Finite fields 2018

A winter course at the Nesin Matematik Köyü

David Pierce

January 22–28, 2018
Edited January 29, 2018
Revised March 20, 2018

Matematik Bölümü
Mimar Sinan Güzel Sanatlar Üniversitesi
mat.msgsu.edu.tr/~dpierce/

Abstract

We do some algebra as it arises from number theory. We prove the basic theorems about finite fields by analogy with some basic theorems about the integers.

Preface

The course was scheduled for 9:00–10:30 in the Langlands *dersliği.* I have given the scheduled dates for the course, Monday to Sunday; but I cancelled Sunday's class so I could go home Saturday night. Of the seven students who attended at least one lecture each, only one was left on Saturday anyway.

This was my third winter teaching a similar course. The present notes are edited from a version prepared from the notes of last year's course for my use during this course. Because all but one or two of the students this year were in graduate programs, I apparently went faster than in previous years, and I added the final section, on automorphism groups, in response. In the beginning, after the Introduction, I skipped ahead to §2.3.

Contents

1 Introduction

- $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are *fields.*
- Every field is a *ring.*
- $\mathbb{Z}$ is a ring that is not a field.
- The ring $\mathbb{Z}_n$ of integers *modulo* n is a field if and only if n is a prime p.
- From $\mathbb{Z}_p$, we shall create new finite fields as $\mathbb{C}$ is created from $\mathbb{R}$.
- An important tool will be the *Euclidean Algorithm.*
- This will be part of an analogy between $\mathbb{Z}$ and the ring $K[X]$ of polynomials in X over a field K.

2 Rings

2.1 Ordered ring of integers

This section is a summary of things the reader may already already aware of. Some terminology may be new.

The **integers** or **whole numbers** compose the structure denoted by

$$\mathbb{Z}$$

for the German $\mathfrak{Zahl}$ “number.” The structure is a **commutative ring** for having the following three properties.

1. $\mathbb{Z}$ is a **commutative monoid** with respect to multiplication, because this is commutative and associative and has an *identity* or neutral element, so that the following equations are *identities*—are universally true—in $\mathbb{Z}$:

$$\left.\begin{aligned} xy &= yx, \\ x(yz) &= (xy)z, \\ x \cdot 1 &= x. \end{aligned}\right\} \tag{2.1}$$

2. $\mathbb{Z}$ is an **abelian group** with respect to addition, because it is a commutative monoid in which every element has a **negative** or additive inverse, so that the following are

identities:

$$\left.\begin{aligned} x+y&=y+x,\\ x+(y+z)&=(x+y)+z,\\ x+0&=x,\\ x-x&=0. \end{aligned}\right\} \tag{2.2}$$

3. Multiplication distributes over addition in $\mathbb{Z}$:

$$x\cdot(y+z)=xy+xz. \tag{2.3}$$

These three properties make $\mathbb{Z}$ a commutative ring. There are non-commutative rings, such as the ring of 2×2 matrices with entries from $\mathbb{Z}$. We shall not be looking at any such rings. Therefore, for us,

$$\boxed{\text{the word } \textbf{ring} \text{ shall always mean commutative ring.}}$$

The ring $\mathbb{Z}$ is **linearly ordered,** because it has a subset, denoted by

$$\mathbb{N}$$

for "number," that

1) is closed under addition and multiplication,
2) does not contain 0,
3) contains the negative of its every nonzero non-element.

The elements of $\mathbb{N}$ are the **positive integers** or **counting-numbers.** The linear ordering of $\mathbb{Z}$ is given by

$$x<y \iff y-x\in\mathbb{N}.$$

As an ordering, or more precisely a *strict* ordering, the relation $<$ is irreflexive and transitive, so that the following formulas are universally true:

$$x\nless x,$$
$$x<y\ \&\ y<z \implies x<z.$$

In addition to being transitive, the associated relation $\leqslant$ is reflexive and antisymmetric:

$$x \leqslant x,$$
$$x \leqslant y \ \& \ y \leqslant x \implies x = y.$$

The ordering is linear, because

$$y \nleqslant x \implies x < y.$$

We shall see other ordered rings. $\mathbb{Z}$ is determined among them by the first of the following three properties of $\mathbb{N}$. To define these properties, we describe $x+1$ as the **successor** of x. Then the following three statements are true about the set $\mathbb{N}$.

1. The only subset that both
   (i) contains 1 and
   (ii) contains the successor of its every element

   is the whole set: this is the **inductive property.**
2. The successor of no element is 1.
3. No element is the successor of more than one element.

These three properties can be called the **Peano Axioms.** They involve only the element 1 and the singulary operation $x \mapsto x + 1$ of succession.

One can prove from the Peano Axioms that, on $\mathbb{N}$,

1) there is a commutative, associative binary operation of addition,
2) there is a commutative, associative binary operation of multiplication that distributes over addition, and
3) there is a linear ordering $<$ such that always $x < x + y$.

One way to do this, but not the only way, is to start by proving the following.

1 Recursion Theorem. *If a set A is given with an element b and a singulary operation f, then there is a unique function g from $\mathbb{N}$ to A, defined by* ***recursion****, so that*

(i) $g(1) = b$, *and*
(ii) $g(n+1) = f(g(n))$ *for all* n *in* $\mathbb{N}$.

Then for example addition is defined by requiring, for each k in $\mathbb{N}$, that
(i) $k+1$ be the successor of k, and
(ii) $k+(n+1)$ be the successor of $k+n$.
The desired properties of addition are proved by induction.

One ultimately proves:

2 Well Ordering Theorem. $\mathbb{N}$ *is* ***well ordered,*** *in the sense that every nonempty subset of* $\mathbb{N}$ *has a least element.*

Proving everything from the Peano Axioms is not easy and is sometimes not well understood. We shall take the proofs for granted.

2.2 Fields

If a and b are counting-numbers, we can form from them the positive **fraction** denoted by

$$\frac{a}{b}$$

or a/b. By formal definition, this is the equivalence class of the pair (a, b) with respect to the equivalence relation $\sim$ given by

$$(a, b) \sim (x, y) \iff ay = bx.$$

One has to prove that $\sim$ is indeed an equivalence relation, in the sense of being reflexive, symmetric, and transitive. The positive fractions, along with their negatives and 0 compose the structure

$$\mathbb{Q}$$

(for "quotient"), which is an ordered ring in the natural way learned in school. Since, moreover, every nonzero fraction a/b has a **reciprocal** or multiplicative inverse, namely b/a, the ordered ring $\mathbb{Q}$ is an ordered **field.** The ring $\mathbb{Z}$ is a sub-ring of $\mathbb{Q}$ when we identify every n in $\mathbb{Z}$ with the fraction $n/1$. The ordering of $\mathbb{Z}$ agrees with that of $\mathbb{Q}$. The elements of $\mathbb{Q}$ are the **rational numbers.**

From the rational numbers, one obtains the ordered field $\mathbb{R}$ of **real numbers** by an infinitary process. One can understand real numbers as Dedekind cuts, or as equivalence classes of Cauchy sequences, or Laurent series (and their negatives) in 10^{-1} or some other base.

We shall not need real numbers as such, but the example of obtaining from $\mathbb{R}$ the field $\mathbb{C}$ of **complex numbers** will be useful. $\mathbb{C}$ is first of all the real vector space $\mathbb{R}\oplus\mathbb{R}\,\mathrm{i}$, which has basis $\{1, \mathrm{i}\,\}$. One defines a commutative, associative operation of multiplication on this space by the rule

$$\mathrm{i}^{\,2}+1=0.$$

Then every nonzero complex number $a+b\,\mathrm{i}$ has the reciprocal $(a-b\,\mathrm{i}\,)/(a^2+b^2)$, so $\mathbb{C}$ is a field. From given finite fields, we shall obtain larger finite fields in a similar way.

2.3 Measurement

The word arithmetic derives from the Greek ἀριθμός, which denotes a *number of things,* as opposed to a pure or abstract number. In ordinary language, a number *of things* means some *whole* number of them, and at least two.

