# Finite fields

**A course at the Nesin Mathematics Village**

David Pierce

January 18–24, 2016
Last edited, February 25, 2016

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

# Preface

The present typeset document is based on a course of lectures at the Nesin Mathematics Village in Şirince, Selçuk, İzmir, January 18–24 (Monday–Sunday), 2016.

My course was one of three on the general theme of linear algebra. The students were mostly advanced undergraduates, though not necessarily from mathematics departments. The daily schedule was as follows.

| | |
|---|---|
| 9:00 | Ali Nesin |
| 11:00 | Ali Nesin |
| 13:00 | lunch, chores, rest |
| 15:00 | Haluk Oral |
| 17:00 | David Pierce |
| 19:00 | dinner |

Thursdays at the Village are free in the summer, but not in the winter. However, on the Friday of the program, Haluk took my hours, so that he could leave the next morning; I took his hours that afternoon. Most students were gone by Sunday afternoon, but I gave a lecture then to the few who remained and wanted to come.

Haluk's course treated basic linear algebra (subspaces, linear independence, etc.); Ali's, modules. Ali suggested on Monday morning that my course

> treat one or several of the following:
>
> 1. Finite fields.

2. Nondegenerate bilinear forms and some classification problems.
3. Tensor product of modules.
4. Jordan normal form.

I chose finite fields, with Jordan normal forms on the last day, since one student asked to hear about eigenvectors.

I started typesetting this document on Thursday morning, after three lectures had been completed. The sources are (1) my handwritten notes, prepared before the lectures, (2) my memory of what happened in the lectures, and sometimes (3) my wishes for enlargements or improvements. Thus the notes are not a precise record of what actually happened, though they should be close. The appendix completes the work on Jordan normal forms started in the last class.

I started the first lecture in Turkish, though writing on the board in English. During the break after fifty minutes, students invited me to speak in English. Since nobody in the class as a whole expressed a preference for Turkish when I asked, I did switch to English, mostly. Sometimes students asked questions, or asked for clarifications, in Turkish.

I originally learned much of the content of the course in a graduate algebra course by Larry Washington that used Hungerford [5] as a reference. I learned elementary number theory through teaching it, in a course based on Burton [1]. In Şirince, for the linear algebra, I had the books of Cemal Koç [7, 8] at hand; for finite fields, I had a look at Lidl and Niederreiter [10].

# Contents

# List of Figures

# Introduction

The main aim of the course is to establish the classification and description of finite fields as follows.

1. The size of every finite field is a power of a prime.
2. For each prime power $p^n$, there is, up to isomorphism, exactly one field, called $\mathbb{F}_{p^n}$, whose size is $p^n$.
3. $\mathbb{F}_{p^m} \subseteq \mathbb{F}_{p^n}$ if and only if $m \mid n$.
4. The group of units of every finite field is cyclic.

There is a similar list on page 24. Our work will require the classication of finite abelian groups (page 42), which uses linear algebra. The full statement of the classification uses the Chinese Remainder Theorem (page 30).

Our work will also involve polynomial rings (in one variable) over fields. Such rings are analogous to the ring $\mathbb{Z}$ of rational integers. An incidental aim then is to analyze three theorems about integers:

1. Euclid's Lemma (pages 14 and 38), namely Proposition 30 of Book VII of Euclid's *Elements*;
2. Bézout's Lemma (pages 30 and 39); and
3. the efficacy of the Euclidean Algorithm (page 29) for finding greatest common divisors: Propositions 1 and 2 of Book VII of the *Elements.*

Among commutative rings that have no infinite descending chains of divisors, that is, no sequences $(a_0, a_1, a_2, \dots)$ such that $a_{n+1} \mid a_n$, but $a_n \nmid a_{n+1}$, the three theorems establish sufficient and necessary conditions for being, respectively,

(1) a unique factorization domain (UFD),

(2) a principal ideal domain (PID), and
(3) a Euclidean domain (ED).

# 1. Monday, January 18

What is a *vector space*? The answer could be

- a definition: "A vector space is $X$";
- an example (or family of examples): "$X$ is a vector space."

## Definitions

By definition, a **vector space** is an *abelian group,* together with a homomorphism from a *field* to the *endomorphism ring* of the group.

An **abelian group** is a pair $(V, +)$, where $+$ is a binary operation called **addition** on the set $V$, and

1) the equations

$$\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}, \qquad \boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) = (\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w}$$

   are *identities* on $V$, that is, $+$ is **commutative** and **associative** on $V$;
2) there is an **identity** for $+$ in $V$, namely an element of $V$ denoted by

$$\mathbf{0}$$

   and called **zero,** such that, for all $\boldsymbol{a}$ in $V$,

$$\boldsymbol{a} + \mathbf{0} = \boldsymbol{a}$$

   (and therefore $\mathbf{0} + \boldsymbol{a} = \boldsymbol{a}$, since $+$ is commutative);

3) each $\boldsymbol{a}$ in $V$ has an **(additive) inverse,** namely an element of $V$ denoted by

$$-\boldsymbol{a}$$

such that

$$-\boldsymbol{a}+\boldsymbol{a}=\mathbf{0}$$

(and therefore $\boldsymbol{a}+(-\boldsymbol{a})=\mathbf{0}$).

Yes, we use the word "identity" in two ways, to describe both (1) an equation that is true for all values of its variables, and also (2) an element of a set that has no effect when combined with others by means of some operation.

We may speak of $V$ as an abelian group if it is clear that we mean $(V,+)$.

An **endomorphism** of $(V,+)$ is a function $f$ from $V$ to itself such that the equation

$$f(\boldsymbol{u}+\boldsymbol{v})=f(\boldsymbol{u})+f(\boldsymbol{v})$$

is an identity on $V$. The set of endomorphisms of $(V,+)$ can be denoted by

$$\operatorname{End}(V,+).$$

This contains the function $\boldsymbol{v}\mapsto\boldsymbol{v}$, called the **identity** on $V$ and abbreviated by

$$\mathrm{id}_V\,.$$

Also $\operatorname{End}(V,+)$ is closed under the binary operations of **composition** and **addition,** defined by the identities

$$(f\circ g)(\boldsymbol{v})=f(g(\boldsymbol{v})),\qquad (f+g)(\boldsymbol{v})=f(\boldsymbol{v})+g(\boldsymbol{v}).$$

Addition in $\operatorname{End}(V,+)$ is different from addition in $V$, but they are related, and they are given the same symbol. The quadruple $(\operatorname{End}(V,+),+,\circ,\mathrm{id}_V)$ is a **ring:** this means

1) $\mathrm{End}(V,+)$ is an abelian group with respect to $+$;
2) composition on $\mathrm{End}(V,+)$ is associative, and $\mathrm{id}_V$ is an identity for it; this means that the structure $(\mathrm{End}(V,+),\circ,\mathrm{id}_V)$ a **monoid,** though I did not use this term in class;
3) the equations

$$(f+g)\circ h = f\circ h + g\circ h, \quad f\circ(g+h) = f\circ g + f\circ h$$

are identities on $\mathrm{End}(V,+)$, that is, composition **distributes** over addition.

The element $\mathrm{id}_V$ of $\mathrm{End}(V,+)$ is the **identity** both of the monoid and of the ring. Some persons' rings are not required to have identities; but our rings are. In our example, composition is the **multiplication** of the ring. In an arbitrary ring, this operation may be denoted by a dot, or by nothing at all, as in $a\cdot b$ or $ab$. Also, the identity of the ring may be denoted by 1. If the multiplication of a ring is commutative, then the ring itself is called a **commutative ring.***

A ring $(K,+,\,\cdot\,,1)$ is a **field** if $(K\smallsetminus\{0\},\,\cdot\,)$ is an abelian group, where 0 is the identity of the group $(K,+)$. In this case, if $\varphi$ is a function from $K$ to $\mathrm{End}(V,+)$ such that the equations

$$\varphi(x+y)=\varphi(x)+\varphi(y), \quad \varphi(xy)=\varphi(x)\circ\varphi(y), \quad \varphi(1)=\mathrm{id}_V$$

are identities, this means $\varphi$ *preserves* addition, multiplication, and the identity, and $\varphi$ is a **homomorphism** from $K$ as a

*The question arose among students in Şirince of why rings are so called. If $\eta$ is a complex number such that $\eta^2+B\eta+C=0$ for some integers $B$ and $C$, then the abelian group $\langle 1,\eta\rangle$, namely $\{x+\eta y\colon (x,y)\in\mathbb{Z}\times\mathbb{Z}\}$, is closed under multiplication. According to Harvey Cohn [2, p. 49], David Hilbert in 1892 introduced the term *number ring* (*Zahlring*) for the group, because $\eta^2$ "circles directly back" to the group as $-B\eta-C$.

ring to $\mathrm{End}(V,+)$ as a ring. In this case, the triple $(V,+,\varphi)$ is a vector space. We may also say that the pair $(V,+)$ is vector space **over** $K$, with respect to $\varphi$. Preservation of the identity by $\varphi$ is important. without it, $\varphi$ could be $x \mapsto 0$, where 0 stands for the element $\boldsymbol{v} \mapsto \mathbf{0}$ of $\mathrm{End}(V,+)$, even if $V$ has more than one element, and so $\mathrm{id}_V$ is different from 0 in $\mathrm{End}(V,+)$. We do not wish to allow $\varphi$ to be $x \mapsto 0$ unless $V = \{\mathbf{0}\}$.

In this last case, when $V = \{\mathbf{0}\}$, then $(V,+)$ is a **trivial group,** $\mathrm{End}(V,+)$ is the **trivial ring,** having a single element, which is both $\boldsymbol{v} \mapsto \mathbf{0}$ and $\mathrm{id}_V$, the zero and the identity. In this case, $V$ is a vector space over every field. In case $V$ is nontrivial, then every $\varphi$ as above must be injective, and therefore it can be understood as an identity, so that $K$ is just a sub-ring of $\mathrm{End}(V,+)$ that happens to be a field.

## Examples

Examples of fields are $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{Q}$.

$\mathbb{Z}$ is a commutative ring, but not a field.

The **counting numbers** are 1, 2, 3, and so on. I use the expression

$$\mathbb{N}$$

to denote the set of these numbers. Here presumably $\mathbb{N}$ stands for the **natural numbers.** Some persons count 0 as a natural number, considering it to belong to $\mathbb{N}$. I prefer to use the notation

$$\omega = \{0,1,2,\dots\} = \{0\} \cup \mathbb{N}.$$

Neither $\mathbb{N}$ nor $\omega$ is a ring.

If $n \in \mathbb{N}$, then by definition

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

where

$$\begin{aligned} [x] = [y] &\iff x \equiv y \pmod{n} \\ &\iff n \mid x - y. \end{aligned}$$

Here "$x \equiv y \pmod{n}$" is read as $x$ **is congruent to** $y$ ***modulo*** $n$.* The definition of $\mathbb{Z}_n$ does not say what $[x]$ *is*; it says only when two instances of $[x]$ are to be counted as the *same.* We can think that

$$\mathbb{Z}_n = \{0, \dots, n-1\} = \{x \in \omega \colon x < n\}, \tag{1.1}$$

since every integer is indeed congruent *modulo* $n$ to an element of this set, but no two elements are congruent to one another. Since, *modulo* $n$,

$$x \equiv x_1 \ \& \ y \equiv y_1 \implies x + y \equiv x_1 + y_1 \ \& \ xy \equiv x_1 y_1,$$

operations of addition and multiplication on $\mathbb{Z}_n$ are well defined by the rules

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

With respect to these operations, $\mathbb{Z}_n$ is a commutative ring, because $\mathbb{Z}$ is a commutative ring. If we use the understanding in (1.1), then for example $[n] = 0$, $[n+1] = 1$, and $[-1] = n-1$.

When is the ring $\mathbb{Z}_n$ a field, that is, when do nonzero elements have (multiplicative) inverses?

---

* *Modulo* is the dative or ablative case (roughly, the Turkish *-e* or *-den hali*) of the Latin noun *modulus.* Writing entirely in Latin, Gauss in the *Disquisitiones Arithmeticae* uses the expression *secundum modulum* $n$, "according to the modulus $n$" [3, p. 2].

- If $n = 1$, then $\mathbb{Z}_n = \{0\}$, which cannot be a field, since $\{0\} \smallsetminus \{0\} = \varnothing$, and every group is nonempty (it contains at least the identity).
- Suppose $n$ is composite, that is, $n = ab$ for some $a$ and $b$ such that $1 < a < n$, so that also $1 < b < n$. If $ax \equiv 1 \pmod{n}$, then

  $$b \equiv bax \equiv nx \equiv 0 \pmod{n},$$

  which is absurd. Thus $\mathbb{Z}_n$ is not a field in this case.
- If $n$ is prime, we shall show $\mathbb{Z}_n$ is a field. The proof will use the following.

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

$$p \mid ab \;\&\; p \nmid a \implies p \mid b.$$

Proving this is an **exercise** (which will be solved on pages 21 and 30). The letter $p$ will always stand for a prime. Because of the Lemma, if $1 \leqslant a < p$, then the endomorphism $x \mapsto ax$ of $(\mathbb{Z}_n, +)$ is injective, since, *modulo* $p$,

$$\begin{aligned} ax \equiv ay &\implies p \mid ax - ay \\ &\implies p \mid a(x - y) \\ &\implies p \mid x - y \qquad \text{[by the Lemma, since } p \nmid a] \\ &\implies x \equiv y. \end{aligned}$$

By the Pigeonhole Principle, since $\mathbb{Z}_p$ is finite, $x \mapsto ax$ must also be surjective. In particular, the congruence

$$ax \equiv 1 \pmod{p}$$

is soluble.

