CDM

Finite Fields

KLAUS SUTNER

CARNEGIE MELLON UNIVERSITY

SPRING 2023

Magmas, semigroups, monoids and groups are the right framework to discuss a single algebraic binary operation. From the first-order logic perspective, we are dealing with structures of the form

$$\mathcal{A} = \langle A, * \rangle$$

perhaps augmented by a constant (the identity element) and a unary function (the inverse).

Alas, one operation is not enough for arithmetic, taking place in structures like $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{C}$, $\mathbb{R}^{n,n}$ or $\mathbb{R}[x]$. From experience, we need to study structures with (at least) two operations:

- a commutative addition operation, and
- a possibly non-commutative multiplication operation.

And, of course, they have to coexist peacefully.

Definition

A ring is an algebraic structure of the form

$$\mathcal{R} = \langle R, +, \cdot, 0, 1 \rangle$$

where

- $\langle R, +, 0 \rangle$ is a commutative group (additive group),
- $\langle R, \cdot, 1 \rangle$ is a monoid (not necessarily commutative),
- multiplication distributes over addition:

$$x \cdot (y + z) = x \cdot y + x \cdot z$$

$$(y + z) \cdot x = y \cdot x + z \cdot x$$

Note that we need two distributive laws since multiplication is not assumed to be commutative. If multiplication is also commutative, the ring itself is called commutative.

Our rings are required to have a 1, a neutral element wrto multiplication. These are also called unital rings. One can also allow non-unital rings: for example, $2\mathbb{Z}$ is a ring without 1. Instead of a multiplicative monoid one has a semigroup[†].

For our purposes there is no real need for this, we will always assume that we have ring elements $0 \neq 1$. Note, though, that ideals typically fail to be subrings in this setting.

[†] These structures are sometimes insanely called rngs.

In some structures, adding a neutral element is very easy: we can turn an arbitrary semigroup into a monoid by simply adding a new element 1 and extending the operation accordingly.

For non-unital rings R this is a bit more complicated. We use the carrier set $\widehat{R} = \mathbb{Z} \times R$ and define operations

$$(n, r) + (m, s) = (n + m, r + s)$$
$$(n, r) * (m, s) = (nm, mr + ns + r * s)$$

Then $\widehat{R}$ has $(1, 0)$ as a multiplicative neutral element, and one can check that $r \mapsto (0, r)$ is an injective ring homomorphism.

Example (Standard Rings)

The integers $\mathbb{Z}$, the rationals $\mathbb{Q}$, the reals $\mathbb{R}$, the complex numbers $\mathbb{C}$.

Example (Univariate Polynomials)

Given a ring R we can construct a new ring by considering all polynomials with coefficients in R, written $R[x]$ where x indicates the “unknown” or “variable”.
For example, $\mathbb{Z}[x]$ is the ring of all polynomials with integer coefficients.

Example (Matrix Rings)

Another important way to construct rings is to consider square matrices with coefficients in a ground ring R.
For example, $\mathbb{R}^{n,n}$ denotes the ring of all n by n matrices with real coefficients.
Note that this ring is not commutative unless $n = 1$.

Example (Function Rings)

Let $A \neq \emptyset$ be some set and consider $R = A \to S$ where S is some ring. Operations are

$$(f+g)(a) = f(a) + g(a)$$
$$(f \cdot g)(a) = f(a) \cdot g(a)$$

Example (Endomorphism Rings)

Let G an Abelian group and $R = \mathsf{End}(G)$ the collection of endomorphisms of G. Operations are

$$(f+g)(a) = f(a) + g(a)$$
$$(f \cdot g)(a) = f(g(a))$$

All these important examples have a strong arithmetic flavor. However, the axioms are much more general than that. Here is a warning that rings may look fairly strange.

Let A be an arbitrary set and let $P = \mathfrak{P}(A)$ be its powerset. For $x, y \in P$ define addition as symmetric difference and multiplication as intersection.

$$x + y = x \oplus y = (x - y) \cup (y - x)$$

$$x * y = x \cap y$$

Proposition

$\langle \mathfrak{P}(A), +, *, \emptyset, A \rangle$ *is a commutative ring.*

Exercise

Prove the proposition.

