Chapter 7

Introduction to finite fields

This chapter provides an introduction to several kinds of abstract algebraic structures, particularly groups, fields, and polynomials. Our primary interest is in finite fields, *i.e.,* fields with a finite number of elements (also called Galois fields). In the next chapter, finite fields will be used to develop Reed-Solomon (RS) codes, the most useful class of algebraic codes. Groups and polynomials provide the requisite background to understand finite fields.

A field is more than just a set of elements: it is a set of elements under two operations, called addition and multiplication, along with a set of properties governing these operations. The addition and multiplication operations also imply inverse operations called subtraction and division. The reader is presumably familiar with several examples of fields, such as the real field $\mathbb{R}$, the complex field $\mathbb{C}$, the field of rational numbers $\mathbb{Q}$, and the binary field $\mathbb{F}_2$.

7.1 Summary

In this section we briefly summarize the results of this chapter. The main body of the chapter will be devoted to defining and explaining these concepts, and to proofs of these results.

For each prime p and positive integer $m \geq 1$, there exists a finite field $\mathbb{F}_{p^m}$ with p^m elements, and there exists no finite field with q elements if q is not a prime power. Any two fields with p^m elements are isomorphic.

The integers modulo p form a prime field $\mathbb{F}_p$ under mod-p addition and multiplication. The polynomials $\mathbb{F}_p[x]$ over $\mathbb{F}_p$ modulo an irreducible polynomial $g(x) \in \mathbb{F}_p[x]$ of degree m form a finite field with p^m elements under mod-$g(x)$ addition and multiplication. For every prime p, there exists at least one irreducible polynomial $g(x) \in \mathbb{F}_p[x]$ of each positive degree $m \geq 1$, so all finite fields may be constructed in this way.

Under addition, $\mathbb{F}_{p^m}$ is isomorphic to the vector space $(\mathbb{F}_p)^m$. Under multiplication, the nonzero elements of $\mathbb{F}_{p^m}$ form a cyclic group $\{1, \alpha, \ldots, \alpha^{p^m-2}\}$ generated by a primitive element $\alpha \in \mathbb{F}_{p^m}$.

The elements of $\mathbb{F}_{p^m}$ are the p^m roots of the polynomial $x^{p^m} - x \in \mathbb{F}_p[x]$. The polynomial $x^{p^m} - x$ is the product of all monic irreducible polynomials $g(x) \in \mathbb{F}_p[x]$ such that $\deg g(x)$ divides m. The roots of a monic irreducible polynomial $g(x) \in \mathbb{F}_p[x]$ form a cyclotomic coset of $\deg g(x)$ elements of $\mathbb{F}_{p^m}$ which is closed under the operation of raising to the pth power.

For every n that divides m, $\mathbb{F}_{p^m}$ contains a subfield with p^n elements.

For further reading on this beautiful subject, see [E. R. Berlekamp, *Algebraic Coding Theory*, Aegean Press, 1984], [R. Lidl and H. Niederreiter, *Introduction to Finite Fields and their Applications*, Cambridge University Press, 1986] or [R. J. McEliece, *Finite Fields for Computer Scientists and Engineers*, Kluwer, 1987], [M. R. Schroeder, *Number Theory in Science and Communication*, Springer, 1986], or indeed any book on finite fields or algebraic coding theory.

7.2 The integers

We begin with a brief review of the familiar factorization properties of the set $\mathbb{Z}$ of integers. We will use these properties immediately in our discussion of cyclic groups and their subgroups and of prime fields. Moreover, we will model our later discussion of the factorization properties of polynomials on the discussion here.

7.2.1 Definitions

An integer n is said to be a *divisor* of an integer i if i is an integer multiple of n; *i.e.,* $i = qn$ for some integer q. Thus all integers are trivially divisors of 0.

The integers that have integer inverses, namely ± 1, are called the *units* of $\mathbb{Z}$. If u is a unit and n is a divisor of i, then un is a divisor of i and n is a divisor of ui. Thus the factorization of an integer can only be unique up to a unit u, and ui has the same divisors as i. We therefore consider only factorizations of positive integers into products of positive integers.

Every nonzero integer i is divisible by 1 and i; these divisors are called trivial. An integer n is said to be a *factor* of an integer i if n is positive and a nontrivial divisor of i. For example, 1 has no nontrivial divisors and thus no factors.

A positive integer that has no nontrivial divisors is called a *prime integer*.

7.2.2 Mod-n arithmetic

Given a positive integer n, every integer i may be uniquely expressed as $i = qn + r$ for some integer remainder r in the interval $0 \le r \le n - 1$ and some integer quotient q. This may be proved by the Euclidean division algorithm, which if $i \ge n$ just subtracts n from i repeatedly until the remainder lies in the desired interval.

The remainder r, denoted by $r = i \bmod n$, is the more important part of this expression. The set of possible mod-n remainders is the set of n integers $R_n = \{0, 1, \ldots, n-1\}$. Evidently n is a divisor of i if and only if $i \bmod n = 0$.

Remainder arithmetic using the mod-n remainder set R_n is called "mod-n arithmetic." The rules for mod-n arithmetic follow from the rules for integer arithmetic as follows. Let $r = i \bmod n$ and $s = j \bmod n$; then, as integers, $r = i - qn$ and $s = j - tn$ for some quotients q and t. Then

$$\begin{aligned} r + s &= i + j - (q + t)n; \\ rs &= ij - (qj + ti)n + qtn^2. \end{aligned}$$

Hence $(r + s) \bmod n = (i + j) \bmod n$ and $rs \bmod n = ij \bmod n$; *i.e.,* the mod-n remainder of the sum or product of two integers is equal to the mod-n remainder of the sum or product of their mod-n remainders, as integers.

The mod-n addition and multiplication rules are therefore defined as follows:

$$\begin{aligned} r \oplus s &= (r+s) \bmod n; \\ r * s &= (rs) \bmod n, \end{aligned}$$

where "r" and "s" denote elements of the remainder set R_n on the left and the corresponding ordinary integers on the right. This makes mod-n arithmetic consistent with ordinary integer arithmetic in the sense expressed in the previous paragraph.

7.2.3 Unique factorization

Given a positive integer i, we may factor i into a unique product of prime factors by simply factoring out primes no greater than i until we arrive at the quotient 1, as the reader has known since grade school. For the time being, we will take this unique factorization property as given. A proof will be given as an exercise after we prove the corresponding property for polynomials.

7.3 Groups

We now introduce groups.

Definition 7.1 *A group is a set of elements $G = \{a, b, c, \ldots\}$ and an operation $\oplus$ for which the following axioms hold:*

- *Closure: for any $a \in G$, $b \in G$, the element $a \oplus b$ is in G.*
- *Associative law: for any $a, b, c \in G$, $(a \oplus b) \oplus c = a \oplus (b \oplus c)$.*
- *Identity: There is an identity element 0 in G for which $a \oplus 0 = 0 \oplus a = a$ for all $a \in G$.*
- *Inverse: For each $a \in G$, there is an inverse $(-a)$ such that $a \oplus (-a) = 0$.*

In general it is not necessary that $a \oplus b = b \oplus a$. A group G for which $a \oplus b = b \oplus a$ for all $a, b \in G$ is called *abelian* or *commutative*. In these notes all groups will be abelian.

In view of the associative law, we may write $(a \oplus b) \oplus c$ as $a \oplus b \oplus c$ without ambiguity. Moreover, in an abelian group the elements a, b, c may be written in any order.

Frequently, the operation in a group is called multiplication, usually represented either by $*$ or juxtaposition. The identity is then denoted by 1 (or e) and the inverse of a by a^{-1}. Additive notation is generally used only for abelian groups, whereas multiplicative notation is used for both abelian and nonabelian groups. Since we consider only abelian groups, we will use additive notation when the nature of the group is unspecified.

As an example, the set of integers $\mathbb{Z}$ with the usual addition operation $+$ forms an abelian group. Also, the real field $\mathbb{R}$ forms an additive abelian group under ordinary addition in which the identity is 0 and the inverse of a is $-a$. More interestingly, as the reader should verify, the nonzero elements of $\mathbb{R}$ form a multiplicative abelian group under ordinary multiplication, in which the identity is 1 and the inverse of a is $a^{-1} = 1/a$. We will see that every field has similar additive and multiplicative group properties.

This example illustrates that the group structure (*i.e.,* the properties stemming from the group operation $\oplus$) may reflect only part of the structure of the given set of elements; *e.g.,* the additive group structure of $\mathbb{R}$ takes no account of the fact that real numbers may also be multiplied, and the multiplicative group structure of $\mathbb{R} - \{0\}$ takes no account of the fact that real numbers may also be added.

We abbreviate $b \oplus (-a)$ for any $a, b \in G$ by $b - a$ and regard "$-$" as an additional operation implicitly defined by the axioms. In an additive group, "$-$" is called subtraction; in a multiplicative group, "$-$" is called division and denoted by / or $\div$.

Because of the inverse operation, cancellation is always permissible; *i.e.,* if $x \oplus a = y \oplus a$, we can add $-a$ to both sides, showing that $x = y$. Similarly, one can move terms from one side of an equation to the other; *i.e.,* $x \oplus a = y$ implies $x = y - a$.

Exercise 1 (Inverses and cancellation)

(a) Verify the following set of implications for arbitrary elements a, b of a group G which is not necessarily abelian:

$$b \oplus a = 0 \;\Rightarrow\; b = -a \;\Rightarrow\; a \oplus b = 0 \;\Rightarrow\; a = -b \;\Rightarrow\; b \oplus a = 0.$$

(b) Use this result to show that the inverse is unique, *i.e.,* that $a \oplus b = 0 \;\Rightarrow\; b = -a$, and that the inverse also works on the left, *i.e.,* $b \oplus a = 0 \;\Rightarrow\; b = -a$. Note that this shows that cancellation is permitted on either the right or the left.

(c) Show that the identity element is unique, *i.e.,* that for $a, b \in G$, $a \oplus b = a \;\Rightarrow\; b = 0$ and $b \oplus a = a \;\Rightarrow\; b = 0$. □

If G has a finite number of elements, $G = \{a_1, a_2, \ldots, a_n\}$, then G is said to be *finite* and $|G| = n$ is said to be the *order* of G. The group operation $\oplus$ may then be specified by an $n \times n$ "addition table" whose entry at row i, column j is $a_i \oplus a_j$. The cancellation property implies that if $a_j \neq a_k$, then $a_i \oplus a_j \neq a_i \oplus a_k$. This means that all elements in any row i of the addition table are distinct; *i.e.,* each row contains each element of G exactly once. Similarly, each column contains each element of G exactly once. Thus the group axioms restrict the group operation $\oplus$ more than might be immediately evident.

7.3.1 Alternative group axioms

The property that a "row of the addition table," namely $a \oplus G = \{a \oplus b \mid b \in G\}$ is just the set of elements of G in a different order (*i.e.,* a *permutation* of G) is a fundamental property of any group G. We will now show that this permutation property may be taken as one of the group axioms. Subsequently we will use this property to prove that certain sets are groups.

Theorem 7.1 (Alternative group axioms) *Let $G = \{a, b, c, \ldots\}$ be a set of elements on which an operation $\oplus$ is defined. Then G is a group under the operation $\oplus$ if and only if the following axioms hold:*

- *Associative law: for any $a, b, c \in G$, $(a \oplus b) \oplus c = a \oplus (b \oplus c)$.*
- *Identity: There is an identity element 0 in G for which $a \oplus 0 = 0 \oplus a = a$ for all $a \in G$.*
- *Permutation property: For each $a \in G$, $a \oplus G = \{a \oplus b \mid b \in G\}$ is a permutation of G.*

Proof. ($\Rightarrow$) If G is a group under $\oplus$, then by the closure property every element $a \oplus b$ is in G. Moreover, the fact that $a \in G$ has an inverse $-a \in G$ implies that every element $b \in G$ may be written as $a \oplus (-a \oplus b) \in a \oplus G$, so every element of G is in $a \oplus G$. Finally, from the cancellation property, $a \oplus b = a \oplus c$ implies $b = c$. Thus the correspondence between G and $a \oplus G$ defined by $b \leftrightarrow a \oplus b$ is one-to-one; *i.e.,* a permutation.