Instead of developing numbers *arithmetically,* ultimately by the Peano Axioms as in §2.1, we can work geometrically. The

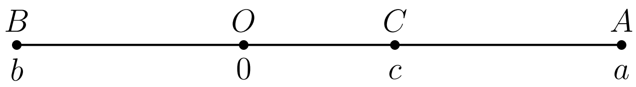

Figure 2.1: Geometric addition

third proposition of Book I of Euclid's *Elements* [3] allows us to add line segments by placing them end to end. By one of the so-called common notions of Euclid, we can just place one segment on top of another, in order to decide whether they are *equal.* Fixing one point as 0 on an infinite straight line, we obtain a binary operation of addition, so that, as in Figure 2.1, if AC and OB are equal as *directed* segments, then

$$a + b = c.$$

commutative, associative operation of addition on the points of the line, so that these points compose an abelian group. If in addition we designate one direction along the line, or one side of 0, as positive, this gives us a linearly ordered abelian group.

To define a commutative, associative operation of *multiplication* on the infinite line that distributes over addition, we need to select a positive point as 1, and we need to pass to a second dimension and define a notion of proportion, so as to prove what is known as Thales's Theorem [10]. Ultimately the straight line becomes the field of real numbers. The idea seems to be due to Descartes [1, 2].

Without going so far, we can recursively define multiplication of real numbers *by* counting-numbers by the rules

$$\begin{aligned} a \cdot 1 &= a, \\ a \cdot (m + 1) &= a \cdot m + a. \end{aligned}$$

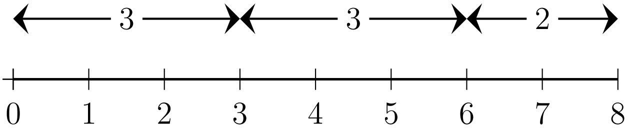

Figure 2.2: Measurement by 3

By induction,

$$a \cdot (k + m) = a \cdot k + a \cdot m, \qquad a \cdot (k \cdot m) = (a \cdot k) \cdot m.$$

In a product $a \cdot n$,

- the real number a is the **multiplicand,** while
- the counting-number n is the **multiplier.**

If we write $a \cdot n$ as b, then we say

- n **divides** b, while
- a **measures** b and is a **submultiple** of b, while b is a **multiple** of a.

When a is itself a counting-number, then

$$a \cdot n = n \cdot a,$$

though again one has to prove this by induction. When we multiply two counting-numbers then, we may forget the distinction between dividing and measuring.

Still the distinction may be useful. When we *measure* 8 by 3, we end up with a remainder of 2, as in Figure 2.2. We can exactly *divide* any number—conceived as a line segment—by 3, as AB is divided into the three equal parts AC, CD, and DB in Figure 2.3.

If again $b = a \cdot n$, we may conceive a as a *number of things,* while n is a pure number, or at least a number of some other kind of thing, but b is a number of the same kind of thing as a.

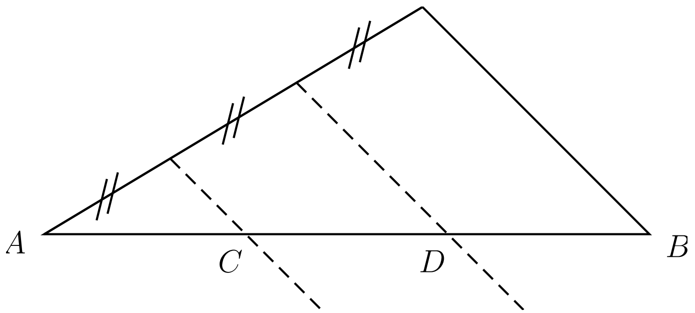

Figure 2.3: Division by 3

Here a might be three apples, while n is a number of children, say four; then b is twelve apples. We can *divide* twelve apples among four children, so that each child ends up with three apples. We can *measure* twelve apples by three apples, ending up with four piles, each consisting of three apples.

Using the Euclidean language of measurement, in a letter to Bernard Frénicle de Bessy (1605–1675) written Thursday, October 18, 1640, Pierre de Fermat (1601–65) describes [4, p. 209] what we now know as Fermat's Theorem (and shall prove on page 38). The emphasis is mine:

> Tout nombre premier **mesure infailliblement** une des puissances $-$ 1 de quelque progression que se soit, et l'exposant de la dite puissance est **sous-multiple** du nombre premier donné $-$ 1; et, après qu'on a trouvé la première puissance qui satisfait à la question, toutes celles dont les exposants sont multiples de l'exposant de la première satisfont tout de même à la question.

In his *Source Book in Mathematics, 1200–1800* [5, p. 28], Struik translates Fermat as saying "is a factor of" instead of measures evenly, "divisor" instead of submultiple.

> Every prime number is always a factor [*mesure infailliblement*] of one of the powers of any progression minus 1, and the exponent of this power is a divisor of the prime number minus 1. After one has found the first power that satisfies the proposition, all those powers of which the exponents are multiples of the exponent of the first power also satisfy the proposition.

Struik misdates the letter as being of October 10, 1640; but such mistakes may be inevitable in such a compilation as his. He modernizes Fermat's language, but this may mean losing the *geometrical* sense of numbers that Fermat inherited from Euclid.

2.4 Division

In $\mathbb{Z}$ or any ring, when an equation

$$ax = b$$

is soluble, one says that a **divides** b, or a is a **divisor** or **factor** of b, but also b is a **multiple** of a. In this case we use the abbreviation

$$a \mid b.$$

I propose to say also that a **measures** b or is a **submultiple** of b. We may refer to the relation $\mid$ itself as **dividing** or **divisibility** or **measuring.** We shall use the following all the time, especially (2.4).

3 Theorem. *In any ring,*

$$\begin{gathered} a \mid 0, \\ 0 \mid a \implies a = 0, \\ a \mid a, \\ a \mid b \ \& \ b \mid c \implies a \mid c, \\ a \mid b \ \& \ a \mid c \implies a \mid bx + cy. \end{gathered} \tag{2.4}$$

The ring $\mathbb{Z}$ is special among rings for having the ordering that makes the following possible.

4 Theorem. *In* $\mathbb{N}$*,*

$$a \mid b \implies a \leqslant b.$$

In particular, measuring (|) *is an ordering of* $\mathbb{N}$ *(though not linear).*

The number of factors of an integer is finite. In particular, of two integers a and b, not both 0, there is a common factor that is *greatest* with respect to the usual linear ordering $<$. This **greatest common divisor** (or factor, or submultiple) is denoted by

$$\gcd(a, b).$$

We shall show that this is greatest with respect to divisibility (|) as well.

The following is also called the Division *Algorithm,* but it is rather a lemma that makes possible the *Euclidean Algorithm* described afterwards. Given two lengths a and b, the former being less, we can subtract a from b until what is left is less than a, again as in Figure 2.2. This is *measuring* b by a, and perhaps it is the performance of an algorithm; the following theorem proves that the algorithm works.

5 Division Theorem. *For all integers a and b, where $a \neq 0$, either $a \mid b$, or else the system*

$$b = ax + y \ \ \& \ \ 0 < y < |a|$$

has a unique solution.

Proof. Assuming $a \nmid b$, let A be the subset

$$\{b - ax \colon x \in \mathbb{Z}\} \cap \mathbb{N}$$

of $\mathbb{N}$. Then A is nonempty (why?); so it has a least element, r, and then for some q,

$$b = aq + r. \tag{2.5}$$

If $r \geqslant |a|$, so that $0 \leqslant r - |a| < r$, then, since also

$$b = a(q \pm 1) + (r - |a|),$$

we have that $r - |a| \in A$, but is less than the least element, which is absurd. Therefore $0 \leqslant r < |a|$. Uniqueness is an exercise. □

In (2.5), r is the **remainder** of b after measurement (or division) by a, and q is the **quotient** of b by a.

Euclid gives the following as a *porism* (a corollary of the proof) of the second proposition of Book VII of the *Elements.* It is a consequence of the possibility of applying the **Euclidean Algorithm,** whereby we can compute the greatest common measure of 942 and 180 as follows.

$$\left.\begin{aligned} 942 &= 180 \cdot 5 + 42 \\ 180 &= 42 \cdot 4 + 12 \\ 42 &= 12 \cdot 3 + 6 \\ 12 &= 6 \cdot 2, \end{aligned}\right\} \tag{2.6}$$

and therefore $\gcd(942, 180) = 6$, since

- 6, measuring 12, measures in turn 42, 180, and 942;
- any common measure of 942 and 180 measures in turn 42, 12, and 6.