So now $\mathbb{Z}_p$ is field, called

$$\mathbb{F}_p.$$

We shall show that there are other finite fields. In particular, for every prime $p$, for every $n$ in $\mathbb{N}$, there will be a unique field called

$$\mathbb{F}_{p^n}$$

of size $p^n$. For example, $\mathbb{F}_4 = \{0, 1, \alpha, \alpha+1\}$, with multiplication defined so that

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

If we write $\beta$ for $\alpha + 1$, the operations are thus:

| $+$ | $0$ | $1$ | $\alpha$ | $\beta$ |
|---|---|---|---|---|
| $0$ | $0$ | $1$ | $\alpha$ | $\beta$ |
| $1$ | $1$ | $0$ | $\beta$ | $\alpha$ |
| $\alpha$ | $\alpha$ | $\beta$ | $0$ | $1$ |
| $\beta$ | $\beta$ | $\alpha$ | $1$ | $0$ |

| $\times$ | $0$ | $1$ | $\alpha$ | $\beta$ |
|---|---|---|---|---|
| $0$ | $0$ | $0$ | $0$ | $0$ |
| $1$ | $0$ | $1$ | $\alpha$ | $\beta$ |
| $\alpha$ | $0$ | $\alpha$ | $\beta$ | $1$ |
| $\beta$ | $0$ | $\beta$ | $1$ | $\alpha$ |

If $K$ is a finite field, then as a group it has the subgroup denoted by

$$\langle 1 \rangle,$$

namely the smallest subgroup that contains 1. Since $K$ is finite, $\langle 1 \rangle$ must be $\mathbb{Z}_n$ for some $n$. This will be a sub-ring of $K$, and so $n$ must be prime, since (in the terms of tomorrow's lecture) every sub-ring of a field must be an *integral domain,* and we have shown in effect that $\mathbb{Z}_n$ is not an integral domain if $n = 1$ or $n$ is composite. The prime $p$ such that $\langle 1 \rangle = \mathbb{Z}_p$, or rather such that

$$\underbrace{1 + \cdots + 1}_{p} = 0$$

in $K$, is called the **characteristic** of the field. There is no such $p$ for the fields $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$, and so these are said to have characteristic 0.

If $K$ is a field, and $n \in \mathbb{N}$, then $K^n$ is a vector space over $K$. The elements of $K^n$ are

$$(x_1, \dots, x_n),$$

where each $x_j$ is in $K$. Addition in $K^n$ is given by

$$(x_1, \dots, x_n) + (y_1, \dots, y_n) = (x_1 + y_1, \dots, x_n + y_n),$$

and each $a$ in $K^n$ determines the endomorphism

$$(x_1, \dots, x_n) \mapsto (ax_1, \dots, ax_n).$$

But what *is* $(x_1, \dots, x_n)$? It is the function on $\{1, \dots, n\}$ that takes the value $x_j$ at each $j$.

If $A$ is an arbitrary set, then we can define

$$K^A = \{\text{functions from } A \text{ to } K\};$$

this is a vector space over $K$, just as $K^n$ is. In particular,

$$K^n = K^{\{1,\dots,n\}}.$$

We cannot define $n$ as $\{1, \dots, n\}$. We can however define

$$n = \{0, \dots, n-1\}.$$

Thus

$$0 = \varnothing, \qquad\qquad n + 1 = n \cup \{n\}.$$

With this understanding, which is the one that I prefer, elements of $K^n$ can be written as

$$(x_0, \dots, x_{n-1}).$$

# 2. Tuesday, January 19

## Products

For any sets $A$ and $B$, we can let

$$B^A = \{\text{functions from } A \text{ to } B\}.$$

If $B$ is a field, then, as we said yesterday, $B^A$ is a vector space. Moreover, if $B$ is itself a vector space over some field, then $B^A$ is also a vector space over that field. If $B$ is a group or a ring, then so is $B^A$.

More generally, if we are given a set $B_a$ for each $a$ in $A$, we define the **product**

$$\prod_{a\in A} B_a$$

to be the set of functions $f$ on $A$ such that, for each $a$ in $A$,

$$f(a) \in B_a.$$

Thus

$$B^A = \prod_{a\in A} B.$$

Also, if $n \in \omega$, we may use the notations

$$\prod_{i\in n} B_i = B_0 \times \cdots \times B_{n-1}, \qquad B^n = \underbrace{B \times \cdots \times B}_{n}.$$

If each $B_a$ is a vector space (over the same field), or a group, or a ring, then so is $\prod_{a\in A} B_a$. In each case, this is because, for example, if $*$ is a binary operation on each $B_a$, then the "same" operation can be defined on $\prod_{a\in A} B_a$ by

$$(f * g)(x) = f(x) * g(x).$$

As we have seen, groups, rings, and vector spaces over a given field are defined by *identities,* namely (as we said on page 10) equations that hold for all possible values of the variables. An identity on $B$ is also an identity on $B^A$. However, among the field axioms, there is the formula

$$\exists y\ (x \neq 0 \implies xy = 1).$$

As an axiom, this formula is understood to hold for all values of its free variable, which is $x$. But the formula is not an equation, and so we cannot automatically expect it to hold in a product of fields. Indeed $K^A$ is not a field when $K$ is, unless $|A| = 1$.

## Integral domains

An **integral domain** is a nontrivial commutative ring with no **zero divisors,** that is, no nonzero elements $a$ and $b$ such that $ab = 0$. Thus integral domains have the axiom

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

which is not an equation. Indeed, in $\mathbb{Z} \times \mathbb{Z}$, we have

$$(1, 0) \cdot (0, 1) = (0, 0),$$

and $(0, 0)$ is the group identity of $\mathbb{Z} \times \mathbb{Z}$; so $(1, 0)$ and $(0, 1)$ are zero divisors.

If $n$ is a composite element of $\mathbb{N}$, we proved yesterday that $\mathbb{Z}_n$ is not a field by showing, in effect, that $\mathbb{Z}_n$ has zero divisors, and these are not invertible. This is true in every ring: a zero divisor cannot be invertible. Thus invertible elements are not zero divisors, and therefore every field is an integral domain. However, the converse fails: $\mathbb{Z}$ is an integral domain that is not a field.

## Finite cyclic groups

As we showed yesterday,

- $\mathbb{Z}_1$ is the trivial ring, which is not a field;
- $\mathbb{Z}_n$ is not a field when $n$ is composite;
- $\mathbb{Z}_p$ is a field when $p$ is prime.

Thus if $n \in \mathbb{N}$ and $\mathbb{Z}_n$ is a field, then $n$ must be prime, by pure logic. If $n$ is composite, so that $n = ab$, where $a$ and $b$ are strictly between 1 and $n$, then $a$ and $b$ are zero divisors in $\mathbb{Z}_n$. We consider the first few groups $\mathbb{Z}_n$.

- $\mathbb{Z}_2$ and $\mathbb{Z}_3$ are fields.
- In $\mathbb{Z}_4$, $1^2 = 1$, $2^2 = 0$, $3^2 = 1$, so 1 and 3 are invertible, but 2 is a zero-divisor and is therefore not invertible.
- $\mathbb{Z}_5$ is a field.
- In $\mathbb{Z}_6$, $1^2 = 1$, $2 \cdot 3 = 0$, $4 \cdot 3 = 0$, $5^2 = 1$; so 1 and 5 are invertible, but 2, 3, and 4 are zero divisors.
- $\mathbb{Z}_7$ is a field.
- In $\mathbb{Z}_8$, $1^2 = 3^3 = 5^2 = 7^2$, but $2 \cdot 4 = 6 \cdot 4 = 0$.
- In $\mathbb{Z}_9$, $1^2 = 2 \cdot 5 = 4 \cdot 7 = 8^2 = 1$, but $3^2 = 6^2 = 0$.

## Groups of units

If $R$ is a commutative ring, we define

$$R^\times = \{\text{invertible elements of } R\}.$$

Here $\times$ is the multiplication sign (and not the letter $x$). Thus if $p$ is prime, then

$$\mathbb{Z}_p{}^\times = \{1, \dots, p-1\};$$

but

$$\mathbb{Z}_4{}^\times = \{1, 3\}, \qquad \mathbb{Z}_6{}^\times = \{1, 5\}, \qquad \mathbb{Z}_8{}^\times = \{1, 3, 5, 7\},$$

$$\mathbb{Z}_9{}^\times = \{1, 2, 4, 5, 7, 8\}, \qquad \mathbb{Z}_{10}{}^\times = \{1, 3, 7, 9\}.$$

In general, $R^\times$ is called the **group of units** of $R$. The set $R^\times$ is indeed a group with respect to multiplication, since each element $a$ by definition has an inverse, $a^{-1}$, and then this too has an inverse, namely $a$, and so $a^{-1} \in R^\times$. Invertible elements of a ring are **units.**

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

$$\mathbb{Z}_n{}^\times = \{k \in \omega \colon k < n \ \& \ \gcd(k, n) = 1\}. \tag{2.1}$$

*Proof.* Assuming $0 \leqslant k < n$, we want to show that, *modulo* $n$,

$$\gcd(k, n) = 1 \iff \exists x\ kx \equiv 1.$$

($\Leftarrow$). Suppose $ka \equiv 1$. Then $n \mid ka - 1$, so

$$\gcd(k, n) \mid ka - 1.$$

But $\gcd(k,n) \mid ka$, so $\gcd(k,n) \mid 1$, and therefore $\gcd(k,n) = 1$.

($\Rightarrow$). Suppose $\gcd(k,n) = 1$. As yesterday, the endomorphism $x \mapsto kx$ of $\mathbb{Z}_n$ is injective, by the more general form of Euclid's Lemma below. Again $x \mapsto kx$ must be surjective by the Pigeonhole Principle, so $kx \equiv 1 \pmod{n}$ is soluble. $\square$

So the proof is completed by the following.

**Euclid's Lemma, generalized.** *For all $m$ in $\mathbb{N}$, for all $a$ and $b$ in $\omega$,*

$$m \mid ab \ \& \ \gcd(m,a) = 1 \implies m \mid b. \tag{2.2}$$

*Proof.* Suppose the claim is false; we shall find a contradiction. If the claim is false, then, because $\mathbb{N}$ is well ordered, there is a *minimal counterexample,* namely some $n$ in $\mathbb{N}$ such that, for all $m$ in $\mathbb{N}$, if $m < n$, then for all $a$ and $b$ in $\omega$, (2.2) holds; but for some $a$ and $b$ in $\omega$,

$$n \mid ab \ \& \ \gcd(n,a) = 1 \ \& \ n \nmid b. \tag{2.3}$$

If $a \geqslant n$, then we can replace $a$ with $a - n$. Thus, since $\omega$ is well ordered, we may assume $a < n$. Similarly we may assume $b < n$. Then $ab < n^2$, and so, for some $m$,

$$mn = ab \ \& \ m < n.$$

In particular,

$$\frac{m}{\gcd(m,a)} \cdot n = \frac{a}{\gcd(m,a)} \cdot b.$$

Therefore in (2.3) we can replace $a$ with $a/\gcd(m,a)$. Thus we may assume $\gcd(m,a) = 1$. Now, (2.2) holds, since $m < n$.

Therefore $m \mid b$. Hence in (2.3) we can replace $b$ with $b/m$. But in this case $n = ab$, so $\gcd(n, a) = a$. Thus $a = 1$, and so $n \mid b$, contradicting (2.3). So there is no minimal counterexample to the original claim. Therefore there is no counterexample at all, and the claim holds. □

By definition,

$$\varphi(n) = \{x \in \omega \colon x < n \ \ \& \ \ \gcd(n, x) = 1\},$$

and so now we have, by (2.1),

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

Here $\varphi$ is the **Euler phi-function.**[*] A few values are as follows.

| $n$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\varphi(n)$ | 1 | 1 | 2 | 2 | 4 | 2 | 6 | 4 | 6 | 4 | 10 | 4 |

In general,

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

[*]This terminology is used by Burton [1, pp. 131–2], who mentions also the alternative terms *indicator* and *totient,* while saying, "the functional notation $\phi(n)$, however, is credited to Gauss." Burton does not say who does the crediting. Kline [6, p. 608] does it, saying, "The notation $\phi(n)$ was introduced by Gauss." Indeed, Article 38 of the *Disquisitiones Arithmeticae* [4, p. 20] is, "PROBLEM. *To find how many positive numbers are smaller than a given positive number A and relatively prime to it.* For brevity we will designate the number of positive numbers which are relatively prime to the given number and smaller than it by the prefix $\phi$. We seek therefore $\phi A$." In the next article, Gauss adjusts the definition to allow $\phi 1$ to be 1 (as we do) and not 0.

From the table,

$$\begin{aligned}
\varphi(6) &= \varphi(2) \cdot \varphi(3),\\
\varphi(10) &= \varphi(2) \cdot \varphi(5),\\
\varphi(12) &= \varphi(3) \cdot \varphi(4).
\end{aligned}$$

We are going to prove that this is not an accident, but φ is **multiplicative,** that is,*

$$\gcd(x, y) = 1 \implies \varphi(xy) = \varphi(x) \cdot \varphi(y).$$

*It may be noted that, for the letter phi, instead of $\phi$ I prefer $\varphi$, as being less like the symbol ∅ for the empty set, and as being more like what I write by hand. For use as a permanently defined constant, I prefer the upright form φ. Most of the students were familiar with the phi-function, though a few were not.

# 3. Wednesday, February 19

## Finite fields

Our ultimate goal is to prove the following four theorems:

1. For every prime $p$, for every $n$ in $\mathbb{N}$, there is a field of size $p^n$.

2. There is a unique such field (up to isomorphism), called $\mathbb{F}_{p^n}$.

3. The group of units of a finite field is cyclic, so

$$\mathbb{F}_{p^n}{}^{\times} \cong \mathbb{Z}_{p^n-1}.$$

4. $\mathbb{F}_{p^m} \subseteq \mathbb{F}_{p^n}$ if and only if—what?

We answer the last question as follows. For every field $L$, if $K$ is a subfield of $L$, then $L$ is a vector space over $K$. In this case, suppose the **dimension** of $L$ over $K$ is $n$:

$$\dim_K L = d.$$

What is the size of $L$?* We know $L$ has a **basis** $(a_i : i < d)$;† this means that every element of $L$ is uniquely of the form

$$\sum_{i<d} x_i a_i,$$

where $x_i \in K$. Thus

$$|L| = |K|^d.$$

In case $K = \mathbb{F}_{p^m}$, we have $|L| = p^{md}$. Thus

$$\mathbb{F}_{p^m} \subseteq \mathbb{F}_{p^n} \implies m \mid n.$$

The converse will be true as well. Thus the finite fields of characteristic $p$ will form a tree, part of which is seen in Figure 3.1. In fact there will be a field $\mathbb{F}_{p^\ell}$ right above $\mathbb{F}_p$ in the tree for every prime $\ell$. Also, for all $a$ and $b$ in $\mathbb{N}$, both $\mathbb{F}_{p^a}$ and $\mathbb{F}_{p^b}$ will be included in $\mathbb{F}_{p^{\gcd(a,b)}}$. Thus the union $\bigcup_{n\in\mathbb{N}} \mathbb{F}_{p^n}$ of all of the finite fields of characteristic $p$ will be a field, namely the *algebraic closure* of $\mathbb{F}_p$. (See page 52.)