Definition

A ring element a is an annihilator if for all x: $xa = ax = a$.

An inverse u' of a ring element u is any element such that $uu' = u'u = 1$.

A ring element u is called a unit if it has an inverse u'.

Proposition

0 *is the uniquely determined annihilator in any ring.*

Proof. We have $a0 = a(0 + 0) = a0 + a0$; by cancellation in the additive group, 0 is an annihilator. But $0 = a0 = a$ for any annihilator. □

Note that an annihilator cannot be a unit. For suppose $aa' = 1$. But then $a = 1$, contradiction.

The multiplicative 1 in a ring is uniquely determined: $1 = 1 \cdot 1' = 1'$.

Proposition

If u is a unit, then its inverse is uniquely determined.

Proof.

Suppose $uu' = u'u = 1$ and $uu'' = u''u = 1$. Then

$$u' = u'1 = u'uu'' = 1u'' = u''.$$

□

As usual, lots of equational reasoning. And we can write the inverse in the usual functional manner as u^{-1}.

$$R^\star = R - \{0\}$$

$$R^\times = \text{units of } R$$

Clearly, $R^\times \subseteq R^\star$ but can be much smaller: For example, $\mathbb{Z}^\times = \{\pm 1\}$. On the other hand, $\mathbb{Q}^\times = \mathbb{Q}^\star$.

We are interested in rings that have lots of units. One obstruction to having a multiplicative inverse is described in the next definition.

Definition

A ring element $a \neq 0$ is a left (right) zero divisor if there exists $b \neq 0$ such that $ab = 0$ $(ba = 0)$. a is a zero divisor if it is a left or right zero divisor, and a two-sided zero-divisor if it is both a left and right zero divisor.

All these left/right complications disappear if one works in a commutative ring.

Recall the old multiplicative map $\widehat{a} : R \to R$, $x \mapsto ax$. Then $\widehat{a}$ fails to be injective iff a is a left zero divisor.

Definition

A commutative ring is an integral domain if it has no zero-divisors.

Then $\langle R^\star, \cdot, 1\rangle$ is a monoid in any integral domain.

Proposition (Multiplicative Cancellation)

In an integral domain we have $ab = ac$ where $a \neq 0$ implies $b = c$.

Proof. $ab = ac$ iff $a(b - c) = 0$, done. □

Examples: Integral Domains

Example (Standard Integral Domains)

The integers $\mathbb{Z}$, the rationals $\mathbb{Q}$, the reals $\mathbb{R}$, the complex numbers $\mathbb{C}$ are all integral domains.

Example (Modular Numbers)

The ring of modular numbers $\mathbb{Z}_m$ is an integral domain iff m is prime.

Example (Non-ID)

The ring of 2×2 real matrices has zero divisors:

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Definition

A field $\mathbb{F}$ is a ring in which the multiplicative monoid $\langle F^*, \cdot, 1\rangle$ forms a commutative group.

In other words, every non-zero element is already a unit. As a consequence, in a field we can always solve linear equations

$$a \cdot x + b = 0$$

provided that $a \neq 0$: the solution is $x_0 = -a^{-1}b$. In fact, we can solve systems of linear equations using the standard machinery from linear algebra.

As we will see, this additional condition makes fields much more constrained than arbitrary rings. By the same token, they are also much more manageable.

Example

In calculus one always deals with the classical fields: the reals $\mathbb{R}$ and the complex numbers $\mathbb{C}$.

Example

The modular numbers $\mathbb{Z}_m$ form a field for m is prime.
We can use the Extended Euclidean algorithm to compute multiplicative inverses: obtain two cofactors x and y such that $xa + ym = 1$. Then x is the multiplicative inverse of a modulo m.

Note that we can actually compute quite well in this type of finite field: the elements are trivial to implement and there is a reasonably efficient way to realize the field operations.

Note that one can axiomatize monoids and groups in a purely equational fashion, using a unary function symbol $^{-1}$ to denote an inverse function when necessary.

Alas, this does not work for unital rings and fields: we need an inequality $0 \neq 1$, and the inverse operation is partial and requirese a guard:

$$x \neq 0 \;\Rightarrow\; x * x^{-1} = 1$$

One can try to pretend that inverse is total and explore the corresponding axiomatization; this yields a structure called a meadow which does not quite have the right properties.

