quantity $\begin{bmatrix} n \\ k \end{bmatrix}_q$ is called the (*Gaussian*) *q-binomial coefficient*. It is not hard to see that (regarding q as a real parameter)

$$\lim_{q \to 1^+} \begin{bmatrix} n \\ k \end{bmatrix}_q = \binom{n}{k}.$$

2.5 Factoring via Berlekamp's algorithm

In Section 2.3, we saw how to factor $x^n - 1$ over $\mathbb{F}_2[x]$. Here, let's generalize both the characteristic and the polynomial we're factoring. Let p be the characteristic; in fact, often $q = p$ but we can allow extensions in the most general case. And suppose $f \in \mathbb{F}_q[x]$ (monic) is given for factorization.

Similar to before, we can reduce to the case when f is squarefree using its formal derivative.

Observation 2.29. *Suppose f is given for factorization in characteristic p. Let $d = \gcd(f, f')$ and suppose $d \neq 1$.*

- *If $d = f$, then $f' = 0$ and hence $f(x) = g(x^p)$ for some g of smaller degree.*
- *If $d \neq f$, then d is a nontrivial factor of f.*

Otherwise $d = 1$ and f has distinct irreducible factors.

In what follows we will take for granted another useful fact, namely that the Chinese Remainder Theorem extends to polynomial rings over $\mathbb{F}_q$. That is, if $f = f_1 \dots f_r$ with distinct irreducible factors, then

$$\mathbb{F}_q[x]/\langle f \rangle = \bigoplus_{i=1}^{r} \mathbb{F}_q[x]/\langle f_i \rangle.$$

The next result parallels the use of idempotents for $q = 2$.

Proposition 2.30. *Suppose $h^q \equiv h \pmod f$. Then $f = \prod_{s \in \mathbb{F}_q} \gcd(f, h - s)$.*

Proof. It is clear that each factor on the right divides f. Since the $h - s$, $s \in \mathbb{F}_q$, are all coprime, the given product divides f.

On the other hand, we have $\prod_{s \in \mathbb{F}_q}(h - s) = h^q - h \equiv 0 \pmod f$. It follows that f divides $\prod_{s \in \mathbb{F}_q} \gcd(f, h - s)$. □

It remains to find polynomials h with $h^q \equiv h \pmod f$. Earlier, we used orbits of the doubling map on $\mathbb{Z}/n\mathbb{Z}$. Here, we make use of an $n \times n$ matrix $Q = (a_{ki})$, indexed $0 \le k, i < n$, whose kth row is the coefficient list of $x^{qk} \pmod f$:

$$x^{qk} \equiv a_{k0} + a_{k1}x + \cdots + a_{k,n-1}x^{n-1} \pmod f.$$

In this way, the desired polynomials h have coefficient lists which are in the left nullspace of $Q - I$. (These are solutions $\mathbf{v}$ to $\mathbf{v}(Q - I) = \mathbf{0}$.) One solution will always be $\mathbf{v}^{(1)} = (1, 0, 0, \dots, 0)$. If the nullity of $Q - I$ is r, there will be q^r such polynomials. On the other hand, there are also q^r possible choices for $(s_1, \dots, s_r)$ to yield unique solutions $h \equiv s_i \pmod{f_i}$ via the CRT.

To implement a factorization algorithm along these lines, it is enough to observe: (1) the $\mathbb{F}_q$-rank of $Q - I$ being $n - 1$ tests for irreducibility of f; and (2) if this rank is less than $n - 1$, a nontrivial factor of f can be obtained. Either recursive application to factors or computation of several gcds with f yields the complete factorization.

Example 2.31. Suppose we wish to factor $f(x) = x^5 + 2x^4 + 2x^3 + 2x^2 + 2x + 2 \in \mathbb{F}_3[x]$. After checking that $\gcd(f, f') = 1$, it turns out that f is squarefree.

To make the matrix Q, we must compute exponents of $x \pmod{3, f}$. Since $x^5 \equiv 1 + x + \cdots + x^4 \pmod f$, we can reduce powers of x by iteratively right-shifting the coefficient

list, and 'carrying' 11111 or 22222, according to the excess ternary digit that falls off after the shift.

$$\begin{array}{cccccccccccc}
 & & & & & {\scriptstyle +1} & 0_1 & 0_1 & 0_1 & 0_1 & 1 & x^4 \\
 & & & & & 1 & 1 & 1 & 1 & 1 & & x^5 \\
 & & & & 1 & 2 & 2 & 2 & 2 & & & x^6 \\
 & & & 2 & 0 & 1 & 1 & 1 & & & & x^7 \\
 & & 1 & 0 & 1 & 2 & 2 & & & & & x^8 \\
 & 2 & 0 & 2 & 0 & 1 & & & & & & x^9 \\
1 & 0 & 1 & 0 & 1 & & & & & & & x^{10} \\
2 & 1 & 2 & 1 & 2 & & & & & & & x^{11} \\
0 & 2 & 0 & 2 & 0 & & & & & & & x^{12}
\end{array}$$

We have

$$Q = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 2 & 2 & 2 & 2 \\ 2 & 0 & 2 & 0 & 1 \\ 1 & 2 & 0 & 2 & 0 \end{bmatrix},$$

where rows are indexed by the first 5 nonnegative powers of x^3, and columns are indexed by coefficients of $1, x, x^2, x^3, x^4$. For the factoring algorithm, we need the left nullspace of $Q - I$. A reduced row-echelon form calculation gives

$$(Q - I)^\top \to \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

where two zero rows have been omitted. Therefore, a basis for the left nullspace is $\{(1,0,0,0,0), (0,2,1,1,0)\}$, corresponding to polynomials $h^{(1)}(x) = 1$ and $h^{(2)}(x) = 2x + x^2 + x^3$.

The next step in the algorithm tells us to compute, for $s \in \mathbb{F}_3$, the gcd of $h^{(2)} - s$ with f.

s	$\gcd(h^{(2)} - s, f)$
0	$x^2 + x + 2$
1	1
2	$x^3 + x^2 + 2x + 1$

We see irreducible factors as indicated, and this gives the required factorization of f.

Exercises

1. Prove that $x^n + a$ cannot be primitive for $n > 1$.

2. Find a primitive quadratic $g(x)$ over $\mathbb{F}_5$ and construct $\mathbb{F}_{25}$ using $g(x)$.

   (a) Reduce the first 12 nonnegative exponents of $x \pmod{5, g(x)}$.

   (b) What happens for the next 12 exponents?

3. Prove the following facts about cyclotomic polynomials.

   (a) $\Phi_N(0) = 1$, if $n \geq 2$.

   (b) If $N = pq$ for distinct primes p, q, then $\Phi_N(1) = 1$.

4. (a) Write a computer program to find, given n, the orbits of the doubling map in $\mathbb{Z}/n\mathbb{Z}$.

   (b) Use your code from (a) to write a factoring procedure for $x^n + 1$ in $\mathbb{F}_2[x]$.

5. Find a polynomial $p(x) \in \mathbb{F}_2[x]$ of degree 14 such that $p(x)^2 \equiv p(x) \pmod{x^{17} - 1}$.

6. (a) Let p be an odd prime. Prove that $g(x) = 1 + x + ... + x^{p-1}$ is irreducible in $\mathbb{F}_2[x]$ if and only if 2 is a primitive root mod p.

   (b) What happens for $g(x) = 1 + x + \cdots + x^{n-1}$ when n is not prime and 2 is a primitive root mod n?

7. Solve problem A4 of the 2011 Putnam competition by using matrix row operations over $\mathbb{F}_2$.

   > For which positive integers n is there an $n \times n$ matrix with integer entries such that every dot product of a row with itself is even, while every dot product of two different rows is odd?

8. Prove that there exists a p-dimensional subspace W of $\mathbb{F}_p^{2p}$ satisfying $W^\perp = W$ if and only if -1 is a square (mod p).

9. Show that $x^9 + x + 1$ is irreducible in $\mathbb{F}_2[x]$ by analyzing the matrix Q in Berlekamp's algorithm.

Chapter 3

Applications

In this chapter, we survey a few of the main applications of polynomials over finite fields.

3.1 Lagrange interpolation and secret sharing

Given a polynomial $f \in \mathbb{F}_q[x]$, its *evaluation map* sends a to $f(a)$ for each $a \in \mathbb{F}_q$. We begin with an observation on the evaluation maps of polynomials.

Theorem 3.1. *Every function $f : \mathbb{F}_q \to \mathbb{F}_q$ is the evaluation map of some polynomial in $\mathbb{F}_q[x]$.*

Proof. Consider the polynomial

$$\delta(x) = 1 - x^{q-1} = \begin{cases} 1 & \text{if } x = 0, \\ 0 & \text{otherwise.} \end{cases}$$

It follows that the job is done by

$$f(x) = \sum_{a \in \mathbb{F}_q} f(a)\delta(x - a). \quad \square$$

The task of computing a polynomial (often with the aim of minimizing the degree) which agrees with a given function on some set of inputs is called *interpolation*.

Let $a_1, \ldots, a_n$ be distinct elements of some field $\mathbb{F}$. Their *Vandermonde matrix* $V = V(a_1, \ldots, a_n) \in \mathbb{F}^{n \times n}$ has ij-entry $V_{ij} = a_j^{i-1}$. (Some authors 'transpose' the definition.) Note: if one of the chosen elements is 0, we use $0^0 = 1$.

An important fact is that V is nonsingular. In fact, the determinant can be computed by row operations and induction.

Proposition 3.2.

$$\det V(a_1, \ldots, a_n) = \prod_{1 \le i < j \le n} (a_i - a_j).$$

Vandermonde matrices are useful for interpolation.

Example 3.3. Suppose we wish to find a polynomial $f(x) \in \mathbb{F}_5[x]$ whose evaluation map sends $(0, 1, 2, 3, 4)$ to $(1, 2, 4, 3, 0)$. Using δ-functions, we have that

$$\delta(x) + 2\delta(x-1) + 4\delta(x-2) + 3\delta(x-3)$$

does the job. Alternatively, we can solve for the coefficient vector $\mathbf{f}$ in $\mathbf{f}V = \mathbf{b}$, where $\mathbf{b} = (1, 2, 4, 3, 0)$ and

$$V = V(0, 1, 2, 3, 4) = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \\ 0 & 1 & 4 & 4 & 1 \\ 0 & 1 & 3 & 2 & 4 \\ 0 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

It is easy to check that $\mathbf{f} = (1, 0, 0, 1, 0)$ is a solution, corresponding to $f(x) = x^3 + 1$.

Polynomial interpolation has a nice cryptographic application to so-called *secret sharing schemes*. In these scenarios, an organization wishes to distribute 'partial keys' to some secret among q individuals. Any one individual has insufficient information to obtain the secret. However, any collection of n or more people can access the secret by using their partial keys together. One very standard way to set this up is to distribute the evaluations $f(0)$, $f(1), \ldots$ to the individuals, and keep a polynomial $f(x)$ unknown of degree $n - 1$. The polynomial (it's coefficient list, say) unlocks the secret.

A polynomial whose evaluation map gives a permutation of the elements of $\mathbb{F}_q$, as in the previous example, is called a *permutation polynomial*. These are useful in transmitting data in applications where some average modulation is to be achieved. For example, in

modulating frequency over North American power lines, it is important to maintain an average frequency of 60 Hz.