($\Leftarrow$) Conversely, if $a \oplus G$ is a permutation of G for every $a \in G$, then (a) the closure property holds; *i.e.,* $a \oplus b \in G$ for all $a, b \in G$; (b) since $0 \in a \oplus G$, there must exist a unique $b \in G$ such that $a \oplus b = 0$, so a has a unique inverse $-a = b$ under $\oplus$. Thus G is a group under $\oplus$. □

The properties of "rows" $a \oplus G$ hold equally for "columns" $G \oplus a$, even when G is nonabelian.

For example, the set $\mathbb{R}^*$ of nonzero elements of the real field $\mathbb{R}$ form an abelian group under real multiplication, because real multiplication is associative and commutative with identity 1, and $\alpha\mathbb{R}^*$ is a permutation of $\mathbb{R}^*$ for any $\alpha \in \mathbb{R}^*$.

Exercise 2 (Invertible subsets).

(a) Let H be a set of elements on which an associative operation $\oplus$ is defined with identity 0, and let G be the subset of elements $h \in H$ which have unique inverses $-h$ such that $h \oplus -h = 0$. Show that G is a group under $\oplus$.

(b) Show that the nonzero elements of the complex field form a group under complex multiplication.

(c) Show that the set of invertible $n \times n$ real matrices forms a (nonabelian) group under real matrix multiplication.

(d) What are the invertible elements of $\mathbb{Z}$ under multiplication? Do they form a group?

7.3.2 Cyclic groups

An important example of a finite abelian group is the set of remainders $R_n = \{0, 1, \ldots, n-1\}$ under mod-n addition, where n is a given positive integer. This group is called "the integers mod n" and is denoted by $\mathbb{Z}_n$. Note that $\mathbb{Z}_1$ is the trivial group $\{0\}$.

A *finite cyclic group* is a finite group G with a particular element $g \in G$, called the *generator*, such that each element of G can be expressed as the sum, $g \oplus \cdots \oplus g$, of some number of repetitions of g.[1] Thus each element of G appears in the sequence of elements $\{g, g \oplus g, g \oplus g \oplus g, \ldots\}$. We denote such an i-fold sum by ig, where i is a positive integer and g is a group element; *i.e.,*

$$1g = g, 2g = g \oplus g, \ldots, ig = \underbrace{g \oplus \cdots \oplus g}_{i \text{ terms}}, \ldots$$

Since g generates G and G includes the identity element 0, we must have $ig = 0$ for some positive integer i. Let n be the smallest such integer; thus $ng = 0$ and $ig \neq 0$ for $1 \leq i \leq n-1$. Adding the sum of j g's for any $j > 0$ to each side of $ig \neq 0$ results in $(i+j)g \neq jg$. Thus the elements $\{1g, 2g, \ldots, ng = 0\}$ must all be different.

[1] Mathematicians say also that an infinite group $G = \{\ldots, -1g, 0g, 1g, 2g, \ldots\}$ generated by a single element g is cyclic; *e.g.,* the group of integers $\mathbb{Z}$ is an infinite cyclic group with generator 1. Although such infinite cyclic groups have the single-generator property of finite cyclic groups, they do not "cycle." Hereafter, "cyclic group" will mean "finite cyclic group."

We can also add jg to both sides of the equality $ng = 0$, yielding $(j+n)g = jg$ for any $j > 0$. Thus for each $i > n$, ig is equal to some earlier element in the sequence, namely $(i-n)g$. The elements $\{1g, 2g, \ldots, ng = 0\}$ therefore constitute all of the distinct elements in G, and the order of G is $|G| = n$. If we define $0g$ to be the identity 0, then the elements of G may be conveniently represented as $G = \{0g = 0, 1g, \ldots, (n-1)g\}$.

Figure 1 illustrates the cyclic structure of G that arises from the relation $(j+n)g = jg$.

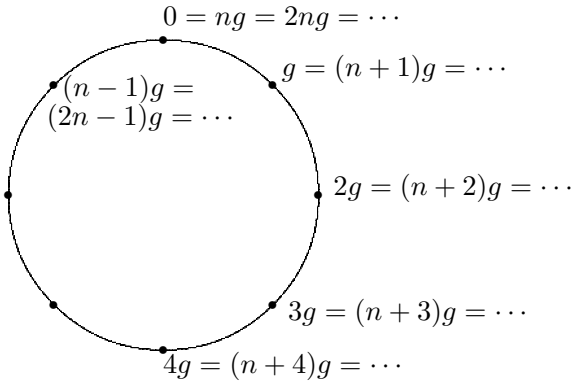

Figure 1. The cyclic structure of a cyclic group: the sequence $\{1g, 2g, \ldots\}$ goes from the group element g up to $ng = 0$, then returns to g and continues to cycle.

Addition in a cyclic group of order n can be understood in terms of mod-n addition. In particular, since $ng = 0$, we also have $2ng = 0, 3ng = 0$, etc. Since any integer i may be uniquely written as $i = qn + r$ where the remainder $r = i \bmod n$ is in the set $R_n = \{0, 1, \ldots, n-1\}$, we have $ig = (qn)g + rg = rg$, where $rg = (i \bmod n)g$ is one of the elements of G. The addition rule of G is thus as follows: for each $0 \le i, j < n$,

$$ig \oplus jg = (i + j \bmod n)g.$$

Evidently $0g$ is the identity, and the inverse of a nonzero element ig is $(n-i)g$.

We thus see that any cyclic group G of order n is essentially identical to $\mathbb{Z}_n$. More precisely, the correspondence $ig \in G \leftrightarrow i \in \mathbb{Z}_n$ is preserved under addition; *i.e.,* $ig \oplus jg \leftrightarrow i \oplus j$ for each $i, j \in \mathbb{Z}_n$. This type of correspondence is called an *isomorphism*. Specifically, two finite groups G and H are *isomorphic* if there exists an invertible[2] function $h : G \to H$ mapping each $\alpha \in G$ into a $\beta = h(\alpha) \in H$ such that $h(\alpha \oplus \alpha') = h(\alpha) \oplus h(\alpha')$, where $\oplus$ denotes the group operation of G on the left and that of H on the right. In summary:

Theorem 7.2 (Cyclic groups) *The elements of a cyclic group G of order n with generator g are $\{0g, 1g, 2g, \ldots, (n-1)g\}$. The addition rule is $ig \oplus jg = (i + j \bmod n)g$, the identity is $0g$, and the inverse of $ig \neq 0g$ is $(n-i)g$. Finally, G is isomorphic to $\mathbb{Z}_n$ under the one-to-one correspondence $ig \leftrightarrow i$.*

Since $\mathbb{Z}_n$ is abelian, it follows that all cyclic groups are abelian.

In multiplicative notation, the elements of a cyclic group G of order n with generator g are denoted by $\{g^0 = 1, g^1, g^2, \ldots, g^{n-1}\}$, the multiplication rule is $g^i * g^j = g^{(i+j \bmod n)}$, the identity is $g^0 = 1$, and the inverse of $g^i \neq 1$ is g^{n-i}. For example, if $\omega = e^{2\pi i/n}$, the set $\{1, \omega, \omega^2, \ldots, \omega^{n-1}\}$ of complex nth roots of unity is a cyclic group under complex multiplication, isomorphic to $\mathbb{Z}_n$.

[2] A function $h : G \to H$ is called invertible if for each $\beta \in H$ there is a unique $\alpha \in G$ such that $\beta = h(\alpha)$. An invertible function is also called a one-to-one correspondence, denoted by $G \leftrightarrow H$.

7.3.3 Subgroups

A subgroup S of a group G is a subset of the elements of the group such that if $a, b \in S$, then $a \oplus b \in S$ and $-a \in S$. A subgroup S therefore includes the identity element of G and the inverse of each element in S. The associative law holds for S since it holds for G. Therefore a subgroup $S \subseteq G$ is itself a group under the group operation of G.

For example, the set of integers $\mathbb{Z}$ is a subgroup of the additive group of $\mathbb{R}$.

If G is abelian, then S must be abelian; however, S may be abelian even if G is nonabelian.

For any $g \in G$, we define the coset (translate) $S \oplus g = \{s \oplus g \mid s \in S\}$. The zero coset $S \oplus 0$ is thus equal to S itself; moreover, by Theorem 7.1, $S \oplus g = S$ whenever $g \in S$.

The following theorem states a more general result:

Lemma 7.3 (Cosets) *Two cosets $S \oplus g$ and $S \oplus h$ are the same if $g - h \in S$, but are disjoint if $g - h \notin S$.*

Proof. If $g - h \in S$, then the elements of $S \oplus h$ include $(g - h) \oplus h = g$ and therefore all elements of $S \oplus g$, so $S \oplus g \subseteq S \oplus h$; similarly $S \oplus h \subseteq S \oplus g$.

On the other hand, if $S \oplus g$ and $S \oplus h$ have any element in common, say $s \oplus g = s' \oplus h$, then $g - h = s' - s \in S$; thus, $g - h \notin S$ implies that $S \oplus g$ and $S \oplus h$ are disjoint. □

It follows that the distinct cosets $S \oplus g$ of a subgroup $S \subseteq G$ form a disjoint partition of G, since every element $g \in G$ lies in some coset, namely $S \oplus g$.

The elements $s \oplus g$ of a coset $S \oplus g$ are all distinct, since $s \oplus g = s' \oplus g$ implies $s = s'$. Therefore if S is finite, then all cosets of S have the same size, namely the size $|S|$ of $S = S \oplus 0$. If G is finite, G is therefore the disjoint union of a finite number $|C|$ of cosets of $S \subseteq G$, each of size $|S|$, so $|G| = |C||S|$. This proves Lagrange's theorem:

Theorem 7.4 (Lagrange) *If S is a subgroup of a finite group G, then $|S|$ divides $|G|$.*

7.3.4 Cyclic subgroups

Given any finite group G and any element $g \in G$, the set of elements generated by g, namely $S(g) = \{g, g \oplus g, \ldots\}$, is a cyclic subgroup of G. The *order* of g is defined as the order $|S(g)|$ of $S(g)$. By Lagrange's theorem, $|S(g)|$ divides $|G|$, and by the cyclic groups theorem, $S(g)$ is isomorphic to $\mathbb{Z}_{|S(g)|}$. (If $g = 0$, then $S(g) = \{0\}$ and $|S(g)| = 1$. We will assume $g \neq 0$.)

As a fundamental example, let G be the cyclic group $\mathbb{Z}_n = \{0, 1, \ldots, n-1\}$, and let $S(m)$ be the cyclic subgroup $\{m, 2m, \ldots\}$ generated by $m \in \mathbb{Z}_n$. Here $im = m \oplus \cdots \oplus m$ is simply the sum of m with itself i times; *i.e.,* $im \in G$ is the ordinary product im mod n. The order $|S(m)|$ of $S(m)$ is the least positive integer k such that $km = 0 \bmod n$; *i.e.,* such that the integer product km is divisible by n. Thus km is the least common multiple of m and n, denoted $\mathrm{lcm}(m, n)$, and $|S(m)| = k = \mathrm{lcm}(m, n)/m$. By elementary number theory, $\mathrm{lcm}(m, n) = mn/\gcd(m, n)$ for any positive integers m, n, so we may alternatively write $|S(m)| = n/\gcd(m, n)$, where $\gcd(m, n)$ denotes the greatest common divisor of m and n. This shows explicitly that $|S(m)|$ divides n.

For example, suppose $n = 10$ and $m = 4$. Then $S(4) = \{4, 8, 2, 6, 0\}$. Thus $|S(4)| = 5$, consistent with $|S(4)| = \mathrm{lcm}(4, 10)/4 = 20/4$ or $|S(4)| = 10/\gcd(4, 10)/4 = 10/2$.