## Polynomial rings

On page 15 we saw $\mathbb{F}_4 = \{0, 1, \alpha, \alpha+1\}$. A better way to write this will be as $\mathbb{F}_2[\alpha]$ or else

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

---

*One student initially proposed the answer $n \cdot |K|$.

†Probably in class I wrote $\{a_i : i < d\}$, but strictly this notation does not establish that the basis has size $d$. The ensuing explanation of a basis relies not just on the *set* of basis elements, but on a function into this set from a set of size $d$. Here the set of size $d$ is $d$ itself (page 16), which is $\{0, \dots, d-1\}$ or $\{x \in \omega : x < d\}$, and the function on it is $i \mapsto a_i$.

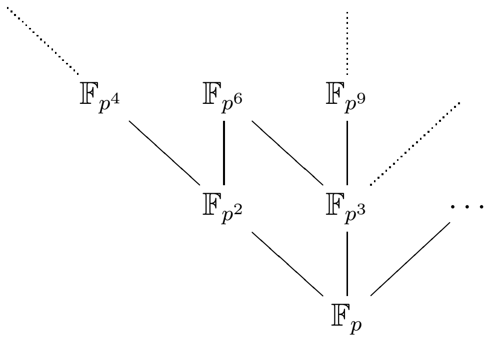


Figure 3.1.: The tree of finite fields of characteristic $p$

For any commutative ring $R$,

$$\begin{aligned} R[X] &= \{\text{polynomials in } X \text{ over } R\} \\ &= \{0\} \cup \left\{ \sum_{i=0}^{n} a_i X^i \colon n \in \omega \ \ \& \ \ a_i \in R \ \ \& \ \ a_n \neq 0 \right\}. \end{aligned}$$

If $f = \sum_{i=0}^{n} a_i X^i$, where $a_n \neq 0$, then we say the **degree** of $f$ is $n$, that is,

$$\deg f = n.$$

Also, by definition,

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

Then for all $f$ and $g$ in $R[X]$,

$$\deg(fg) \leqslant \deg f + \deg g,$$

with equality if $R$ is an integral domain. In this case, $R[X]$ too must be an integral domain.

We are interested mainly in the case where $R$ is a field $K$. If $f \in K[X]$, we define

$$K[X]/(f) = \{[g] \colon g \in K[X]\},$$

where

$$[g] = [h] \iff f \mid g - h.$$

In the same way, if $n \in \mathbb{N}$, we can write

$$\mathbb{Z}/(n) = \mathbb{Z}_n,$$

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

$$[x] = [y] \iff n \mid x - y.$$

Indeed, we are going to be investigating an analogy between $\mathbb{Z}$ and $K[X]$. (See the remark on page 13 about how we do not say what $[x]$ is in itself.)

In $K[X]/(f)$, if we write $[X]$ as $\alpha$ (that is, $[g] = \alpha$ when $g$ is $X$), then for any $g$ in $K[X]$,

$$[g] = g(\alpha).$$

Assuming $\deg f = n > 0$, by *division* (see below) we have that, for every $g$ in $K[X]$, there is $g_1$ in $K[X]$ such that

$$f \mid g - g_1, \qquad \deg g_1 < n.$$

Here $g_1$ is the **remainder** after dividing $g$ by $f$. Thus

$$K[X]/(f) = \{g(\alpha) \colon g \in K[X] \ \& \ \deg g < n\}.$$

We may denote this set also by

$$K[\alpha].$$

In fact it is a ring in a fairly obvious way. For, we may assume that $f$ is **monic,** that is, $f = X^n - g$ for some $g$ having degree less than $n$. Then multiplication in $K[\alpha]$ is as we expect, except that $f(\alpha) = 0$, so that $\alpha^n$ is replaced with $g(\alpha)$ whenever it comes up. We shall say more on this tomorrow.

If $f = gh$, where $0 < \deg g < n$, then also $0 < \deg h < n$, and so $g(\alpha)$ and $g(\alpha)$ are nonzero, although their product is $0$. Thus $K[\alpha]$ has zero divisors in this case. We say $f$ is **reducible.**

Otherwise $f$ is **irreducible,** and then $K[\alpha]$ is an integral domain. In this case, if $K$ is finite, then $K[\alpha]$ is actually a field, for the same reason that $\mathbb{Z}_p$ was shown to be a field on page 15: multiplication by a nonzero element is an injective endomorphism of the additive group, so it is also surjective. (By a different argument, $K[\alpha]$ will be a field, even if $K$ is infinite; see page 50.)

Thus, to prove the first of the four theorems above about finite fields, for each prime $p$, for each $n$ in $\mathbb{N}$, we shall show that there is an irreducible polynomial over $\mathbb{F}_p$ of degree $n$.

For example, $X^2 + X + 1$ is irreducible over $\mathbb{F}_2$, because for all $a$ and $b$ in $\mathbb{F}_2$,

$$(X - a)(X - b) = X^2 - (a + b)X + ab,$$

but

$$\begin{aligned} a = b &\implies a + b = 0, \\ a \neq b &\implies ab = 0, \end{aligned}$$

and so

$$(X - a)(X - b) \neq X^2 + X + 1.$$

Similarly, $X^3 + X + 1$ is irreducible over $\mathbb{F}_2$: showing this is an **exercise.**

## The Euclidean algorithm

The chief analogy between $\mathbb{Z}$ and $K[X]$ lies in the similarity of the functions $x \mapsto |x|$ on $\mathbb{Z}$ and $f \mapsto \deg f$ on $K[X]$. Each one allows us to perform **division** effectively. Indeed, for all $a$ in $\mathbb{Z} \smallsetminus \{0\}$, for all $b$ in $\mathbb{Z}$, we have

$$b = ax + y$$

for some unique $x$ and $y$ in $\mathbb{Z}$ such that also

$$0 \leqslant y < |a| .$$

Similarly, for all $f$ in $K[X] \smallsetminus \{0\}$, for all $g$ in $K[X]$, we have

$$g = fh + r$$

for some unique $h$ and $r$ in $K[X]$ such that also

$$\deg r < \deg f.$$

This means that, in both $\mathbb{Z}$ and $K[X]$, the **Euclidean algorithm** can be used to find greatest common divisors. For example, in $\mathbb{Z}$, we have $\gcd(13, 8) = 1$, because

$$\begin{aligned} 13 &= 8 + 5, \\ 8 &= 5 + 3, \\ 5 &= 3 + 2, \\ 3 &= 2 + 1. \end{aligned}$$

Reversing the steps, we find

$$\begin{aligned} 1 &= 3 - 2 \\ &= 3 - (5 - 3) = 3 \cdot 2 - 5 \\ &= (8 - 5) \cdot 2 - 5 = 8 \cdot 2 - 5 \cdot 3 \\ &= 8 \cdot 2 - (13 - 8) \cdot 3 = 8 \cdot 5 - 13 \cdot 3. \end{aligned}$$

In the same way, for all $a$ and $b$ in $\mathbb{Z}$ that are not both 0, the equation

$$ax + by = \gcd(a, b) \tag{3.1}$$

is soluble in $\mathbb{Z}$. This result can be called **Bézout's Lemma.*** This now gives us an easy proof of the general form of Euclid's Lemma on page 21. We assume $n \mid ab$ and $\gcd(n, a) = 1$. Then for some $x$ and $y$ we have

$$\begin{aligned} nx + ay &= 1, \\ bnx + aby &= b, \end{aligned}$$

and so $n \mid b$ since $n \mid bnx$ and $n \mid aby$.

A reason for giving the other proof is that solubility of (3.1) is actually stronger than Euclid's Lemma: there are rings in which the latter is true, but (3.1) is not always soluble. We shall see an example tomorrow (page 40).

## The Chinese Remainder Theorem

For any $k$ and $m$ in $\mathbb{N}$, there is a well-defined ring homomorphism $x \mapsto (x, x)$ from $\mathbb{Z}_{km}$ to $\mathbb{Z}_k \times \mathbb{Z}_m$. If $\gcd(k, m) = 1$,

*The term is used in *Wikipedia,* but I have not found clear historical justification for it. Bézout's relevant concern seems to have been polynomials. According to Morris Kline [6, p. 608] (who writes Bézout's name without the accent), "Bezout's idea [sketched first in 1764] was that by multiplying $f(x, y)$ and $g(x, y)$ by suitable polynomials, $F(x)$ and $G(x)$ respectively, he could form $R(y) = F(x)f(x, y) + G(x)g(x, y)$." Now, an "idea" is not necessarily a theorem; but perhaps Kline made a mistake, and $F$ and $G$ were supposed to be polynomials in $x$ *and* $y$. This way, unless $f$ and $g$ have a common factor in $x$, the $x$ can be eliminated by performing the Euclidean algorithm in $K(y)[x]$, then clearing fractions.

this homomorphism is injective: this was shown in Ali Nesin's class earlier today. In this case, it follows that

$$\mathbb{Z}_{km}{}^{\times} \cong \mathbb{Z}_{k}{}^{\times} \times \mathbb{Z}_{m}{}^{\times},$$

and so

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

What is the inverse of the isomorphism $x \mapsto (x, x)$? It is given by the **Chinese Remainder Theorem.** For all $a$ and $b$ in $\mathbb{Z}$, we know that the congruences

$$x \equiv a \pmod{k}, \qquad x \equiv b \pmod{m} \tag{3.2}$$

are equivalent to a single congruence

$$x \equiv \underline{\hspace{2cm}} \pmod{km}.$$

Indeed, we just let $x$ be the inverse image of $(a, b)$ under $x \mapsto (x, x)$. We expect

$$x \equiv a\underline{\hspace{0.6cm}} + b\underline{\hspace{0.6cm}} \pmod{km}.$$

The multiple of $a$ should disappear *modulo* $k$, and the multiple of $b$ should disappear *modulo* $m$. Thus we expect

$$x \equiv am\underline{\hspace{0.3cm}} + bk\underline{\hspace{0.3cm}} \pmod{km}.$$

But then the multiple of $a$ should be $a$ *modulo* $k$, and the multiple of $b$ should be $b$ *modulo* $m$. Therefore the congruences (3.2) are equivalent to

$$x \equiv amc + bkd \pmod{km},$$

where

$$mc \equiv 1 \pmod{k}, \qquad kd \equiv 1 \pmod{m}.$$

$$
\begin{array}{cc|c|cccc|}
 & & \multicolumn{4}{c}{\mathbb{Z}_4} \\
\cline{3-7}
 & & & 0 & 1 & 2 & 3 \\
\cline{2-7}
 & 0 & & 0 & 9 & 6 & 3 \\
\mathbb{Z}_3 & 1 & & 4 & 1 & 10 & 7 \\
 & 2 & & 8 & 5 & 2 & 11 \\
\cline{2-7}
\end{array}
$$

Figure 3.2.: The Chinese Remainder Theorem in a table

Such $c$ and $d$ exist since $\gcd(k, m) = 1$, so that

$$m \in \mathbb{Z}_k{}^\times, \qquad k \in \mathbb{Z}_m{}^\times.$$

The inverse of $x \mapsto (x, x)$ is now

$$(y, z) \mapsto mcy + kdz.$$

One can read the values of this inverse off a table, as in Figure 3.2, where one can read directly that the solution to

$$x \equiv 1 \pmod 3, \qquad x \equiv 2 \pmod 4$$

is 10.

There is also a well-defined homomorphism

$$(y, z) \mapsto my + kz$$

from $\mathbb{Z}_k \times \mathbb{Z}_m$ to $\mathbb{Z}_{km}$ as *groups,* again assuming $\gcd(k, m) = 1$. There are group endomorphisms $x \mapsto cx$ of $\mathbb{Z}_k$ and $x \mapsto dx$ of $\mathbb{Z}_m$. These fit into a commutative diagram as in Figure 3.3. We can therefore conclude that $(x, y) \mapsto mx + ky$ is an *iso*morphism from $\mathbb{Z}_k \times \mathbb{Z}_m$ as a group to $\mathbb{Z}_{km}$.

This means for example that every element of $\mathbb{Z}_{12}$ is, in exactly one way, the difference of an element of $\langle 3 \rangle$ and an element of $\langle 4 \rangle$; see Figure 3.4.

$$
\begin{tikzcd}
\mathbb{Z}_k \times \mathbb{Z}_m \arrow[dd, "(y,z) \longmapsto (cy,dz)"'] & \\
 & \mathbb{Z}_{km} \arrow[ul, "(x,x) \longmapsfrom x"'] \\
\mathbb{Z}_k \times \mathbb{Z}_m \arrow[ur, "(y,z) \longmapsto my+kz"'] &
\end{tikzcd}
$$

Figure 3.3.: The Chinese Remainder Theorem in three isomorphisms

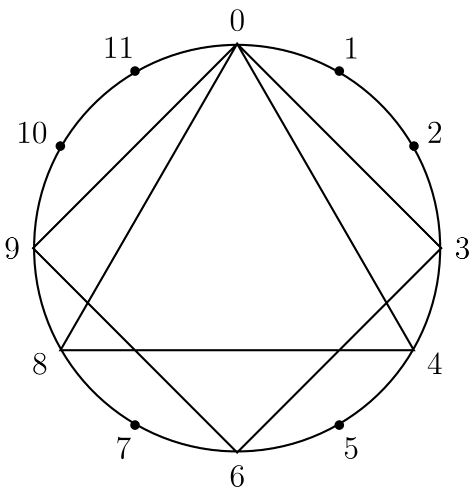


Figure 3.4.: $\mathbb{Z}_{12} = \langle 3 \rangle \times \langle 4 \rangle \cong \mathbb{Z}_4 \times \mathbb{Z}_3$

# 4. Thursday, January 21

## Endomorphism rings

We have observed (page 10) that, for every abelian group $(V, +)$, there is an **endomorphism ring,** namely

$$(\mathrm{End}(V, +), +, \circ, \mathrm{id}_V).$$

If the abelian group $V$ is trivial, then the endomorphism ring $\mathrm{End}(V, +)$ is trivial. If $\mathrm{End}(V, +)$ is not trivial, then $V$ is a vector space over any sub-ring of $\mathrm{End}(V, +)$ that happens to be a field. Also $V$ is a **module** over any sub-ring of $\mathrm{End}(V, +)$ at all. ($V$ is a **left** module over a sub-ring of $\mathrm{End}(V, +)$, if the elements of $\mathrm{End}(V, +)$ are written on the left of their arguments, or more precisely if $(f \circ g)(x)$ means $f(g(x))$ and not $g(f(x))$.)