Permutation polynomials are closed under horizontal and vertical translations, and also under nonzero scalings.

Proposition 3.4. *If $f(x)$ is a permutation polynomial over $\mathbb{F}_q$, then so is $af(x-b)+c$ for any $a \in \mathbb{F}_q^\times$, and any $b, c \in \mathbb{F}_q$.*

See also §11 of Lidl and Pilz, and the more comprehensive reference *Finite Fields*, by Lidl and Niederreiter. In that latter book, a table of permutation polynomials is given.

3.2 Linear homogeneous recurrences and M-sequences

A card trick. A magician has a partial deck of 32 cards, containing A-7 and Q in each of the 4 suits. The magician shows the cards in a seemingly random order, but can always tell the top card by looking at the bottom card. Without having a good memory, how does the magician do this?

Encode each card with a binary string of length 5, say $d_0d_1d_2d_3d_4$, where $d_0 = 1$ if and only if the suit is black, $d_1 = 1$ if and only if the suit is pointy, and with the other three bits giving the binary representation of the rank (000 $\leftrightarrow$ Q, 001 $\leftrightarrow$ A, and so on). For instance,

$$01110 \leftrightarrow \overbrace{01}^{\text{suit}}\ \underbrace{110}_{\text{rank}} \leftrightarrow 6\heartsuit.$$

The magician has arranged the cards in sequence so that they correspond to windows of size 5 of the terms of a linear recurrence, where $s_0 = s_1 = s_2 = s_3 = 0$, $s_4 = 1$, and

$$s_{n+5} = s_{n+2} + s_n, \quad n \geq 0$$

So, for example, the magician can use the fingers from one hand to determine that the card following $6\heartsuit \leftrightarrow \mathbf{01}110$ is $\emptyset 1110\underline{1} \leftrightarrow 5\spadesuit$. Here, the new (underlined) bit is obtained as the (mod 2) sum of the zeroth and second (bold) bits.

It turns out that this binary sequence is periodic with period $31 = 2^5 - 1$, and the windows exhaust all nonzero binary strings of length 5! A full period of the sequence is given below.

We see 6♡ as indicated, and in general all 32 cards except Q♢.

$$000010010110011111000110\underline{11110}1010000...$$

To understand why this trick works, we must have a more detailed look at recursively defined sequences over finite fields.

A kth order *linear recurring sequence* (s_n) over $\mathbb{F}_q$ is a sequence in $\mathbb{F}_q$ whose terms satisfy

$$s_{n+k} = a_{k-1}s_{n+k-1} + \cdots + a_1 s_{n+1} + a_0 s_n + a, \qquad (*)$$

where $a, a_i \in \mathbb{F}_q$. Usually, initial values $s_0, s_1, \ldots, s_{k-1}$ are given; then, the case $n = 0$ in $(*)$ explicitly computes the next term s_k.

Since linear recurring sequences depend only on the previous k terms, and since there are only q^k possible k-tuples, these sequences must (eventually) repeat. We are interested in the *least period*, which is the smallest positive integer P so that $s_{n+P} = s_n$ for all sufficiently large n. We have $P \leq q^k$.

When $a = 0$, the recurrence relation $(*)$ and the sequence are called *homogeneous*. This is the case of primary interest for us. Here, the least period is $\leq q^k - 1$.

Definition 3.5. A homogeneous k-term linear recurring sequence over $\mathbb{F}_q$ with maximum period $q^k - 1$ is called an *M-sequence*.

The *characteristic polynomial* of $s_{n+k} - a_{k-1}s_{n+k-1} - \cdots - a_1 s_{n+1} - a_0 s_n$ is

$$f(x) = x^k - a_{k-1}x^{k-1} - \cdots - a_1 x - a_0,$$

and a closed form for s_n results in the usual way from the zeros of f and the initial conditions.

Example 3.6. Consider the Fibonacci sequence (mod p), where $F_0 = 0$, $F_1 = 1$, and for $n \geq 2$, $F_n \equiv F_{n-1} + F_{n-2} \pmod{p}$. For $p = 2$, the sequence proceeds as $0, 1, 1, 0, 1, 1, 0, 1, 1, \ldots$. It is periodic with least period 3 (an M-sequence). The characteristic polynomial $x^2 - x - 1$ is irreducible (primitive, in fact) and has zeros $\alpha, \alpha + 1$. So

$$F_n = \alpha^n + (\alpha + 1)^n$$

is a closed form expression for F_n. Check the initial conditions. Note that, even though F_n is presented as an identity in the extension $\mathbb{F}_4 = \mathbb{F}_2(\alpha)$, the values are necessarily in $\mathbb{F}_2$.

On the other hand, in characteristic 5, we compute the first few terms of (F_n) as

$$0, 1, 1, 2, 3, 0, 3, 3, 1, 4, 0, 4, 4, 3, 2, 0, 2, 2, 4, 1, \ldots$$

and the sequence then repeats with least period 20! This shows that the least period (although it is $\leq q^k - 1 = 24$) need not be a divisor of $q^k - 1$. This is the case when the characteristic polynomial is irreducible; however $x^2 - x - 1 = (x - 3)^2$ in $\mathbb{F}_5[x]$.

Theorem 3.7. *Suppose $\{s_n\}$ is a homogeneous k-term linear recurring sequence over $\mathbb{F}_q$ with irreducible characteristic polynomial $f(x)$. Then $\{s_n\}$ is periodic and its least period equals the order of $f(x) \in \mathbb{F}_q[x]$.*

Corollary 3.8. *A homogeneous k-term linear recurring sequence over $\mathbb{F}_q$ is an M-sequence if and only if its characteristic polynomial is primitive.*

The *generating function* for $\{s_n\}$ is

$$G(x) = \sum_{n=0}^{\infty} s_n x^n \in \mathbb{F}_q[[x]].$$

For periodic sequences, we can express their generating functions as rational functions. This is because

$$G(x) = h(x) + \sum_{n=0}^{\infty} s_{n+P} x^{n+P} = h(x) + x^P G(x),$$

where $s_{n+P} = s_n$ for all n and $h(x) = \sum_{n=0}^{P-1} s_n x^n$. We see that the denominator of $G(x)$ as a 'lowest terms' rational function divides $x^P - 1$. Seen another way, the denominator is actually a close relative of the characteristic polynomial.

Theorem 3.9. *The generating function for a homogeneous k-term linear recurring sequence $\{s_n\}$ with characteristic polynomial $f(x)$ is*

$$G(x) = \frac{h(x)}{f^*(x)},$$

where $f^*(x) = x^k f(1/x)$ is the reciprocal polynomial *of f and $h(x)$ is a polynomial computable from the initial conditions $s_0, s_1, \ldots, s_{k-1}$.*

Example 3.10. Let $q = 3$ and consider the recurrence $s_{n+3} = s_{n+1} - s_n$, where $s_0 = s_1 = 0$ and $s_2 = 1$. The generating function for $\{s_n\}$ satisfies

$$G(x) = x^2 + \sum_{n=0}^{\infty} s_{n+3} x^{n+3} = x^2 + (x^2 - x^3)G(x),$$

or as a rational function

$$G(x) = \frac{x^2}{1 - x^2 + x^3}.$$

Remarks. We need $f(0) \neq 0$ to define $f^*(x)$. The coefficient list of f^* is simply the reverse of that of f. The zeros of $f^*(x)$ are the reciprocals of the zeros of f.

Please refer to §33 of Lidl and Pilz for detailed proofs and, in general, for more on linear recurrences.

3.3 Orthogonal arrays and finite planes

An *orthogonal array* $\mathrm{OA}(t, k, n)$ is an $n^t \times k$ array with symbols from an alphabet of size n having the property that

> when restricted to any t columns, each of the n^t possible words appears in exactly one row.

By crossing out columns, it is clear from the definition that the existence of $\mathrm{OA}(t, k, n)$ implies the existence of $\mathrm{OA}(t, l, n)$ for $t \leq l < k$ as well.

Example 3.11. Here is an $\mathrm{OA}(2, 3, 2)$.

000
110
101
011

Theorem 3.12. *Let q be a prime power and suppose $1 \leq t < q$. Then there exists an $OA(t, q, q)$, and hence an $OA(t, k, q)$ for any k, $t \leq k \leq q$.*

Proof. List the elements of $\mathbb{F}_q$ as $e_1, \ldots, e_q$ (in any order) and list the polynomials of degree $< t$ in $\mathbb{F}_q[x]$ by $f_1, \ldots, f_{q^t}$. Note there are q^t such polynomials by choosing coefficients in $a_0 + a_1 x + \cdots + a_{t-1}x^{t-1}$ arbitrarily. Define the $q^t \times q$ matrix A by

$$A_{ij} = f_i(e_j).$$

To show this is an $\mathrm{OA}(t, q, q)$, it suffices to prove that any $q^t \times t$ submatrix of A has no two distinct rows, say indexed by i, i', that are identical. But since $f_i - f_{i'}$ is a polynomial of degree at most $t - 1$, it can have at most $t - 1$ zeros in $\mathbb{F}_q$. □

Orthogonal arrays are useful in information-based applications. Two applications which show the diversity of end-uses are to software testing (where all t-wise combinations of inputs are covered economically) and numerical integration (where a function can be averaged on a t-wise balanced 'mesh' in k dimensions).

For $t = 2$, orthogonal arrays connect to some important combinatorial structures. The existence of an $\mathrm{OA}(2, k, n)$ is equivalent to a set of $k-2$ 'mutually orthogonal latin squares' of order n. Each square is an $n \times n$ array of n symbols in which every row and column is a permutation of the symbols, and any two squares, when superimposed, contain all n^2 ordered pairs in the n^2 entries. A set of two MOLS of order 3 is shown in Figure 3.1 at left.

Orthogonal arrays with $k = n + 1$ are extremal in the sense that k is as large as possible. In this case, the level sets of the latin squares discussed above produce an 'affine plane' of order n on n^2 points. In a little more detial, an affine plane is a set system $(X, \mathcal{L})$, where X is a set of points and $\mathcal{L}$ is a set of lines, satisfying the usual incidence axioms of Euclidean geometry (any two points on exactly one line, together with Euclid's the parallel postulate). Finite affine planes are known to exist for $n = q$ a prime power. In particular, an affine plane of order q can be constructed with points $\mathbb{F}_q^2$ and lines given by all affine translates of 1-dimensional subspaces. This construction yields $q + 1$ parallel classes of q lines, each covering q points. Notice that

$$q(q+1)\binom{q}{2} = \binom{q^2}{2},$$

where the left side counts the pairs of points covered by lines, and the right side counts the total number of pairs of points.

The construction is illustrated for $q = 3$ in Figure 3.1 at right.

1	2	3
2	3	1
3	1	2

1	2	3
3	1	2
2	3	1

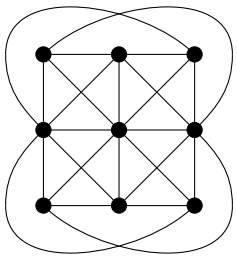


Figure 3.1: Orthogonal latin squares and affine plane of order 3

The projective extension of an affine plane is known as a projective plane. For an affine

plane on n^2 points, its associated projective plane has n^2+n+1 points. Finite planes are used in the construction of combinatorial block designs, which in turn have applications to the design of statistical experiments and to scheduling problems.