Now when does $S(m) = \mathbb{Z}_n$? This occurs if and only if $\gcd(m, n) = 1$; *i.e.,* if and only if m is relatively prime to n. In short, m generates $\mathbb{Z}_n$ and has order $|S(m)| = n$ if and only if m and n are relatively prime. The number of integers in the set $\{0, 1, \ldots, n-1\}$ that have order n is called the *Euler number* $\phi(n)$.

For example, in $\mathbb{Z}_{10}$ the integers that are relatively prime to 10 are $\{1, 3, 7, 9\}$, so $\phi(10) = 4$. The order of the other elements of $\mathbb{Z}_{10}$ are as follows:

- 0 is the only element of order 1, and $S(0) = \{0\}$.
- 5 is the only element of order 2, and $S(5) = \{0, 5\}$.
- $\{2, 4, 6, 8\}$ have order 5, and $S(2) = S(4) = S(6) = S(8) = \{0, 2, 4, 6, 8\}$.

In general, $\mathbb{Z}_n$ has a cyclic subgroup S_d of order d for each positive integer d that divides n, including 1 and n. S_d consists of $\{0, n/d, 2n/d, \ldots, (d-1)n/d\}$, and is isomorphic to $\mathbb{Z}_d$. S_d thus contains $\phi(d)$ elements that are relatively prime to d, each of which has order d and generates S_d. The remaining elements of S_d belong also to smaller cyclic subgroups.

For example, $\mathbb{Z}_{10}$ has a subgroup $S_5 = \{0, 2, 4, 6, 8\}$ with 5 elements. Four of these elements, namely $\{2, 4, 6, 8\}$, are relatively prime to 5 and generate S_5. The remaining element of S_5, namely 0, has order 1.

Since every element of $\mathbb{Z}_n$ has some definite order d that divides n, we have

$$n = \sum_{d:\, d|n} \phi(d). \tag{7.1}$$

The notation $d : d|n$ means the set of positive integers d, including 1 and n, that divide n. All Euler numbers may be determined recursively from this expression. For example, $\phi(1) = 1, \phi(2) = 2 - \phi(1) = 1, \phi(3) = 3 - \phi(1) = 2, \phi(4) = 4 - \phi(1) - \phi(2) = 2, \ldots$.

Exercise 3. Show that $\phi(n) \geq 1$ for all $n \geq 1$. [Hint: Find the order of 1 in $\mathbb{Z}_n$.] □

Since every cyclic group G of size n is isomorphic to $\mathbb{Z}_n$, these results apply to every cyclic group. In particular, every cyclic group G of size n has $\phi(n)$ generators that generate G, which are called the *primitive elements* of G. G also contains one cyclic subgroup of size d for each d that divides n.

Exercise 4. Show that every subgroup of $\mathbb{Z}_n$ is cyclic. [Hint: Let s be the smallest nonzero element in a subgroup $S \subseteq \mathbb{Z}_n$, and compare S to the subgroup generated by s.] □

7.4 Fields

Definition 7.2 *A field is a set $\mathbb{F}$ of at least two elements, with two operations $\oplus$ and $*$, for which the following axioms are satisfied:*

- *The set $\mathbb{F}$ forms an abelian group (whose identity is called 0) under the operation $\oplus$.*
- *The set $\mathbb{F}^* = \mathbb{F} - \{0\} = \{a \in \mathbb{F}, a \neq 0\}$ forms an abelian group (whose identity is called 1) under the operation $*$.*
- *Distributive law: For all $a, b, c \in \mathbb{F}$, $(a \oplus b) * c = (a * c) \oplus (b * c)$.*

The operation $\oplus$ is called addition (and often denoted by $+$), and the operation $*$ is called multiplication (and often denoted by juxtaposition). As in ordinary arithmetic, we often omit the parentheses around a product of elements, using the convention "multiplication before addition;" *e.g.,* we interpret $a \oplus b * c$ as $a \oplus (b * c)$.

The reader may verify that $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Q}$ and $\mathbb{F}_2$ each form a field according to this definition under conventional addition and multiplication.

Exercise 5. Show that for any element $a \in \mathbb{F}$, $a * 0 = 0$. □

7.4.1 Prime fields

A fundamental example of a finite (Galois) field is the set $\mathbb{F}_p$ of mod-p remainders, where p is a given prime number. Here, as in $\mathbb{Z}_p$, the set of elements is $R_p = \{0, 1, \cdots, p-1\}$, and the operation $\oplus$ is mod-p addition. The multiplicative operation $*$ is mod-p multiplication; *i.e.,* multiply integers as usual and then take the remainder after division by p.

Theorem 7.5 (Prime fields) *For every prime p, the set $R_p = \{0, 1, \cdots, p-1\}$ forms a field (denoted by $\mathbb{F}_p$) under mod-p addition and multiplication.*

Proof. We have already seen that the elements of $\mathbb{F}_p$ form an abelian group under addition modulo p, namely the cyclic group $\mathbb{Z}_p$.

The associative and commutative properties of multiplication mod p follow from the corresponding properties of ordinary multiplication; the distributive law follows from the corresponding property for ordinary addition and multiplication. The multiplicative identity is 1.

To see that the nonzero elements $\mathbb{F}_p^* = \mathbb{F}_p - \{0\}$ form a group under multiplication, we use Theorem 7.1. By unique factorization, the product of two nonzero integers $a, b < p$ cannot equal 0 mod p. Therefore the nonzero elements $\mathbb{F}_p^*$ are closed under multiplication mod p. Also, for $a, b, c \in \mathbb{F}_p^*$ and $b \neq c$ we have $a(b-c) \bmod p \neq 0$. Thus $ab \neq ac \bmod p$, which implies $a * b \neq a * c$. Consequently there are no zeroes or repetitions in the set of $p-1$ elements $\{a*1, a*2, \ldots, a*(p-1)\}$, which means they must be a permutation of $\mathbb{F}_p^*$. □

We next show that $\mathbb{F}_p$ is essentially the only field with p elements. More precisely, we show that all fields with p elements are isomorphic. Two fields $\mathbb{F}$ and $\mathbb{G}$ are isomorphic if there is an invertible function $h : \mathbb{F} \to \mathbb{G}$ mapping each $\alpha \in \mathbb{F}$ into a $\beta = h(\alpha) \in \mathbb{G}$ such that $h(\alpha \oplus \alpha') = h(\alpha) \oplus h(\alpha')$ and $h(\alpha * \alpha') = h(\alpha) * h(\alpha')$. Less formally, $\mathbb{F}$ and $\mathbb{G}$ are isomorphic if there is a one-to-one correspondence $\mathbb{F} \leftrightarrow \mathbb{G}$ that translates the addition and multiplication tables of $\mathbb{F}$ to those of $\mathbb{G}$ and *vice versa.*

Let $\mathbb{F}$ be any field with a prime p number of elements. By the field axioms, $\mathbb{F}$ has an additive identity 0 and multiplicative identity 1. Consider the additive cyclic subgroup generated by 1, namely $S(1) = \{1, 1 \oplus 1, \ldots\}$. By Lagrange's theorem, the order of $S(1)$ divides $|\mathbb{F}| = p$, and therefore must be equal to 1 or p. But $1 \oplus 1 \neq 1$, else $1 = 0$, so 1 must have order p. In other words, $S(1) = \mathbb{F}$, and the additive group of $\mathbb{F}$ is isomorphic to that of $\mathbb{Z}_p$. We may therefore denote the elements of $\mathbb{F}$ by $\{0, 1, 2, \ldots, p-1\}$, and use mod-p addition as the addition rule.

The only remaining question is whether this correspondence $\mathbb{F} \leftrightarrow \mathbb{Z}_p$ under addition extends to multiplication. The distributive law shows that it does: $j * i$ is the sum of j terms each equal to i, so $j * i = (ji \bmod p)$. Therefore, in summary:

Theorem 7.6 (Prime field uniqueness) *Every field $\mathbb{F}$ with a prime number p of elements is isomorphic to $\mathbb{F}_p$ via the correspondence* $\underbrace{1 \oplus \cdots \oplus 1}_{i \text{ terms}} \in \mathbb{F} \leftrightarrow i \in \mathbb{F}_p$.

In view of this elementary isomorphism, we will denote any field with a prime number p of elements by $\mathbb{F}_p$.

It is important to note that the set $\mathbb{Z}_n$ of integers mod n does not form a field if n is not prime. The reason is that $n = ab$ for some positive integers $a, b < n \in \mathbb{Z}_n$; thus $ab = 0 \bmod n$, so the set of nonzero elements of $\mathbb{Z}_n$ is not closed under multiplication mod n.

However, we will see shortly that there do exist finite fields with non-prime numbers of elements that use other rules for addition and multiplication.

7.4.2 The prime subfield of a finite field

A subfield $\mathbb{G}$ of a field $\mathbb{F}$ is a subset of the field that is itself a field under the operations of $\mathbb{F}$. For example, the real field $\mathbb{R}$ is a subfield of the complex field $\mathbb{C}$. We now show that every finite field $\mathbb{F}_q$ has a subfield that is isomorphic to a prime field $\mathbb{F}_p$.

Let $\mathbb{F}_q$ be a finite field with q elements. By the field axioms, $\mathbb{F}_q$ has an additive identity 0 and a multiplicative identity 1.

Consider the cyclic subgroup of the additive group of $\mathbb{F}_q$ that is generated by 1, namely $S(1) = \{1, 1 \oplus 1, \ldots\}$. Let $n = |S(1)|$. By the cyclic group theorem, $S(1)$ is isomorphic to $\mathbb{Z}_n$, and its elements may be denoted by $\{0, 1, 2, \ldots, n-1\}$, with mod-n addition.

By the distributive law in $\mathbb{F}_q$, the product $i * j$ (in $\mathbb{F}_q$) of two nonzero elements in $S(1)$ is simply the sum of ij ones, which is an element of $S(1)$, namely $ij \bmod n$. Since this is a product of nonzero elements of $\mathbb{F}_q$, by the field axioms $ij \bmod n$ must be nonzero for all nonzero i, j. This will be true if and only if n is a prime number p.

Thus $S(1)$ forms a subfield of $\mathbb{F}_q$ with a prime number p of elements. By the prime field theorem of the previous subsection, $S(1)$ is isomorphic to $\mathbb{F}_p$. Thus the elements of $S(1)$, which are called the *integers* of $\mathbb{F}_q$, may be denoted by $\mathbb{F}_p = \{0, 1, \ldots, p-1\}$, and the addition and multiplication rules of $\mathbb{F}_q$ reduce to mod-p addition and multiplication in $\mathbb{F}_p$.

The prime p is called the *characteristic* of $\mathbb{F}_q$. Since the p-fold sum of the identity 1 with itself is 0, the p-fold sum of every field element $\beta \in \mathbb{F}_q$ with itself is 0: $p\beta = 0$.

In summary:

Theorem 7.7 (Prime subfields) *The integers $\{1, 1 \oplus 1, \ldots\}$ of any finite field $\mathbb{F}_q$ form a subfield $\mathbb{F}_p \subseteq \mathbb{F}_q$ with a prime number p of elements, where p is the characteristic of $\mathbb{F}_q$.*

7.5 Polynomials

We now consider polynomials over $\mathbb{F}_p$, namely polynomials whose coefficients lie in $\mathbb{F}_p$ and for which polynomial addition and multiplication is performed in $\mathbb{F}_p$. We will see that the factorization properties of polynomials are similar to those of the integers, and that the analogue to mod-n arithmetic is arithmetic modulo a polynomial $f(x)$.