6 Theorem. *The greatest common measure of two counting-numbers is measured by their every common measure.*

Proof. Given two unequal counting-numbers, we continually replace the larger by its remainder after measurement by the smaller, if this remainder exists. Thus if the original numbers are a_1 and a_2, the former being greater, if possible we let

$$a_1 = a_2x_1 + a_3 \;\;\&\;\; a_2 > a_3,$$

and in general, as far as possible,

$$a_k = a_{k+1}x_k + a_{k+2} \;\;\&\;\; a_{k+1} > a_{k+2}.$$

Every common measure of a_1 and a_2 is a common measure of a_3, a_4, and so on, by (2.4). Since the counting-numbers are well ordered, the strictly decreasing sequence of a_k must terminate in some a_n, so that

$$\begin{aligned} a_{n-2} &= a_{n-1}x_{n-2} + a_n, \\ a_{n-1} &= a_nx_{n-1}. \end{aligned}$$

Then a_n measures a_{n-1}, and a_{n-2}, and so on up to a_1. Thus a_n is a common measure of a_2 and a_1, and, since it is measured by every common measure, it must have been the greatest of these. □

Evidently we can apply the Euclidean Algorithm to arbitary integers, as long as one of them is not 0. As we have proved it, Theorem 6 has the next result as a porism. Thus from the computations (2.6), working backwards we have

$$\begin{aligned}6 &= 42 - 12 \cdot 3 = 42 - (180 - 42 \cdot 4) \cdot 3\\ &= 42 \cdot 13 - 180 \cdot 3 = (942 - 180 \cdot 5) \cdot 13 - 180 \cdot 3\\ &= 942 \cdot 13 - 180 \cdot 68.\end{aligned}$$

7 Bézout's Lemma. *In $\mathbb{Z}$, when a and b are not both 0, the equation*

$$\gcd(a, b) = ax + by.$$

is soluble.

We can also prove Theorem 6 as a corollary of Bézout's Lemma, since there is a non-constructive proof of Bézout's Lemma, not relying on our first proof of Theorem 6.

Abstract proof of Bézout's Lemma. The set

$$\{ax + by\colon (x, y) \in \mathbb{Z} \times \mathbb{Z}\} \tag{2.7}$$

of integers has positive elements, such as $|a| + |b|$; therefore the set has a least positive element. Let this be $as + bt$ or d. If now r is the remainder of a after measurement by d, and q the quotient, then

$$r = a - dq = a \cdot (1 - sq) - btq.$$

Then r belongs to the set in (2.7), so $r = 0$ by minimality of d. Thus $d \mid a$, and in the same way, $d \mid b$. Then d must be $\gcd(a, b)$. □

There is a third way to prove Theorem 6. It is the method that Landau uses in his *Elementary Number Theory* [8, p. 14], and it will be useful to us. Since the product ab of any two counting-numbers a and b is a common multiple of them, a and b have a **least common multiple,** denoted by

$$\operatorname{lcm}(a,b).$$

8 Theorem. *The least common multiple of any two counting-numbers measures every common multiple.*

Proof. The remainder of a common multiple of a and b after measurement by $\operatorname{lcm}(a,b)$ is itself a common multiple of a and b, so it must be 0. □

9 Corollary. *For any counting-numbers a and b, their every common measure measures $ab/\operatorname{lcm}(a,b)$, which itself is $\operatorname{gcd}(a,b)$.*

Proof. Note first that $ab/\operatorname{lcm}(a,b)$ is a common measure of a and b, since for example $\operatorname{lcm}(a,b)/b$ is an integer and

$$\frac{ab}{\operatorname{lcm}(a,b)}\cdot\frac{\operatorname{lcm}(a,b)}{b}=a.$$

For any common measure c of a and b, ab/c is a common multiple of a and b, so

$$\operatorname{lcm}(a,b)\mid\frac{ab}{c},\qquad\qquad c\mid\frac{ab}{\operatorname{lcm}(a,b)}.$$

Letting $c=\operatorname{gcd}(a,b)$, by Theorem 4 we conclude that the common measure $ab/\operatorname{lcm}(a,b)$ of a and b is greater than the greatest common measure, and so the two measures are equal. □

3 Analogy

3.1 Congruence

Gauss defined the terminology and notation of *modulus* and *congruence* at the beginning of the *Disquisitiones Arithmeticae* [6, 7]. In $\mathbb{Z}$, with respect to a given element n of $\mathbb{N}$ as a **modulus** (“little measure” in Latin), we define the relation of **congruence** by

$$a \equiv b \iff n \mid a - b.$$

If we need to make n explicit, instead of just $a \equiv b$ we write

$$a \equiv b \pmod{n}.$$

Gauss does not use the following notation; but, assuming we have verified that congruence *modulo* n is indeed an equivalence relation, we may denote the congruence class of any k by $[k]$, and then we may define

$$\mathbb{Z}_n = \{[x] \colon x \in \mathbb{Z}\},$$

the set of congruence classes *modulo* n.

We prove now that $\mathbb{Z}_n$ is a ring with precisely n distinct elements, namely $[0]$, …, $[n-1]$.

10 Theorem. *For all n in $\mathbb{N}$,*

$$\mathbb{Z}_n = \{[x] \colon 0 \leqslant x < n\},$$
$$|\mathbb{Z}_n| = n.$$

Proof. The first equation follows from the Division Theorem. For the second equation, if $0 \leqslant a < b < n$, then $0 < b - a < n$, so $n \nmid b - a$, and so $a \not\equiv b \pmod{n}$. □

11 Theorem. *For all n in $\mathbb{N}$, for all a, b, c, and d in $\mathbb{Z}$, if,* modulo n,

$$a \equiv c \ \ \& \ \ b \equiv d,$$

then

$$a + b \equiv c + d \ \ \& \ \ ab \equiv cd.$$

This means the following definitions are valid on $\mathbb{Z}_n$:

$$[x] + [y] = [x + y], \qquad\qquad [x] \cdot [y] = [xy].$$

Then the identies (2.1), (2.2), and (2.3) that make $\mathbb{Z}$ a ring must be true in $\mathbb{Z}_n$ as well, so this is a ring.

I said on page 8 that one can define addition and multiplication on $\mathbb{N}$ by means of the Recursion Theorem. However, the inductive property of $\mathbb{N}$ alone suffices for the definition, as Landau shows implicitly in *Foundations of Analysis* [9, I, §§2, 4]. If we accept Landau's proofs, then, since $\mathbb{Z}_n$ also has the inductive property (and we shall use it in proving Fermat's Theorem on page 38), we automatically obtain Theorem 11.

3.2 Prime numbers

In any ring, the elements with reciprocals are called **units.** The units of a ring compose an abelian group with respect to multiplication, and if the ring is R, the group of units can be denoted by

$$R^\times,$$

where the superscript × is a multiplication sign. The ring R is a field if and only if $R^\times = R \smallsetminus \{0\}$.

Two integers are called **prime to one another** (or *relatively prime,* or *co-prime*) if their greatest common divisor is 1.

By (2.4), *modulo* n,

$$x \equiv y \implies \gcd(x,n) = \gcd(y,n).$$

Thus the following makes sense.

12 Theorem. ${\mathbb{Z}_n}^\times = \{[x] \in \mathbb{Z}_n \colon \gcd(x,n) = 1\}$.

Proof. If $[a] \in {\mathbb{Z}_n}^\times$, so that, *modulo* n, the congruence

$$ax \equiv 1 \tag{3.1}$$

is soluble, then

$$\gcd(a,n) = 1 \tag{3.2}$$

since $\gcd(a,n)$,

- measuring a, measures ax;
- measuring n, measures $ax - 1$.

Conversely, if (3.2), then by Bézout's Lemma, the equation

$$ax + ny = 1$$

is soluble, so (3.1) is soluble. □

By definition, the **Euler phi-function** on $\mathbb{N}$ is such that, for every n greater than 1, $\varphi(n)$ is the number of counting numbers less than n that are prime to n. Symbolically,

$$\varphi(n) = |\{x \in \mathbb{N} \colon x \leqslant n \ \& \ \gcd(x,n) = 1\}|\,.$$

We can replace the requirement $x \leqslant n$ with $x < n$ unless $n = 1$. By Theorems 10 and 12,

$$\varphi(n) = \left|\mathbb{Z}_n{}^{\times}\right|. \tag{3.3}$$

We shall show how to calculate this number in Theorems 16 and 19.