Exercises

1. (a) Let p be an odd prime. Find a polynomial $f(x)$ in $\mathbb{F}_p[x]$ so that

$$f(x) = \begin{cases} 1 & \text{if } x \text{ is a square in } \mathbb{F}_p, \\ -1 & \text{otherwise.} \end{cases}$$

   (b) Find and simplify a polynomial $f(x) \in \mathbb{F}_5[x]$ with $(0,1,2,3,4) \mapsto (1,2,1,2,0)$.

2. Find a quadratic in $\mathbb{F}_{13}[x]$ using only three of the following pairs $(a, f(a))$:

   $(0,4),(1,3),(2,6),(3,0),(4,11),(5,0),(6,6),(7,3),(8,4),(9,9),(10,5),(11,5),(12,9)$.

3. Show that $x^5 \pm 2x^2$ are two permutation polynomials over $\mathbb{F}_7$, and use them to construct 588 different permutation polynomials over $\mathbb{F}_7$.

4. Compute the generating function for the linear recurring sequence in $\mathbb{F}_3$ defined by $s_0 = s_2 = 1$, $s_1 = 0$, and

$$s_{n+3} = s_{n+2} - s_{n+1} + s_n, \quad n \geq 0.$$

5. Calculate the first five terms of $(1 - 2x + x^3)^{-1}$ in $\mathbb{F}_7[[x]]$.

6. Prove that in an $\mathrm{OA}(t,k,n)$, the restriction to any $s < t$ columns contains each of the n^s words exactly n^{t-s} times each. This justifies why the parameter t is often called the *strength* of the OA.

7. Construct an $\mathrm{OA}(2,3,3)$ and an $\mathrm{OA}(3,4,2)$.

8. Extend the polynomial construction of an $\mathrm{OA}(t,q,q)$ to produce an $\mathrm{OA}(t,q+1,q)$. (*Hint*: Consider the leading coefficient of polynomials.)

9. Construct the affine plane of order 4 on $\mathbb{F}_4^2$.

Part II

Coding Theory

Chapter 4

Codes and Hamming Distance

4.1 Introduction

Set-up:
$\mathbb{A}$: a nonempty set called the *alphabet*
$\mathbb{A}^n$: tuples or *words* of length n over $\mathbb{A}$ (with $\mathbb{A}^0 = \{\text{empty string}\}$)
$\mathbb{A}^* = \cup_{n \geq 0} \mathbb{A}^n$: the set of words of finite length over $\mathbb{A}$

A *code* is simply a subset $C \subseteq \mathbb{A}^*$. Elements of C are *codewords*.

Focus is placed on comparing symbols in different codewords. A "good" code can be used to detect (or correct) errors, in the sense that codewords perturbed by small errors are no longer codewords (or are far from other codewords). Errors might occur when information is transmitted over a noisy channel (e.g. radio waves over the air) or when media gets damaged (e.g. a scratched DVD) or when information is ambiguous (e.g. converting handwriting to digital text).

We are interested exclusively in finite alphabets. Normally $\mathbb{A} = \mathbb{F}_q$, with $q = 2$ (binary) frequently taking centre stage. Apart from a brief encouragement for the reader to investigate prefix codes or insertion/deletion codes, we are also only interested in the case when $C \subseteq \mathbb{F}_q^n$ for a fixed n. These are sometimes called *block q-ary* codes of *length n*.

Example 4.1. Let $C = \{000000, 111000, 000111, 111111\}$. This is a block binary code of length 6. If sender and receiver agree on C ahead of time, note that transmission of one of the four codewords is robust against "mild" errors.

Definition 4.2. For $u, v \in \mathbb{A}^n$, the *Hamming distance* from u to v is

$$d(u, v) = |\{i : u_i \neq v_i\}|,$$

the number of positions in which v differs from u.

Fact 4.3. *Hamming distance d is a metric on $\mathbb{A}^n$:*

- $d(u, v) = 0$ *iff* $u = v$;
- $d(u, v) = d(v, u)$ *for all* u, v;
- $d(u, w) \leq d(u, v) + d(v, w)$ *for all* $u, v, w \in \mathbb{A}^n$.

The first two properties above are obvious. The last item is known as the *triangle inequality*.

Proof of triangle inequality: If A, B, C denote the sets of positions (indices) where, respectively, u and w differ, u and v differ, v and w differ, then it is clear that $A \subseteq B \cup C$. So

$$d(u, w) = |A| \leq |B \cup C| \leq |B| + |C| = d(u, v) + d(v, w).$$

□

This idea of considering sets of positions is useful. With a group structure on $\mathbb{A}$, observe $u - v$ has zero entries in positions where u and v agree, and nonzero entries where they disagree.

For $w \in \mathbb{F}_q^n$, its *support* is $\mathrm{supp}(w) = \{i : w_i \neq 0\}$ and its *weight* is $\mathrm{wt}(w) = |\mathrm{supp}(w)|$. Note that, using the addition structure on $\mathbb{F}_q^n$, we have $d(u, v) = \mathrm{wt}(u - v)$.

4.2 Balls, errors, minimum distance

Definition 4.4. For $u \in \mathbb{F}_q^n$ and $t \geq 0$, the *ball* of radius t centred at u is

$$B_t(u) = \{v : d(u, v) \leq t\}.$$

Alternatively, $B_t(u) = \{u + w : w \in \mathbb{F}_q^n, \mathrm{wt}(w) \leq t\}$, showing that $|B_t(u)| = |B_t|$ is independent of the centre u.

Fact 4.5. *For integers t,*

$$|B_t| = \sum_{i=0}^{t} \binom{n}{i} (q-1)^i.$$

Example 4.6. In $\mathbb{F}_3^4$, $B_1(0) = \{0000, 1000, 2000, 0100, 0200, 0010, 0020, 0001, 0002\}$.

The Hamming distance and its balls are designed nicely for error detection and correction.

Definition 4.7. A code C *detects the error pattern* $e \in \mathbb{F}_q^n$ if $u + e \notin C$ for all $u \in C$. We say that C is *t-error-detecting* if t is the maximum integer such that C corrects all error patterns of nonzero weight $\leq t$.

(It helps to agree that any code automatically detects error pattern $e = 0$.) In terms of balls, C is t-error-detecting if and only if $B_t(u) \cap C = \{u\}$ for all $u \in C$, but that this fails for balls of radius $t + 1$.

Example 4.8. The code $C = \{000000, 111000, 000111, 111111\} \subseteq \mathbb{F}_2^6$ is 2-error-detecting. Some, but not all, error patterns of weight 3 get detected. For instance, C detects $e = 100110$.

Definition 4.9. A code C *corrects the error pattern* $e \in \mathbb{F}_q^n$ if $d(u + e, u) < d(u + e, v)$ for all $u \neq v \in C$. Say that C is *t-error-correcting* if t is the maximum integer such that C corrects all error patterns of weight $\leq t$.

In other words, with a t-error-correcting code, up to t errors (but no more) still leave a perturbed word closer to the original codeword than to any other codeword. This means the receiver can assume (and be right) that the smallest possible number of errors occurred when comparing received words to codewords. This is known as "maximum likelihood decoding", where a received word is corrected to the nearest codeword.

Example (cont'd). The code C from before is at least 1-error-correcting, since wlog 100000 is closer to 000000 than any other codeword. But, say,

$$d(110000, 000000) = 2 \not< 1 = d(110000, 111000).$$

In terms of balls, we can symmetrize the definition with respect to u and v.

Proposition 4.10. *A code C is $\geq t$-error-correcting if and only if $B_t(u) \cap B_t(v) = \emptyset$ for all $u \neq v \in C$.*

Proof. Consider an error pattern e of weight $\leq t$ so that $u + e \in B_t(u)$. Suppose C fails to correct e. Then there exists $v \in C$ such that $d(u + e, v) \leq d(u + e, u) \leq t$. It follows that $u + e \in B_t(v)$. In particular, $B_t(u) \cap B_t(v) \neq \emptyset$.

Conversely, suppose $B_t(u) \cap B_t(v) \neq \emptyset$. We assert the existence of $w \in B_t(u) \cap B_t(v)$ so that $d(w, u) \geq d(w, v)$. The conclusion is then that C fails to detect $e = w - u$, which of course has weight $\leq t$ since $w \in B_t(u)$.

It remains to justify the existence of w. Pick $w \in B_t(u) \cap B_t(v)$ so that $d(w, v)$ is minimized. If the property does not already hold for w, we may pick a position, say the ith, such that $w_i = u_i$ but $w_i \neq v_i$. If we replace w_i by v_i in w, the resulting word $\tilde{w}$ is closer to v but still in the intersection of balls. This is a contradiction. □

The capability of a code to both detect and correct errors can be summarized with one very important parameter.

Definition 4.11. The *minimum distance* of $C \subseteq \mathbb{A}^n$ is

$$d_{\min}(C) = \min\{d(u, v) : u \neq v \in C\}.$$

Theorem 4.12.

- *C is t-error-detecting iff $d_{\min}(C) = t + 1$.*
- *C is t-error-correcting iff $d_{\min}(C) = 2t + 1$ or $2t + 2$.*

4.3 Bounds on code sizes

The basic problem in coding theory is to maximize $|C|$ subject to a given length n, alphabet size q, and required minimum distance $\geq d$. Here are some naive bounds on $|C|$ in terms of the other parameters.

Theorem 4.13 (Hamming bound)**.** *If $C \subseteq \mathbb{F}_q^n$ with $d_{\min}(C) = d$, then*

$$|C| \leq \frac{|\mathbb{F}_q^n|}{|B_{\frac{d-1}{2}}|} = \frac{q^n}{\sum_{i=0}^{(d-1)/2} \binom{n}{i}(q-1)^i}.$$

Proof. For $d = 2t + 1$ or $2t + 2$, we have seen that C is t-error-correcting. Hence $B_t(u)$, $u \in C$, are disjoint balls. It follows that

$$|\mathbb{F}_q^n| \geq \left| \bigcup_{u \in C} B_t(u) \right| = \sum_{u \in C} |B_t(u)| = |C| \cdot |B_t|.$$

□

Example 4.14. Any binary code of length 6 and minimum distance 3 (i.e. $t = 1$) has size

$$|C| \leq \frac{2^6}{\binom{6}{0} + \binom{6}{1}} = \frac{64}{7} \approx 9.14.$$

Therefore, by integrality, $|C| \leq 9$.

Next is a dual lower bound.