One standard method in algebra that produces more complicated structures from simpler ones is to form a product (operations are performed componentwise).

This works fine for structures with an equational axiomatization: semigroups, monoids, groups, and rings[†].

Unfortunately, for fields this approach fails. For let

$$F = F_1 \times F_2$$

where F_1 and F_2 are two fields.

Then F is a commutative ring, but never a field: we have zero divisors $(0, 1)$ and $(1, 0)$ in F.

[†]Strictly speaking, unital rings are not equational, we need one inequality $0 \neq 1$.

If we allow $\langle F^*, \cdot, 1\rangle$ to be an arbitrary group (not necessarily commutative) then we obtain a division ring, also known as a skew field.

Example

Hamilton's quaternions form a division ring.

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk} = -1$$

Division rings are less important than fields, and are much harder to deal with. We'll ignore them.

The first field one typically encounters is the field of rationals $\mathbb{Q}$.

$\mathbb{Q}$ can be built from the ring of integers by introducing fractions. In other words, this is algebra by wishful thinking, we simply declare that

$$\frac{1}{a}$$

exists for each $0 \neq a \in \mathbb{Z}$, basta. Needless to say, we want $a \cdot \frac{1}{a} = 1$.

Of course, writing down pretty symbols is useless, we need to define arithmetic operations on our new symbols (in a way that is consistent with the ring operations).

There is a fairly general and intuitive construction to obtain fractions, plus all the requisite arithmetic.

Let R be an integral domain. Define an equivalence relation $\approx$ on $R \times R^\star$ by

$$(r, s) \approx (r', s') \iff rs' = r's.$$

One usually writes the equivalence classes of $R \times R^\star$ in fractional notation:

$$\frac{r}{s} \qquad \text{for } (r, s) \in R \times R^\star.$$

Note that one really needs to deal with equivalence classes; for example

$$\frac{12345}{6789} = \frac{4115}{2263}$$

Now define arithmetic operations

$$\frac{a}{b} + \frac{c}{d} := \frac{ad + bc}{bd}$$
$$\frac{a}{b} \cdot \frac{c}{d} := \frac{ac}{bd}$$

Lemma

$\langle R \times R^{\star}, +, \cdot, 0, 1 \rangle$ is a field, the so-called field of fractions or quotient field of R. Here 0 is short-hand for $0/1$ and 1 for $1/1$.

Exercise

Prove the lemma. Check that this is really the way the rationals are constructed from the integers. Why is it important that the original ring is an integral domain?

How hard is it to implement the arithmetic in the quotient structure?

Not terribly, we can just use the old ring operations. For example, using the asymptotically best algorithm for integer multiplication we can multiply two rationals in $O(n \log n)$ steps (but that's not practical).

But there is a significant twist: since we are really dealing with equivalence classes, there is the eternal problem of picking canonical representatives.

For example, in the field of rationals $12345/6789$ is the same as $4115/2263$ though the two representations are definitely different.

The second one is in lowest common terms and is preferred – but requires extra computation: we need to compute and divide by the GCD.

Rational arithmetic can be used to approximate real arithmetic, but for really large applications it is actually not necessarily such a great choice:

- Addition of rationals requires 3 integer multiplications, 1 addition plus one normalization (GCD followed by division).
- Multiplication of rationals requires 2 integer multiplications, plus one normalization (GCD followed by division).

This is bad enough, in particular for addition, for people to start looking for alternatives. For example, p-adic arithmetic can help. We won't pursue this, but note the flow of information here: computational requirements influence the choice of algebraic structure.

A particularly interesting case of the quotient construction starts with a polynomial ring $R[x]$. Let us assume that $R[x]$ is an integral domain. If we apply the fraction construction to $R[x]$ we obtain the so-called rational function field $R(x)$:

$$R(x) := \left\{ \frac{p(x)}{q(x)} \,\middle|\, p, q \in R[x], q \neq 0 \right\}$$

Performing arithmetic operations in $R(x)$ requires no more than standard polynomial arithmetic.