In any ring R, two nonzero elements a and b are **associates** if the equation $ax = b$ is soluble by a unit. Being associated this way is an equivalence relation. A nonzero non-unit is called **irreducible** if its only factors are units and associates. Thus if a nonzero element π of a ring R is irreducible, this means

$$\begin{gathered} \pi \notin R^{\times}, \\ \pi = ab \;\; \& \;\; b \notin R^{\times} \implies a \in R^{\times}. \end{gathered}$$

Euclid refers to each positive irreducible of $\mathbb{Z}$ as *πρῶτος* “first,” or in Anglicized Latin **prime,** because the irreducibles in $\mathbb{N}$ are first in measuring (|). Here it is worthwhile to recall from page 10 that 1 is not an *ἀριθμός*. By the well-ordering of $\mathbb{N}$, along with its relation to measuring given by Theorem 4, every element has a **prime factorization,** meaning it can be written as a product of primes. Even 1 is the product of the empty set of primes.

13 Theorem. *The ring $\mathbb{Z}_n$ is a field if and only if n is prime.*

Proof. There are three possibilities for n.

1. Having but a single element, $\mathbb{Z}_1$ is not a field, since $1 \neq 0$ in every field.

2. If $a > 1$ and $b > 1$, then, *modulo* ab, $a \not\equiv 0$, but $ab \equiv 0$, so a has no inverse in $\mathbb{Z}_{ab}$, and this cannot be a field.

3. For all primes p, by Theorem 12,

$$\mathbb{Z}_p{}^{\times} = \{[x] \in \mathbb{Z}_p \colon 0 < x < p\} = \mathbb{Z}_p \smallsetminus \{0\}. \qquad \square$$

To emphasize that it is a field, we shall write $\mathbb{Z}_p$ as

$$\mathbb{F}_p.$$

The following is Proposition 30 of Book VII of the *Elements.*

14 Euclid's Lemma. *In $\mathbb{Z}$, for all primes p,*

$$p \mid ab \;\;\&\;\; p \nmid a \implies p \mid b. \tag{3.4}$$

Proof from Theorem 13. We can rewrite (3.4) in terms of congruence *modulo* p:

$$ab \equiv 0 \;\;\&\;\; a \not\equiv 0 \implies b \equiv 0. \tag{3.5}$$

This is true since $\mathbb{F}_p$ is a field. □

Proof from Bézout's Lemma. By hypothesis, $\gcd(p, a) = 1$, so by Bézout's Lemma we can solve the equations

$$\begin{gathered} px + ay = 1, \\ pbx + aby = b. \end{gathered}$$

Since p divides the left member of the latter equation, it must divide b. □

Before continuing, let us note some applications of Euclid's Lemma.

15 Fundamental Theorem of Arithmetic. *Prime factorizations of counting-numbers are unique.*

Proof. Suppose $p_1 \cdots p_m = q_1 \cdots q_n$, where each p_i or q_j is prime, and

$$p_1 \leqslant \ldots \leqslant p_m, \qquad\qquad q_1 \leqslant \ldots \leqslant q_n.$$

Then $p_1 \mid q_1 \cdots q_n$, so by Euclid's Lemma, p_1 divides one of the q_j and is therefore equal to it. Likewise, q_1 is equal to some p_i. Then

$$p_1 = q_j \geqslant q_1 = p_i \geqslant p_1,$$

so $p_1 = q_1$. Now $p_2 = q_2$, and so forth, and $m = n$. □

In some rings, irreducible factorizations exist, but they are not unique. Thus for example when we define

$$\mathbb{Z}[\sqrt{-5}] = \{x + y\sqrt{-5} \colon (x, y) \in \mathbb{Z} \times \mathbb{Z}\} \tag{3.6}$$

this is a sub-ring of $\mathbb{C}$ in which

$$2 \cdot 3 = 6 = (1 + \sqrt{-5})(1 - \sqrt{-5}).$$

The two different factorizations of 6 are irreducible, as one can prove by means of the multiplicative function $z \mapsto |z|^2$ from $\mathbb{Z}[\sqrt{-5}]$ to $\mathbb{N} \cup \{0\}$. However, none of the irreducible factors 2, 3, and $1 \pm \sqrt{-5}$ is prime. Apparently the term "ring" comes from examples like $\mathbb{Z}[\sqrt{-5}]$, where, in squaring $\sqrt{-5}$, we *circle back* to $\mathbb{Z}$.

16 Theorem. *If p is prime and $n \in \mathbb{N}$, then*

$$\varphi(p^n) = p^n - p^{n-1}.$$

Proof. We use that

$$\varphi(p^n) = p^n - |\{[x] \in \mathbb{Z}_{p^n} \colon \gcd(p^n, x) > 1\}|.$$

By Euclid's Lemma, the only prime factor of p^n is p, so

$$\gcd(p^n, x) > 1 \iff p \mid x.$$

This yields the claim. □

By Theorem 8, numbers are congruent *modulo* k and m if and only if they are congruent *modulo* $\operatorname{lcm}(k, m)$. This gives uniqueness in the following.

17 Chinese Remainder Theorem. *If* $\gcd(k, m) = 1$*, then every system*

$$z \equiv a \pmod{k}, \qquad\qquad z \equiv b \pmod{m}$$

has a solution that is unique modulo km*.*

Proof. The desired solution is

$$z \equiv amy + bkx \pmod{km},$$

where $kx + my = 1$. □

18 Corollary. *If* $\gcd(k, m) = 1$*, then the ring homomorphism*

$$[x] \mapsto ([x], [x]) \tag{3·7}$$

from $\mathbb{Z}_{km}$ *to* $\mathbb{Z}_k \times \mathbb{Z}_m$ *is an isomorphism.*

19 Theorem. *If* $\gcd(k, m) = 1$*, then*

$$\varphi(km) = \varphi(k) \cdot \varphi(m).$$

Proof. Restricted to ${\mathbb{Z}_k}^\times$, The isomorphism in (3·7) is surjective onto ${\mathbb{Z}_k}^\times \times {\mathbb{Z}_m}^\times$ by Euclid's Lemma: if $\gcd(k, x)$ and $\gcd(m, x)$ are both 1, then $\gcd(km, x)$ must be 1. □

The property (3.5) of $\mathbb{F}_p$ that is analogous with (3.4) can be written as

$$xy = 0 \ \& \ x \neq 0 \implies y = 0$$

for an arbitrary field. A *ring* where this holds along with $0 \neq 1$ is called an **integral domain.** All fields are integral domains, but $\mathbb{Z}$ is an integral domain that is not a field. The example $\mathbb{Z}_4$ is not even an integral domain.

20 Theorem. *Every finite integral domain is a field.*

Proof. Let R be a finite integral domain and $a \in R \smallsetminus \{0\}$. The function $x \mapsto ax$ having domain $R \smallsetminus \{0\}$ has range included in $R \smallsetminus \{0\}$, since if $b \neq 0$ in R, then $ab \neq 0$. Similarly, the function is injective:

$$ab = ac \implies a(b - c) = 0 \implies b - c = 0 \implies b = c.$$

By the Pigeonhole Principle, the function must be surjective. In particular, the equation $ax = 1$ must have a solution. □

By analogy with the rules for an integral domain, and especially considering Euclid's Lemma, we say a nonzero element π of an arbitrary ring R is **prime** if

$$\begin{gathered} \pi \notin R^{\times}, \\ \pi \mid xy \ \& \ \pi \nmid x \implies \pi \mid y. \end{gathered}$$

In these terms, Euclid's Lemma is that positive irreducibles of $\mathbb{Z}$ are prime and that, when p is such, then $\mathbb{Z}_p$ is an integral domain. There is a partial converse that holds more generally:

21 Theorem. *In an integral domain, primes are irreducible.*

Proof. In an integral domain, suppose a prime π is equal to ab. Then $\pi \mid ab$, so we may assume $\pi \mid a$, and so there is x such that $\pi x = a$. Then

$$\begin{gathered}\pi x b = ab = \pi,\\ xb = 1\end{gathered}$$

since we are in an integral domain. In particular, b is a unit. □

The proof fails for arbitrary rings. Thus in $\mathbb{Z}_6$, 2 is prime, since the multiples of 2 are 0, 2, and 4, and these are products as follows:

$$\begin{aligned}0&=2\cdot 3\\ &=3\cdot 4\\ &=0x,\end{aligned}\qquad \begin{aligned}2&=2\cdot 4\\ &=4\cdot 5,\end{aligned}\qquad \begin{aligned}4&=2\cdot 2\\ &=4\cdot 4\\ &=4\cdot 5,\end{aligned}$$

and 2 measures a factor in each case. However, since $2 = 2\cdot 4$, it is not irreducible.