Theorem 4.15 (Gilbert-Varshamov bound)**.** *There is a code $C \subseteq \mathbb{F}_q^n$ with $d_{\min}(C) \geq d$ and*

$$|C| \geq \frac{|\mathbb{F}_q^n|}{|B_{d-1}|} = \frac{q^n}{\sum_{i=0}^{d-1} \binom{n}{i}(q-1)^i}.$$

Proof. Initialize $U_1 = \mathbb{F}_q^n$ as the set of "available" words. For $i \geq 1$, choose any $u_i \in U_i$, include it in C, and put $U_{i+1} = U_i \setminus B_{d-1}(u_i)$. It is clear that

$$|U_{i+1}| \geq |U_1| - \left| \bigcup_{j \leq i} B_{d-1}(u_j) \right| \geq q^n - i|B_{d-1}|.$$

So the process can continue until (at least) the right hand side becomes non-positive; that is, until $i \geq q^n/|B_{d-1}|$. By construction, the words u_i chosen are at minimum pairwise distance $\geq d$, since if $d(u_i, u_j) \leq d-1$ for $j < i$, then u_i was removed from being available at step j. □

Example 4.16. There is a binary code of length 6 and minimum distance ≥ 3 guaranteed by the GV bound of size at least

$$\frac{2^6}{\binom{6}{0} + \binom{6}{1} + \binom{6}{2}} = \frac{64}{22} \approx 2.9.$$

Rounding up, we can attain three codewords. But note that applying the algorithm rather easily gets $|C| \geq 4$.

step # i	$\lvert U_i \rvert$	u_i
1	64	000000
2	42	111000
3	26	000111
4	≥ 4	111111
5	0	

In fact, there is such a code of size 5. At step 3 there is an intelligent choice: 100110. One gets

$$C = \{000000, 111000, 100110, 110011, 011111\}.$$

Theorem 4.17 (Singleton bound)**.** *If $C \subseteq \mathbb{F}_q^n$ with $d_{\min}(C) = d$, then*

$$|C| \leq q^{n-d+1}.$$

Proof. Deleting the last $d-1$ positions from each codeword in C yields a family of $|C|$ distinct words in $\mathbb{F}_q^{n-(d-1)}$. □

It is not hard to see that the Singleton bound is already implied by the Hamming bound for $q = 2$. But in general it may be the stronger bound.

When equality holds in the Hamming bound, C is called *perfect*. We'll see these more later on.

When equality holds in the Singleton bound, C is called *maximum distance separable*, or an "MDS code". It is easy to see that an MDS code with parameters q, n, d is equivalent to an $\mathrm{OA}(n-d+1, n, q)$.

Exercises

1. Suppose the code $C = \{000000, 000111, 111000, 111111\}$ is used over a binary channel (alphabet $\mathbb{F}_2$). Suppose the "reliability" is $p = 0.9$, so that the probability that any bit is changed is 0.1.

   For each codeword $w \in C$, compute the probability that w was sent, given that the word 000110 is received. (Your final answer should consist of four probabilities which sum to 1.)

2. If a value t does not appear as a distance in C, argue that C detects every error pattern of weight exactly t. On the other hand, argue that missing distances do not help for error correction.

3. Show that a binary code can be used to correct any combination of $d_{\min}(C) - 1$ *erasure* errors, in which the affected symbols are received as '$\star$' instead of '0' or '1'.

4. Let $A(n, d)$ denote the maximum size of a binary code of length n and minimum distance d. Prove that, for odd d, we have $A(n, d) = A(n + 1, d + 1)$.

5. (a) Let H_w be the set of words in $\mathbb{F}_2^n$ with weight w. For $u \in H_w$, compute $|B_t(u) \cap H_w|$.

   (b) A *constant weight* code C has the property that all of its codewords have the same weight. Using (a), state and prove an analog of the Hamming bound for constant weight binary codes of length n, minimum distance d, and weight w.

6. A q-ary code has *constant composition* if each of its words has the same number of occurrences of each symbol. Find optimal constant composition ternary codes of length 6, composition $[3, 2, 1]$, and each possible minimum distance.

7. Find a constant weight binary code of largest possible minimum distance, with length 13, size 13, and such that every codeword has weight 4. (*Hint*: Use a $(13, 4, 1)$-difference set.)

8. The *information rate* of a q-ary code C of length n is defined to be $\frac{1}{n} \log_q |C|$. Prove that the information rate of a 1 error-correcting code is at most $1 - \frac{1}{n} \log_q(n(q-1)+1)$.

9. Take $\mathbb{A} = \mathbb{Z}/q\mathbb{Z} = \{0, 1, \ldots, q-1\}$ and define $[x] = \min(x, q - x)$ for $x \in \mathbb{A}$. Define a function $\lambda : \mathbb{A}^n \times \mathbb{A}^n \to \mathbb{Z}$ by

$$\lambda(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} [x_i - y_i].$$

   Prove that λ is a metric.

10. (a) Prove that, for fixed q and d, the Hamming bound is better than the Singleton bound for sufficiently large n.

    (b) With $q = 3$ and $d = 6$, for what value of n does the Hamming bound start to be better than the Singleton bound?

11. Research the *Reed-Soloman codes* and investigate their performance with respect to the Singleton bound.

Chapter 5

Linear Codes

5.1 Preliminaries

Definition 5.1. A code $C \subseteq \mathbb{F}_q^n$ is *linear* if it forms a subspace of $\mathbb{F}_q^n$. That is, linear codes are nonempty with

- $u + v \in C$ whenever $u, v \in C$ and
- $\alpha u \in C$ whenever $\alpha \in \mathbb{F}_q$ and $u \in C$.

Note that if q is prime, the second condition can actually be dropped, since

$$ku = \overbrace{u + \cdots + u}^{k \text{ times}}.$$

Fact 5.2. *In a linear code C, its minimum distance equals the least weight of a nonzero codeword.*

Proof. Suppose $u \in C$ has least nonzero weight. Then $d_{\min}(C) \leq d(0, u) = \mathrm{wt}(u)$. On the other hand, if $v, w \in C$ satisfy $d(v, w) = d_{\min}(C)$, then $v - w$ is a nonzero codeword by linearity, and $\mathrm{wt}(u) \leq \mathrm{wt}(v - w) = d_{\min}(C)$. □

Remark. In both computational and theoretical settings, it is helpful to know that a code is linear for computing $d_{\min}(C)$.

Naturally, the notions of linear independence, span, basis and dimension are used for linear codes.

Example 5.3. $C = \{000000, 111000, 000111, 111111\} \subset \mathbb{F}_2^6$ is linear, of dimension 2, and any two of the three nonzero codewords form a basis for C.

We usually present a linear code simply by supplying a basis for it. Concretely finding a basis for $C = \langle S \rangle$ given a generating set S can be done in two slightly different ways. You have already seen this in a first linear algebra course.

Row method

Make a matrix whose rows are the vectors in S, row-reduce to RREF, and take the nonzero rows as a basis.

Advantage: The resulting basis has 'pivots'. This gives a certificate for linear independence and also allows for explicit reading of information bits.
Disadvantage: This basis has possibly no relationship to the given vectors, and so it is not obvious that it spans the same code.

Example 5.4. Let $S = \{123, 314, 111, 104\} \subset \mathbb{F}_5^3$. We have

$$\begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 4 \\ 1 & 1 & 1 \\ 1 & 0 & 4 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

It follows that $\langle S \rangle$ has basis $\{104, 012\}$ and dimension 2.

Column method

Make a matrix whose columns are the vectors in S, row-reduce to RREF, and select the corresponding pivot columns in the original matrix.

Advantage: The resulting basis is a subset of the given set of vectors.
Disadvantage: It is possibly not obvious on inspection that this set of vectors is linearly independent.

Example 5.5. With the same set S, We have

$$\begin{bmatrix} 1 & 3 & 1 & 1 \\ 2 & 1 & 1 & 0 \\ 3 & 4 & 1 & 4 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 3 & 0 & 4 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

It follows that $\langle S \rangle$ has basis $\{123, 111\} \subset S$.

We normally reserve the parameter k for the dimension of a linear code.

Definition 5.6. A *generator matrix* for a linear code $C \subseteq \mathbb{F}_q^n$ of dimension k is any $k \times n$ matrix G whose rowspace equals C. Alternatively, the rows of G form a basis for C.

Given a generator matrix G, we have

$$C = \{wG : w \in \mathbb{F}_q^k\}.$$

It is easy to see that, after permutation of the columns if necessary, we can choose a generator matrix G for C which is in *standard form* $G = [\ I \mid X\ ]$. In that case,

$$C = \{(w, wX) : w \in \mathbb{F}_q^k\},$$

where codewords are a concatenation of 'information bits' w with 'check bits' wX.

5.2 Duals and parity check matrices

The dual of a linear code C, denoted $C^\perp$, is the usual vector space orthogonal complement:

$$C^\perp = \{v \in \mathbb{F}_q^n : u \cdot v = 0 \text{ for all } u \in C\}.$$

Although $C \cap C^\perp \neq \{0\}$ in general, it is true that $\dim(C^\perp) = n - k$ when $\dim(C) = k$.

Example 5.7. The linear code over $\mathbb{F}_4 = \{0, 1, \alpha, \beta\}$ with generator matrix

$$G = \begin{bmatrix} 1 & 0 & \alpha & \beta \\ 0 & 1 & \beta & \alpha \end{bmatrix}$$

is self-dual.

Definition 5.8. A *parity check matrix*, usually denoted H, for a linear code $C \subseteq \mathbb{F}_q^n$ is any generator matrix for the dual code $C^\perp$.

Fact 5.9. *A linear code $C \subseteq \mathbb{F}_q^n$ is the null space of its parity check matrix:*

$$C = \{u \in \mathbb{F}_q^n : Hu^\top = 0^\top\}.$$

We have $H \in \mathbb{F}_q^{(n-k)\times n}$ and $GH^\top = O$. Also, $H^\top$ can be computed as the 'nullspace matrix' of G.

Proposition 5.10. *If G is in standard form, say $[\ I_k \mid X\ ]$, then $H = [-X^\top | \, I_{n-k}\,]$.*

Example 5.11 (The cup game)**.** A magician displays a rectangular grid of cups, some up and some down. A volunteer flips one cup without the magician looking. The magician is able to quickly identify which cup is flipped because in his initial arrangement, every row and every column has an even number of cups in the 'up' position. After the volunteer flips some cup, (only) that row and that column have changed parity.

Let's represent this game in the case of a 3×3 arrangement of cups. Each arrangement corresponds to a 3×3 binary matrix, which for convenienece we write in vectorized form (as concatenation of rows); for instance

$$\begin{array}{|c|c|c|}\hline 0 & 1 & 1 \\ \hline 1 & 0 & 1 \\ \hline 0 & 0 & 0 \\ \hline \end{array} \mapsto 011101000.$$

The magician's special arrangements in which every row and column have an even number of '1's comprise a linear code $C \subseteq \mathbb{F}_2^9$. Here are two descriptions of C. First, we have four basis words in C which correspond to the standard basis for upper-left 2×2 submatrices. Check bits are added to the right and below. This produces a generator matrix for C as

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

Columns 1, 2, 4 and 5 of G form a sub-identity matrix I_4. Transpose the other columns (and negate) to build the corresponding parity-check matrix

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

The description of C via H consists of relations, namely that the first row, second row, first column, second column, and overall sum is even.

For $w \in \mathbb{F}_q^n$, its *syndrome* with respect to C is $Hw^\top \in \mathbb{F}_q^{n-k}$. Syndromes are in correspondence with cosets of (i.e. affine translates of) C in $\mathbb{F}_q^n$.

If word $w \in \mathbb{F}_q^n$ is received, we can compute its syndrome $Hw^\top$ and identify a least weight representative $e = w^*$ for the associated coset $e + C = w + C$. We can make a good guess that the sent word is $u = w - e \in C$. This is called *syndrome decoding*.