A nonzero polynomial $f(x)$ of degree m over a field $\mathbb{F}$ is an expression of the form

$$f(x) = f_0 + f_1 x + f_2 x^2 + \cdots + f_m x^m,$$

where $f_i \in \mathbb{F}$, $0 \le i \le m$, and $f_m \neq 0$. We say that $\deg f(x) = m$. The symbol x represents an indeterminate (or "placeholder"), *not* an element of $\mathbb{F}$; *i.e.,* two polynomials are different if and only if their coefficients are different[3]. The nonzero polynomials of degree 0 are simply the nonzero field elements $f_0 \in \mathbb{F}$. There is also a special *zero polynomial* $f(x) = 0$ whose degree is defined by convention as $\deg 0 = -\infty$; we will explain the reason for this convention shortly. The set of all polynomials over $\mathbb{F}$ in an indeterminate x is denoted by $\mathbb{F}[x]$.

The rules for adding, subtracting or multiplying polynomials are the same over a general field $\mathbb{F}$ as over the real field $\mathbb{R}$, except that coefficient operations are in $\mathbb{F}$. In particular, addition and subtraction are performed componentwise. For multiplication, the coefficients of a polynomial product $f(x) = h(x)g(x)$ are determined by convolution:

$$f_i = \sum_{j=0}^{i} h_j g_{i-j}.$$

If two nonzero polynomials are multiplied, then their degrees add; *i.e.,* $\deg(h(x)g(x)) = \deg h(x) + \deg g(x)$. The convention $\deg 0 = -\infty$ ensures that this formula continues to hold when $h(x)$ or $g(x)$ is the zero polynomial.

The set $\mathbb{F}[x]$ has many of the properties of a field. It is evidently an abelian group under addition whose identity is the zero polynomial $0 \in \mathbb{F}[x]$. It is closed under multiplication, which is both associative and commutative and which distributes over addition. It has a multiplicative identity $1 \in \mathbb{F}[x]$, and the cancellation law holds.

However, in general we cannot divide evenly by a nonzero polynomial, since a polynomial $f(x)$ with $\deg f(x) > 0$ has no multiplicative inverse. Therefore $\mathbb{F}[x]$ is a *ring*,[4] not a field, like the ring of integers $\mathbb{Z}$. We now develop a series of properties of $\mathbb{F}[x]$ that resemble those of $\mathbb{Z}$.

[3] Over the real field $\mathbb{R}$, a polynomial $f(x)$ is sometimes regarded as a function $f : \mathbb{R} \to \mathbb{R}$. This alternative viewpoint makes little difference in the real case, since two polynomials over $\mathbb{R}$ are different if and only if the corresponding polynomial functions are different. However, over finite fields it is important to maintain the distinction. For example, over $\mathbb{F}_2$ the polynomial functions x and x^2 both map $0 \to 0, 1 \to 1$, yet the polynomials x and x^2 are different.

[4] The axioms of a ring are similar to those for a field, except that there is no multiplicative inverse. For example, $\mathbb{Z}$ and $\mathbb{Z}_n$ (for n not a prime) are rings. In fact, $\mathbb{Z}$ and $\mathbb{F}[x]$ are integer domains, which are the nicest kind of rings. An integer domain is a ring with commutative multiplication and a multiplicative identity 1 such that the nonzero elements are closed under multiplication.

Exercise 6. Show that an integer domain with a finite number of elements must be a finite field. [Hint: consider its cyclic multiplicative subgroups.] □

7.5.1 Definitions

A polynomial $g(x)$ is said to be a *divisor* of an polynomial $f(x)$ if $f(x)$ is a polynomial multiple of $g(x)$; *i.e.,* $f(x) = q(x)g(x)$ for some polynomial $q(x)$. Thus all polynomials are trivially divisors of the zero polynomial 0.

The polynomials that have polynomial inverses are the nonzero degree-0 polynomials $\beta \in \mathbb{F}^* = \mathbb{F} - \{0\}$. These are called the *units* of $\mathbb{F}[x]$. If $u(x)$ is a unit polynomial and $g(x)$ is a divisor of $f(x)$, then $u(x)g(x)$ is a divisor of $f(x)$ and $g(x)$ is a divisor of $u(x)f(x)$. Thus the factorization of a polynomial can be unique only up to a unit polynomial $u(x)$, and $u(x)f(x)$ has the same divisors as $f(x)$.

A monic polynomial is a nonzero polynomial $f(x)$ of degree m with high-order coefficient f_m equal to 1; *i.e.,* $f(x) = f_0 + f_1x + f_2x^2 + \cdots + x^m$. Every nonzero polynomial $g(x)$ may be written as the product $g(x) = g_m f(x)$ of a monic polynomial $f(x)$ of the same degree with a unit polynomial $u(x) = g_m$, and the product of two monic polynomials is monic. We may therefore consider only factorizations of monic polynomials into products of monic polynomials.

Every nonzero polynomial $f(x)$ is divisible by 1 and $f(x)$; these divisors are called trivial. A polynomial $g(x)$ is said to be a *factor* of a polynomial $f(x)$ if $g(x)$ is monic and a nontrivial divisor of $f(x)$. Thus the degree of any factor $g(x)$ of $f(x)$ satisfies $1 \leq \deg g(x) < \deg f(x)$.

A polynomial $g(x)$ of degree 1 or more that has no factors is called an *irreducible polynomial*, and a monic irreducible polynomial is called a *prime polynomial*. Our goal now is to show that every monic polynomial has a unique factorization into prime polynomial factors.

7.5.2 Mod-$g(x)$ arithmetic

Given a monic polynomial $g(x)$ of degree m, every polynomial $f(x)$ may be expressed as $f(x) = q(x)g(x)+r(x)$ for some polynomial remainder $r(x)$ such that $\deg r(x) < m$ and some polynomial quotient $q(x)$. This may be proved by the Euclidean long division algorithm of high school, with component operations in $\mathbb{F}$; *i.e.,* divide $g(x)$ into $f(x)$ by long division, high-degree terms first, stopping when the degree of the remainder is less than that of $g(x)$. The following exercise shows that the resulting quotient $q(x)$ and remainder $r(x)$ are unique.

Exercise 7 (Euclidean division algorithm).

(a) For the set $\mathbb{F}[x]$ of polynomials over any field $\mathbb{F}$, show that the distributive law holds: $(f_1(x) + f_2(x))h(x) = f_1(x)h(x) + f_2(x)h(x)$.

(b) Use the distributive law to show that for any given $f(x)$ and $g(x)$ in $\mathbb{F}[x]$, there is a unique $q(x)$ and $r(x)$ with $\deg r(x) < \deg g(x)$ such that $f(x) = q(x)g(x) + r(x)$. □

The remainder polynomial $r(x)$, denoted by $r(x) = f(x) \bmod g(x)$, is the more important part of this decomposition. The set of all possible remainder polynomials is the set $R_{\mathbb{F},m} = \{r_0 + r_1x + \cdots + r_{m-1}x^{m-1} \mid r_j \in \mathbb{F}, 0 \leq j \leq m-1\}$, whose size is $|R_{\mathbb{F},m}| = |\mathbb{F}|^m$. Evidently $g(x)$ is a divisor of $f(x)$ if and only if $f(x) \bmod g(x) = 0$.

Remainder arithmetic using the remainder set $R_{\mathbb{F},m}$ is called "mod-$g(x)$ arithmetic." The rules for mod-$g(x)$ arithmetic follow from the rules for polynomial arithmetic as follows. Let $r(x) = f(x) \bmod g(x)$ and $s(x) = h(x) \bmod g(x)$; then, as polynomials, $r(x) = f(x) - q(x)g(x)$ and $s(x) = h(x) - t(x)g(x)$ for some quotient polynomials $q(x)$ and $t(x)$. Then

$$\begin{aligned} f(x) + h(x) &= r(x) + s(x) - (q(x) + t(x))g(x); \\ f(x)h(x) &= r(x)s(x) - (q(x)s(x) + t(x)r(x))g(x) + q(x)t(x)g^2(x). \end{aligned}$$

Hence $(f(x) + h(x)) \bmod g(x) = (r(x) + s(x)) \bmod g(x)$ and $f(x)h(x) \bmod g(x) = r(x)s(x) \bmod g(x)$. In other words, the mod-$g(x)$ remainder of the sum or product of two polynomials is equal to the mod-$g(x)$ remainder of the sum or product of their mod-$g(x)$ remainders.

The mod-$g(x)$ addition and multiplication rules are therefore defined as follows:

$$\begin{aligned} r(x) \oplus s(x) &= (r(x) + s(x)) \bmod g(x); \\ r(x) * s(x) &= (r(x)s(x)) \bmod g(x), \end{aligned}$$

where "$r(x)$" and "$s(x)$" denote elements of the remainder set $R_{\mathbb{F},m}$ on the left and the corresponding ordinary polynomials on the right. This makes mod-$g(x)$ arithmetic consistent with ordinary polynomial arithmetic in the sense of the previous paragraph.

Note that the mod-$g(x)$ addition rule is just componentwise addition of coefficients in $\mathbb{F}$. In this sense the additive groups of $R_{\mathbb{F},m}$ and of the vector space $\mathbb{F}^m$ of m-tuples over $\mathbb{F}$ are isomorphic.

7.5.3 Unique factorization

By definition, every monic polynomial $f(x)$ is either irreducible or can be factored into a product of monic polynomial factors, each of lower degree. In turn, if a factor is not irreducible, it can be factored further. Since factor degrees are decreasing but bounded below by 1, we must eventually arrive at a product of monic irreducible (prime) polynomials. The following theorem shows that there is only one such set of prime polynomial factors, regardless of the order in which the polynomial is factored.

Theorem 7.8 (Unique factorization of polynomials) *Over any field $\mathbb{F}$, every monic polynomial $f(x) \in \mathbb{F}[x]$ of degree $m \geq 1$ may be written in the form*

$$f(x) = \prod_{i=1}^{k} a_i(x),$$

where each $a_i(x), 1 \leq i \leq k$, is a prime polynomial in $\mathbb{F}[x]$. This factorization is unique, up to the order of the factors.

Proof. We have already shown that $f(x)$ may be factored in this way, so we need only prove uniqueness. Thus assume hypothetically that the theorem is false and let m be the smallest degree such that there exists a degree-m monic polynomial $f(x)$ with more than one such factorization,

$$f(x) = a_1(x) \cdots a_k(x) = b_1(x) \cdots b_j(x); \qquad j, k \geq 1, \tag{7.2}$$

where $a_1(x), \ldots, a_k(x)$ and $b_1(x), \ldots, b_j(x)$ are prime polynomials. We will show that this implies a polynomial $f'(x)$ with degree less than m with non-unique factorization, and this contradiction will prove the theorem. Now $a_1(x)$ cannot appear on the right side of (7.2), else it could be factored out for an immediate contradiction. Similarly, $b_1(x)$ cannot appear on the left. Without loss of generality, assume $\deg b_1(x) \leq \deg a_1(x)$. By the Euclidean division algorithm, $a_1(x) = q(x)b_1(x) + r(x)$. Since $a_1(x)$ is irreducible, $r(x) \neq 0$ and $0 \leq \deg r(x) < \deg b_1(x) \leq \deg a_1(x)$.

Thus $r(x)$ has a prime factorization $r(x) = \beta r_1(x) \cdots r_n(x)$, where β is the high-order coefficient of $r(x)$, and $b_1(x)$ is not a divisor of any of the $r_i(x)$, since it has greater degree. Substituting into (7.2), we have

$$(q(x)b_1(x) + \beta r_1(x) \cdots r_n(x))a_2(x) \cdots a_k(x) = b_1(x) \cdots b_j(x),$$

or, defining $f'(x) = r_1(x) \cdots r_n(x)a_2(x) \cdots a_k(x)$ and rearranging terms,

$$f'(x) = r_1(x) \cdots r_n(x)a_2(x) \cdots a_k(x) = \beta^{-1}b_1(x)(b_2(x) \cdots b_j(x) - q(x)a_2(x) \cdots a_k(x)).$$

Now $f'(x)$ is monic, because it is a product of monic polynomials; it has degree less than $f(x)$, since $\deg r(x) < \deg a_1(x)$; and it has two different factorizations, with $b_1(x)$ a factor in one but not a divisor of any of the factors in the other; contradiction. □

Exercise 8. Following this proof, prove unique factorization for the integers $\mathbb{Z}$. □

7.5.4 Enumerating prime polynomials

The prime polynomials in $\mathbb{F}[x]$ are analogous to the prime numbers in $\mathbb{Z}$. One way to enumerate the prime polynomials is to use an analogue of the sieve of Eratosthenes. For integers, this method goes as follows: Start with a list of all integers greater than 1. The first integer on the list is 2, which is prime. Erase all multiples of 2 (even integers). The next remaining integer is 3, which must be the next prime. Erase all multiples of 3. The next remaining integer is 5, which must be the next prime. Erase all multiples of 5. And so forth.