What is $\mathrm{End}(\mathbb{Z}_n, +)$? Suppose $\varphi \in \mathrm{End}(\mathbb{Z}_n, +)$. If $0 \leqslant k < n$, then

$$\varphi(k) = \varphi(\underbrace{1 + \cdots + 1}_{k}) = \underbrace{\varphi(1) + \cdots + \varphi(1)}_{k} = \varphi(1) \cdot k.$$

Thus $\varphi$ is determined by $\varphi(1)$. Conversely, if $a \in \mathbb{Z}_n$, we can define $\varphi_a$ in $\mathrm{End}(\mathbb{Z}_n, +)$ by

$$\varphi_a(x) = ax.$$

Thus $\varphi = \varphi_{\varphi(1)}$. We have moreover

$$\varphi_{ab} = \varphi_a \circ \varphi_b, \qquad \varphi_{a+b} = \varphi_a + \varphi_b.$$

Since, finally, $\varphi_1 = \mathrm{id}_{\mathbb{Z}_n}$, we can conclude

$$x \longmapsto \varphi_x \colon (\mathbb{Z}_n, +, \,\cdot\,, 1) \xrightarrow{\cong} (\mathrm{End}(\mathbb{Z}_n, +), +, \circ, \mathrm{id}_{\mathbb{Z}_n}).$$

In this way, the ring structure of $\mathbb{Z}_n$ is determined by the group structure.

## Automorphism groups

An **automorphism** is an endomorphism that is invertible as an endomorphism. In other words, it is an isomorphism from a structure to itself. (The letters AU of "auto-" in "automorphism" are related to *EF* in the Turkish *efendi,* since the latter—which was taken into English as "effendi"—comes from the Greek αὐθέντης, source also of the English "authentic" and a compound of αὐτο- and ἕντης.) The automorphisms of $(\mathbb{Z}_n, +)$, for example, compose a set denoted by

$$\mathrm{Aut}(\mathbb{Z}_n, +).$$

Then by what we have just shown,

$$\mathrm{Aut}(\mathbb{Z}_n, +) \cong {\mathbb{Z}_n}^{\times}.$$

We are going to show

$$\mathrm{Aut}(\mathbb{Z}_p) \cong \mathbb{Z}_{p-1}.$$

Meanwhile, what is $\mathrm{Aut}(\mathbb{Z}_{10})$? We have

$$\begin{gathered}{\mathbb{Z}_{10}}^{\times} = \{1, 3, 7, 9\} = \{1, 3, -3, -1\},\\ 3^2 = -1, \quad 3^3 = -3,\end{gathered}$$

and therefore ${\mathbb{Z}_{10}}^\times \cong \mathbb{Z}_4$. However,

$$ {\mathbb{Z}_{30}}^\times = \{1, 7, 11, 13, -13, -11, -7, -1\} $$

(the number of elements is correct since $\varphi(30) = \varphi(2) \cdot \varphi(3) \cdot \varphi(5) = 2 \cdot 4 = 8$). We find

$$ \begin{gathered} 7^2 = -11, \quad 7^3 = 13, \quad 7^4 = 1, \\ \langle 7 \rangle \cong \mathbb{Z}_4, \\ {\mathbb{Z}_{30}}^\times = \langle 7, -1 \rangle \cong \mathbb{Z}_4 \times \mathbb{Z}_2. \end{gathered} $$

It is also the case that ${\mathbb{Z}_{30}}^\times = \langle 7, -7 \rangle$. However, $\langle 7 \rangle \cap \langle -7 \rangle = \langle -11 \rangle \cong \mathbb{Z}_2$, so ${\mathbb{Z}_{30}}^\times$ is not $\langle 7 \rangle \times \langle -7 \rangle$: this product is too big. The main point to observe is that ${\mathbb{Z}_{30}}^\times$ is not cyclic.

For any group $(G, \cdot\,)$, abelian or not, if $G$ contains $a$ and $b$, and $ab = ba$, and $\langle a \rangle \cap \langle b \rangle$ is trivial, then $\langle a, b \rangle \cong \langle a \rangle \times \langle b \rangle$. (In class I stated this only for abelian groups, written additively; but it is not hard to prove in any case.)

As an **exercise,** write ${\mathbb{Z}_{24}}^\times$ as a product of finite cyclic groups. In fact

$$ {\mathbb{Z}_{24}}^\times = \langle 5, 7, -1 \rangle \cong {\mathbb{Z}_2}^3. $$

It is a theorem (which we are not going to prove in its entirety) that ${\mathbb{Z}_n}^\times$ is cyclic if and only if $n$ is 2, 4, $p^k$, or $2p^k$ for some $k$ in $\mathbb{N}$ and some odd prime $p$.

Suppose $\gcd(k, m) = 1$. Then $k \in {\mathbb{Z}_m}^\times$ and $m \in {\mathbb{Z}_k}^\times$, so $c$ and $d$ exist as before such that

$$ mc \equiv 1 \pmod{k}, \qquad kd \equiv 1 \pmod{m}. $$

Then $c \in {\mathbb{Z}_k}^\times$ and $d \in {\mathbb{Z}_m}^\times$, so $(x, y) \mapsto (cx, dy)$ is indeed an automorphism of $\mathbb{Z}_k \times \mathbb{Z}_m$. Then $x \mapsto (cx, dx)$ is an isomorphism from $\mathbb{Z}_{km}$ to $\mathbb{Z}_k \times \mathbb{Z}_m$. As an **exercise,** verify our earlier claim that the inverse of this isomorphism is $(x, y) \mapsto mx + ky$.

If $K$ is a field and $f \in K[X]$, we defined

$$K[X]/(f) = \{[g] \colon g \in K[X]\},$$

where

$$\begin{aligned} [g] = [h] &\iff f \mid g - h \\ &\iff g \equiv h \mod f. \end{aligned}$$

We let $\alpha = [X]$ and wrote $K[X]/(f)$ as $K[\alpha]$, noting $g(\alpha) = [g]$. We observed that, if $\deg f = n > 0$, then

$$K[\alpha] = \{g(\alpha) \colon g \in K[X] \ \& \ \deg g < n\}.$$

In particular, $K[\alpha]$ has basis $\{1, \alpha, \dots, \alpha^{n-1}\}$ as a vector space over $K$. This is true, even if $n = 0$, in the sense that the basis is empty in this case. For, if $\deg f = 0$, this means $f \in K^{\times}$. Moreover,

$$K^{\times} = K[X]^{\times}. \tag{4.1}$$

Thus $f$ is invertible, which means it divides everything in $K[X]$, and so $[g] = [h]$ for all $g$ and $h$ in $K[X]$. In particular, $K[\alpha] = \{0\} = \{\alpha\}$.

There is a final case to consider. If $f = 0$, then

$$[g] = [h] \iff g = h.$$

Thus $K[X]/(0) \cong K[X]$.

## Irreducibles

By (4.1), for all $f$ in $K[X]$,

$$\deg f = 0 \iff f \in K[X]^{\times}.$$

This allows us to rewrite the definition of irreducible polynomial in a form that makes sense for any commutative ring. If $R$ is such a ring, an element $\pi$ is **irreducible** if $\pi \neq 0$ and $\pi \notin R^{\times}$ and

$$\pi = ab \ \& \ a \notin R^{\times} \implies b \in R^{\times}.$$

By this definition, the irreducibles of $\mathbb{Z}$ are just $\pm p$. An element $\pi$ of $R$ is **prime** if $\pi \neq 0$ and $\pi \notin R^{\times}$ and $\pi$ satisfies Euclid's Lemma, that is,

$$\pi \mid ab \ \& \ \pi \nmid a \implies \pi \nmid b.$$

Then we can reformulate an earlier statement (from page 14) as follows:

**Euclid's Lemma.** *In $\mathbb{Z}$, all irreducibles are prime.*

Euclid's Lemma ensures that irreducible factorizations are unique, when they exist at all. Thus if

$$p_0 \cdots p_{m-1} = q_0 \cdots q_{n-1},$$

where $p_i$ and $q_j$ are positive irreducibles in $\mathbb{Z}$, then

$$p_0 \mid q_0 \cdots q_{n-1},$$

and so $p_0 \mid q_j$ for some $j$, and then $p_0 = q_j$. In this way $m = n$, and for some permutation $\sigma$ of $m$,

$$p_0 \cdots p_{m-1} = q_{\sigma(0)} \cdots q_{\sigma(m-1)}.$$

The converse of Euclid's Lemma is more generally true:

**Theorem.** *In every integral domain, all primes are irreducible.*

*Proof.* Suppose $\pi$ is prime and $\pi = ab$. Then $\pi \mid ab$, so we may assume $\pi \mid a$. For some $c$ then, $c\pi = a$. Hence

$$bc\pi = ab = \pi,$$

so $bc = 1$ in an integral domain. In this case, $c$ is invertible. $\square$

In integral domain whose every nonzero element is uniquely the product of irreducibles is called a **unique factorization domain** or UFD. Here the uniqueness is up to replacement by *associates*; two elements of a ring are **associates** if they divide one another, like $a$ and $-a$ in $\mathbb{Z}$. Euclid's Lemma is that $\mathbb{Z}$ is a UFD. Any two elements of a UFD that are not both 0 have a greatest common divisor.* Indeed, if two elements $a$ and $b$ are respectively

$$\prod_{i<n} \pi^{a(i)}, \qquad \prod_{i<n} \pi^{b(i)},$$

where the $\pi_i$ are irreducible, then

$$\gcd(a,b) = \prod_{i<n} \pi^{\min(a(i),b(i))}.$$

We also proved something stronger for $\mathbb{Z}$ (page 30):

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

$$ax + by = \gcd(a,b) \tag{4.2}$$

*is soluble.*

*In class I did not define associates or show that greatest common divisors exist in UFDs.

This is stronger than Euclid's Lemma, since in $\mathbb{Z}[X]$, all irreducibles are prime, and in fact $\mathbb{Z}[X]$ is a UFD; but $\gcd(X, 2) = 1$, and the equation

$$2u + Xt = 1$$

is insoluble. A UFD in which Bézout's Lemma holds is a **principal ideal domain** or PID.

As earlier (page 29), we can derive Bézout's Lemma from the following:

**Theorem** (division algorithm)**.** *On $\mathbb{Z}$, the function $x \mapsto |x|$ on $\mathbb{Z}$ has the following properties.*

1. *Its range is well ordered.*
2. *For every nonzero $a$, for every $b$, for some $c$,*

$$a \mid b - c, \qquad |c| < |a|.$$

Any commutative ring with the properties of $\mathbb{Z}$ in the theorem is a **Euclidean domain** or ED. Thus if $K$ is a field, then $K[X]$ is an ED. Some PIDs are not EDs, though this is hard to prove.

As a reminder that not all EDs are PIDs, one may wish to prove Bézout's Lemma for $\mathbb{Z}$ without relying on the Euclidean algorithm, just as we originally proved Euclid's Lemma on page 21. If $a$ and $b$ are not both 0, one can show that the least positive element of $\{ax + by\colon (x, y) \in \mathbb{Z}^2\}$ must be $\gcd(a, b)$.

Nonetheless, we may understand **Bézout's Lemma** as the theorem that equation (4.2) is always soluble in any Euclidean domain, particularly $K[X]$ when $K$ is a field.

# 5. Saturday, January 23

About finite fields, we know that

1) the size of each of them is a power of a prime, and
2) if $L$ is a finite field with subfield $K$, then for some $n$ in $\mathbb{N}$,
$$|L| = |K|^n\,.$$

We are now going to show that $L^\times$ is cyclic.

## Finite abelian groups

On Thursday we looked at ${\mathbb{Z}_n}^\times$ when $n$ is 10 and 24. Also,

$$\begin{aligned}{\mathbb{Z}_{28}}^\times &= \{x \in \mathbb{Z}\colon -14 < x \leqslant 14 \;\;\&\;\; \gcd(28, x) = 1\}\\ &= \{\pm1, \pm3, \pm5, \pm9, \pm11, \pm13\},\end{aligned}$$

which is confirmed by

$$\varphi(28) = \varphi(4) \cdot \varphi(7) = 2 \cdot 6 = 12.$$

We compute

$$\begin{array}{c|cccccc} k & 1 & 2 & 3 & 4 & 5 & 6 & \\ \hline 3^k & 3 & 9 & -1 & -3 & -9 & 1 & \pmod{28} \end{array},$$

so that $\langle 3\rangle \cong \mathbb{Z}_6$. Since $13^2 = 169$ and $6 \cdot 28 = 168$, we have

$$\begin{gathered} 13^2 \equiv 1 \pmod{28},\\ \langle 13\rangle \cap \langle 3\rangle = \{1\},\\ \langle 13, 3\rangle = \langle 13\rangle \times \langle 3\rangle,\\ {\mathbb{Z}_{28}}^\times = \langle 13\rangle \times \langle 3\rangle \cong \mathbb{Z}_2 \times \mathbb{Z}_6. \end{gathered}$$

We use here the more general result that, if $(G, \cdot)$ is an arbitrary group with elements $a$ and $b$ such that $ab = ba$, then there is a well-defined homomorphism $(x, y) \mapsto xy$ or $(a^i, b^j) \mapsto a^i b^j$ from $\langle a \rangle \times \langle b \rangle$ to $G$ (**exercise**). If, further, $\langle a \rangle \cap \langle b \rangle = \{1\}$, then the homomorphism is injective, since if $a^i b^j = 1$, then $a^i = b^{-j}$, so both of these are in $\langle a \rangle \cap \langle b \rangle$, and therefore they are 1. In this case

$$\langle a \rangle \times \langle b \rangle \cong \langle a, b \rangle.$$

We can write this as an equation if it is understood that $a$ and $b$ indeed belong to the same group, so that $\langle a \rangle \times \langle b \rangle$ is the so-called **internal direct product.***

We know $\mathbb{Z}_6 \cong \mathbb{Z}_2 \times \mathbb{Z}_3$, and in fact $\mathbb{Z}_6$ is the internal direct product $\langle 3 \rangle \times \langle 2 \rangle$. Thus

$$\begin{gathered}
{\mathbb{Z}_{28}}^{\times} \cong \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_3, \\
{\mathbb{Z}_{28}}^{\times} \cong \langle 13 \rangle \times \langle 3^3 \rangle \times \langle 3^2 \rangle \cong \langle 13 \rangle \times \langle -1 \rangle \times \langle 9 \rangle.
\end{gathered}$$

**Classification of finite abelian groups.** *Every finite abelian group is isomorphic to a product*

$$\mathbb{Z}_{k(0)} \times \cdots \times \mathbb{Z}_{k(n-1)}$$

*for some $n$ in $\omega$, for some $k(i)$ in $\mathbb{N} \setminus \{1\}$ such that*

$$k(0) \mid k(1) \ \ \& \ \ \cdots \ \ \& \ \ k(n-2) \mid k(n-1).$$

*(If $n = 0$, the product is the trivial product, which is the trivial group.) By the Chinese Remainder Theorem, every finite abelian group is therefore isomorphic to a product*

$$\mathbb{Z}_{q(0)} \times \cdots \times \mathbb{Z}_{q(m-1)}$$

*I did not use this terminology in class, but I stated the theorem because of a question from a student.