Example 5.12. Consider the linear code $C = \{0000, 1011, 0101, 1110\} \subset \mathbb{F}_2^4$. Its cosets are listed as rows in the following array, where least weight representative 'coset leaders' begin each row.

$$\begin{array}{cccc} 0000 & 1011 & 0101 & 1110 \\ 1000 & 0011 & 1101 & 0110 \\ 0100 & 1111 & \underline{0001} & 1010 \\ 0010 & 1001 & 0111 & 1100 \end{array}$$

Observe that the third coset contains a second weight-one word (underlined); consequently, this code fails to correct errors in positions 2 or 4.

5.3 Minimum distance for linear codes

Recall that, for a linear code C, its minimum distance is the least weight of a nonzero codeword. In terms of a parity check matrix H,

$$d_{\min}(C) = \min\{\mathrm{wt}(u) : Hu^\top = 0^\top, u \neq 0\}.$$

Definition 5.13. For a matrix $A \in \mathbb{F}^{m \times n}$ its *spark* is the least integer k such that there exists a set of k linearly dependent columns in A. (If A has linearly independent columns, we could define its spark as ∞.)

It is easy to see that the minimum distance of a linear code C equals the spark of its parity check matrix: $d_{\min}(C) = \mathrm{spark}(H)$.

Here is a 'random' construction of matrices with elements in $\mathbb{F}_q$ having a lower bound on their spark.

Proposition 5.14. *Let $d \leq m$ be positive integers. There exists a matrix in $\mathbb{F}_q^{m\times n}$ with spark at least d if*

$$q^m > \sum_{i=0}^{d-2} \binom{n-1}{i} (q-1)^i.$$

Proof. We build such a matrix column by column. To get started, the $m \times m$ identity matrix has spark at least d. Now, suppose an $m \times r$ matrix of spark $\geq d$ has been built for some $r \geq m$. It can be extended by an additional column vector in $\mathbb{F}_q^m$ provided that vector is not a linear combination of $d-2$ or fewer of the already chosen vectors. That is, we can extend if

$$q^m > \sum_{i=0}^{d-2} \binom{r}{i} (q-1)^i.$$

As a polynomial in r, the left side is increasing for $r \geq d$. So if the given inequality holds (with $n-1$ in place of r), then our matrix can be enlarged to have n columns. □

With $m = n - k$, it follows that we can build the columns of a parity check matrix H affording minimum distance d provided

$$q^{n-k} > \sum_{i=0}^{d-2} \binom{n-1}{i} (q-1)^i.$$

Corollary 5.15 (Gilbert-Varshamov for linear codes)**.** *There exists a linear code $C \subseteq \mathbb{F}_q^n$ with minimum distance $\geq d$ and size*

$$|C| \geq \frac{q^{n-1}}{\sum_{i=0}^{d-2} \binom{n-1}{i}(q-1)^i}.$$

Remark. Note that this denominator is the size of the $(d-2)$-ball in dimension $n-1$. Let's call this b_{d-2}^{n-1}.

Proof. We can take $|C|$ to be the power of q satisfying $q^{n-1} \leq |C| \cdot b_{d-2}^{n-1} < q^n$. □

Example 5.16. Consider $\mathbb{F}_2^{15}$ with a prescribed minimum distance $d = 4$. Since

$$\binom{14}{0} + \binom{14}{1} + \binom{14}{2} = 106 < 2^7,$$

it follows that there exists a code with dimension 8 $(= 15 - 7)$ and minimum distance 4.

Example 5.17. With $k = n - 2$, the largest ball we can take is with radius $d - 2 = 1$. This gives $1 + (n-1)(q-1) < q^2$. This is satisfied for $n = q + 1$ with

$$H = \left[\begin{array}{c|cc} \mathbf{1} & 1 & 0 \\ \mathbb{F}_q^\times & 0 & 1 \end{array}\right],$$

where the second row of H begins with a listing of the nonzero field elements. Check that, in H, no column is a multiple of any other. So $\text{spark}(H) = 3$.

Exercises

1. Let X be an abelian group, written additively. Suppose d is a metric on X such that $d(x_1, x_2) = d(x + x_1, x + x_2)$ for all $x, x_1, x_2 \in X$. Define $M_{\geq k}$ as the maximum size of a subset $Y \subseteq X$ where $d(y_1, y_2) \geq k$ for all $y_1 \neq y_2$ in Y. Define $M_{<k}$ analogously. Prove that
$$(M_{\geq k})(M_{<k}) \leq |X|.$$

2. Let $C \subseteq \mathbb{F}_q^n$ be a linear code.
   (a) Show that the error patterns which C detects are precisely the words in its complement $\mathbb{F}_q^n \setminus C$.
   (b) Show that the number of error patterns which C can correct is at most $q^n/|C|$.

3. Let $S = \{11000, 01111, 11110, 01010\} \subset \mathbb{F}_2^5$ and put $C = \langle S \rangle$.
   (a) Find both a generator matrix and a parity check matrix for C.
   (b) What is the dimension of C and of $C^\perp$, the dual code (subspace)?
   (c) Determine the minimum distance of C.

4. Let C be the set of all zero-sum vectors in $\mathbb{F}_q^n$. Show that C is linear and repeat (a) through (c) above.

5. Suppose there exists a binary linear code of length n, dimension k, and minimum distance d. Prove that $n \geq \sum_{i=0}^{k-1} \left\lceil \frac{d}{2^i} \right\rceil$.

6. For the linear code $C = \{u_1u_2u_3u_4 \in \mathbb{F}_3^4 : u_1 + u_2 = u_3 + u_4 = 0\}$, find the least weight representatives for each coset.

7. Apply the 'linear code' version of the Gilbert-Varshamov bound to find a binary linear code C of length 9, dimension 2, and distance 5. Can you find a larger non-linear code?

8. (a) Find a matrix $H \in \mathbb{F}_3^{5\times 7}$ having spark (smallest size of a linearly dependent set of columns) equal to 5.

   (b) Show that the linear code resulting from (a) is optimal with respect to the Hamming bound.

9. Show that $\text{spark}(H) > k$ if and only if every $k \times k$ principal submatrix of $H^\top H$ is nonsingular.

10. (a) Show that there is a set of $n = 1 + \lfloor\sqrt{2q}\rfloor$ threewise independent vectors in $\mathbb{F}_q^3$; that is, a $3 \times n$ matrix with spark at least 4.

    (b) Do a bit better than the bound in (a) for $q = 5$.

    (c) What are the parameters of the linear code resulting from your answer to (b)?

11. The *weight enumerator* of a binary code C of length n is

$$A(x, y) = \sum_{i=0}^{n} w_i x^i y^{n-i},$$

where w_i is the number of words in C of weight i.

    (a) For the linear code $C = \{0000, 1111\}$ and its dual $C^\perp$, find the weight enumerators $A(x, y)$ and $A^\perp(x, y)$.

    (b) Verify in this case the *MacWilliams Identity*, namely that

$$A^\perp(x, y) = \frac{1}{2^k} A(y - x, x + y).$$

Chapter 6

Perfect Codes

Recall that a code $C \subseteq \mathbb{F}_q^n$ is *perfect* if equality holds in the Hamming bound. That is, C is perfect if it has size $|C| = q^n/|B_t|$, where $t = \lfloor (d_{\min}(C) - 1)/2 \rfloor$.

There are some trivial examples. First, $\mathbb{F}_q^n$ is perfect for any n with $t = 0$ (note that $|B_0| = 1$). At the other extreme, a code consisting of just one word is perfect if we take $t = \infty$ (so that $|B_t| = q^n$). Finally, the two-word binary code $\{\mathbf{0}, \mathbf{1}\}$ is perfect for odd length $n = 2t + 1$. It is a well-known binomial identity that $|B_t| = \sum_{i=0}^{t} \binom{n}{i} = 2^{n-1}$ in this case.

6.1 The Hamming codes

Let $r \geq 2$. A *Hamming code* is a linear code $C \subseteq \mathbb{F}_q^n$, where $n = (q^r - 1)/(q - 1)$, with a parity check matrix consisting of n columns of height r which form representatives for the 1-dimensional subspaces (lines through $\mathbf{0}$) in $\mathbb{F}_q^r$.

When $q = 2$, we simply have $n = 2^r - 1$ and the columns are all nonzero binary r-tuples. When $q > 2$, the columns are all possible "direction vectors", which can be assumed normalized of the form $(\underbrace{0, \ldots, 0}_{r-1-i}, 1, \underbrace{*, \ldots, *}_{i})^\top$. There are q^i choices for each i, and $n = \sum_{i=0}^{r-1} q^i$ columns in total.

Example 6.1. Suppose $q = 2$. For $r = 2$, we have

$$H = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

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

For $r = 3$, we have $n = 2^3 - 1 = 7$,

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

and a the resulting Hamming code is sometimes called C_7. (Check that $|C_7| = 16$.) Note that the lower right 2×3 submatrix appeared above in the case $r = 2$.

Example 6.2. Consider the Hamming code with $q = 3$, $r = 3$. It has length $n = 1 + 3 + 3^2 = 13$ and parity check matrix

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 2 & 1 & 1 & 1 & 0 \\ 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 & 1 \end{bmatrix}.$$

We see that $\dim C = n - r$, since H is $r \times n$. Therefore, $|C| = q^{n-r}$.

Regarding the minimum distance, observe that no nonzero linear combination of 2 columns of H is ever $\mathbf{0}$ (by construction). It follows that $d_{\min}(C) \geq 3$. On the other hand, plenty of triples of columns of H combine to $\mathbf{0}$. So in fact $d_{\min}(C) = 3$. That is, C is 1-error-correcting.

In fact, Hamming codes are as large as possible with this minimum distance.

Theorem 6.3. *The Hamming codes are perfect 1-error-correcting q-ary codes of length $n = (q^r - 1)/(q - 1)$ and dimension $n - r$.*

Proof. The relevant balls have size $|B_1| = 1 + (q - 1)n = q^r$. Therefore,

$$|C| = q^{n-r} = \frac{q^n}{|B_1|},$$

achieving equality in the Hamming bound. □

6.2 The Golay codes

Let $N \in \{0,1\}^{11\times 11}$ defined by

$$N_{ij} = \begin{cases} 1 & \text{if } i-j \text{ is a nonzero quadratic residue (mod 11),} \\ 0 & \text{otherwise.} \end{cases}$$

More explicitly, this is the 'back-circulant' matrix

$$N = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

There are several elegant structural properties of N. The first row of N is the characteristic vector of the set Q of quadratic residues in $\mathbb{F}_{11}$, indexed from 0. This set Q is a 'difference set': any nonzero element of $\mathbb{F}_{11}$ appears equally often (twice) as a difference between two elements of Q.

As a related fact, our matrix N is also an incidence matrix for the $(11,5,2)$ symmetric design, or biplane. A diagram of this structure is given in Figure 6.2.

Binary Golay Code

Let B be the 12×12 matrix over $\mathbb{F}_2$ constructed as

$$B = \begin{bmatrix} 0 & \mathbf{1} \\ \mathbf{1}^\top & J-N \end{bmatrix},$$

where J is the 11×11 all ones matrix and $\mathbf{1}$ is the 1×11 row of all ones. Note that $B = B^\top$, since the same is true for N. Also, from the biplane model of N, we obtain $B^2 = BB^\top \equiv I \pmod 2$ after a small calculation.