Similarly, to find the prime polynomials in $\mathbb{F}_2[x]$, for example, first list all polynomials of degree 1 or more in $\mathbb{F}_2[x]$ in order of degree. (Note that all nonzero polynomials in $\mathbb{F}_2[x]$ are monic.) No degree-1 polynomial can have a factor, so the two degree-1 polynomials, x and $x + 1$, are both prime. Next, erase all degree-2 multiples of x and $x + 1$, namely

$$\begin{aligned} x^2 &= x * x; \\ x^2 + x &= x * (x + 1); \\ x^2 + 1 &= (x + 1) * (x + 1) \end{aligned}$$

from the list of four degree-2 polynomials. This leaves one prime degree-2 polynomial, namely $x^2 + x + 1$. Next, erase all degree-3 multiples of x, $x + 1$, and $x^2 + x + 1$ from the list of eight degree-3 polynomials, namely the six polynomials

$$\begin{aligned} x^3 &= x * x * x; \\ x^3 + x^2 &= (x + 1) * x * x; \\ x^3 + x &= (x + 1) * (x + 1) * x; \\ x^3 + x^2 + x &= x * (x^2 + x + 1); \\ x^3 + 1 &= (x + 1) * (x^2 + x + 1); \\ x^3 + x^2 + x + 1 &= (x + 1) * (x + 1) * (x + 1). \end{aligned}$$

The remaining two polynomials, namely $x^3 + x^2 + 1$ and $x^3 + x + 1$, must therefore be prime.

Exercise 9. Find all prime polynomials in $\mathbb{F}_2[x]$ of degrees 4 and 5. [Hint: There are three prime polynomials in $\mathbb{F}_2[x]$ of degree 4 and six of degree 5.] □

Continuing in this way, we may list all prime polynomials in $\mathbb{F}_2[x]$ up to any desired degree.

It turns out that the number $N(m)$ of prime polynomials of $\mathbb{F}_2[x]$ of degree m is $N(m) = 2, 1, 2, 3, 6, 9, 18, 30, 56, 99, \ldots$ for $m = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, \ldots$. (In Section 7.9 we will give a simpler method to compute $N(m)$, and will show that $N(m) > 0$ for all m.)

A similar sieve algorithm may be used to find the prime polynomials in $\mathbb{F}[x]$ over any finite field $\mathbb{F}$. The algorithm starts with a listing of the monic polynomials ordered by degree, and successively erases the multiples of lower-degree prime polynomials.

7.6 A construction of a field with p^m elements

We now show how to construct a field with p^m elements for any prime integer p and positive integer $m \geq 1$. Its elements will be the set $R_{\mathbb{F},m}$ of remainder polynomials of degree less than m, and multiplication will be defined modulo an irreducible polynomial $g(x)$ of degree m. We will subsequently show that that every finite field is isomorphic to a finite field that is constructed in this way.

The construction assumes the existence of a prime polynomial $g(x) \in \mathbb{F}_p[x]$ of degree m. The proof that such a polynomial exists for all prime p and $m \geq 1$ will be deferred until later. The field that we construct will be denoted by $\mathbb{F}_{g(x)}$.

The set of elements of $\mathbb{F}_{g(x)}$ will be taken to be the mod-$g(x)$ remainder set $R_{\mathbb{F}_p,m} = \{r_0 + r_1 x + \cdots + r_{m-1}x^{m-1} \mid r_j \in \mathbb{F}_p, 0 \leq j \leq m-1\}$, whose size is $|R_{\mathbb{F}_p,m}| = p^m$.

The addition and multiplication rules will be taken to be those of mod-$g(x)$ arithmetic. We must show that the axioms of a field are satisfied with these definitions.

The associative, commutative and distributive laws for mod-$g(x)$ arithmetic follow from the corresponding laws for ordinary polynomial arithmetic.

Mod-$g(x)$ addition of two remainder polynomials in $\mathbb{F}_{g(x)}$ yields a remainder polynomial of degree $< m$ in $\mathbb{F}_{g(x)}$. $\mathbb{F}_{g(x)}$ evidently forms an abelian group under mod-$g(x)$ addition. (As already mentioned, this group is isomorphic to the additive group of $(\mathbb{F}_p)^m$.)

Mod-$g(x)$ multiplication of two remainder polynomials $r(x), s(x)$ yields the remainder polynomial $t(x) = r(x)s(x) \bmod g(x)$. The following exercise shows that the nonzero elements of $\mathbb{F}_{g(x)}$ form an abelian group under mod-$g(x)$ multiplication:

Exercise 10. Let $g(x)$ be a prime polynomial of degree m, and let $r(x), s(x), t(x)$ be polynomials in $\mathbb{F}_{g(x)}$.

(a) Prove the distributive law, *i.e.*, $(r(x)+s(x))*t(x) = r(x)*t(x)+s(x)*t(x)$. [Hint: Express each product as a remainder using the Euclidean division algorithm.]

(b) For $r(x) \neq 0$, show that $r(x) * s(x) \neq r(x) * t(x)$ if $s(x) \neq t(x)$.

(c) For $r(x) \neq 0$, show that as $s(x)$ runs through all nonzero polynomials in $\mathbb{F}_{g(x)}$, the product $r(x) * s(x)$ also runs through all nonzero polynomials in $\mathbb{F}_{g(x)}$.

(d) Using part (c) and Theorem 7.1, show that the nonzero elements of $\mathbb{F}_{g(x)}$ form an abelian group under mod-$g(x)$ multiplication. □

Since we have verified the three field axioms, we have proved:

Theorem 7.9 (Construction of $\mathbb{F}_{g(x)}$) *If $g(x)$ is an prime polynomial of degree m over a prime field $\mathbb{F}_p$, then the set of remainder polynomials $R_{\mathbb{F}_p,m}$ with mod-$g(x)$ arithmetic forms a finite field $\mathbb{F}_{g(x)}$ with p^m elements.*

Example 1. Let us construct a finite field with $2^2 = 4$ elements using the prime degree-2 polynomial $g(x) = x^2 + x + 1 \in \mathbb{F}_2[x]$.

There are four remainder polynomials mod $x^2 + x + 1$, namely $\{0, 1, x, x + 1\}$. Addition is componentwise mod 2. For multiplication, note that $x*x = x+1$ since x^2 mod $(x^2+x+1) = x+1$. Also $x * x * x = x * (x + 1) = 1$ since x^3 mod $(x^2 + x + 1) = 1$. The three nonzero elements $\{1, x, x + 1\}$ thus form a cyclic group under mod-$g(x)$ multiplication, which verifies the second field axiom for this example.

The complete mod-$g(x)$ addition and multiplication tables are as follows:

$\oplus$	0	1	x	$x+1$
0	0	1	x	$x+1$
1	1	0	$x+1$	x
x	x	$x+1$	0	1
$x+1$	$x+1$	x	1	0

$*$	0	1	x	$x+1$
0	0	0	0	0
1	0	1	x	$x+1$
x	0	x	$x+1$	1
$1+x$	0	$x+1$	1	x

7.7 The multiplicative group of $\mathbb{F}_q^*$ is cyclic

In this section we consider an arbitrary finite field $\mathbb{F}_q$ with q elements. By the second field axiom, the set $\mathbb{F}_q^*$ of all $q - 1$ nonzero elements must form a finite abelian group under multiplication. In this section we will show that this group is actually cyclic.

We start by showing that every element of $\mathbb{F}_q^*$ is a root of the polynomial $x^{q-1} - 1 \in \mathbb{F}_q[x]$. Thus we first need to discuss roots of polynomials over arbitrary fields.

7.7.1 Roots of polynomials

Let $\mathbb{F}[x]$ be the set of polynomials over an arbitrary field $\mathbb{F}$. If $f(x) \in \mathbb{F}[x]$ has a degree-1 factor $x - \alpha$ for some $\alpha \in \mathbb{F}$, then α is called a *root* of $f(x)$.

Since any $f(x)$ may be uniquely expressed as $f(x) = q(x)(x-\alpha)+\beta$ for some quotient $q(x)$ and some $\beta \in \mathbb{F}$ (*i.e.,* for some remainder $r(x) = \beta$ of degree less than 1), it follows that $f(\alpha) = \beta$. Therefore α is a root of $f(x)$ if and only if $f(\alpha) = 0$— *i.e.,* if and only if α is a root of the polynomial equation $f(x) = 0$.

By degree additivity, the degree of a polynomial $f(x)$ is equal to the sum of the degrees of its prime factors, which are unique by unique factorization. Therefore a polynomial of degree m can have at most m degree-1 factors. This yields what is sometimes called the fundamental theorem of algebra:

Theorem 7.10 (Fundamental theorem of algebra) *Over any field $\mathbb{F}$, a monic polynomial $f(x) \in \mathbb{F}[x]$ of degree m can have no more than m roots in $\mathbb{F}$. If it does have m roots $\{\beta_1, \ldots, \beta_m\}$, then the unique factorization of $f(x)$ is $f(x) = (x - \beta_1) \cdots (x - \beta_m)$.* □

Since the polynomial $x^n - 1$ can have at most n roots in $\mathbb{F}$, we have an important corollary:

Theorem 7.11 (Cyclic multiplicative subgroups) *In any field $\mathbb{F}$, the multiplicative group $\mathbb{F}^*$ of nonzero elements has at most one cyclic subgroup of any given order n. If such a subgroup exists, then its elements $\{1, \beta, \ldots, \beta^{n-1}\}$ satisfy*

$$x^n - 1 = (x - 1)(x - \beta) \cdots (x - \beta^{n-1}). \qquad \square$$

For example, the complex multiplicative group $\mathbb{C}^*$ has precisely one cyclic subgroup of each finite size n, consisting of the n complex nth roots of unity. The real multiplicative group $\mathbb{R}^*$ has cyclic subgroups of size 1 ($\{1\}$) and 2 ($\{\pm 1\}$), but none of any larger size.

Exercise 11. For $1 \le j \le n$, the jth *elementary symmetric function* $\sigma_j(S)$ of a set S of n elements of a field $\mathbb{F}$ is the sum of all $\binom{n}{j}$ products of j distinct elements of S. In particular, $\sigma_1(S)$ is the sum of all elements of S, and $\sigma_n(S)$ is the product of all elements of S.

(a) Show that if $S = \{1, \beta, \ldots, \beta^{n-1}\}$ is a cyclic subgroup of $\mathbb{F}^*$, then $\sigma_j(S) = 0$ for $1 \le j \le n-1$ and $\sigma_n(S) = (-1)^{n+1}$. In particular,

$$\sum_{j=0}^{n-1} \beta^j = 0, \quad \text{if } n > 1; \qquad \prod_{j=0}^{n-1} \beta^j = (-1)^{n+1}.$$

Verify for $S = \{\pm 1, \pm i\}$ (the four complex 4th roots of unity).

(b) Prove that for any odd prime integer p,

$$(p-1)! = 1 \cdot 2 \cdot 3 \cdots (p-1) = -1 \bmod p.$$

Verify for $p = 3, 5$ and 7. □

7.7.2 Factoring $x^q - x$ over $\mathbb{F}_q$

For any $\beta \in \mathbb{F}_q^*$, consider the cyclic subgroup $S(\beta) = \{1, \beta, \beta^2, \beta^3, \ldots\}$ of $\mathbb{F}_q^*$ generated by β. The size $|S(\beta)|$ of this subgroup is called the multiplicative order of β.