3.3 Polynomials

It will now be useful to define

$$\omega = \{0,1,2,\dots\} = \{0\} \cup \mathbb{N} = \{x \in \mathbb{Z} \colon x \geqslant 0\},$$

the set of **natural numbers** or **non-negative integers.** (Here ω is the minuscule Greek omega.)

For any field K, we can form an infinite-dimensional vector-space over K having the formal basis $\{X^n \colon n \in \omega\}$. An arbitrary element f of this space can be written as

$$\sum_{n\in\omega} f_n X^n,$$

where $f_n = 0$ for all but finitely many n. The least n such that $f_n \neq 0$ is called the **degree** of f and can be denoted by

$$\deg(f).$$

Then $f_{\deg(f)}$ is the **leading coefficient** of f. For completeness, we may define

$$\deg(0) = -\infty.$$

We can define a multiplication on the whole space by first setting

$$X^k \cdot X^m = X^{k+m},$$

then extending to ensure commutativity, associativity, and distributivity. This gives

$$fg = \sum_{n \in \omega} \sum_{i+j=n} f_i g_j X^n.$$

This operation makes the space into the ring

$$K[X]$$

of **polynomials in X over K**. There is strong analogy between $K[X]$ and $\mathbb{Z}$. The ring $K[X]$ is not ordered, but the degrees of its elements are.

22 Division Theorem (for polynomials)**.** *For all fields K, for all f in $K[X]$ and g in $K[X] \smallsetminus \{0\}$, there are some q and r in $K[X]$ such that*

$$f = gq + r \;\;\&\;\; \deg(r) < \deg(g).$$

Moreover, $\deg(r)$ *is uniquely determined by f and g.*

Proof. Follow the pattern of the proof of Theorem 5, letting r be an element of $\{f - g \cdot \xi \colon \xi \in K[X]\}$ having minimal degree. If $\deg(r) \geqslant \deg(g)$, then the set also contains

$$r - \frac{r_{\deg(r)}}{g_{\deg(g)}} \cdot g,$$

whose degree is strictly less than $\deg(r)$. □

Of two polynomials, at least one of which is not 0, there is a common divisor of maximal degree. We can find such a common divisor by performing the Euclidean Algorithm; and then every common divisor will divide it, as follows.

23 Bézout's Lemma (for polynomials)**.** *For all fields K, for all f and g in $K[X]$, not both 0, for every element h of the set*

$$\{f \cdot \xi + g \cdot \eta \colon (\xi, \eta) \in K[X] \times K[X]\}$$

having minimal degree,
(i) $h \mid f$ and $h \mid g$;
(ii) if $k \mid f$ and $k \mid g$, then $k \mid h$.

The polynomial h in the theorem is a greatest common divisor of f and g and is unique up to multiplication by a nonzero scalar. As we could prove Euclid's Lemma from Bézout's Lemma in the original case of integers, so we can do it for polynomials:

24 Euclid's Lemma (for polynomials)**.** *For every field K, every irreducible of $K[X]$ is prime.*

3.4 Quotients

The theory of congruence carries over to an arbitrary ring R. If a is a nonzero element of this ring, two elements b and c can be called congruent *modulo* a if $a \mid b-c$. The set of congruence classes of elements of R is then denoted by

$$R/(a). \tag{3.8}$$

Thus $\mathbb{Z}_n$ becomes $\mathbb{Z}/(n)$.

It is not important for us, but the notation in (3.8) can be analyzed, and (a) is the set of elements of R that are congruent to 0; this set is just $\{ax \colon c \in R\}$. Thus, *modulo* a,

$$b \equiv c \iff b-c \in (a).$$

Then $R/(a)$ is the **quotient** of R by (a), which itself is called an **ideal** of R, because it is closed under addition and under multiplication by elements of R. More precisely, (a) is a **principal ideal,** because it consists of the multiples of a single element. In any ring, such as $\mathbb{Z}$ or $K[X]$ where K is a field, in which Bézout's Lemma is true, all ideals are principal; but in the ring $\mathbb{Z}[\sqrt{-5}]$ defined in (3.6) on page 25, when we define

$$\begin{aligned}&(1+\sqrt{-5}, 1-\sqrt{-5})\\ &\qquad= \{(1+\sqrt{-5})x + (1-\sqrt{-5})y \colon\\ &\qquad\qquad (x,y) \in \mathbb{Z}[\sqrt{-5}] \times \mathbb{Z}[\sqrt{-5}]\},\end{aligned}$$

this is a non-principal ideal, as again one can show by means of $z \mapsto |z|$.

Corresponding to Theorem 13, by the same proof, we have the following.

25 Theorem. *For all fields K, for all nonzero f in $K[X]$,*

$$K[X]/(f) \text{ is a field} \iff f \text{ is irreducible.}$$

We looked at the example of $\mathbb{C}$ on page 10; now we can write

$$\mathbb{C} = \mathbb{R}[X]/(X^2+1).$$

However, over $\mathbb{F}_2$,

$$X^2+1 = (X^2-1) = (X+1)(X-1) = (X+1)^2,$$

so it is not irreducible. Nonetheless, X^2+X+1 is irreducible over $\mathbb{F}_2$, since it can factorize only as $(X-a)(X-b)$, where a and b are zeros of the polynomial; and there are no such zeros in $\mathbb{F}_2$. If we denote the congruence class of X in $\mathbb{F}_2[X]$ *modulo* X^2+X+1 by α, then

$$\alpha^2+\alpha+1=0.$$

Thus $\alpha^2 = \alpha+1$. We can write

$$\mathbb{F}_2(\alpha) = \mathbb{F}_2[X]/(X^2+X+1).$$

Here X^2+X+1 is a **minimal polynomial** of α over $\mathbb{F}_2$, since α is a root of it, but not of any nonzero polynomial of less degree, since if α is also a root of f, then α is a root of $\gcd(X^2+X+1, f)$, so this is not 1, so it must be X^2+X+1, since this is irreducible.

As a vector space over $\mathbb{F}_2$, the field is $\mathbb{F}_2 \oplus \mathbb{F}_2\alpha$. Addition and multiplication in $\mathbb{F}_2(\alpha)$ are as in Figure 3.1. Note that $\mathbb{F}_2[\alpha]^\times \cong \mathbb{Z}_3$. Also

$$(X-\alpha)(X-\alpha-1) = X^2+X+1,$$

$+$	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

Figure 3.1: Arithmetic in $\mathbb{F}_2[\alpha]$ where $\alpha^2 = \alpha + 1$

and thus

$$\begin{aligned}\prod_{t\in\mathbb{F}_2[\alpha]}(X-t) &= X(X-1)(X^2+X+1)\\ &= X(X^3-1) = X^4 - X.\end{aligned}$$

Still over $\mathbb{F}_2$, there are four polynomials of degree 3 that do not have the factor X, namely

$$X^3+1,\quad X^3+X+1,\quad X^3+X^2+1,\quad X^3+X^2+X+1.$$

The first and last have factor $X+1$; the middle two do not, since 1 is not a zero of them, and therefore they are irreducible. We can now understand $\mathbb{F}_2/(X^3+X+1)$ as $\mathbb{F}_2(\beta)$, where

$$\beta^3 = \beta + 1.$$

As a vector space, the field is $\mathbb{F}_2 \oplus \mathbb{F}_2\beta \oplus \mathbb{F}_2[\beta^2]$, with eight elements. To understand multiplication in the field, we may observe that the powers of β are thus:

k	0	1	2	3	4	5	6	7
β^k	1	β	β^2	$\beta+1$	$\beta^2+\beta$	$\beta^2+\beta+1$	β^2+1	1

Hence an abbreviated multiplication table can be written out as in Figure 3.2.