*for some $m$ in $\omega$, for some powers $q(j)$ of primes. The latter factorization is unique, and therefore the former is unique.*

*Proof.* We shall prove only the existence of the former factorization, because that is all that we shall need; the rest is an **exercise.** Every finite abelian group $(G,+)$ is $\langle a_0,\dots,a_{n-1}\rangle$, that is,

$$\{a_0x_0+\dots+a_{n-1}x_{n-1}\colon \boldsymbol{x}\in\mathbb{Z}^n\},$$

for some $a_i$ in $G$. Here possibly $G=\{a_0,\dots,a_{n-1}\}$ as a set; but instead we could require $n$ to be minimal such that $G=\langle a_0,\dots,a_{n-1}\rangle$ for some $a_i$ in $G$. In any case, there is an **epimorphism** (a surjective homomorphism)

$$\boldsymbol{x}\mapsto\sum_{i<n}a_ix_i$$

from $\mathbb{Z}^n$ onto $G$. Let $N$ be the kernel of this epimorphism; then by the (first) **isomorphism theorem** (learned in Ali Nesin's class),

$$G\cong\mathbb{Z}^n/N.$$

We can think of the elements of $\mathbb{Z}^n$ as $1\times n$ matrices, which are **row vectors.** Then every *automorphism* (invertible endomorphism, as on page 35) of $\mathbb{Z}^n$ can be understood as $\boldsymbol{x}\mapsto\boldsymbol{x}\cdot Q$ for some $n\times n$ invertible matrix $Q$ over $\mathbb{Z}$. The set of such matrices (or else the automorphisms that they represent) is denoted by

$$\mathrm{GL}_n(\mathbb{Z}),$$

where GL stands for "general linear [group]." Then the function

$$\boldsymbol{x}\cdot Q\mapsto\sum_{i<n}a_ix_i,$$

$$\begin{pmatrix} k(0) & 0 & \dots & 0 \\ 0 & k(1) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & k(n-1) \end{pmatrix}$$

Figure 5.1.: A diagonal matrix

namely $\boldsymbol{x} \mapsto \boldsymbol{x} \cdot Q^{-1}$ followed by $\boldsymbol{x} \mapsto \sum_{i<n} a_i x_i$, is an epimorphism from $\mathbb{Z}^n$ onto $G$ with kernel

$$\{\boldsymbol{x} \cdot Q \colon \boldsymbol{x} \in N\}.$$

Thus $G \cong \mathbb{Z}^n / \{\boldsymbol{x} \cdot Q \colon \boldsymbol{x} \in N\}$. We are going to find $Q$ so that $\{\boldsymbol{x} \cdot Q \colon \boldsymbol{x} \in N\}$ is generated by the rows of an $n \times n$ diagonal matrix as in Figure 5.1; we can write the matrix also as

$$\operatorname{diag}(k(0), \dots, k(n-1)).$$

We shall be able to require further that

$$k(0) \mid \dots \mid k(n-1);$$

and in this case $G$ will be as desired. (Some of the $k(i)$ at the end might be 0.)

The first step is to show that any subgroup of $\mathbb{Z}^n$ is generated by $n$ elements. Suppose $s \geqslant n$, and some subgroup of $\mathbb{Z}^n$ is generated by $s$ elements. Let those $s$ elements by the rows of an $s \times n$ matrix $A$ over $\mathbb{Z}$. Then we can write the group as

$$\langle A \rangle.$$

For all $P$ in $\mathrm{GL}_s(\mathbb{Z})$, the matrix product $PA$ is obtained from $A$ by a sequence of **elementary row operations,** namely

$$\left(\begin{array}{cccc} a_0 & * & \dots & * \\ 0 & a_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \dots & 0 & a_{n-1} \\ \hline 0 & \dots & \dots & 0 \\ \vdots & & & \vdots \\ 0 & \dots & \dots & 0 \end{array}\right)$$

Figure 5.2.: An upper triangular matrix

1) adding a multiple of one row to another,
2) interchanging two rows,
3) multiplying a row by an element of $\mathbb{Z}^\times$, which is $\{\pm 1\}$.

Thus
$$\langle PA \rangle = \langle A \rangle.$$

For some $P$ in $\mathrm{GL}_s(\mathbb{Z})$, $PA$ will be an upper triangular matrix, having the form in Figure 5.2. For example, by elementary row operations, we obtain

$$\begin{pmatrix} 8 & 14 \\ 4 & 13 \\ 6 & 9 \end{pmatrix}, \begin{pmatrix} 0 & -12 \\ 4 & 13 \\ 2 & -4 \end{pmatrix}, \begin{pmatrix} 0 & -12 \\ 0 & 21 \\ 2 & -4 \end{pmatrix},$$
$$\begin{pmatrix} 0 & -12 \\ 0 & -3 \\ 2 & -4 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & -3 \\ 2 & -4 \end{pmatrix}, \begin{pmatrix} 2 & -4 \\ 0 & -3 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 2 & 2 \\ 0 & 3 \\ 0 & 0 \end{pmatrix},$$

and so
$$\langle (8, 14), (4, 13), (6, 9) \rangle = \langle (2, 2), (0, 3) \rangle.$$

The general procedure is the following.

1) If the first column of $A$ is $(a_0 \ \cdots \ a_{n-1})^T$ (that is, the column vector that is the transpose of the row vector $(a_0 \ \ldots \ a_{n-1})$), then, using the Euclidean algorithm, reduce this to

$$(\gcd(a_0, \ldots, a_{n-1}) \ 0 \ \cdots \ 0)^T.$$

2) If the second column of the resulting matrix is $(b_0 \ \cdots \ b_{n-1})^T$, reduce this to

$$(b_0 \ \gcd(b_1, \ldots, b_{n-1}) \ 0 \ \cdots \ 0)^T.$$

3) Treat the third column similarly, and so on.

This shows that any subgroup of $\mathbb{Z}^n$ generated by $s$ elements is generated by $n$ elements. The same goes for subgroups generated by infinitely many elements, since the foregoing procedure applies to $\infty \times n$ matrices.

In the second and final step, we may assume $N$ is generated by the rows of an $n \times n$ matrix $A$. If $P$ and $Q$ are in $\mathrm{GL}_n(\mathbb{Z})$, we say that $PAQ$ is **similar** to $A$. We want to show that $A$ is similar to a diagonal matrix as above. Let $\ell(0)$ be the least positive entry of any matrix similar to $A$. Then

$$A \sim \left(\begin{array}{c|c} \ell(0) & * \\ \hline * & * \end{array}\right) \sim \left(\begin{array}{c|c} \ell(0) & \mathbf{0} \\ \hline \mathbf{0} & A_1 \end{array}\right)$$

since otherwise the first row or column can be given a positive entry that is less than $\ell(0)$. Now let $\ell(1)$ be the least positive entry of any such matrix $A_1$. Then

$$A \sim \left(\begin{array}{cc|c} \ell(0) & 0 & \mathbf{0} \\ 0 & \ell(1) & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & A_2 \end{array}\right),$$

and so on. Thus

$$A \sim \operatorname{diag}(\ell(0), \ell(1), \dots, \ell(n-1)).$$

Finally we obtain $A$ as desired by observing

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \sim \begin{pmatrix} a & b \\ 0 & b \end{pmatrix} \sim \begin{pmatrix} \gcd(a,b) & 0 \\ bx & by \end{pmatrix} \sim \begin{pmatrix} \gcd(a,b) & 0 \\ 0 & by \end{pmatrix},$$

and $\gcd(a,b) \mid by$. □

## Units of finite fields

**Theorem.** *The group of units of every finite field is cyclic.*

*Proof.* Let $K$ be a finite field. Then by the classification of finite abelian groups, $K^\times$ is isomorphic to a product

$$\mathbb{Z}_{k(0)} \times \dots \times \mathbb{Z}_{k(n-1)}$$

for some $n$ in $\omega$, for some $k(i)$ in $\mathbb{N} \setminus \{1\}$ such that

$$k(0) \mid \ldots \mid k(n-1).$$

If $n = 0$, the product is the trivial group, which is cyclic, and so we are done. (The trivial group is the group of units of $\mathbb{F}_2$.) We may henceforth assume $n > 0$. For every $(x_0, \dots, x_{n-1})$ in $\mathbb{Z}_{k(0)} \times \dots \times \mathbb{Z}_{k(n-1)}$, writing $k(n-1)$ as $k$, we have

$$k \cdot (x_0, \dots, x_{n-1}) = (kx_0, \dots, kx_{n-1}) = (0, \dots, 0) = 0,$$

since $k(i) \mid k$. Therefore, for every $a$ in $K^\times$ we have

$$a^k = 1.$$

Thus every element of $K^\times$ is a zero of the polynomial

$$X^k - 1.$$

Over any field, for every $d$ in $\omega$, every polynomial of degree $d$ has at most $d$ zeroes. We prove this by induction. A polynomial of degree 0 is a nonzero constant, so it has 0 zeroes. Suppose $\deg f = d + 1$ and $f(\alpha) = 0$. By division,

$$f = (X - \alpha) \cdot g + h,$$

where $g$ is a polynomial of degree $d$, and $\deg h < 1$, so $h$ is constant. But then

$$0 = f(\alpha) = 0 \cdot g(\alpha) + h = h.$$

Thus $f = (X - \alpha) \cdot g$. Suppose $\beta$ is a zero of $f$ different from $\alpha$. Then

$$(\beta - \alpha) \cdot g(\beta) = 0,$$

so $g(\beta) = 0$ since a field is an integral domain. So the zeroes of $f$, other than $\alpha$, are zeroes of $g$. If $g$ has at most $d$ zeroes, then $f$ must have at most $d + 1$ zeroes. This completes the induction.

We now have

$$\begin{gathered}\left|K^\times\right| \leqslant k = k(n-1) \leqslant k(0) \cdots k(n-1) = \left|K^\times\right|,\\ k(n-1) = k(0) \cdots k(n-1),\end{gathered}$$

and so $n = 1$ (since we assume $k(0) > 1$). Thus

$$K^\times \cong \mathbb{Z}_k = \mathbb{Z}_{|K|-1}. \qquad \square$$

## Existence of finite fields

If $K$ is a finite field of size $p^n$, then we now know that the elements of $K^\times$ are precisely the zeroes of the polynomial $X^{p^n-1}-1$ in any field that includes $K$. Hence the elements of $K$ are precisely the zeroes of the polynomial

$$X^{p^n}-X$$

in any field that includes $K$. This means the elements of $K$ are the fixed points of the function $x \mapsto x^{p^n}$. Now, in $\mathbb{Z}$ we have

$$(a+b)^{p^n} \equiv a^{p^n}+b^{p^n} \pmod{p},$$

since, for example,

$$\begin{aligned}(a+b)^p = a^p + pa^{p-1}b + \binom{p}{2}a^{p-2}b^2 + \dots + pab^{p-1} + b^p \\ \equiv a^p + b^p \pmod{p};\end{aligned}$$

the general result is left as an **exercise.** In a field of characteristic $p$ then, we have

$$(x+y)^{p^n} = x^{p^n}+y^{p^n}, \qquad (xy)^{p^n} = x^{p^n}\cdot y^{p^n}, \qquad 1^{p^n}=1.$$

Thus $x \mapsto x^{p^n}$ is an endomorphism of the field.

For any endomorphism $\varphi$ of a field, the set $\{x\colon \varphi(x)=x\}$ of fixed points is a subfield: this too is an **exercise.** Thus, to show that there is finite field of size $p^n$, we need only show that there is a field of characteristic $p$ that contains $p^n$ zeroes of $X^{p^n}-X$; these zeroes will then constitute the desired subfield. An example of a field containing $p^n$ zeroes of $X^{p^n}-X$ is the *algebraic closure* of $\mathbb{F}_p$.

A field is **algebraically closed** if the only irreducible polynomials over it are **linear** (that is, of degree 1). An example is $\mathbb{C}$, though we are not going to prove this; the result does follow from the fact that $\mathbb{R}$ is **real closed,** that is, every polynomial function that takes both positive and negative values has a zero.

We have seen that if $f$ is irreducible over a field $K$, then $f$ has a zero in $K[X]/(f)$, namely the element called $[X]$. We showed on page 28 that $K[X]/(f)$ is a field if $K$ is finite. In fact it is a field in any case. For if $f \nmid g$, then, since $f$ is irreducible and therefore prime, we have $\gcd(f, g) = 1$, and so, by Bézout's Lemma (in the sense of page 40), or by the Euclidean algorithm, we can find polynomials $h$ and $k$ such that

$$fh + gk = 1.$$

In $K[X]/(f)$ then, $[g][k] = 1$, so $[g]$ is invertible. We could have proved similarly that $\mathbb{F}_p$ is a field.

Starting with $\mathbb{F}_p$, we can obtain a **chain** or **tower**

$$K_0 \subseteq K_1 \subseteq K_2 \subseteq \cdots$$

of fields, where $K_0 = \mathbb{F}_p$ and for each $n$ in $\omega$, for some irreducible $f_n$ over $K_n$,

$$K_{n+1} = K_n[X]/(f_n).$$

Then $\bigcup_{n\in\omega} K_n$ is a field, and if we have chosen the $f_n$ properly, that field will be algebraically closed. It will then be the algebraic closure

$$\mathbb{F}_p{}^{\mathrm{alg}}$$

of $\mathbb{F}_p$, in the sense described below. Inside this field, we can find each $\mathbb{F}_{p^n}$, as above.