By the cyclic group theorem, $\beta^{|S(\beta)|} = 1$, and by Lagrange's theorem, $|S(\beta)|$ must divide $|\mathbb{F}_q^*| = q - 1$. It follows that $\beta^{q-1} = 1$ for all $\beta \in \mathbb{F}_q^*$.

In other words, every $\beta \in \mathbb{F}_q^*$ is a root of the polynomial equation $x^{q-1} = 1$, or equivalently of the polynomial $x^{q-1} - 1 \in \mathbb{F}_q[x]$. By the polynomial roots theorem, $x^{q-1} - 1$ can have at most $q - 1$ roots in $\mathbb{F}_q$, so these are all the roots of $x^{q-1} - 1$. Thus $x^{q-1} - 1$ factors into the product of the degree-1 polynomials $x - \beta$ for all $\beta \in \mathbb{F}_q^*$. Moreover, since $0 \in \mathbb{F}_q$ is a root of the polynomial x and $x(x^{q-1} - 1) = x^q - x$, the polynomial $x^q - x$ factors into the product of the degree-1 polynomials $x - \beta$ for all $\beta \in \mathbb{F}_q$.

To summarize:

Theorem 7.12 *In a finite field $\mathbb{F}_q$ with q elements, every nonzero field element $\beta \in \mathbb{F}_q$ satisfies $\beta^{q-1} = 1$ and has a multiplicative order $|S(\beta)|$ that divides $q - 1$. The nonzero elements of $\mathbb{F}_q$ are the $q - 1$ distinct roots of the polynomial $x^{q-1} - 1 \in \mathbb{F}_q[x]$; i.e.,*

$$x^{q-1} - 1 = \prod_{\beta \in \mathbb{F}_q^*} (x - \beta). \tag{7.3}$$

The elements of $\mathbb{F}_q$ are the q distinct roots of the polynomial $x^q - x \in \mathbb{F}_q[x]$; i.e.,

$$x^q - x = \prod_{\beta \in \mathbb{F}_q} (x - \beta). \tag{7.4}$$

Exercise 12.

(a) Verify (7.3) for the prime field $\mathbb{F}_5$.

(b) Verify (7.3) for the field $\mathbb{F}_4$ that was constructed in Example 1. [Hint: use a symbol other than x for the indeterminate in (7.3).] □

7.7.3 Every finite field has a primitive element

A *primitive element* of a finite field $\mathbb{F}_q$ is an element α whose multiplicative order $|S(\alpha)|$ equals $q-1$. If α is a primitive element, then the cyclic group $\{1, \alpha, \alpha^2, \ldots, \alpha^{q-2}\}$ is a set of $q-1$ distinct nonzero elements of $\mathbb{F}_q$, which therefore must be all the nonzero elements. Thus if we can show that $\mathbb{F}_q$ has at least one primitive element, we will have shown that its nonzero elements $\mathbb{F}_q^*$ form a cyclic group under multiplication of size $q-1$.

By Lagrange's theorem, the multiplicative order $|S(\beta)|$ of each nonzero element $\beta \in \mathbb{F}_q^*$ divides $q-1$. Therefore the size d of each cyclic subgroup of $\mathbb{F}_q^*$ divides $q-1$. As we have seen, the number of elements in a cyclic group or subgroup of size d that have order d is the Euler number $\phi(d)$. Since by the cyclic subgroups theorem $\mathbb{F}_q^*$ has at most one cyclic subgroup of each size d, the number of elements in $\mathbb{F}_q^*$ with order less than $q-1$ is at most

$$\sum_{d:\, d|(q-1),\ d \neq q-1} \phi(d).$$

But since the Euler numbers satisfy the relationship (7.1), which in this case is

$$q-1 = \sum_{d:\, d|(q-1)} \phi(d),$$

we conclude that there must be at least $\phi(q-1)$ elements of $\mathbb{F}_q^*$ with order $q-1$. Indeed, since $\mathbb{F}_q^*$ has at most $\phi(q-1)$ elements of order $q-1$, all inequalities must be satisfied with equality; *i.e.,* $\mathbb{F}_q^*$ has precisely $\phi(d)$ elements of order d for each divisor d of $q-1$.

We saw in Exercise 3 that $\phi(q-1) \geq 1$, so a primitive element α of order $q-1$ exists. Thus $\mathbb{F}_q^*$ is cyclic and has one cyclic subgroup of each order d that divides $q-1$. This proves the following theorem:

Theorem 7.13 (Primitive elements) *Given any field $\mathbb{F}_q$ with q elements, the nonzero elements of $\mathbb{F}_q$ form a multiplicative cyclic group $\mathbb{F}_q^* = \{1, \alpha, \alpha^2, \ldots, \alpha^{q-2}\}$. Consequently $\mathbb{F}_q^*$ has $\phi(d) \geq 1$ elements of multiplicative order d for every d that divides $q-1$, and no elements of any other order. In particular, $\mathbb{F}_q^*$ has $\phi(q-1) \geq 1$ primitive elements.*

Henceforth we will usually write the elements of a finite field $\mathbb{F}_q$ as $\{0, 1, \alpha, \alpha^2, \ldots, \alpha^{q-2}\}$, where α denotes a primitive element. For $\mathbb{F}_{g(x)}$, denoting a field element β as a power of α rather than as a remainder polynomial helps to avoid confusion when we consider polynomials in β.

Example 2. The prime field $\mathbb{F}_5$ has $\phi(1) = 1$ element of order 1 (the element 1), $\phi(2) = 1$ element of order 2 (namely 4 = -1), and $\phi(4) = 2$ primitive elements of order 4 (namely, 2 and 3). We can therefore write $\mathbb{F}_5 = \{0, 1, 2, 2^2, 2^3\}$, since $2^2 = 4$ and $2^3 = 3 \bmod 5$. □

Example 3. A field $\mathbb{F}_{16} = \{0, 1, \alpha, \ldots, \alpha^{14}\}$ with 16 elements has

- $\phi(1) = 1$ element of order 1 (the element 1);
- $\phi(3) = 2$ elements of order 3 (α^5 and α^{10});
- $\phi(5) = 4$ elements of order 5 ($\alpha^3, \alpha^6, \alpha^9, \alpha^{12}$), and
- $\phi(15) = 8$ primitive elements of order 15 ($\alpha, \alpha^2, \alpha^4, \alpha^7, \alpha^8, \alpha^{11}, \alpha^{13}, \alpha^{14}$). □

The "logarithmic" representation of the nonzero elements of $\mathbb{F}_q$ as distinct powers of a primitive element α is obviously highly convenient for multiplication and division. Multiplication in $\mathbb{F}_q$ is often carried out by using such a "log table" to convert a polynomial $f(x) \in \mathbb{F}_q$ to the exponent i such that $f(x) = \alpha^i$, and then using an inverse "antilog table" to convert back after adding or subtracting exponents. (Note that the zero element can be included in this scheme if we define $0 = \alpha^{-\infty}$.)

7.8 Every finite field is isomorphic to a field $\mathbb{F}_{g(x)}$

We now wish to show that every finite field $\mathbb{F}_q$ is isomorphic to a field $\mathbb{F}_{g(x)}$ of the type that we have previously constructed. In particular, this will show that the number of elements of a finite field must be $q = p^m$, a prime power.

The development relies on the properties of minimal polynomials, which are the factors that appear in the unique factorization of $x^q - x$ over the prime subfield $\mathbb{F}_p$ of $\mathbb{F}_q$.

7.8.1 Factoring $x^q - x$ into minimal polynomials over $\mathbb{F}_p$

Again, consider any field $\mathbb{F}_q$ with q elements. We have seen in Theorem 7.12 that the polynomial $x^q - x \in \mathbb{F}_q[x]$ factors completely into q deqree-1 factors $x - \beta \in \mathbb{F}_q[x], \beta \in \mathbb{F}_q$.

We have also seen that if $\mathbb{F}_q$ has characteristic p, then $\mathbb{F}_q$ has a prime subfield $\mathbb{F}_p$ with p elements. The prime subfield $\mathbb{F}_p$ contains the integers of $\mathbb{F}_q$, which include $\{0, \pm 1\}$. Therefore we may regard $x^q - x$ alternatively as a polynomial in $\mathbb{F}_p[x]$.

By unique factorization, $x^q - x$ factors over $\mathbb{F}_p$ into a unique product of prime polynomials $g_i(x) \in \mathbb{F}_p[x]$:

$$x^q - x = \prod_i g_i(x). \tag{7.5}$$

Since each coefficient of $g_i(x)$ is an element of $\mathbb{F}_p \subseteq \mathbb{F}_q$, it is also an element of $\mathbb{F}_q$, so $g_i(x)$ is also a monic polynomial in $\mathbb{F}_q[x]$. We therefore have the following two factorizations of $x^q - x$ in $\mathbb{F}_q[x]$:

$$x^q - x = \prod_{\beta \in \mathbb{F}_q} (x - \beta) = \prod_i g_i(x). \tag{7.6}$$

Since the first factorization is the unique prime factorization, it follows that each monic polynomial $g_i(x)$ of degree greater than 1 must be reducible over $\mathbb{F}_q$, and must factor into a product of degree-1 monic polynomials; *i.e.,*

$$g_i(x) = \prod_{j=1}^{\deg g_i(x)} (x - \beta_{ij}). \tag{7.7}$$

The prime polynomials $g_i(x)$ are called the *minimal polynomials* of $\mathbb{F}_q$. Since each $\beta \in \mathbb{F}_q$ appears exactly once on the left side of (7.6), it also appears as a factor in exactly one minimal polynomial in (7.7). Thus the elements of $\mathbb{F}_q$ are partitioned into disjoint sets $\{\beta_{i1}, \ldots, \beta_{ik}\}$ where $k = \deg g_i(x)$, and each $\beta \in \mathbb{F}_q$ is a root of exactly one minimal polynomial of $\mathbb{F}_q$, called the minimal polynomial of β.

The key property of the minimal polynomial of β is the following:

Lemma 7.14 *Let $g(x)$ be the minimal polynomial of any given $\beta \in \mathbb{F}_q$. Then $g(x)$ is the monic polynomial of least degree in $\mathbb{F}_p[x]$ such that $g(\beta) = 0$. Moreover, for any $f(x) \in \mathbb{F}_p[x]$, $f(\beta) = 0$ if and only if $g(x)$ divides $f(x)$.*

Proof: Let $h(x) \in \mathbb{F}_p[x]$ be a monic polynomial of least degree such that $h(\beta) = 0$. Using the Euclidean division algorithm, $g(x) = q(x)h(x) + r(x)$ where $\deg r(x) < \deg h(x)$. Since $h(\beta) = g(\beta) = 0$, we must have $r(\beta) = 0$. By the smallest degree property of $h(x)$, this implies that $r(x) = 0$, so $h(x)$ divides $g(x)$. But since $g(x)$ is irreducible, $h(x)$ cannot have degree less than $g(x)$; *i.e.,* $\deg h(x) = \deg g(x)$. Moreover, since both $h(x)$ and $g(x)$ are monic, this implies that $h(x) = g(x)$. Thus $g(x)$ is the monic polynomial of least degree in $\mathbb{F}_p[x]$ such that $g(\beta) = 0$.