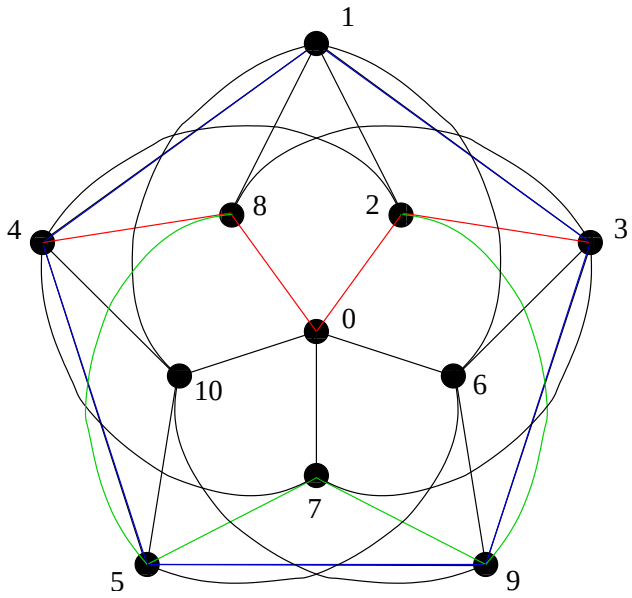


Figure 6.1: The (11,5,2) biplane

Definition 6.4. The extended binary Golay code, C_{24}, is the linear code with generator matrix $G = [I_{12} \mid B]$.

It follows easily from this definition that and the facts above that C_{24} is self-dual ($GG^\top = O$) with length 24 and dimension 12. We now consider the minimum distance.

Theorem 6.5. *The extended Golay code C_{24} has minimum distance equal to* 8.

Proof. We begin by proving that the weight of any codeword $u = xG \in C_{24}$ is a multiple of 4. This is done by induction on the weight of x. Since every row of G has weight 8 or 12, the result follows for $\mathrm{wt}(x) \leq 1$.

Now, suppose for some $t \geq 1$ that $\mathrm{wt}(u = xG) \equiv 0 \pmod 4$ whenever $\mathrm{wt}(x) = t$. Let v be another row of G. Then

$$\mathrm{wt}(u + v) = \mathrm{wt}(u) + \mathrm{wt}(v) - 2(u \cdot v) \equiv 0 \pmod 4,$$

since any two different rows of G are orthogonal.

Since most rows of G have weight 8, it remains to prove that C_{24} has no codewords of weight 4. Suppose for contradiction that $v \in C_{24}$ has weight 4. Since $B^2 = I$, it follows

that $[I \mid B]$ and $[B \mid I]$ are each generator matrices for C_{24}. Therefore

$$v = [w_1 \mid w_2] = w_1[I \mid B] = w_2[B \mid I]$$

for some $w_1, w_2 \neq 0$. Now neither of the two halves of v can be identically zero. This is because of the identity matrices in the generator matrices and also since $w_1, w_2 \neq 0$. Further, if either half of v contained only one 1, this would imply that v equalled a row of either $[I \mid B]$ or $[B \mid I]$, but each row has weight at least eight. Therefore each half of v must contain exactly two ones. This implies that $\mathrm{wt}(w_1) = \mathrm{wt}(w_2) = 2$, but the sum of two rows of B has weight at least 4. Therefore $\mathrm{wt}(v) = \mathrm{wt}(w_1) + \mathrm{wt}(w_1 B) > 2 + 4 > 4$, a contradiction. We have shown no $v \in C_{24}$ has weight 4 and so $d_{\min}(C_{24}) = 8$. □

The *binary Golay Code* C_{23} is obtained by a small operation from C_{24}.

Definition 6.6. *Puncturing a code* means removing the entry in a common position from every codeword. For example, puncturing with respect to the last position of C gives $C' = \{u_1 \cdots u_{n-1} : u_1 \cdots u_n \in C\}$.

Now C_{23} is obtained by puncturing C_{24}. It doesn't matter which bit is removed; equivalent codes will result. In particular, C_{23} retains dimension $k = 12$. Since some words of weight 8 in C_{24} have been reduced to weight 7 after puncturing, we get $d_{min}(C_{23}) = 7$.

Theorem 6.7. *The binary Golay code C_{23} has length 23, dimension 12, and minimum distance 7. Since $|B_3| = 1 + \binom{23}{1} + \binom{23}{2} + \binom{23}{3} = 2^{11}$, this is a perfect code.*

Ternary Golay Code

There is a ternary perfect code which is a cousin of C_{23}. Let's work from N just as before, except we order the rows of N so that it is forward-circulant (instead of back-circulant as presented earlier). With a different labelling, N is also a point-line incidence matrix for the biplane.

Now take the matrix N' over $\mathbb{F}_3$, where $N' = 2N + 2I - J$ (here N is interpreted as a zero-one matrix over $\mathbb{F}_3$). This is just the matrix obtained from N by placing 1s on the

main diagonal and changing all other "0"s to "−1"s. Next, put

$$B' = \begin{bmatrix} -1 & -\mathbf{1} \\ -\mathbf{1}^\top & N' \end{bmatrix} = \begin{bmatrix} - & - & - & - & - & - & - & - & - & - & - & - \\ - & 1 & 1 & - & 1 & 1 & 1 & - & - & - & 1 & - \\ - & - & 1 & 1 & - & 1 & 1 & 1 & - & - & - & 1 \\ - & 1 & - & 1 & 1 & - & 1 & 1 & 1 & - & - & - \\ - & - & 1 & - & 1 & 1 & - & 1 & 1 & 1 & - & - \\ - & - & - & 1 & - & 1 & 1 & - & 1 & 1 & 1 & - \\ - & - & - & - & 1 & - & 1 & 1 & - & 1 & 1 & 1 \\ - & 1 & - & - & - & 1 & - & 1 & 1 & - & 1 & 1 \\ - & 1 & 1 & - & - & - & 1 & - & 1 & 1 & - & 1 \\ - & 1 & 1 & 1 & - & - & - & 1 & - & 1 & 1 & - \\ - & - & 1 & 1 & 1 & - & - & - & 1 & - & 1 & 1 \\ - & 1 & - & 1 & 1 & 1 & - & - & - & 1 & - & 1 \end{bmatrix}.$$

This B' is a 12×12 'Hadamard matrix', which implies that $(B')(B')^\top \equiv O \pmod 3$. Using this property, it is not hard to show that the rowspace of B' over $\mathbb{F}_3$, which we call the *extended ternary Golay code* and denote by C_{12}, is self-dual with $d_{min}(C_{12}) = 6$. (Similar to the analysis for C_{24}, we show that weights of codewords in C_{12} are multiples of three, and then rule out the case of weight three using structure of a Hadamard matrix.)

The (ordinary) *ternary Golay code* C_{11} is obtained by puncturing C_{12}.

Theorem 6.8. *The ternary Golay code C_{11} has length 11, dimension 6, and minimum distance 5. Since $|B_2| = 1 + 2\binom{11}{1} + 2^2\binom{11}{2} = 3^5$, this is a perfect code.*

6.3 Classification

So far, we have seen three types of perfect codes:

- the trivial codes $C = \mathbb{F}_q^n$ and $C = \{00\ldots0, 11\ldots1\}$ in $\mathbb{F}_2^{2k+1}$.
- the Hamming codes of length $n = (q^r - 1)/(q - 1)$, which are 1-error-correcting
- the Golay codes C_{11} and C_{23}, which are 2 and 3-error-correcting, respectively.

It transpires that there are no multiple-error-correcting perfect linear codes besides the two Golay codes. The following classification was proved in stages by Perko, Tietäväinen and van Lint.

Theorem 6.9 (Classification of Perfect Codes)**.** *For $t > 1$, the only nontrivial perfect t-error-correcting codes have the parameters of either C_{11} or C_{23}.*

The proof of Theorem 6.9 is quite technical, so let's instead have a look at some elementary aspects of the problem.

First, a perfect code cannot have an even minimum distance: Let C be a code with minimum distance $d_{\min}(C) = 2t+2$. Let $v \in C$ and change v in $t+1$ places to obtain a new word z. Therefore $d(v, z) = t+1$. Let $u \in C$, with $u \neq v$. Then $d(u, v) \leq d(u, z) + d(z, v)$ or $d(u, z) \geq d(u, v) - d(z, v) \geq 2t + 2 - (t + 1) = t + 1$. This implies that z has distance at least $t + 1$ from every codeword in C. Therefore z is in no ball of radius t centred at the codewords in C. This contradicts C being perfect.

Also, recall that for a perfect q-ary code of length n and minimum distance $2t + 1$, we must have $|B_t| \mid q^n$. This rather strong condition is necessary simply for the Hamming bound to be an integer.

Next, there is an even stronger constraint on perfect codes involving polynomials. The origins of this come from the theory of association schemes.

Definition 6.10. For parameters q, n, t, the *Lloyd polynomial* is

$$L_t(x) = \sum_{i=0}^{t} (-1)^i (q-1)^{t-i} \binom{n-x}{t-i} \binom{x-1}{i}.$$

Note $L_t(x)$ is a polynomial of degree t in x, since each term of the sum is a polynomial of degree t having the same sign $(-1)^t$.

Theorem 6.11 (Lloyd's Theorem)**.** *If C is a q-ary perfect code of length n with $d_{\min}(C) = 2t + 1$, then Lloyd's polynomial $L_t(x)$ has t distinct integral zeros among $1, 2, \ldots, n-1$.*

With an elementary argument, we can use Lloyd's theorem to rule out perfect codes in the case $q = 2$, $t = 2$ (binary codes of minimum distance 5). On the other hand, it is noteworthy that $|B_2| = 1 + n + \binom{n}{2}$ sometimes does divide 2^n. For instance when $n = 90$, $1+90+\binom{90}{2} = 2^{12}$. (That is, a naïve 'number-theoretic' argument fails to kill these codes.)

Corollary 6.12. *There are no nontrivial perfect 2-error-correcting binary codes.*

Proof. First, we compute

$$L_2(x) = 2x^2 - 2(n+1)x + 1 + n + \binom{n}{2}.$$

The constant coefficient is $|B_2|$, which must be a power of 2. So we may assume, by Lloyd's Theorem, that $(x-2^a)(x-2^b) = x^2-(n+1)x+2^s$ for some integers s, a, b with $0 \le a < b$ and $s = a+b$. Comparing coefficients of x, we see $2^a + 2^b = n+1$. Then a calculation gives

$$(2^{a+1}+2^{b+1}-1)^2 = (2n+1)^2 = 8|B_2| - 7 = 2^{a+b+4} - 7.$$

Suppose $a, b \ge 2$. Modulo 16, the left side is $(-1)^2 = 1$ and the right side is $-7 = 9$. So it follows that $a = 1$ and

$$(2^{b+1}+3)^2 = 2^{b+5} - 7.$$

It is simple to verify that the only solution is $b = 2$, for which $n = 2^1 + 2^2 - 1 = 5$. This leads to the trivial repetition code $\{00000, 11111\}$. □

In a similar spirit, the full proof of Theorem 6.9 involves a careful analysis of the integrality condition on zeros of $L_t(x)$.