For one more example, we note that X^2+1 is irreducible over $\mathbb{F}_3$, since it has no zero there. Then $\mathbb{F}_3[X]/(X^2+1) = \mathbb{F}_3(\gamma)$, where

$$\gamma^2 = -1,$$

and so $\gamma^4 = 1$. However, $\gamma + 1$ has the following powers.

k	0	1	2	3	4
$(\gamma+1)^k$	1	$\gamma+1$	$-\gamma$	$1-\gamma$	-1
$(\gamma+1)^{4+k}$	-1	$-(\gamma+1)$	γ	$\gamma-1$	1

Writing $\gamma+1$ as δ, we have $\gamma = \delta - 1$, so

$$\begin{gathered} 0 = (\delta-1)^2 + 1 = \delta^2 + \delta - 1, \\ \mathbb{F}_3(\gamma) = \mathbb{F}_3(\delta) = \mathbb{F}_3[X]/(X^2+X-1). \end{gathered}$$

In fact

$$\begin{aligned} X^8 - 1 &= (X-1)(X+1)(X^2+1)(X^4+1) \\ &\quad (X-1)(X+1)(X^2+1)(X^2+X-1)(X^2-X-1) \end{aligned}$$

over $\mathbb{F}_3$. Any of the three irreducible polynomials X^2+1 and $X^2 \pm X - 1$ will give us $\mathbb{F}_9$.

If f in $K[X]$ is irreducible of degree n, then, as a vector space over K, $K[X]/(f)$ is isomorphic to the space with basis

$$\{X^k \colon 0 \leqslant k < n\}.$$

In particular, in case K is $\mathbb{F}_p$, the space has finite size p^n; see Theorem 26. We shall show that such irreducibles f exist for all positive n, and moreover, finite fields arise in no other way.

$\times$	β	β^2	$\beta+1$	$\beta^2+\beta$	$\beta^2+\beta+1$	β^2+1
β	β^2	$\beta+1$	$\beta^2+\beta$	$\beta^2+\beta+1$	β^2+1	1
β^2	$\beta+1$	$\beta^2+\beta$	$\beta^2+\beta+1$	β^2+1	1	β
$\beta+1$	$\beta^2+\beta$	$\beta^2+\beta+1$	β^2+1	1	β	β^2
$\beta^2+\beta$	$\beta^2+\beta+1$	β^2+1	1	β	β^2	$\beta+1$
$\beta^2+\beta+1$	β^2+1	1	β	β^2	$\beta+1$	$\beta^2+\beta$
β^2+1	1	β	β^2	$\beta+1$	$\beta^2+\beta$	$\beta^2+\beta+1$

Figure 3.2: Multiplication in $\mathbb{F}_2[\beta]$ where $\beta^3 = \beta + 1$

4 Characterization

4.1 Characteristic

Every commutative ring R contains all of the sums

$$\underbrace{1+\cdots+1}_{n},$$

where $n \in \mathbb{N}$. If the sum is 0 for some n, then the least such n is called the **characteristic** of R, or

$$\operatorname{char}(R).$$

If R is an integral domain, then $\operatorname{char}(R)$ must be a prime p. In this case, we may suppose $\mathbb{F}_p \subseteq R$; and then R is a vector space over $\mathbb{F}_p$.

26 Theorem. *The size of every finite field is a prime power.*

Proof. If K is a finite field, then, by what we have just seen, K is a vector space over some $\mathbb{F}_p$. As such, K has a basis of some finite size n; and then, as at the end of the last section, $|K| = p^n$. □

We are going to prove Fermat's Theorem by induction, as mentioned on page 21. By recursive definition,

$$0! = 1, \qquad (n+1)! = n! \cdot (n+1).$$

If $0 \leqslant k \leqslant n$, then by non-recursive definition,

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}.$$

This is the number of k-element subsets of a set of size n. If one accepts this, then $\binom{n}{k}$ is automatically a counting-number. Alternatively, one may observe

$$\binom{n}{0} = 1, \qquad \binom{n}{n} = 1,$$

while if $0 \leqslant k < n$, then by computation

$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1},$$

so by induction, each $\binom{n}{k}$ is a counting-number. One may also prove by induction the **Binomial Theorem,**

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k,$$

and note that the coefficients must be counting-numbers.

27 Lemma. *When $0 < k < p$, then*

$$p \,\Bigg|\, \binom{p}{k}.$$

Proof. The "intuitive" proof is that, if A is a proper nonempty subset of $\mathbb{Z}_p$, and σ is the permutation $x \mapsto x+1$ of this set, then the sets $\sigma^n[A]$ all differ as n itself ranges over $\mathbb{Z}_p$; thus the number of subsets of $\mathbb{Z}_p$ having the size of A must be divisible by p.

The more formal proof is that, first of all, since $\binom{p}{k}$ is a whole number,

$$k! \cdot (p-k)! \mid (p-1)! \cdot p.$$

By Euclid's Lemma, since $0 < k < p$ and p is prime,

$$p \nmid k! \cdot (p-k)!.$$

By the Fundamental Theorem (page 24), we must therefore have

$$k! \cdot (p-k)! \mid (p-1)!. \qquad \square$$

28 Theorem. *In any field having characteristic p,*

$$(x+y)^p = x^p + y^p.$$

Proof. This is an immediate consequence of the Binomial Theorem and Lemma 27. □

29 Fermat's Theorem. *For all primes p and all a in $\mathbb{Z}$, if $p \nmid a$, then*

$$a^{p-1} \equiv 1 \pmod{p}.$$

Proof. The claim is that, in $\mathbb{F}_p$,

$$a \neq 0 \implies a^{p-1} = 1,$$

or equivalently, for all a in $\mathbb{F}_p$,

$$a^p = a. \tag{4.1}$$

We can prove this by induction. Immediately (4.1) holds when $a = 1$. Suppose (4.1) is true when $a = b$, so $b^p = b$. By Theorem 28,

$$(b+1)^p = b^p + 1 = b + 1$$

in $\mathbb{F}_p$, so (4.1) is true when $a = b + 1$. By induction, (4.1) is true for all a in $\mathbb{F}_p$. □

We can also derive Fermat's Theorem as a special case of Euler's Theorem below.

The **order** of a finite group is just its size. The **order** of an element a is the least non-negative exponent n such that $a^n = 1$ in the group; but this is just the order of the subgroup $\{a^n \colon n \in \mathbb{Z}\}$, which is denoted by

$$\langle a \rangle,$$

of the original group. So the order of a is $|\langle a \rangle|$; we just write this as

$$|a| .$$

30 Lagrange's Theorem. *In any finite group, the order of any element divides the order of the group.*

Proof. Let G be the group and a the element, and let $n = |a|$. We can construct a matrix

$$\begin{pmatrix} 1 & a & a^2 & \cdots & a^{n-1} \\ b_1 & b_1 a & b_1 a^2 & \cdots & b_1 a^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ b_{k-1} & b_{k-1} a & b_{k-1} a^2 & \cdots & b_{k-1} a^{n-1} \end{pmatrix}$$

with entries from G, line by line, left to right within each line, so that every element of G appears exactly once. In particular, $|G| = nk$. □

31 Euler's Theorem. *For all n in $\mathbb{N}$, for all a in $\mathbb{Z}$, if*

$$\gcd(a, n) = 1,$$

then

$$a^{\varphi(n)} \equiv 1 \pmod{n}.$$

Proof. By Theorem 12 and (3.3), the claim is a consequence of Lagrange's Theorem. □

4.2 Units of finite fields

If K is a field included in a ring R, and $a \in R$, then there is a homomorphism

$$f \mapsto f(a)$$

from $K[X]$ to R that is uniquely determined by the requirement that $f(a)$ be a when f is X. If $f(a) = 0$, then a is called a **root** or **zero** of f. By Fermat's Theorem, every element of $\mathbb{F}_p$ is a root of the polynomial

$$X^p - X.$$

Using Lagrange's Theorem, we generalized Fermat's Theorem to Euler's Theorem. A similar generalization will be that, if K is a finite field of size q, then every element of K is a root of the polynomial

$$X^q - X.$$

We shall use this to construct and characterize finite fields.