Actually we have to show also* that the polynomial $X^{p^n} - X$ has $p^n$ *distinct* zeroes in $\mathbb{F}_p{}^{\mathrm{alg}}$. For example, the polynomial $X^{p^n} - 1$ has only one zero in a field of characteristic $p$, since there the polynomial factorizes as $(X-1)^{p^n}$. If $X^{p^n} - X$ has fewer than $p^n$ distinct zeroes in $\mathbb{F}_p{}^{\mathrm{alg}}$, then for one of those zeroes, say $\alpha$, for some polynomial $g$ of degree $p^n - 2$, we must have

$$X^{p^n} - X = (X - \alpha)^2 \cdot g.$$

In this case, by *formal differentiation,* since we are in a field of characteristic $p$,

$$\begin{aligned} -1 &= 2(X-\alpha) \cdot g + (X-\alpha)^2 \cdot g' \\ &= (X-\alpha) \cdot \big(2g + (X-\alpha) \cdot g'\big), \end{aligned} \tag{5.1}$$

which is absurd. Here **formal differentiation** over a field $K$ is the endomorphism $f \mapsto f'$ of $K[X]$ as a vector space over $K$ given by

$$\left(\sum_{i=0}^{n} a_i X^i\right)' = \sum_{i=1}^{n} i a_i X^{i-1}.$$

One shows that

$$(fg)' = f'g + fg', \tag{5.2}$$

which is what we used to establish (5.1). One can prove (5.2) by induction on $\deg g$, or directly from the equation

$$\left(\sum_i a_i X^i\right) \cdot \sum_j b_j X^j = \sum_k \left(\sum_{i+j=k} a_i b_j\right) X^k$$

*I did not do this in class.

## Uniqueness of finite fields

We establish the uniqueness of $\mathbb{F}_{p^n}$ as follows. If $f$ is irreducible over a field $K$, but has a zero $\alpha$ in some larger field $L$, we let

$$K[\alpha]$$

denote the smallest sub-ring of $L$ that includes $K \cup \{\alpha\}$. In fact this sub-ring will be $\{g(\alpha)\colon g \in K[X]\}$, which is the image of the ring $K[X]$ under the homomorphism $g \mapsto g(\alpha)$. Also the kernel of this homomorphism is $(f)$, that is, $\{g \in K[X]\colon f \mid g\}$. Thus, by the isomorphism theorem,

$$K[X]/(f) \cong K[\alpha].$$

Hence $K[\alpha]$ is a field. Using this, one shows that the algebraically closed field $\bigcup_{n\in\omega} K_n$ constructed above **embeds** in every algebraically closed field that includes $\mathbb{F}_p$. (An embedding is a **monomorphism,** that is, an injective homomorphism.) This is the sense in which the field $\bigcup_{n\in\omega} K_n$ is the **algebraic closure** of $\mathbb{F}_p$. There is only one algebraic closure, up to isomorphism, called ${\mathbb{F}_p}^{\mathrm{alg}}$ as above; its uniqueness proves the uniqueness of each $\mathbb{F}_{p^n}$.

# 6. Sunday, January 24

## Endomorphisms of groups and vector spaces

We said yesterday that the automorphisms of $\mathbb{Z}^n$ are the functions

$$\boldsymbol{x} \mapsto \boldsymbol{x} \cdot Q,$$

where $\boldsymbol{x}$ is an element of $\mathbb{Z}^n$ considered as a row vector, and $Q$ is an invertible $n \times n$ matrix, that is,

$$Q \in \mathrm{GL}_n(\mathbb{Z}).$$

Why is this? If $i < n$, let

$$\boldsymbol{e}^i = (e^i_0, \dots, e^i_{n-1}), \qquad \text{where} \qquad e^i_j = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

Then

$$\boldsymbol{x} = \sum_{i<n} x_i \boldsymbol{e}^i$$

for all $\boldsymbol{x}$ in $\mathbb{Z}^n$. Thus if $\varphi$ in $\mathrm{End}(\mathbb{Z}^n)$, then

$$\varphi(\boldsymbol{x}) = \sum_{i<n} \varphi(x_i \boldsymbol{e}^i) = \sum_{i<n} x_i \cdot \varphi(\boldsymbol{e}^i) = \boldsymbol{x} \cdot \begin{pmatrix} \varphi(\boldsymbol{e}^0) \\ \hline \vdots \\ \hline \varphi(\boldsymbol{e}^{n-1}) \end{pmatrix}.$$

If $\varphi \in \mathrm{Aut}(\mathbb{Z}^n)$, this means the $n \times n$ matrix whose rows are the $\varphi(\boldsymbol{e}^i)$ must be invertible.

Here we have been considering $\mathbb{Z}^n$ as an abelian group, not a ring: as $(\mathbb{Z}^n, +)$, not $(\mathbb{Z}^n, +, \cdot, 1)$.

We know that $\mathbb{R}^n$ is a vector space over $\mathbb{R}$, which means there is a homomorphism from $\mathbb{R}$ to $\operatorname{End}(\mathbb{R}^n, +)$. As a vector space over $\mathbb{R}$, $\mathbb{R}^n$ has the endomorphism ring

$$\mathrm{M}_n(\mathbb{R}),$$

namely the ring of $n \times n$ matrices over $\mathbb{R}$, by the same reasoning as for $\mathbb{Z}^n$. However, as a group, $\mathbb{R}^n$ has a larger endomorphism ring, unless $n = 0$. For, as a group, $\mathbb{R}$ is a vector space over $\mathbb{Q}$, in the sense that

$$\frac{k}{m} \cdot x = y \iff k \cdot x = m \cdot y$$

for all $x$ and $y$ in $\mathbb{R}$, all $k$ in $\mathbb{Z}$, and all $m$ in $\mathbb{N}$. But $\mathbb{R}$ then has a basis over $\mathbb{Q}$ (by the Axiom of Choice, as discussed in Ali Nesin's class), in fact an uncountable basis of size $|\mathbb{R}|$; and then $\operatorname{End}(\mathbb{R}, +, \mathbb{Q})$ will have size $\left|2^{\mathbb{R}}\right|$.

## The characteristic polynomial of a matrix

$\mathbb{R}$ embeds in $\mathrm{M}_n(\mathbb{R})$ under

$$x \mapsto x \cdot \mathrm{I}_n,$$

where $\mathrm{I}_n$ is the $n \times n$ identity matrix, that is,

$$\mathrm{I}_n = \begin{pmatrix} \boldsymbol{e}^0 \\ \hline \vdots \\ \hline \boldsymbol{e}^{n-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{pmatrix}.$$

The same is true when $\mathbb{R}$ is replaced with an arbitrary field $K$. Let $A \in \mathrm{M}_n(K)$. It may be that for some nonzero $\boldsymbol{v}$ in $K^n$, for some $\lambda$ in $K$,

$$\boldsymbol{v} \cdot A = \lambda \cdot \boldsymbol{v}.$$

This means

$$\mathbf{0} = \boldsymbol{v} \cdot (\lambda \cdot \mathrm{I}_n - A).$$

Since we assume $\boldsymbol{v} \neq \mathbf{0}$, the matrix $\lambda \cdot \mathrm{I}_n - A$ must not be invertible, and so we must have $\det(\lambda \cdot \mathrm{I}_n - A) = 0$. The polynomial

$$\det(X \cdot \mathrm{I}_n - A)$$

in $X$ of degree $n$ over $K$ is the **characteristic polynomial** of $A$. Its zeroes are the **characteristic values** or **eigenvalues** of $A$. If $\lambda$ is an eigenvalue of $A$, then the solution space of the equation

$$\mathbf{0} = \boldsymbol{x} \cdot (\lambda \cdot \mathrm{I}_n - A)$$

is the **eigenspace** associated with $\lambda$, and its nonzero elements are the **eigenvectors** (or **characteristic vectors**) associated with $\lambda$.

## Diagonalizable matrices

For example, if

$$A = \begin{pmatrix} -1 & -5 \\ -5 & -1 \end{pmatrix},$$

then

$$\begin{aligned} \det(X \cdot \mathrm{I}_2 - A) &= \det \begin{pmatrix} X+1 & 5 \\ 5 & X+1 \end{pmatrix} \\ &= (X+1)^2 - 25 = X^2 + 2X - 24 = (X+6)(X-4), \end{aligned}$$

so the eigenvalues of $A$ are $-6$ and $4$. The corresponding eigenspaces are as follows.

- $\lambda = -6$:

$$\begin{aligned}
\mathbf{0} = \boldsymbol{x} \cdot \begin{pmatrix} -5 & 5 \\ 5 & -5 \end{pmatrix} &\iff \mathbf{0} = \boldsymbol{x} \cdot \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix} \\
&\iff (0,0) = (x_0 - x_1, 0) \\
&\iff x_0 = x_1 \\
&\iff \boldsymbol{x} = t \cdot (1,1) \text{ for some } t \text{ in } K.
\end{aligned}$$

- $\lambda = 4$:

$$\begin{aligned}
\mathbf{0} = \boldsymbol{x} \cdot \begin{pmatrix} 5 & 5 \\ 5 & 5 \end{pmatrix} &\iff \mathbf{0} = \boldsymbol{x} \cdot \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \\
&\iff x_0 = -x_1 \\
&\iff \boldsymbol{x} = t \cdot (1,-1) \text{ for some } t \text{ in } K.
\end{aligned}$$

Letting the eigenvectors $(1,1)$ and $(1,-1)$ be the rows of a matrix, we obtain

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \cdot A = \begin{pmatrix} -6 & -6 \\ 4 & -4 \end{pmatrix} = \begin{pmatrix} -6 & 0 \\ 0 & 4 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

$$\begin{pmatrix} -6 & 0 \\ 0 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \cdot A \cdot \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^{-1}.$$

In general, if $A$ is an $n \times n$ matrix with $n$ distinct eigenvalues $\lambda_0, \ldots, \lambda_{n-1}$, and $\boldsymbol{v}_0, \ldots, \boldsymbol{v}_{n-1}$ are corresponding eigenvectors, then

$$PA = \operatorname{diag}(\lambda_0, \ldots, \lambda_{n-1}) \cdot P,$$

where

$$P = \left(\begin{array}{c} \boldsymbol{v}_0 \\ \hline \vdots \\ \hline \boldsymbol{v}_{n-1} \end{array}\right).$$

We can conclude that $PAP^{-1}$ is a diagonal matrix whose diagonal entries are the eigenvalues of $A$, provided $A$ is invertible. It is invertible by the following.

**Theorem.** *Eigenvectors corresponding to distinct eigenvalues of any matrix are linearly independent.*

*Proof.* We prove the claim by induction on the number of eigenvectors. The empty set of eigenvectors is trivially linearly independent. Suppose $\boldsymbol{v}_i$ is the eigenvector corresponding to $\lambda_i$, and

$$\sum_{i<n+1} x^i \boldsymbol{v}_i = \mathbf{0}. \tag{6.1}$$

Multiplying by the matrix gives

$$\sum_{i<n+1} \lambda_i x^i \boldsymbol{v}_i = \mathbf{0}.$$

But just multiplying by $\lambda_n$ gives

$$\sum_{i<n+1} \lambda_n x^i \boldsymbol{v}_i = \mathbf{0}.$$

By subtraction, one term is cancelled, and we have

$$\sum_{i<n} (\lambda_n - \lambda_i) x^i \boldsymbol{v}_i = \mathbf{0}.$$

If $n$ eigenvectors corresponding to distinct eigenvalues must be linearly independent, then we can conclude $(\lambda_n - \lambda_i)x^i = 0$ when $i < n$, so $x^i = 0$ when $i < n$, and then also $x^n = 0$ by (6.1); so $n + 1$ eigenvectors corresponding to distinct eigenvalues must be linearly independent. □

## A nondiagonalizable matrix

Possibly an $n \times n$ matrix does not have $n$ distinct eigenvectors. Still, if the sum of the dimensions of the eigenspaces is $n$, $A$ will be **diagonalizable,** just as in the case where there are $n$ distinct eigenvectors.

However, not every square matrix is diagonalizable, even over an algebraically closed field. Suppose for example

$$A = \begin{pmatrix} 8 & 4 \\ -9 & -4 \end{pmatrix}.$$

Then

$$\begin{aligned} \det(X \cdot \mathrm{I}_2 - A) &= \det \begin{pmatrix} X-8 & -4 \\ 9 & X+4 \end{pmatrix} \\ &= (X-8)(X+4) + 36 = X^2 - 4X + 4 = (X-2)^2, \end{aligned}$$

so $A$ has the unique eigenvalue 2. Moreover,

$$\begin{aligned} \mathbf{0} = \boldsymbol{x} \cdot (2\,\mathrm{I}_2 - A) &\iff \mathbf{0} = \boldsymbol{x} \cdot \begin{pmatrix} -6 & -4 \\ 9 & 6 \end{pmatrix} \\ &\iff \mathbf{0} = \boldsymbol{x} \cdot \begin{pmatrix} 2 & 2 \\ -3 & -3 \end{pmatrix} \\ &\iff \mathbf{0} = \boldsymbol{x} \cdot \begin{pmatrix} 2 & 0 \\ -3 & 0 \end{pmatrix} \\ &\iff 2x_0 = 3x_1 \\ &\iff \boldsymbol{x} = t \cdot (3,2) \text{ for some } t \text{ in } K, \end{aligned}$$

so the unique eigenspace is one-dimensional, spanned by $(3,2)$.