Exercises

1. Verify that $\{00\cdots0, 11\cdots1\}$ is perfect for odd lengths $n = 2k+1$.
2. Use C_{24} to construct a Steiner system $S(5, 8, 24)$. Conclude the existence of $S(4, 7, 23)$ and $S(3, 6, 22)$.
3. Count the codewords of weight 7 in the Golay code C_{23}. (*Hint*: Start by proving that every word of weight 4 in $\mathbb{F}_2^{23}$ is distance 3 from exactly one codeword.)
4. Show that puncturing C_{24} at different positions results in equivalent codes.
5. Complete the proof that $d_{min}(C_{12}) = 6$, and conclude that C_{11} is a perfect 2-error-correcting ternary code.
6. Let n, t be fixed integers with $n > t \ge 1$. Prove that perfect t-error-correcting q-ary codes of length n fail to exist for sufficiently large primes q.
7. Show that the constant coefficient of the Lloyd polynomial $L_t(x)$ equals $|B_t|$.
8. Define the *Krawtchouk polynomials*

$$K_i(x) = \sum_{j=0}^{i} (-1)^j (q-1)^{i-j} \binom{x}{j} \binom{n-x}{i-j}.$$

   It turns out that $L_t(x) = \sum_{i=0}^{t} K_i(x)$. Verify this in the case $q = t = 2$.

9. Lloyd's theorem is proved using a certain algebra of matrices, which we develop here in the specific case of $\mathbb{F}_2^3 = \{000, 001, 010, 011, 100, 101, 110, 111\}$.

   For $t = 0, 1, 2, 3$, let A_t be the 8×8 zero-one matrix whose rows and columns are indexed by $\mathbb{F}_2^3$ (say in the order as above), where

   $$A_t(u, v) = \begin{cases} 1 & \text{if } d(u, v) = t, \\ 0 & \text{otherwise.} \end{cases}$$

   It turns out that the subspace of matrices $\mathfrak{A} := \operatorname{span}_{\mathbb{R}}(\{A_0, A_1, A_2, A_3\})$ they generate is closed under matrix multiplication.

   (a) Check that each A_i is symmetric and conclude that $\mathfrak{A}$ is commutative.

   (b) Express $A_1 A_2$ in the basis $\{A_0, A_1, A_2, A_3\}$.

   (c) Express A_2^2 in the basis $\{A_0, A_1, A_2, A_3\}$.

Chapter 7

Cyclic Codes

7.1 Introduction and classification

Definition 7.1. Let $w = a_0 a_1 \dots a_{n-1} \in \mathbb{F}_q^n$. The *cyclic shift* of w is

$$\sigma(w) = a_{n-1} a_0 a_1 \dots a_{n-2}.$$

A code $C \subseteq \mathbb{F}_q^n$ is *cyclic* if $\sigma(w) \in C$ whenever $w \in C$.

For example, one presentation C_7 of the binary Hamming code of length 7 is cyclic. We have

$$C_7 = \{0000000, \overline{1101000}, \overline{0010111}, 1111111\},$$

where the lines represent that all 7 cyclic shifts of these codewords are included.

It is easy to see that $\sigma : \mathbb{F}_q^n \to \mathbb{F}_q^n$ is a linear transformation. Thus, a linear code is cyclic if and only if $\sigma(w_i) \in C$ for each w_i in a generating set for C.

It is convenient to associate $w = a_0 a_1 \dots a_{n-1} \in \mathbb{F}_q^n$ with $w(x) = a_0 + a_1 x + \cdots + a_{n-1} x^{n-1} \in \mathbb{F}_q[x]$. Note $[(u+v)](x) = u(x) + v(x)$, where arithmetic on coefficients is of course done in $\mathbb{F}_q$. This polynomial correspondence also has nice properties with respect to cyclic shifts:

$$\begin{aligned} [\sigma(w)](x) &= a_{n-1} + a_0 x + a_1 x^2 + \cdots + a_{n-2} x^{n-1} \\ &= x w(x) - a_{n-1}(x^n - 1) \\ &\equiv x w(x) \pmod{x^n - 1}. \end{aligned}$$

So in this section, let's identify $\mathbb{F}_q^n$ with the vector space $\mathbb{F}_q[x]/\langle x^n - 1\rangle$. The latter has the advantage of a ring (algebra) structure. As usual, arithmetic of polynomials takes place mod $x^n - 1$.

Lemma 7.2. *Let $C \subseteq \mathbb{F}_q[x]/\langle x^n - 1\rangle$ be a linear cyclic code. For any $v(x) \in C$ and any $a(x) \in \mathbb{F}_q[x]$, we have $a(x)v(x) \in C$.*

Proof. We have

$$\begin{aligned} a(x)v(x) &= a_0 v(x) + a_1 x v(x) + a_2 x^2 v(x) + \dots \\ &\equiv a_0 v + a_1 \sigma(v) + a_2 \sigma^2(v) + \dots \pmod{x^n - 1}, \end{aligned}$$

which belongs to C since C is linear and cyclic. □

Definition 7.3. A (actually **the**) *generator polynomial* $g(x)$ of a nontrivial linear cyclic code C is a nonzero monic polynomial of minimum degree in C. If $C = \{0\}$, we take $g(x) = x^n - 1$ or $g(x) = 0$ as a special case.

It is not hard to see using linearity and the division algorithm that the generator polynomial is unique. (Refer to the proof below.)

Theorem 7.4. *Let C be a linear cyclic code with generator polynomial $g(x)$. Then $w(x) \in C$ if and only if $g(x) \mid w(x)$.*

Proof. The "if" part was proved in Lemma 7.2. For the "only if" part, use the division algorithm. Let $w(x) = q(x)g(x) + r(x)$, with $\deg(r) < \deg(g)$. Now $r(x) = w(x) - q(x)g(x) \in C$ by linearity, so since g is of minimum degree we must have $r(x) \equiv 0$. □

It follows that C is the ideal generated by $g(x)$. The third isomorphism theorem for rings lets us identify ideals in the quotient $F_q[x]/\langle x^n - 1\rangle$ with ideals in $\mathbb{F}_q[x]$ containing $\langle x^n - 1\rangle$, and these are precisely the ideals of the form $\langle g(x)\rangle$ for some divisor $g(x) \mid x^n - 1$.

Corollary 7.5. *If $C = \langle g(x)\rangle$ is a linear cyclic code of length n over $\mathbb{F}_q$, then $g(x)$ divides $x^n - 1$ in $\mathbb{F}_q[x]$.*

From the above, we can easily determine a basis for linear cyclic codes (viewed as subspaces only and forgetting that they are ideals).

Proposition 7.6. *Suppose* $C = \langle g(x) \rangle$, *where* $\deg(g) = t$. *Then*

$$\{g(x), xg(x), \ldots, x^{n-t-1}g(x)\}$$

corresponds with a basis for C *as a subspace of* $\mathbb{F}_q^n$. *It follows that* $\dim(C) = n - t$.

Proof. The given polynomials are clearly independent. Now any codeword $v_0 v_1 \ldots v_{n-1}$ in C must have unique last $n-t$ coordinates $(v_t \ldots v_{n-1})$, for otherwise a contradiction to $\deg(g) = t$ would result. Therefore $|C| \leq q^{n-t}$, and so the given set of $n - t$ polynomials span C. □

Example 7.7. In the Hamming code C_7 from before, we have $g(x) = 1 + x + x^3$ as the generator polynomial for C. We can verify all of the above by noting

$$1 + x^7 = (1 + x)(1 + x + x^3)(1 + x^2 + x^3)$$

in $\mathbb{F}_2[x]$. (This factorization shows another possible $g(x)$ for an equivalent code. What is this code?) Note also that $\dim(C) = 4 = n - \deg(g)$. A generator matrix G for C has rows corresponding to $g(x), xg(x), x^2g(x), x^3g(x)$.

We see that for linear cyclic codes, the generator polynomial encodes the information of the generator matrix. What about the parity check matrix H? Recall that for linear codes C, one has $w = w_0 w_1 \ldots w_{n-1} \in C$ if and only if $wH^\top = 0$. In our present context, $w \in C$ if and only if

$$w(x) = w_0 + w_1 x + \cdots + w_{n-1}x^{n-1} \equiv 0 \pmod{g(x)}.$$

So it follows that a parity check matrix for $C = \langle g(x) \rangle$ is given by

$$H^\top = \begin{bmatrix} x^0 & (\text{mod } g(x)) \\ & \vdots \\ x^i & (\text{mod } g(x)) \\ & \vdots \\ x^{n-1} & (\text{mod } g(x)) \end{bmatrix},$$

where by this we mean that the columns of H are coefficient lists of x^i, each reduced modulo $g(x)$. Note that since $\deg(g(x)) = n - k$, the columns have height $n - k$. And there is a sub-identity matrix I_{n-k} in H from the first $n - k$ nonnegative powers of x. It follows that the rows of H are linearly independent.

Definition 7.8. The *check polynomial* of a linear cyclic code $C = \langle g(x) \rangle$ of length n is the polynomial

$$h(x) = \frac{x^n - 1}{g(x)}.$$

Observe that the check polynomial "tests" for membership in C similarly to the parity check matrix: $w(x) \in C$ if and only if

$$w(x)h(x) = (x^n - 1)\frac{w(x)}{g(x)} = 0,$$

since the quotient $w(x)/g(x)$ has no remainder precisely when $w(x) \in C$.

The check polynomial of a linear cyclic code is also related to the generator matrix for the dual code (which is also cyclic). We leave the proof of the following to the reader.

Theorem 7.9. *If C is a nontrivial linear cyclic code with check polynomial $h(x)$ of degree k, then $C^\perp$ is also linear and cyclic with generator polynomial $h^*(x) := x^k h(x^{-1})$, the reciprocal polynomial of h.*

Proof. Let $g(x) = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1}$ be the generator polynomial of C. Let $h(x) = b_0 + b_1 x + \cdots + b_k x^k$ be its check polynomial, so that $h^*(x) = b_k + b_{k-1}x + \cdots + b_0 x^k$. The dot product of words corresponding with $g(x)$ and $h^*(x)$ is $a_0 b_k + a_1 b_{k-1} + \cdots + a_k b_0$, which is simply the coefficient of x^k in $g(x)h(x) = x^n - 1$. It follows that h^* is orthogonal to g; the rest of the proof follows by linear and cyclic extension. □

Example 7.10. The check polynomial for $C = \langle 1 + x + x^3 \rangle \subset \mathbb{F}_2^7$ is $h(x) = (1 + x)(1 + x^2 + x^3)$. So the generator polynomial for $C^\perp$ is

$$x^4(1 + x^{-1})(1 + x^{-2} + x^{-3}) = 1 + x^2 + x^3 + x^4.$$

As we've seen, classifying all $\mathbb{F}_q$-ary linear cyclic codes of a given length n amounts to factoring $x^n - 1$ in $\mathbb{F}_q[x]$. Here are a few more details about the correspondence between linear cyclic codes of length n and factors of $x^n - 1$.