32 Theorem. *For any field K, for any f in $K[X] \smallsetminus \{0\}$,*

$$|\{x \in K \colon f(x) = 0\}| \leqslant \deg(f).$$

Proof. Suppose $f(\alpha) = 0$. By the Division Theorem for polynomials (Theorem 22), for some g in $K[X]$,

$$f = (X - \alpha) \cdot g. \tag{4.2}$$

If $\beta \neq \alpha$ and $f(\beta) = 0$, then $g(\beta) = 0$. Thus f can have at most one more zero than g has. If $\deg(g) = 0$, then g has no zeros. By induction, the proof is complete. □

33 Lemma. *Suppose an abelian group (written additively) has elements g and h of orders a and b respectively.*

1. *If* $\gcd(a,b)=1$, *then* $g+h$ *has order* ab.
2. *In general,* $\gcd(a,b)\cdot g+h$ *has order* $\operatorname{lcm}(a,b)$.

Proof. We have $ab(g+h)=bag+abh=0+0=0$. Conversely, if $n(g+h)=0$, then ng, which is $-nh$, belongs to both $\langle g\rangle$ and $\langle h\rangle$. If $\gcd(a,b)=1$, then by Lagrange's Theorem, $\langle a\rangle$ and $\langle b\rangle$ have only the trivial subgroup in common, so n must be a common multiple of a and b, and here this just means $ab\mid n$. This yields the first claim. For the more general claim, we note that $\gcd(a,b)g$ has order $a/\gcd(a,b)$, which is prime to b, and the product of the two orders is $\operatorname{lcm}(a,b)$. □

34 **Theorem.** *For every finite field K, for some nonzero element α of K,*

$$K^\times=\langle\alpha\rangle.$$

Proof. Suppose $|K|=q$, and let α be an element of $K^\times$ of maximal order. If $\beta\in K^\times$, then by Lemma 33 the group has an element of order $\operatorname{lcm}(|\alpha|,|\beta|)$. This must be $|\alpha|$ by its own maximality, and so $|\beta|$ must divide $|\alpha|$. If this order is m, then we now have

$$K^\times=\{x\in K\colon x^m=1\}.$$

By Theorem 32, the size $q-1$ of this set is at most m. Also $m\mid q-1$, so $m=q-1$, and this yields the claim. □

4.3 Splitting fields

We noted on page 34 that if f in $K[X]$ is irreducible of degree n, then the field $K[X]/(f)$ has dimension n as a vector space over K. Moreover, the congruence class of X in the field must be a root of f. If K is included in some field L that contains a root α of f, we denote by

$$K(\alpha)$$

the smallest subfield of L that includes K and contains a. Then $K(a)$ and $K[X]/(f)$ are isomorphic as vector spaces and indeed as fields, and there is an isomorphism from $K(\alpha)$ to $K[X]/(f)$ that is the identity on K and that takes α to the congruence class of X.

If now K and α are as in Theorem 34, and $\operatorname{char}(K) = p$, while $q = p^n$, then

$$K = \mathbb{F}_p(\alpha). \tag{4.3}$$

Moreover, α must be a root of an irreducible factor of $X^{q-1} - 1$ that has degree n.

We shall show that there is always such a factor, for every p and n.

For an arbitrary field K, for any n in $\mathbb{N}$, for any polynomial f of degree n over K, a **splitting field** of f over K is a field extending K containing elements α_k such that

$$f = \prod_{k<n} (X - \alpha_k).$$

35 Theorem. *Splitting fields always exist.*

Proof. In the notation just used, we have seen that f has a root α in some field $K(\alpha)$. Over this field, as in the proof of Theorem 32, there is a polynomial g such that (4.2) holds. Here g may not be irreducible, but it has an irreducible factor, which has a root β in some field $K(\alpha, \beta)$. Then $f = (X - \alpha) \cdot (X - \beta) \cdot h$ for some h, and so forth. □

If, in its complete factorization in a splitting field, a polynomial has no repeated factor, the polynomial is called **separable.** To test for separability, we can take formal derivatives. The derivative f' of a polynomial f is just what one expects from the rules of calculus. By formal definition, the map $\xi \mapsto \xi'$ from $K[X]$ to itself is the unique linear transformation that takes X^n to nX^{n-1} when $n \in \mathbb{N}$, but takes 1 to 0. The multiplication rule

$$(fg)' = f'g + fg'$$

holds when f and g are powers of X and therefore, by linearity, for all f and g.

36 Theorem. *A polynomial is separable if and only if it is prime to its derivative.*

Proof. We can write a separable polynomial as $(X - \alpha)^2 \cdot g$, and $X - \alpha$ is a common factor of this and its derivative.

However, if α is not a root of f, then it is not a root of the derivative of $(X - \alpha) \cdot f$, which is $f + (X - \alpha) \cdot f'$. Thus a separable polynomial can share no roots with its derivative. □

Over a field of characteristic p, a polynomial $X^p - a$ is not separable, since its derivative is 0. If b is a root of the polynomial, then

$$(X - b)^p = X^p - b^p = X^p - a$$

by Theorem 28, so the polynomial has a unique root. Thus the function $x \mapsto x^p$ from the field to itself is injective. If the field is finite, then the function must also be surjective, by the Pigeonhole Principle as on page 27. In this case, $X^p - a$ is never irreducible.

37 Theorem. *For each prime p and counting-number n, there is a field having size p^n.*

Proof. Let $q = p^n$. By Theorem 36, the polynomial $X^q - X$ is separable, since its derivative, -1, is a unit. Let L be a splitting field of the polynomial, and let

$$K = \{x \in L \colon x^q = x\}.$$

Then $|K| = q$. Moreover, K is a field by Theorem 28, since if K contains α and β, then

$$\begin{gathered}(\alpha \pm \beta)^q = \alpha^q \pm \beta^q = \alpha \pm \beta,\\ (\alpha\beta)^q = \alpha^q\beta^q = \alpha\beta,\end{gathered}$$

and also, if $\alpha \neq 0$,

$$\alpha^{q-2}\alpha = 1,$$

so K is closed under addition, subtraction, multiplication, and inversion. □

4.4 Isomorphism

By what we saw at the beginning of the last section, if α and β are both roots of the same irreducible polynomial f over K, then the map from $K(\alpha)$ to $K(\beta)$ that is identical on K and takes α to β must be an isomorphism.

We also saw that any finite field can be written as in (4.3), as $\mathbb{F}_p(\alpha)$. The field having size q, let f be the irreducible factor of $X^q - X$ of which α is a root. Any other field of size q contains a root β of f, and then the field itself must be $\mathbb{F}_p(\beta)$, and this is isomorphic to $\mathbb{F}_p(\alpha)$.

Thus we have:

38 Theorem. *For every prime power, all fields of that size are isomorphic.*

"The" field of size q can be called

$$\mathbb{F}_q.$$

39 Theorem. *The finite fields of characteristic* p *are ordered by inclusion as the exponents of their sizes are ordered by divisibility. Thus for all* m *and* t *in* $\mathbb{N}$,

$$\mathbb{F}_{p^m} \subseteq \mathbb{F}_{p^t} \iff m \mid t.$$

Proof. By the same proof as when $K = \mathbb{F}_p$, if K and L are two finite fields such that $K \subseteq L$, then for some n in $\mathbb{N}$,

$$|L| = |K|^n .$$

In particular, if $|K| = p^m$, then $|L| = p^t$ for some t, where $m \mid t$.

Suppose conversely $m \mid t$. Then $t = mk$ for some k, and so

$$p^t - 1 = (p^m)^k - 1 = (p^m - 1) \cdot n,$$

where

$$n = (p^m)^{k-1} + \cdots + 1.$$

Now

$$\begin{aligned} X^{p^t} - X &= X \cdot (X^{p^t-1} - 1) \\ &= X \cdot \bigl((X^{p^m-1})^n - 1\bigr) \\ &= X \cdot (X^{p^m-1} - 1) \cdot \bigl((X^{p^m-1})^{n-1} + \cdots + 1\bigr) \\ &= X \cdot (X^{p^m-1} - 1) \cdot \bigl((X^{p^m-1})^{n-1} + \cdots + 1\bigr) \\ &= (X^{p^m} - X) \cdot \bigl((X^{p^m-1})^{n-1} + \cdots + 1\bigr), \end{aligned}$$

so by the proof of Theorem 37, $\mathbb{F}_{p^m} \subseteq \mathbb{F}_{p^t}$. □

5 Galois Theory

On a field of characteristic p the function $x \mapsto x^p$ discussed on page 43 is an **endomorphism** (an injective homomorphism from the field into itself) by Theorem 28. If the field is finite, as discussed before, the function is therefore an **automorphism** (an isomorphism of the field with itself, or a surjective endomorphism), called the **Frobenius Automorphism.** Let us denote it by $\boxed{\sigma.}$ By Fermat's Theorem, σ is the identity on $\mathbb{F}_p$. Indeed, this field can have no nontrivial automorphism, since the elements of the field are just finite sums of the element 1. In any case, σ belongs to the group

$$\operatorname{Aut}(\mathbb{F}_{p^n})$$

of automorphisms of $\mathbb{F}_{p^n}$.