Now let $f(x)$ be any polynomial in $\mathbb{F}_p[x]$ that satisfies $f(\beta) = 0$. By Euclidean division, $f(x) = q(x)g(x) + r(x)$ with $\deg r(x) < \deg g(x)$. Thus $r(\beta) = f(\beta) = 0$. Since $\deg r(x) < \deg g(x)$, $r(\beta) = 0$ if and only if $r(x) = 0$; *i.e.,* if and only if $g(x)$ divides $f(x)$. □

Example 1 (cont.). Again consider the field $\mathbb{F}_4$ of Example 1, whose elements we now write as $\{0, 1, \alpha, \alpha^2\}$, where α may be taken as x or $x + 1$. This field has characteristic 2. The prime factorization of the binary polynomial $x^4 - x = x^4 + x \in \mathbb{F}_2[x]$ is

$$x^4 + x = x(x + 1)(x^2 + x + 1),$$

so the minimal polynomials of $\mathbb{F}_4$ are x, $x + 1$ and $x^2 + x + 1$. The elements 0 and $1 \in \mathbb{F}_4$ are the roots of x and $x + 1$, respectively. From (7.7), the other two elements of $\mathbb{F}_4$, namely α and α^2, must be roots of $x^2 + x + 1 \in \mathbb{F}_2[x]$. We verify that

$$x^2 + x + 1 = (x + \alpha)(x + \alpha^2)$$

since $\alpha + \alpha^2 = 1$ and $\alpha * \alpha^2 = \alpha^3 = 1$. □

7.8.2 Valuation maps, minimal polynomials and subfields

Given a field $\mathbb{F}_q$ with prime subfield $\mathbb{F}_p$, we now consider evaluating a nonzero polynomial $f(x) = \sum_i f_i x^i \in \mathbb{F}_p[x]$ at an element $\beta \in \mathbb{F}_q$ to give a value

$$f(\beta) = \sum_{i=0}^{\deg f(x)} f_i \beta^i$$

in $\mathbb{F}_q$, where f_i is taken as an element of $\mathbb{F}_q$ for the purposes of this evaluation. The value of the zero polynomial at any β is 0.

The value $f(\beta)$ depends on both the polynomial $f(x)$ and the field element $\beta \in \mathbb{F}_q$. Rather than regarding $f(\beta)$ as a function of β, as the notation suggests, we will regard $f(\beta)$ as a function of the polynomial $f(x) \in \mathbb{F}_p[x]$ for a fixed β. In other words, we consider the map $m_\beta : \mathbb{F}_p[x] \to \mathbb{F}_q$ that is defined by $m_\beta(f(x)) = f(\beta)$.

The set of values $m_\beta(\mathbb{F}_p[x])$ of this map as $f(x)$ ranges over polynomials in $\mathbb{F}_p[x]$ is by definition the subset of elements $\mathbb{G}_\beta \subseteq \mathbb{F}_q$ that can be expressed as linear combinations over $\mathbb{F}_p$ of powers of β. We will show that $\mathbb{G}_\beta$ forms a subfield of $\mathbb{F}_q$ that is isomorphic to the polynomial remainder field $\mathbb{F}_{g(x)}$, where $g(x)$ is the minimal polynomial of β, namely the monic polynomial of least degree such that $g(\beta) = 0$.

We observe that the map $m_\beta : \mathbb{F}_p[x] \to \mathbb{F}_q$ preserves addition and multiplication; *i.e.*, $m_\beta(f_1(x) + f_2(x)) = m_\beta(f_1(x)) + m_\beta(f_2(x))$ since both sides equal $f_1(\beta) + f_2(\beta)$, and $m_\beta(f_1(x)f_2(x)) = m_\beta(f_1(x))m_\beta(f_2(x))$ since both sides equal $f_1(\beta)f_2(\beta)$.

We can now prove the desired isomorphism between the fields $\mathbb{F}_{g(x)}$ and $\mathbb{G}_\beta$:

Theorem 7.15 (Subfields generated by $\beta \in \mathbb{F}_q$) *For any $\beta \in \mathbb{F}_q$, let $g(x)$ be the minimal polynomial of β. Then the set of all linear combinations $\mathbb{G}_\beta = \{f(\beta) = \sum_i f_i\beta^i, f(x) \in \mathbb{F}_p[x]\}$ over $\mathbb{F}_p$ of powers of β is equal to the set $\{r(\beta), r(x) \in R_{\mathbb{F}_p,m}\}$ of values of remainder polynomials $r(x) \in R_{\mathbb{F}_p,m}$, and $\mathbb{G}_\beta$ is a field which is isomorphic to the field $\mathbb{F}_{g(x)}$ under the correspondence $r(\beta) \in \mathbb{G}_\beta \leftrightarrow r(x) \in R_{\mathbb{F}_p,m}$.*

Proof. We first verify that the correspondence $m_\beta : R_{\mathbb{F}_p,m} \to \mathbb{G}_\beta$ is one-to-one (invertible). First, if $f(\beta)$ is any element of $\mathbb{G}_\beta$, then by Euclidean division we can write $f(x) = q(x)g(x)+r(x)$ where $r(x) \in R_{\mathbb{F}_p,m}$, and then $f(\beta) = q(\beta)g(\beta)+r(\beta) = r(\beta)$, so $f(\beta) = r(\beta)$ for some remainder polynomial $r(x)$. Thus $m_\beta(R_{\mathbb{F}_p,m}) = m_\beta(\mathbb{F}_p[x]) = \mathbb{G}_\beta$. On the other hand, no two remainder polynomials $r(x), s(x)$ with degrees less than m can evaluate to the same element of $\mathbb{G}_\beta$, because if $r(\beta) = s(\beta)$, then $r(x) - s(x)$ is a nonzero polynomial of degree less than $g(x)$ that evaluates to 0, contradiction.

Now, as we have already seen, $m_\beta(r(x) + s(x)) = m_\beta(r(x)) + m_\beta(s(x))$ and $m_\beta(r(x)s(x)) = m_\beta(r(x))m_\beta(s(x))$, which verifies that this correspondence is an isomorphism. □

We remark that $\mathbb{G}_\beta$ may be viewed as the smallest subfield of $\mathbb{F}_q$ containing the element β, because any subfield containing β must also contain all powers of β and all linear combinations of powers over $\mathbb{F}_p$.

7.8.3 Isomorphism theorems

We have shown that every finite field $\mathbb{F}_q$ contains a primitive element α. In this case, the subfield $\mathbb{G}_\alpha$ consisting of all linear combinations over $\mathbb{F}_p$ of powers of α must evidently be the whole field $\mathbb{F}_q$. Thus we obtain our main theorem:

Theorem 7.16 (Every finite field is isomorphic to a field $\mathbb{F}_{g(x)}$) *Every finite field $\mathbb{F}_q$ of characteristic p with q elements is isomorphic to a polynomial remainder field $\mathbb{F}_{g(x)}$, where $g(x)$ is a prime polynomial in $\mathbb{F}_p[x]$ of degree m. Hence $q = p^m$ for some positive integer m.*

Exercise 13. For which integers $q, 1 \le q \le 12$, does a finite field $\mathbb{F}_q$ exist? □

Finally, we wish to show that all fields with p^m elements are isomorphic. The following lemma shows that every prime polynomial $g(x)$ of degree m (we are still assuming that there exists at least one) is a minimal polynomial of every field with p^m elements:

Lemma 7.17 *Every prime polynomial $g(x) \in \mathbb{F}_p[x]$ of degree m divides $x^{p^m} - x$.*

Proof. If $g(x)$ is a prime polynomial in $\mathbb{F}_p[x]$ of degree m, then the set $R_{\mathbb{F}_p,m}$ with mod-$g(x)$ arithmetic forms a field $\mathbb{F}_{g(x)}$ with p^m elements. The remainder polynomial $x \in R_{\mathbb{F}_p,m}$ is a field element $\beta \in \mathbb{F}_{g(x)}$. Evidently $g(\beta) = 0$, but $r(\beta) \neq 0$ if $\deg r(x) < m$; therefore $g(x)$ is the minimal polynomial of β. Since $\beta^{p^m-1} = 1$, β is a root of $x^{p^m-1} - 1$. This implies that $g(x)$ divides $x^{p^m-1} - 1$, and thus also $x^{p^m} - x$. □

Consequently every field of size p^m includes m elements whose minimal polynomial is $g(x)$. Therefore by the same construction as above, we can prove:

Theorem 7.18 (All finite fields of the same size are isomorphic) *For any prime polynomial $g(x) \in \mathbb{F}_p[x]$ of degree m, every field of p^m elements is isomorphic to the polynomial remainder field $\mathbb{F}_{g(x)}$.*

7.8.4 More on the factorization of $x^{p^m} - x$

We can now obtain further information on the factorization of $x^q - x$. In view of Theorem 7.16, we now set $q = p^m$.

We first show that the set of roots of a minimal polynomial $g_i(x) \in \mathbb{F}_p[x]$ is closed under the operation of taking the pth power. This follows from the curious but important fact that over a field $\mathbb{F}$ of characteristic p, taking the pth power is a linear operation. For example, when $p = 2$, squaring is linear because

$$(\alpha + \beta)^2 = \alpha^2 + \alpha\beta + \alpha\beta + \beta^2 = \alpha^2 + \beta^2.$$

More generally, over any field $\mathbb{F}$,

$$(\alpha + \beta)^p = \sum_{j=0}^{p} \binom{p}{j} \alpha^j \beta^{p-j},$$

where $\binom{p}{j}\alpha^j\beta^{p-j}$ denotes the sum of $\binom{p}{j}$ terms equal to $\alpha^j\beta^{p-j}$. If $\mathbb{F}$ has characteristic p, then the integer $\binom{p}{j} = p!/(j!)((n-j)!)$ may be reduced mod p. Now $p!$ contains a factor of p, but for $1 \le j \le p-1$, $j!$ and $(n-j)!$ do not contain a factor of p. Therefore $\binom{p}{j} = 0 \bmod p$ for $1 \le j \le p-1$, and

$$(\alpha + \beta)^p = \alpha^p + \beta^p.$$

By taking the pth power n times, we may extend this result as follows:

Lemma 7.19 (Linearity of taking the p^nth power) *Over any field $\mathbb{F}$ of characteristic p, for any $n \ge 1$, taking the p^nth power is linear; i.e.,*

$$(\alpha + \beta)^{p^n} = \alpha^{p^n} + \beta^{p^n}.$$

Note that if $\mathbb{F}$ has $q = p^m$ elements, then $\beta^{p^m} = \beta$ for all $\beta \in \mathbb{F}$, so this lemma becomes repetitive for $n \ge m$.

Exercise 14. Using this lemma, prove that if $f(x) = \sum_{i=0}^{m} f_i x^i$, then

$$f^{p^n}(x) = (f_0 + f_1 x + f_2 x^2 + \cdots + f_m x^m)^{p^n} = f_0^{p^n} + f_1^{p^n} x^{p^n} + f_2^{p^n} x^{2p^n} + \cdots + f_m^{p^n} x^{mp^n}.$$

□

This result yields a useful test for whether a polynomial $f(x) \in \mathbb{F}[x]$ is in $\mathbb{F}_p[x]$ or not, and a useful formula in case it is:

Lemma 7.20 (Prime subfield polynomials) *For any field $\mathbb{F}$ of characteristic p and any $f(x) \in \mathbb{F}[x]$, $f^p(x) = f(x^p)$ if and only if $f(x) \in \mathbb{F}_p[x]$; i.e., if and only if all coefficients f_i are in the prime subfield $\mathbb{F}_p \subseteq \mathbb{F}$.*

Proof. By Exercise 14, we have

$$f^p(x) = (f_0 + f_1 x + f_2 x^2 + \cdots + f_n x^n)^p = f_0^p + f_1^p x^p + f_2^p x^{2p} + \cdots + f_n^p x^{np}.$$

Now the elements of $\mathbb{F}$ that are in $\mathbb{F}_p$ are precisely the p roots of the polynomial $x^p - x$; thus $\beta^p = \beta$ if and only if $\beta \in \mathbb{F}_p$. Thus the right side of this equation simplifies to $f(x^p)$ if and only if $f_i \in \mathbb{F}_p$ for all i. □

Exercise 15. Prove that a positive integer n is prime if and only if $(x - a)^n = x^n - a \bmod n$ for every integer a that is relatively prime to n.[5] □

Using Lemma 7.20, we now show that the roots of a minimal polynomial are a *cyclotomic coset* of the form $\{\beta, \beta^p, \beta^{p^2}, \ldots\}$:

Theorem 7.21 (Roots of minimal polynomials) *Let $g(x)$ be a minimal polynomial of a finite field $\mathbb{F}$ with p^m elements. Then the roots of $g(x)$ are a set of the form $\{\beta, \beta^p, \beta^{p^2}, \ldots, \beta^{p^{n-1}}\}$, where n is a divisor of m. Moreover, $g(x)$ divides $x^{p^n} - x$.*

Proof. Let β be any root of $g(x)$. Since $g(x) \in \mathbb{F}_p[x]$, Lemma 7.20 shows that $g(x^p) = g^p(x)$. Therefore $g(\beta^p) = g^p(\beta) = 0$. Thus β^p is also a root of $g(x)$. Iterating, $\beta^{p^2}, \beta^{p^3}, \ldots, \beta^{p^i}, \ldots$ are all roots of $g(x)$. Because $\mathbb{F}$ is finite, these roots cannot all be distinct. Therefore let n be the smallest integer such that $\beta^{p^n} = \beta$. Thus $\beta^{p^j} \neq \beta$ for $1 \leq j < n$. This implies that $\beta^{p^j} \neq \beta^{p^{j+k}}$ for $0 \leq j < n$, $1 \leq k < n$; *i.e.,* all elements of the set $\{\beta, \beta^p, \beta^{p^2}, \ldots, \beta^{p^{n-1}}\}$ are distinct. Thus $\beta, \beta^p, \beta^{p^2}, \ldots$ is a cyclic sequence and $\beta^{p^j} = \beta$ if and only if n is a divisor of j. Since $\beta^{p^m} = \beta$, we see that n must divide m.