Proposition 7.11. *Suppose $C_1, C_2 \subseteq \mathbb{F}_q[x]/\langle x^n - 1 \rangle$ are linear cyclic codes with generator polynomials g_1, g_2, respectively. Then $C_1 \subseteq C_2$ if and only if $g_2 \mid g_1$.*

Corollary 7.12. *If C_1 and C_2 have generator polynomials g_1 and g_2, respectively, then*

- *$C_1 + C_2$ has generator polynomial $\gcd(g_1, g_2)$;*
- *$C_1 \cap C_2$ has generator polynomial $\operatorname{lcm}(g_1, g_2)$.*

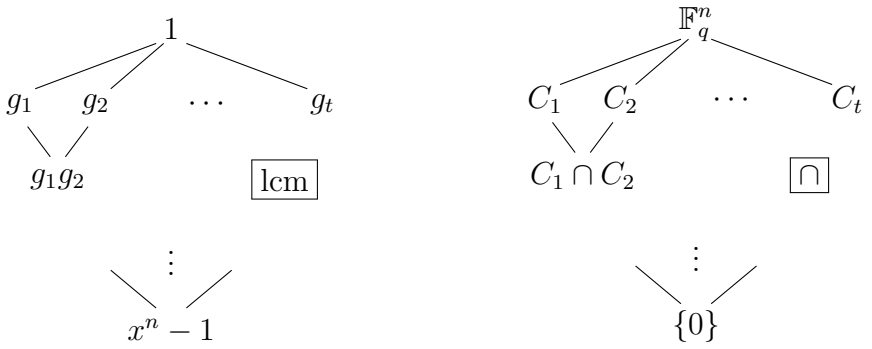


Figure 7.1: The lattices of polynomial divisors and cyclic codes

For instance, the first of these claims just follows from the fact that $C_1 + C_2 = \{u_1 + u_2 : u_i \in C_i\}$ is the smallest (linear and cyclic) code containing both C_1 and C_2 and that $\gcd(g_1, g_2)$ is the largest monic polynomial dividing both g_1 and g_2.

With $x^n - 1 = g_1 g_2 \dots g_t$ as a product of irreducible polynomials in $\mathbb{F}_q[x]$, we obtain the following order structure of $\mathbb{F}_q$-ary linear cyclic codes of length n.

Here, $C_i = \langle g_i(x) \rangle$ are the maximal proper linear cyclic codes.

7.2 BCH codes

So far, we have not said anything about the minimum distance of cyclic codes. The goal of this section is to present a special class of linear cyclic codes with a decent minimum distance.

Let $\mathbb{F}_{q^r}$ be the finite field of order q^r. Recall that, as vector spaces over the ground field, $\mathbb{F}_{q^r} \cong \mathbb{F}_q^r$. Let β be a generator of $\mathbb{F}_{q^r}^\times$, and consider the minimal polynomial $m_\beta(x) \in \mathbb{F}_q[x]$.

For instance, we have seen that if $f(x)$ is primitive and $\beta = x \in \mathbb{F}_q[x]/\langle f(x) \rangle$, then β is a generator with $m_\beta(x) = f(x)$. By the division algorithm, $m_\alpha(x)$ divides any other polynomial $p(x) \in \mathbb{F}_{q^r}[x]$ with $p(\alpha) = 0$. In particular, with $n = q^r - 1$, we have $m_\alpha(x) \mid x^n - 1$.

Definition 7.13. Let $n = q^r - 1$ and $2 \le d \le n$. The *BCH code* for these parameters is

the cyclic code with generator polynomial

$$g(x) = \mathrm{lcm}(m_\beta(x), m_{\beta^2}(x), \ldots, m_{\beta^{d-1}}(x)),$$

where β is a primitive element of $\mathbb{F}_{q^r}$.

Note that any two minimal polynomials are either equal or coprime; so the lcm is actually just a concise way of multiplying the distinct polynomials in the list.

It should be noted that there are generalizations of this definition. There are also important special cases. When $r = 1$, so that $n = q - 1$, these BCH codes are better known as *Reed-Soloman codes*. They are apparently used in DVD encoding and QR codes.

Let's focus on one other interesting special case. Put $q = 2$, $r \geq 3$ and $d = 5$. This binary BCH code has length $n = 2^r - 1$. Since $m_\beta(x) = m_{\beta^2}(x) = m_{\beta^4}(x)$, it follows that the generator polynomial is simply $g(x) = m_\beta(x)m_{\beta^3}(x)$. This leads to codimension $2r$.

Later, we will justify that the minimum distance of such a code is indeed ≥ 5.

Example 7.14. Let $r = 4$ and write $\mathbb{F}_{16} = \mathbb{F}_2[x]/\langle 1 + x + x^4 \rangle$. The field element $\beta = x$ has minimal polynomial $1 + x + x^4$. Suppose $\beta^3 = x^3$ has minimal polynomial $a_0 + a_1 x + a^2 x^2 + \ldots$. Collecting common terms in the expression

$$0 = a_0 + a_1 x^3 + a_2 x^6 + a_3 x^9 + \cdots = a_0 + a_1 x^3 + a_2(x^3 + x^2) + a_3(x^3 + x) + \ldots,$$

we see that $m_{\beta^3}(x) = 1 + x + x^2 + x^3 + x^4$. Multiplying, the generator polynomial of the length 15, codimension 8 BCH code is $g(x) = m_\beta(x)m_{\beta^3}(x) = 1 + x^4 + x^6 + x^7 + x^8$.

A parity check matrix for a general BCH code can be constructed by stacking appropriate powers of β:

$$H = \begin{bmatrix} 1 & \beta & \beta^2 & \cdots & \beta^{n-1} \\ 1 & \beta^2 & \beta^4 & \cdots & \beta^{2(n-1)} \\ & & \vdots & & \\ 1 & \beta^{d-1} & \beta^{2(d-1)} & \cdots & \beta^{(d-1)(n-1)} \end{bmatrix}.$$

This is because $Hw = 0$ if and only if $w(\beta) = w(\beta^2) = \cdots = w(\beta^{d-1}) = 0$, where we are identifying $w \in \mathbb{F}_q^n$ with the polynomial $w(x) \in \mathbb{F}_q[x]/\langle x^n - 1 \rangle$ in the usual way.

Disclaimer: Recall that the minimal polynomials $m_{\beta^i}(x)$, $i = 1, \ldots, d-1$, in general contain repetition. It follows that the above H has dependent rows, although they still generate the dual code $\langle g(x) \rangle^\perp$.

Example (cont'd). Weeding out repetition, we have

$$H = \begin{bmatrix} 1 & \beta & \beta^2 & \cdots & \beta^{14} \\ 1 & \beta^3 & \beta^6 & \cdots & \beta^{42} \end{bmatrix} \in \mathbb{F}_{16}^{2\times 15}.$$

Concretely, as a matrix in $\mathbb{F}_2^{8\times 15}$,

$$H = \left[\begin{array}{c|c|c|c} 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & \cdots \\ 0 & 0 & 1 & \cdots \\ 0 & 0 & 0 & \cdots \\ \hline 1 & 0 & 0 & \cdots \\ 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & \cdots \\ 0 & 1 & 1 & \cdots \end{array}\right],$$

where for instance the $(2,3)$-block entry follows from $\beta^6 = \beta^2 + \beta^3 \in \mathbb{F}_{16}$.

Theorem 7.15. *The parameter d of a BCH code is a lower bound on minimum distance.*

Proof. We show that H has spark at least d. For this, it is enough to use the full $(d-1)$-rowed version of H in $\mathbb{F}_{q^r}$ and check that determinants of all square submatrices of order $d-1$ are nonzero. Let $S \subseteq \{0, 1, \ldots, n-1\}$ be a restriction of the column indices, where $|S| = d-1$. We have

$$\det H|_S = \begin{vmatrix} \beta^{i_1} & \beta^{i_2} & \cdots & \beta^{i_{d-1}} \\ \beta^{2i_1} & \beta^{2i_2} & \cdots & \beta^{2i_{d-1}} \\ & \vdots & & \\ \beta^{(d-1)i_1} & \beta^{(d-1)i_2} & \cdots & \beta^{(d-1)i_{d-1}} \end{vmatrix} = \beta^{\sum S} \det V(\beta^i : i \in S)$$

in terms of the Vandermonde matrix. Recall the determinant of V is, by Proposition 3.2,

$$\det V = \prod_{i<j\in S} (\beta^i - \beta^j) \neq 0$$

since β is a generator of $\mathbb{F}_{q^r}$ and $n = q^r - 1$. It follows that the minimum distance is at least d. □

There is a cute explicit verification of this in the case $q = 2$, $d = 5$. See the exercises.

Corollary 7.16. *Let C be the cyclic code of length $2^r - 1$ with generator polynomial $g(x) = m_\beta(x)m_{\beta^3}(x)$. Then $d_{min}(C) \geq 5$.*

Exercises

1. (a) Let C be a binary cyclic code of length p, where p is prime. Show that $|C| \equiv 0, 1$ or $2 \pmod{p}$.

   (b) How does this change when we impose linearity on C?

   (c) How does this change for the ternary alphabet?

2. Find, with proof, the smallest length of a binary linear cyclic code with generator polynomial $x^7 + x + 1$.

3. Consider $C = \langle x^5 + 2x^3 + x^2 + 2x + 2 \rangle \subseteq \mathbb{F}_3[x]/\langle x^{11} - 1 \rangle$.

   (a) Find a generator matrix for C.

   (b) Find a check polynomial for $C^\perp$.

4. Let $n = 17$, and let ζ be a primitive nth root of unity for $\mathbb{F}_2$. The group $\mathbb{F}_{17}^\times$ can be partitioned as $Q \cup \overline{Q}$, the set of quadratic residues and nonresidues, respectively. The *quadratic residue code* C has generator polynomial

$$g(x) = \prod_{i \in Q} (x - \zeta^i).$$

   (Incidentally, this C has minimum distance 5.)

   (a) Express g as a polynomial in $\mathbb{F}_2[x]$. (There are two possible answers.)

   (b) What is the dimension of the cyclic code $\langle g \rangle$?

   (c) What is a check polynomial h for this code, based on the g from (a)?

5. Show that $C = \langle g(x) \rangle$ is self-dual as a linear code if and only if $g(x)g^*(x) = x^n - 1$.

6. What are the binary BCH codes with $r = 3$? Consider each allowable distance.

7. Prove that the binary BCH code with $d = 3$ is the Hamming code of length $n = 2^r - 1$.

8. (a) Prove Corollary 7.16 by arguing directly that

$$\begin{aligned} \beta^s + \beta^t + \beta^u &= 0, \\ \beta^{3s} + \beta^{3t} + \beta^{3u} &= 0, \end{aligned}$$

   has no solutions for distinct $s, t, u \pmod{2^r - 1}$.

(b) Repeat (a) with extra terms β^v and β^{3v}.
(*Hint*: Let $\gamma_1 = \beta^{s-t}$, $\gamma_2 = \beta^{t-u}$, $\gamma_3 = \beta^{u-s}$, and obtain the relations

$$\begin{aligned} \gamma_1\gamma_2\gamma_3 &= 1, \\ \gamma_1 + \gamma_2 + \gamma_3 &= \gamma_1^{-1} + \gamma_2^{-1} + \gamma_3^{-1}.) \end{aligned}$$

(c) Conclude that the spark of H is at least 5.

Bibliography

[1] T.W. Hungerford, *Algebra*, Springer-Verlag GTM 73, New York, 1974.

[2] R. Lidl and G. Pilz, *Applied Abstract Albebra*, 2nd ed., Springer-Verlag UTM, New York, 1998.

[3] J.H. van Lint, *Introduction to Coding Theory*, Springer-Verlag GTM 86, New York, 1982.