40 Theorem. *For all primes p, for all n in $\mathbb{N}$,*

$$\operatorname{Aut}(\mathbb{F}_{p^n}) = \{\sigma^i \colon 0 \leqslant i < n\} \cong \mathbb{Z}_n.$$

Proof. We know

$$\mathbb{F}_{p^n}{}^{\times} = \langle\alpha\rangle \tag{5.1}$$

for some α in $\mathbb{F}_{p^n}{}^{\times}$. In this case we have

$$\mathbb{F}_{p^n} = \mathbb{F}_p(\alpha).$$

The minimal polynomial f of α over $\mathbb{F}_{p^n}$ is

$$\sum_{k=0}^{n} b_k X^k$$

for some b_k in $\mathbb{F}_{p^n}$. Since $f(\alpha) = 0$, we have also

$$0 = f(\alpha)^p = \sum_{k=0}^{n} {b_k}^p (X^k)^p = \sum_{k=0}^{n} b_k (X^p)^k = f(\alpha^p)$$

by Fermat's Theorem. In the same way, all elements of the set $\{\alpha^{p^i} : 0 \leqslant i < n\}$ are roots of f. Then the set has n distinct elements; for if $0 \leqslant i \leqslant j < n$ and

$$\alpha^{p^i} = \alpha^{p^j},$$

then

$$\alpha^{p^{n+i-j}} = \alpha^{p^n} = \alpha,$$

and so α belongs to $\mathbb{F}_{p^{n+i-j}}$, which it does not, by (5.1), unless $i = j$. If $\tau \in \mathrm{Aut}(\mathbb{F}_p)$, then as with the Frobenius automorphism, writing x^τ for $\tau(x)$ we have

$$0 = f(\alpha)^\tau = f(\alpha^\tau),$$

so $\alpha^\tau = \alpha^{p^i} = \alpha^{\sigma^i}$ for some i. Since every nonzero element of $\mathbb{F}_{p^n}$ is a power of α, we can conclude $\tau = \sigma^i$. □

We define

$$\mathbb{F}_p{}^{\mathrm{alg}} = \bigcup_{n \in \mathbb{N}} \mathbb{F}_{p^n},$$

where $\mathbb{F}_{p^m} \subseteq \mathbb{F}_{p^t}$ when $m \mid t$ as in Theorem 39. Then $\mathbb{F}_p{}^{\mathrm{alg}}$ is a field, the **algebraic closure** of F, since it is the smallest field including $\mathbb{F}_p$ in which every polynomial splits. We have now a map

$$\zeta \mapsto (z_n \colon n \in \mathbb{N})$$

from $\operatorname{Aut}({\mathbb{F}_p}^{\mathrm{alg}})$ to the product

$$\prod_{n\in\mathbb{N}} \mathbb{Z}_n$$

of groups given by

$$\zeta \restriction \mathbb{F}_{p^n} = \sigma^{z_n}.$$

Thus the z_n meet the condition

$$m \mid t \implies z_m \equiv z_t \pmod{m}.$$

We can build up such z_n step by step, by the Chinese Remainder Theorem. Indeed, suppose

$$z_k \equiv z_m \pmod{d}, \tag{5.2}$$

where

$$d = \gcd(k, m).$$

We can let

$$z \equiv z_k \cdot \frac{m}{d} \cdot y + z_m \cdot \frac{k}{d} \cdot x \pmod{\operatorname{lcm}(k, m)}, \tag{5.3}$$

where

$$kx + my = d. \tag{5.4}$$

Then

$$z \equiv z_k \mod \frac{k}{d}, \qquad z \equiv z_m \mod \frac{m}{d}. \tag{5.5}$$

Moreover, since $d \mid \operatorname{lcm}(k, m)$, by (5.3) and (5.4) we have

$$z \equiv z_k \cdot \frac{m}{d} \cdot y + z_k \cdot \frac{k}{d} \cdot x \equiv z_k \pmod{d}, \tag{5.6}$$

and in the same way $z \equiv z_m \pmod{d}$. This and (5.5) give

$$z \equiv z_k \pmod{k}, \qquad z \equiv z_m \pmod{m}.$$

Thus we can let

$$z_{\operatorname{lcm}(k,m)} = z.$$

Now we know that $\operatorname{Aut}({\mathbb{F}_p}^{\mathrm{alg}})$ is isomorphic to the product

$$\prod_{\ell \text{ prime}} \mathbb{Z}_{(\ell)},$$

where $\mathbb{Z}_{(\ell)}$ is the subgroup of

$$\prod_{n \in \mathbb{N}} \mathbb{Z}_{\ell^n}$$

consisting of

$$(a_n \colon n \in \mathbb{N})$$

such that

$$i \leqslant j \implies a_i \equiv a_j \pmod{\ell^i}.$$

Under this condition, we can recursively define

$$b_0 = a_1, \qquad b_n = \frac{a_{n+1} - a_n}{\ell^n}.$$

In this case

$$a_n = \sum_{k=0}^{n-1} b_k \ell^k.$$

We can now write the group element as

$$\sum_{k=0}^{\infty} b_k \ell^k,$$

because this reflects the group structure:

$$\sum_{k=0}^{\infty} b_k \ell^k + \sum_{k=0}^{\infty} c_k \ell^k = \sum_{k=0}^{\infty} d_k \ell^k$$

if and only if, for all n in $\mathbb{N}$,

$$\sum_{k=0}^{n-1} b_k \ell^k + \sum_{k=0}^{n-1} c_k \ell^k \equiv \sum_{k=0}^{n-1} d_k \ell^k \pmod{\ell^n}.$$

Bibliography

[1] René Descartes. *The Geometry of René Descartes.* Dover Publications, New York, 1954. Translated from the French and Latin by David Eugene Smith and Marcia L. Latham, with a facsimile of the first edition of 1637.

[2] René Descartes. *La Géométrie.* Jacques Gabay, Sceaux, France, 1991. Reprint of Hermann edition of 1886.

[3] Euclid. *The Thirteen Books of Euclid's* Elements. Dover Publications, New York, 1956. Translated from the text of Heiberg with introduction and commentary by Thomas L. Heath. In three volumes. Republication of the second edition of 1925. First edition 1908.

[4] Pierre de Fermat. *Oeuvres de Fermat. Tome Deuxième. Correspondance.* Gautiers-Villars et Fils, Paris, 1894. Edited by Paul Tannery and Charles Henry. `archive.org/details/oeuvresdefermat02ferm`, accessed November 23, 2017.

[5] Pierre de Fermat. Letter to Bernard Frénicle de Bessy, October 18, 1640. In D. J. Struik, editor, *A Source Book in Mathematics 1200–1800*, Princeton Paperbacks, pages 27–29. Princeton University Press, Princeton, NJ, 1986. Reprint of the 1969 edition.

[6] Carl Friedrich Gauss. *Disquisitiones Arithmeticae.* Carl Friedrich Gauss Werke. Gerh. Fleischer Jun., Leipzig, 1801. Electronic version of the original Latin text from Goettingen State and University Library.

[7] Carl Friedrich Gauss. *Disquisitiones Arithmeticae.* Springer-Verlag, New York, 1986. Translated into English by Arthur A. Clarke, revised by William C. Waterhouse.

[8] Edmund Landau. *Elementary Number Theory.* Chelsea Publishing, New York, 1958. Originally part of *Vorlesungen über Zahlentheorie* (Leipzig, 1927). Translated by J. E. Goodman.

[9] Edmund Landau. *Foundations of Analysis. The Arithmetic of Whole, Rational, Irrational and Complex Numbers.* Chelsea Publishing, New York, third edition, 1966. Translated by F. Steinhardt; first edition 1951; first German publication, 1929.

[10] David Pierce. Thales and the nine-point conic. *The De Morgan Gazette*, 8(4):27–78, 2016. `http://education.lms.ac.uk/2016/12/thales-and-the-nine-point-conic/`, accessed June 1, 2017.