Finally, we show that these roots are all of the roots of $g(x)$; *i.e.,* $\deg g(x) = n$ and

$$g(x) = \prod_{i=0}^{n-1} (x - \beta^{p^i}).$$

The right side of this equation is a monic polynomial $h(x) \in \mathbb{F}[x]$ of degree n. Since the roots of $h(x)$ are roots of $g(x)$, $h(x)$ must divide $g(x)$ in $\mathbb{F}[x]$. Now, using Lemma 7.20, we can prove that $h(x)$ is actually a polynomial in $\mathbb{F}_p[x]$, because

$$h^p(x) = \prod_{i=0}^{n-1} (x - \beta^{p^i})^p = \prod_{i=0}^{n-1} (x^p - \beta^{p^{i+1}}) = \prod_{i=0}^{n-1} (x^p - \beta^{p^i}) = h(x^p),$$

where we use the linearity of taking the pth power and the fact that $\beta^{p^n} = \beta$. Therefore, since $g(x)$ has no factors in $\mathbb{F}_p[x]$, $g(x)$ must actually be equal to $h(x)$.

Finally, since the roots of $g(x)$ all satisfy $\beta^{p^n} = \beta$, they are all roots of the polynomial $x^{p^n} - x$, which implies that $g(x)$ divides $x^{p^n} - x$. □

[5] This is the basis of the polynomial-time primality test of [Agrawal, Kayal and Saxena, 2002].

This theorem has some important implications. First, the degree n of a minimal polynomial $g(x)$ of a finite field $\mathbb{F}$ with p^m elements must be a divisor of m. Second, the subfield $\mathbb{G}_\beta$ of $\mathbb{F}$ generated by a root β of $g(x)$ must have p^n elements. Third, $x^{p^n} - x$ divides $x^{p^m} - x$, since the elements of $\mathbb{G}_\beta$ are all the roots of $x^{p^n} - x$ and are also roots of $x^{p^m} - x$.

Conversely, let $g(x)$ be any prime polynomial in $\mathbb{F}_p[x]$ of degree n. Then there is a finite field generated by $g(x)$ with p^n elements. This proves that $g(x)$ divides $x^{p^n} - x$, and thus $g(x)$ divides $x^{p^m} - x$ for every multiple m of n. Thus the divisors of $x^{p^m} - x$ include every prime polynomial in $\mathbb{F}_p[x]$ whose degree n divides m.

Moreover, $x^{p^m} - x$ has no repeated factors. We proved this earlier assuming the existence of a field $\mathbb{F}$ with p^m elements; however, we desire a proof that does not make this assumption. The following exercise yields such a proof.

Exercise 16 ($x^{p^m} - x$ has no repeated factors). The formal derivative of a degree-n polynomial $f(x) \in \mathbb{F}_p[x]$ is defined as

$$f'(x) = \sum_{j=1}^{n} (j \bmod p) f_j x^{j-1}$$

(a) Show that if $f(x) = g(x)h(x)$, then $f'(x) = g'(x)h(x) + g(x)h'(x)$.

(b) Show that an prime polynomial $g(x)$ is a repeated divisor of $f(x)$ if and only if $g(x)$ is a divisor of both $f(x)$ and $f'(x)$.

(c) Show that $x^{p^m} - x$ has no repeated prime factors over $\mathbb{F}_p$. □

Now we can conclude our discussion of the factorization of $x^{p^m} - x$ as follows:

Theorem 7.22 (Factors of $x^{p^m} - x$) *The polynomial $x^{p^m} - x$ factors over $\mathbb{F}_p$ into the product of the prime polynomials in $\mathbb{F}_p[x]$ whose degrees divide m, with no repetitions.* □

For example, over $\mathbb{F}_2$, we have

$$\begin{aligned}
x^2 + x &= x(x+1); \\
x^4 + x &= x(x+1)(x^2+x+1); \\
x^8 + x &= x(x+1)(x^3+x^2+1)(x^3+x+1); \\
x^{16} + x &= x(x+1)(x^2+x+1)(x^4+x^3+1)(x^4+x^3+x^2+x+1)(x^4+x+1).
\end{aligned}$$

Exercise 17. Find all prime polynomials $g(x) \in \mathbb{F}_3[x]$ of degree 1 and 2 over the ternary field $\mathbb{F}_3$. Show that the product of these polynomials is $x^9 - x = x^9 + 2x$. Explain, with reference to $\mathbb{F}_9$. □

7.9 Finite fields $\mathbb{F}_{p^m}$ exist for all prime p and $m \geq 1$

At last we can prove that for every prime p and positive integer m there exists a prime polynomial $g(x) \in \mathbb{F}_p[x]$ of degree m. This will prove the existence of a finite field $\mathbb{F}_{g(x)}$ with p^m elements.

Using the factorization of Theorem 7.22, we will show that there do not exist enough prime polynomials of degree less than m that their product could have degree p^m.

Let $N(n)$ denote the number of prime polynomials over $\mathbb{F}_p$ of degree n. The product of these polynomials has degree $nN(n)$. Since $x^{p^m} - x$ is the product of these polynomials for all divisors n of m, and there are no repeated factors, its degree p^m is equal to

$$p^m = \sum_{n:\, n|m} nN(n) \tag{7.8}$$

This formula may be solved recursively for each $N(m)$, starting with $N(1) = p$.

Exercise 18. Calculate $N(m)$ for $p = 2$ for $m = 1$ to 10. Check your results against those stated in Section 7.5.4. □

Now we are in a position to prove the desired theorem:

Theorem 7.23 (Existence of prime polynomials) *Let $N(m)$ be the number of prime polynomials in $\mathbb{F}_p[x]$ of degree m, which is given recursively by (7.8). For every prime p and positive integer m, $N(m) > 0$.*

Proof. Note first that $nN(n) \leq p^n$. Thus

$$p^m \;\leq\; mN(m) + \sum_{n<m:\, n|m} p^n \;\leq\; mN(m) + (m/2)p^{m/2},$$

where we have upperbounded the number of terms in the sum by $m/2$ and upperbounded each term by $p^{m/2}$, since the largest divisor of m other than m is at most $m/2$. Thus

$$mN(m) \geq p^m - (m/2)p^{m/2} = p^{m/2}(p^{m/2} - m/2).$$

The quantity $p^{m/2} - m/2$ is positive for $p = 2, m = 2$, and is increasing in both p and m. Thus $mN(m)$ is positive for all prime p and all $m \geq 2$. Moreover $N(1) = p$. □

Since a finite field $\mathbb{F}_{g(x)}$ with p^m elements can be constructed from any prime polynomial $g(x) \in \mathbb{F}_p[x]$ of degree m, this implies:

Theorem 7.24 (Existence of finite fields) *For every prime p and positive integer m, there exists a finite field with p^m elements.*

Moreover, for each n that divides m, there exists a unique subfield $\mathbb{G}$ with p^n elements, namely the roots of the polynomial $x^{p^n} - x$:

Theorem 7.25 (Existence of finite subfields) *Every finite field with p^m elements has a subfield with p^n elements for each positive integer n that divides m.*

In summary, the factorization of $x^{p^m} - x$ into minimal polynomials partitions the elements of $\mathbb{F}_{p^m}$ into cyclotomic cosets whose properties are determined by their minimal polynomials. The roots of $g(x)$ have multiplicative order k if $g(x)$ divides $x^k - 1$ and does not divide $x^j - 1$ for $j < k$. Moreover, the roots of $g(x)$ are elements of the subfield with p^n elements if and only if $g(x)$ divides $x^{p^n} - x$, or equivalently if their order k divides $p^n - 1$.

Example 3 (cont.) Over $\mathbb{F}_2$, the polynomial $x^{16} + x$ factors as follows:

$$x^{16} + x = x(x+1)(x^2+x+1)(x^4+x^3+1)(x^4+x^3+x^2+x+1)(x^4+x+1).$$

Moreover, $x^3+1 = (x+1)(x^2+x+1)$ and $x^5+1 = (x+1)(x^4+x^3+x^2+x+1)$. The primitive elements are thus the roots of x^4+x+1 and x^4+x^3+1. If we choose a root of x^4+x+1 as α, then $\mathbb{F}_{16} = \{0, 1, \alpha, \ldots, \alpha^{14}\}$ partitions into cyclotomic cosets as follows:

- One zero element (0), minimal polynomial x;
- One element of order 1 (1), minimal polynomial $x+1$;
- Two elements of order 3 (α^5, α^{10}), minimal polynomial x^2+x+1;
- Four elements of order 5 ($\alpha^3, \alpha^6, \alpha^9, \alpha^{12}$), minimal polynomial $x^4+x^3+x^2+x+1$;
- Four elements of order 15 ($\alpha, \alpha^2, \alpha^4, \alpha^8$), minimal polynomial x^4+x+1;
- Four elements of order 15 ($\alpha^7, \alpha^{14}, \alpha^{13}, \alpha^{11}$), minimal polynomial x^4+x^3+1.

$\mathbb{F}_{16}$ has a prime subfield $\mathbb{F}_2$ consisting of the elements whose minimal polynomials divide x^2+x, namely 0 and 1. It also has a subfield $\mathbb{F}_4$ consisting of the elements whose minimal polynomials divide x^4+x, namely $\{0, 1, \alpha^5, \alpha^{10}\}$. Alternatively, $\mathbb{F}_4^*$ consists of the three elements of $\mathbb{F}_{16}^*$ whose multiplicative orders divide 3. □

Exercise 19 (construction of $\mathbb{F}_{32}$).

(a) Find the prime polynomials in $\mathbb{F}_2[x]$ of degree 5, and determine which have primitive roots.

(b) For some minimal polynomial $g(x)$ with a primitive root α, construct a field $\mathbb{F}_{g(x)}$ with 32 elements. Give a table with the elements partitioned into cyclotomic cosets as above. Specify the minimal polynomial and the multiplicative order of each nonzero element. Identify the subfields of $\mathbb{F}_{g(x)}$.

(c) Show how to do multiplication and division in $\mathbb{F}_{g(x)}$ using this "log table." Discuss the rules for multiplication and division in $\mathbb{F}_{g(x)}$ when one of the field elements involved is the zero element $0 \in \mathbb{F}_{g(x)}$.

(d) [Optional] If you know something about maximum-length shift-register (MLSR) sequences, show that there exists a correspondence between the "log table" given above and a certain MLSR sequence of length 31. □