But then

$$
\begin{aligned}
& (3,2) = \boldsymbol{x} \cdot (2\,\mathrm{I}_2 - A) \\
\iff\ & (3,2) = \boldsymbol{x} \cdot \begin{pmatrix} -6 & -4 \\ 9 & 6 \end{pmatrix} \\
\iff\ & (-1,-1) = \boldsymbol{x} \cdot \begin{pmatrix} 2 & 2 \\ -3 & -3 \end{pmatrix} \\
\iff\ & (-1,0) = \boldsymbol{x} \cdot \begin{pmatrix} 2 & 0 \\ -3 & 0 \end{pmatrix} \\
\iff\ & -1 = 2x_0 - 3x_1 \\
\iff\ & \boldsymbol{x} = t \cdot (3,2) - (\tfrac{1}{2}, 0) \text{ for some } t \text{ in } K.
\end{aligned}
$$

One solution here is $(1,1)$. Call this $\boldsymbol{c}$, and let $(3,2) = \boldsymbol{b}$. Then

$$
\mathbf{0} = 2\boldsymbol{b} - \boldsymbol{b} \cdot A, \qquad\qquad \boldsymbol{b} = 2\boldsymbol{c} - \boldsymbol{c} \cdot A,
$$

that is,

$$
\boldsymbol{b} \cdot A = 2\boldsymbol{b}, \qquad\qquad \boldsymbol{c} \cdot A = 2\boldsymbol{c} - \boldsymbol{b},
$$

and so

$$
\begin{gathered}
\begin{pmatrix} \boldsymbol{c} \\ \hline \boldsymbol{b} \end{pmatrix} \cdot A = \begin{pmatrix} 2\boldsymbol{c} - \boldsymbol{b} \\ \hline 2\boldsymbol{b} \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} \boldsymbol{c} \\ \hline \boldsymbol{b} \end{pmatrix}, \\
\begin{pmatrix} \boldsymbol{c} \\ \hline \boldsymbol{b} \end{pmatrix} \cdot A \cdot \begin{pmatrix} \boldsymbol{c} \\ \hline \boldsymbol{b} \end{pmatrix}^{-1} = \begin{pmatrix} 2 & -1 \\ 0 & 2 \end{pmatrix}.
\end{gathered}
$$

We also have

$$
\begin{pmatrix} -\boldsymbol{c} \\ \hline \boldsymbol{b} \end{pmatrix} \cdot A \cdot \begin{pmatrix} -\boldsymbol{c} \\ \hline \boldsymbol{b} \end{pmatrix}^{-1} = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}.
$$

The general claim, which we are not proving, is that, for every square matrix $A$ over an algebraically closed field, for some invertible $P$,

$$PAP^{-1} = \operatorname{diag}(B_0, \ldots, B_{k-1}),$$

where each $B_i$ is not necessarily a number, but a square matrix, of the form

$$\begin{pmatrix} \lambda & 1 & 0 & \ldots & 0 \\ 0 & \lambda & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & 1 \\ 0 & \ldots & \ldots & 0 & \lambda \end{pmatrix},$$

where $\lambda$ is an eigenvalue of $A$. The matrix $\operatorname{diag}(B_0, \ldots, B_{k-1})$ is the **Jordan normal form** of $A$. Note that the same eigenvalue may appear in more than one of the matrices $B_i$.

# A. Jordan Normal Form

We are going to show that every square matrix over an algebraically closed field is "almost" diagonalizable, in the sense of having the Jordan normal form described above. We shall show moreover that the procedure used above for finding a Jordan normal form works generally. The presentation is based mainly on that of Lang [9].

## Polynomial functions of matrices

Although $K$ is a field, the ring $\mathrm{M}_n(K)$ is not commutative. However, it has commutative sub-rings. Indeed, for every $A$ in $\mathrm{M}_n(K)$, there is smallest sub-ring of $\mathrm{M}_n(K)$ that contains $A$. We may denote this sub-ring by

$$K[A].$$

This ring is commutative. As a vector space over $K$, $K[A]$ is spanned by $\mathrm{I}_n$ and the positive powers of $A$. Indeed, for any $f$ in $K[X]$, for any $A$ in $\mathrm{M}_n(K)$, there is a well-defined matrix $f(A)$ in $\mathrm{M}_n(K)$. In particular, if $f = \sum_{i=0}^{n} b_i X^i$, then

$$f(A) = \sum_{i=0}^{n} b_i A^i.$$

Then

$$K[A] = \big\{f(A) \colon f \in K[X]\big\}.$$

If $f(A)$ is the zero matrix, we may say $A$ is a **zero** of $f$. Note however that $K[A]$ may have zero divisors. Indeed, $A$ itself may be a zero divisor, since for example

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}^2 = 0$$

(where the last 0 is the zero matrix).

In proving that the group of units of every finite field is cyclic, we showed that a polynomial has no more zeros than its degree; but these zeroes belong to a *field* over which the polynomial is defined. If $f \in K[X]$, and $K$ is a subfield of a field $L$, then the number of zeroes of $f$ in $L$ is no greater than $\deg f$.

We know that $K$ embeds in $\mathrm{M}_n(K)$ under $x \mapsto x \cdot \mathrm{I}_n$; but a polynomial over $K$ can have any number of zeroes in $\mathrm{M}_n(K)$ or even in a sub-ring $K[A]$. Indeed, if $A = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$ as above, then $\beta \cdot A$ belongs to $K[A]$ for all $\beta$ in $K$, and $\beta \cdot A$ is a zero of $X^2$.

## Cayley–Hamilton Theorem

Given a square matrix over some field, we are going to want to know that the matrix is a zero of some nonzero polynomial. The following gives us this.

**Theorem** (Cayley–Hamilton)**.** *Over any field, every matrix is a zero of its characteristic polynomial.*

*First proof.* Let $K$ be a field, let $A \in \mathrm{M}_n(K)$, and let

$$f = \det(X \cdot \mathrm{I}_n - A), \tag{A.1}$$

$$\begin{pmatrix} a_0^0 & 0 & \ldots\ldots\ldots & & 0 \\ a_0^1 & a_1^1 & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & 0 \\ a_0^{n-1} & \ldots\ldots\ldots & & & a_{n-1}^{n-1} \end{pmatrix}$$

Figure A.1.: A lower triangular matrix

the characteristic polynomial of $A$. We want to show $f(A) = 0$. It so happens that $\det(A \cdot \mathrm{I}_n - A) = \det(0) = 0$. Thus we want to show

$$f(A) = \det(A \cdot \mathrm{I}_n - A).$$

This does *not* follow immediately from (A.1) by replacing $X$ with the matrix $A$, because the multiplications symbolized in the terms $X \cdot \mathrm{I}_n$ and $A \cdot \mathrm{I}_n$ are two different operations, namely scalar multiplication and matrix multiplication respectively.

Since the determinant function is multiplicative, for every $P$ in $\mathrm{GL}_n(\mathbb{Z})$,

$$\begin{aligned} \det(X \cdot \mathrm{I}_n - A) &= \det\bigl(P \cdot (X \cdot \mathrm{I}_n - A) \cdot P^{-1}\bigr) \\ &= \det(X \cdot \mathrm{I}_n - PAP^{-1}). \end{aligned}$$

For some $P$, $PAP^{-1}$ is a **lower triangular matrix,** as we shall show presently, so that we may assume $A$ is as in Figure A.1. The characteristic polynomial of $A$ is now

$$\prod_{i<n} (X - a_i^i).$$

We have to show

$$\prod_{i<n} (A - a_i^i \, \mathrm{I}_n) = 0.$$

The product here is independent of the order of the factors. If $j < n$, then row $j$ of the product is

$$\boldsymbol{e}^j \cdot \prod_{i<n}(A - a^i_i \,\mathrm{I}_n).$$

However, since $A$ is the matrix in Figure A.1, we have

$$\boldsymbol{e}^j \cdot (A - a^j_j \,\mathrm{I}_n) = \boldsymbol{e}^j \cdot A - a^j_j \boldsymbol{e}^j = \sum_{i<j} a^j_i \boldsymbol{e}^i.$$

By induction then,

$$\boldsymbol{e}^j \cdot \prod_{i \leqslant j}(A - a^i_i \,\mathrm{I}_n) = \boldsymbol{0};$$

for if this is true when $j < k$, then

$$\boldsymbol{e}^k \cdot \prod_{i \leqslant k}(A - a^i_i \,\mathrm{I}_n) = \sum_{i<k} a^k_i \boldsymbol{e}^i \cdot \prod_{i<k}(A - a^i_i \,\mathrm{I}_n) = \boldsymbol{0}.$$

It remains to find, for arbitrary $B$ in $\mathrm{M}_n(K)$, an invertible matrix $P$ such that $PBP^{-1}$ is lower triangular. We shall use induction. Replacing $K$ with its algebraic closure if necessary, we know there is a basis $(\boldsymbol{v}^i \colon i < n)$ of $K^n$ such that $\boldsymbol{v}^0$ is an eigenvector. Let $C$ and $D$ in $\mathrm{M}_n(K)$ be such that

$$\sum_{i<n} x_i \boldsymbol{v}^i \cdot C = x_0 \boldsymbol{v}^0, \qquad \sum_{i<n} x_i \boldsymbol{v}^i \cdot D = \sum_{0<i<n} x_i \boldsymbol{v}^i.$$

Then $\boldsymbol{x} \mapsto \boldsymbol{x} \cdot BC$ is a homomorphism from $K^n$ to $\mathrm{span}(\boldsymbol{v}^0)$, and $\boldsymbol{x} \mapsto \boldsymbol{x} \cdot BD$ is an endomorphism of $\mathrm{span}(\boldsymbol{v}^i \colon 0 < i < n)$, and

$$\boldsymbol{x} \cdot B = \boldsymbol{x} \cdot BC + \boldsymbol{x} \cdot BD. \tag{A.2}$$

As an inductive hypothesis, we assume that $\operatorname{span}(\boldsymbol{v}^i \colon 0 < i < n)$ has a basis $(\boldsymbol{w}^i \colon 0 < i < n)$ such that

$$\boldsymbol{w}^j \cdot BD \in \operatorname{span}(\boldsymbol{w}^i \colon 0 < i \leqslant j)$$

whenever $0 < j < n$. Then $(\boldsymbol{v}^0, \boldsymbol{w}^1, \dots, \boldsymbol{w}^{n-1})$ is a basis of $K^n$, and

$$\boldsymbol{v}^0 \cdot B \in \operatorname{span}(\boldsymbol{v}^0)$$

since $\boldsymbol{v}^0$ is an eigenvector of $B$, and by (A.2),

$$\boldsymbol{w}^j \cdot B \in \operatorname{span}(\boldsymbol{v}^0, \boldsymbol{w}^0, \dots, \boldsymbol{w}^j)$$

whenever $0 < j < n$. This completes the induction. If we now define $a_i^j$ by

$$\boldsymbol{v}^0 \cdot B = a_0^0 \boldsymbol{v}^0, \qquad \boldsymbol{w}^j \cdot B = a_0^j \boldsymbol{v}^0 + a_1^j \boldsymbol{w}^1 + \dots + a_j^j \boldsymbol{w}^j,$$

and we let $P$ be the matrix whose $n$ rows are $(\boldsymbol{v}^0, \boldsymbol{w}^1, \dots, \boldsymbol{w}^{n-1})$, then, with $A$ as in Figure A.1, we have $PB = AP$, so $P$ is as desired.

□

*Second proof.* For every $B$ in $\mathrm{M}_n(K)$, there is a matrix in $\mathrm{M}_n(K)$ called the **adjoint** of $B$, and denoted by $\operatorname{adj}(B)$, such that

$$\det B \cdot \mathrm{I}_n = B \cdot \operatorname{adj}(B). \tag{A.3}$$

This is all we need to know about the adjoint; but in fact, if

$$B = (b_j^i)_{j \in n}^{i \in n},$$

so that

$$\det B = \sum_{\sigma \in \operatorname{Sym}(n)} \operatorname{sgn}(\sigma) \cdot \prod_{i < n} b_{\sigma(i)}^i,$$

then

$$\operatorname{adj}(B) = \Bigl((-1)^{i+j} \cdot \det\bigl((b^k_\ell)^{k \in n \smallsetminus \{i\}}_{\ell \in n \smallsetminus \{j\}}\bigr)\Bigr)^{j \in n}_{i \in n}.$$

From (A.3), if we let $B = X \cdot \mathrm{I}_n - A$, then we have

$$\det(X \cdot \mathrm{I}_n - A) \cdot \mathrm{I}_n = (X \cdot \mathrm{I}_n - A) \cdot \operatorname{adj}(B). \tag{A.4}$$

The two members of this equation are elements of $\mathrm{M}_n(K[X])$; that is, they are matrices with polynomial entries. We can write out the left-hand member as a sum

$$X^m \cdot C_m + X^{m-1} \cdot C_{m-1} + \cdots + X \cdot C_1 + C_0, \tag{A.5}$$

where $C_i \in \mathrm{M}_n(K)$ and, as it happens, $m = n$. We are going to perform a variant of the division algorithm (page 40). Again, the sum in (A.5) is equal to either member of (A.4). if we abbreviate the sum by $C$, then we have

$$C = (X \cdot \mathrm{I}_n - A) \cdot X^{m-1} \cdot C_m + D, \tag{A.6}$$

where $D$ stands for

$$X^{m-1} \cdot (A \cdot C_m + C_{m-1}) + X^{m-2} \cdot C_{m-2} + \cdots + X \cdot C_1 + C_0.$$

If we replace $X$ with $A$, either in $C$ or $D$, by (A.6) we get the same result. By induction, for some $Q$ in $\mathrm{M}_n(K[X])$ and $R$ in $\mathrm{M}_n(K)$,

$$C = (X \cdot \mathrm{I}_n - A) \cdot Q + R,$$

where $R$ is the result of replacing $X$ in $C$ with $A$. Moreover, $Q$ and $R$ are unique. Comparing with (A.4), we conclude that $Q = \operatorname{adj}(B)$ and, what is the point for us, $R = 0$. □

## Direct sums

Suppose $V$ is a vector space over $K$, and $n \in \mathbb{N}$, and for each $j$ in $n$, $V_j$ is a subspace of $V$. If the homomorphism

$$(v_i : i < n) \mapsto \sum_{i<n} v_i$$

from $\prod_{i<n} V_i$ to $V$ is surjective, then $V$ is the **sum** of the subspaces $V_i$, and we may write

$$V = V_0 + \cdots + V_{n-1} = \sum_{i<n} V_i.$$

If, further, the homomorphism is injective, then $V$ is the **direct sum** of the $V_i$, and we may write

$$V = V_0 \oplus \cdots \oplus V_{n-1} = \bigoplus_{i<n} V_i.$$

Given $A$ in $\mathrm{M}_n(K)$, we shall understand by

$$\ker A$$

the kernel of the endomorphism $\boldsymbol{x} \mapsto \boldsymbol{x} \cdot A$ of $K^n$. Suppose $A$ has characteristic polynomial $f$, and $K$ is algebraically closed. We may suppose that the zeroes of $f$ in $K$ are given to us in some order, as the entries of a list $(\lambda_j : j < m)$. Then

$$f = \prod_{j<m} (X - \lambda_j)^{r_j}$$

for some $r_j$ in $\mathbb{N}$. Now, $\ker\bigl(f(A)\bigr) = K^n$ by the Cayley–Hamilton Theorem. We are going to show

$$K^n = \bigoplus_{j<m} \ker\left({B_j}^{r_j}\right), \tag{A.7}$$

where

$$B_j = A - \lambda_j \cdot \mathrm{I}_n\,.$$

The result is easier to prove if stated more generally, as follows.

**Theorem.** *If $m \in \mathbb{N}$, and $(f_i \colon i < m) \in K[X]^m$, and each $f_i$ is prime to each of the others, then for all $A$ in $\mathrm{M}_n(K)$,*

$$\ker\left(\prod_{i<m} f_i(A)\right) = \bigoplus_{i<m} \ker(f_i(A)). \tag{A.8}$$

*Proof.* The case when $m = 1$ is trivial. Suppose $m = 2$. By Bézout's Lemma (in the sense of page 40), for some $g_0$ and $g_1$ in $K[X]$,

$$f_0 \cdot g_0 + f_1 \cdot g_1 = 1.$$

Therefore

$$f_0(A) \cdot g_0(A) + f_1(A) \cdot g_1(A) = \mathrm{I}_n, \tag{A.9}$$

and so, for all $\boldsymbol{v}$ in $K^n$,

$$\boldsymbol{v} = \boldsymbol{v} \cdot f_0(A) \cdot g_0(A) + \boldsymbol{v} \cdot f_1(A) \cdot g_1(A). \tag{A.10}$$

Now, if we should have

$$\boldsymbol{v} \in \ker(f_0(A) \cdot f_1(A)),$$

then we can conclude

$$\boldsymbol{v} \cdot f_0(A) \in \ker(f_1(A)), \qquad \boldsymbol{v} \cdot f_1(A) \in \ker(f_0(A)).$$

Comparing with (A.10) shows

$$\ker(f_0(A) \cdot f_1(A)) = \ker(f_1(A)) + \ker(f_0(A)).$$

Now, suppose

$$\boldsymbol{v} = \boldsymbol{u} + \boldsymbol{w},$$

where

$$\boldsymbol{u} \in \ker(f_1(A)), \qquad \boldsymbol{w} \in \ker(f_0(A)).$$

Then

$$\boldsymbol{v} \cdot f_0(A) \cdot g_0(A) = \boldsymbol{u} \cdot f_0(A) \cdot g_0(A);$$

but also, applying (A.9) to $\boldsymbol{u}$ yields

$$\boldsymbol{u} = \boldsymbol{u} \cdot f_0(A) \cdot g_0(A),$$

and so

$$\boldsymbol{v} \cdot f_0(A) \cdot g_0(A) = \boldsymbol{u}.$$

This is enough to establish (A.8) in case $m = 2$. The general case follows by induction. Indeed, under the hypothesis that each $f_i$ is prime to each of the others, it must be prime to the product of the others. In particular, if $m \geqslant 3$ we have $\gcd\left(\prod_{i<m-1} f_i, f_{m-1}\right) = 1$, so

$$\ker\left(\prod_{i<m} f_i(A)\right) = \ker\left(\prod_{i<m-1} f_i(A)\right) \oplus \ker(f_{m-1}(A)),$$

and the inductive hypothesis will take care of $\prod_{i<m-1} f_i(A)$. □

## Cyclic spaces

Assuming again $A \in \mathrm{M}_n(K)$, let $\boldsymbol{v}_0$ be an eigenvector of $A$ corresponding to an eigenvalue $\lambda$. If we let

$$B_\lambda = A - \lambda \cdot \mathrm{I}_n, \tag{A.11}$$

then $\boldsymbol{v}_0 \in \ker B_\lambda \smallsetminus \{\mathbf{0}\}$. We continue recursively. Given a vector $\boldsymbol{v}_i$ in $K^n$, if possible we let $\boldsymbol{v}_{i+1}$ be a solution of

$$\boldsymbol{v}_i = \boldsymbol{x} \cdot B_\lambda.$$

Then by (A.11),

$$\lambda \cdot \boldsymbol{v}_{i+1} + \boldsymbol{v}_i = \boldsymbol{v}_{i+1} \cdot A,$$

but also, by induction,

$$\boldsymbol{v}_i \in \ker({B_\lambda}^{i+1}).$$

If $\boldsymbol{v}_{r-1}$ exists, let

$$P = \begin{pmatrix} \boldsymbol{v}_{r-1} \\ \hline \vdots \\ \hline \boldsymbol{v}_1 \\ \hline \boldsymbol{v}_0 \end{pmatrix}.$$

Then

$$PA = \begin{pmatrix} \lambda\boldsymbol{v}_{r-1} + \boldsymbol{v}_{r-2} \\ \hline \vdots \\ \hline \lambda\boldsymbol{v}_1 + \boldsymbol{v}_0 \\ \hline \lambda\boldsymbol{v}_0 \end{pmatrix} = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \lambda & 1 \\ 0 & \dots & \dots & 0 & \lambda \end{pmatrix} P. \qquad \text{(A.12)}$$

We show next that the rows of $P$ are linearly independent, at least under an additional assumption, which will turn out later always to hold.

**Theorem.** *Suppose* $\boldsymbol{v} \in K^n$, $B \in \mathrm{M}_n(K)$, *and*

$$\mathbf{0} = \boldsymbol{v} \cdot B^s \qquad \text{(A.13)}$$

*for some $s$ in $\mathbb{N}$. If $r$ is the* least *such $s$, then the $r$-tuple*

$$(\boldsymbol{v}, \boldsymbol{v} \cdot B, \dots, \boldsymbol{v} \cdot B^{r-1})$$

*is linearly independent over $K$.*

*Proof.* Suppose (A.13) holds, but the $s$-tuple $(\boldsymbol{v} \cdot B^i : i < s)$ is linearly dependent over $K$. This means

$$\mathbf{0} = c_0\boldsymbol{v} + c_1\boldsymbol{v} \cdot B + \dots + c_{s-1}\boldsymbol{v} \cdot B^{s-1} \tag{A.14}$$

for some $\boldsymbol{c}$ in $K^s \smallsetminus \{\mathbf{0}\}$. We can write (A.14) as

$$\mathbf{0} = \boldsymbol{v} \cdot f(B),$$

where

$$f = c_0 + c_1X + \dots + c_{s-1}X^{s-1}.$$

Let $g = \gcd(X^s, f)$. By Bézout's Lemma,

$$\mathbf{0} = \boldsymbol{v} \cdot g(B).$$

But $g$ must be $X^r$ for some $r$ that is less than $s$. □

If $B \in \mathrm{M}_n(K)$, and $V$ is a subspace of $K^n$, we let $VB$ be the image of $V$ under the endomorphism $\boldsymbol{x} \mapsto \boldsymbol{x} \cdot B$ of $K^n$. Thus

$$VB = \{\boldsymbol{x} \cdot B : x \in V\}.$$

If $VB \subseteq V$, let us say $V$ is **$B$-invariant.** For example, $V$ is $B$-invariant if it is spanned by $\{\boldsymbol{v} \cdot B^k : k \in \omega\}$ for some $\boldsymbol{v}$ in $V \smallsetminus \{\mathbf{0}\}$. If $V$ is thus, and also $\boldsymbol{v} \cdot B^s = \mathbf{0}$ for some $s$ in $\mathbb{N}$, let us say that $V$ is **$B$-cyclic.** If $r$ is the least $s$, then by the last theorem, $(\boldsymbol{v} \cdot B^i : i < r)$ is a basis of the $B$-cyclic space.

**Theorem.** *For all $B$ in $\mathrm{M}_n(K)$, for all $s$ in $\mathbb{N}$, $\ker(B^s)$ is the direct sum of $B$-cyclic subspaces.*

*Proof.* We shall prove that every $B$-invariant subspace of $\ker(B^s)$ is the direct sum of $B$-cyclic subspaces. We use induction on the dimension of the subspace. If the dimension is 0, the claim is vacuously true. Suppose $V$ is a $B$-invariant subspace of $\ker(B^s)$ having positive dimension. Then

$$VB \subseteq V \cap \ker(B^{s-1}) \subseteq V.$$

If $r$ is the *least* $s$ for which $V \subseteq \ker(B^s)$, then $V \cap \ker(B^{r-1}) \subset V$. This shows

$$VB \subset V.$$

As an inductive hypothesis, we assume

$$VB = \bigoplus_{i<m} W_i, \tag{A.15}$$

where each $W_i$ is $B$-cyclic. Then for some $(\boldsymbol{w}_i \colon i < m)$ in $V^m$, for some $(r_i \colon i < m)$ in $\mathbb{N}^m$,

$$W_i = \operatorname{span}(\boldsymbol{w}_i, \boldsymbol{w}_i \cdot B, \dots, \boldsymbol{w}_i \cdot B^{r_i-1}), \quad \boldsymbol{0} = \boldsymbol{w}_i \cdot B^{r_i}. \tag{A.16}$$

For some $\boldsymbol{v}_i$ in $V$,

$$\boldsymbol{w}_i = \boldsymbol{v}_i \cdot B. \tag{A.17}$$

By the last theorem, we know that the tuple

$$(\boldsymbol{v}_i, \boldsymbol{w}_i, \boldsymbol{w}_i \cdot B, \dots, \boldsymbol{w}_i \cdot B^{r_i-1}),$$

which is

$$(\boldsymbol{v}_i, \boldsymbol{v}_i \cdot B, \dots, \boldsymbol{v}_i \cdot B^{r_i}),$$

is linearly independent. Let $V_i$ be the $B$-cyclic space that it is a basis of. We shall show that the sum of the $V_i$ is direct.

An arbitrary element of $V_i$ is $\boldsymbol{v}_i \cdot f_i(B)$ for some $f_i$ in $K[X]$ such that

$$\deg f_i \leqslant r_i. \tag{A.18}$$

Suppose

$$\mathbf{0} = \sum_{i<m} \boldsymbol{v}_i \cdot f_i(B). \tag{A.19}$$

Then by (A.17),

$$\mathbf{0} = \sum_{i<m} \boldsymbol{w}_i \cdot f_i(B).$$

But then by (A.15),

$$\mathbf{0} = \boldsymbol{w}_i \cdot f_i(B),$$

so by (A.18), and (A.16), and the previous theorem,

$$f_i = c_i X^{r_i}$$

for some $c_i$ in $K$. In this case, (A.19) can be written as

$$\mathbf{0} = \sum_{i<m} c_i \boldsymbol{w}_i \cdot B^{r_i - 1},$$

which implies that each $c_i$ is 0. Thus $f_i = 0$.

Now we can let

$$V' = \bigoplus_{i<m} V_i.$$

Then $V' \subseteq V$. By construction, $V_i B = W_i$, so

$$V'B = W = VB.$$

Therefore $V \smallsetminus V' \subseteq \ker B$, and so

$$V = V' + \ker B.$$

Each element of $\ker B$ constitutes a basis of a one-dimensional $B$-cyclic space. Then $V$ is the direct sum of some of these spaces, along with the $V_i$, as desired. □

In the notation of (A.7), there is $(n_j : j < m)$ in $\mathbb{N}^m$ such that there are elements

- $\bigl((\boldsymbol{v}_{j,k} : k < n_j) : j < m\bigr)$ of $\prod_{j<m} \ker({B_j}^{r_j})^{n_j}$,
- $\bigl((s_{j,k} : k < n_j) : j < m\bigr)$ of $\prod_{j<m} \mathbb{N}^{n_j}$, and
- $\bigl((V_{j,k} : k < n_j) : j < m\bigr)$ of

$$\prod_{j<m} \{\text{subspaces of } \ker({B_j}^{r_j})\}^{n_j}$$

such that, for each $j$ in $m$,

- $\ker(B_j)$ has the basis $(\boldsymbol{v}_{j,k} \cdot {B_j}^{s_{j,k}-1} : k < n_j)$,
- $\ker({B_j}^{r_j}) = \bigoplus_{k<n_j} V_{j,k}$, and
- for each $k$ in $n_j$, $V_{j,k}$ has the basis $(\boldsymbol{v}_{j,k} \cdot {B_j}^{\ell} : \ell < s_{j,k})$.

Now we may let

$$P = \left(\begin{array}{c} P_0 \\ \hline \vdots \\ \hline P_{m-1} \end{array}\right),$$

where, for each $j$ in $m$,

$$P_j = \left(\begin{array}{c} P_{j,0} \\ \hline \vdots \\ \hline P_{j,n_j-1} \end{array}\right),$$

where, for each $k$ in $n_j$,

$$P_{j,k} = \left(\begin{array}{c} \boldsymbol{v}_{j,k} \cdot {B_j}^{s_{j,k}-1} \\ \hline \vdots \\ \hline \boldsymbol{v}_{j,k} \cdot B_j \\ \hline \boldsymbol{v}_{j,k} \end{array}\right).$$

Then $PAP^{-1}$ is a **Jordan normal form** for $A$. Indeed, by the considerations yielding (A.12),

$$PAP^{-1} = \operatorname{diag}(\Lambda_0, \dots, \Lambda_{m-1}),$$

where, for each $j$ in $m$,

$$\Lambda_j = \operatorname{diag}(\Lambda_{j,0}, \dots, \Lambda_{j,n_j-1}),$$

where, for each $k$ in $n_j$,

$$\Lambda_{j,k} = \begin{pmatrix} \lambda_j & 1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & 1 \\ 0 & \dots & \dots & 0 & \lambda_j \end{pmatrix},$$

which is in $\mathrm{M}_{s_{j,k}}(K)$.

# Bibliography


[1] David M. Burton. *Elementary Number Theory.* McGraw-Hill, Boston, sixth edition, 2007.

[2] Harvey Cohn. *Advanced Number Theory.* Dover, New York, 1980. Corrected republication of 1962 edition.

[3] 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.

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

[5] Thomas W. Hungerford. *Algebra*, volume 73 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, 1980. Reprint of the 1974 original.

[6] Morris Kline. *Mathematical Thought from Ancient to Modern Times.* Oxford University Press, New York, 1972.

[7] Cemal Koç. Linear algebra I. Printed by the Department of Mathematics, Middle East Technical University, 1998. No ISBN.

[8] Cemal Koç. Topics in linear algebra. Printed by the Department of Mathematics, Middle East Technical University, 2010. No ISBN.

[9] Serge Lang. *Linear Algebra.* Addison-Wesley, Reading, Massachusetts, second edition, 1970. World Student Series edition, second printing, 1972.

[10] Rudolf Lidl and Harald Niederreiter. *Introduction to finite fields and their applications.* Cambridge University Press, Cambridge, 1986.