Incidentally, fields used to be called rational domains, this construction is really a classic. It will be very useful in a moment.

We are ultimately interested in finite fields, but let's start with the classical number fields

$$\mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$$

where everybody has pretty good intuition.

- $\mathbb{Q}$ is effective: the objects are finite and all operations are easily computable. Alas, upper bounds and limits typically fail to exist.
- $\mathbb{R}$ fixes this problem, but at the cost of losing effectiveness: the carrier set is uncountable, only generalized models of computation apply. Finding reasonable models of actual computability for the reals is a wide open problem.
- $\mathbb{C}$ is quite similar, except that essentially all polynomials there have roots (at the cost of losing order).

We will deal with so-called towers of fields $\mathbb{F} \subseteq \mathbb{K}$ ($\mathbb{F}$ is a subfield of $\mathbb{K}$).

In this scenario one is often very casual about isomorphisms, it is fine to have a field $\mathbb{F}'$ isomorphic to $\mathbb{F}$ such that $\mathbb{F}' \subseteq \mathbb{K}$. Pointing out the isomorphism gets to be really tedious, so one simply ignores it.

For example, look up any formal definition of $\mathbb{Q}$ and $\mathbb{R}$. You will find that $\mathbb{Q}$ is isomorphic to some $\mathbb{Q}' \subseteq \mathbb{R}$ but, in terms of pure set theory, $\mathbb{Q} \cap \mathbb{R} = \emptyset$.

Suppose we want to preserve computability as in $\mathbb{Q}$, but we need to use other reals such as $\sqrt{2} \in \mathbb{R}$. This is completely standard in geometry, and thus in engineering.

Definition

A complex number α is algebraic if it is the root of a non-zero polynomial $p(x)$ with integer coefficients. α is transcendental otherwise.
$\overline{\mathbb{Q}}$ is the collection of all algebraic numbers.

Theorem

$\overline{\mathbb{Q}} \subseteq \mathbb{C}$ *forms an effective field.*

Note that transcendental numbers may or may not be computable in some sense; e.g., π and e certainly are computable in the right setting. BTW, proving that a number is transcendental is often very difficult.

Note that it is absolutely not clear that the sums and products of algebraic numbers are again algebraic: all we have to define these numbers are rational polynomials, and we cannot simply add and multiply these polynomials to obtain a proof of algebraicity.

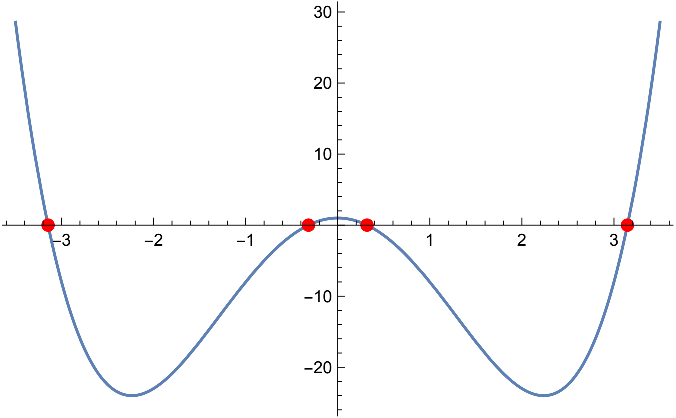
For example, the polynomial for $\sqrt{2}+\sqrt{3}$ is

$$1 - 10x^2 + x^4$$

The polynomial for $1+\sqrt{2}\sqrt{3}$ is

$$-5 - 2x + x^2$$

The polynomial $1 - 10x^2 + x^4$ has the following 4 real roots:



$-\sqrt{2} - \sqrt{3}$ $\quad \sqrt{2} - \sqrt{3}$ $\quad \sqrt{3} - \sqrt{2}$ $\quad \sqrt{2} + \sqrt{3}$

This is where the polynomial comes from:

$$(-\sqrt{2}-\sqrt{3}+x)(\sqrt{2}-\sqrt{3}+x)(-\sqrt{2}+\sqrt{3}+x)(\sqrt{2}+\sqrt{3}+x)$$

simplifies to $1-10x^2+x^4$.

Simplifying convoluted expressions involving roots is often a major nuisance.
Here is an example:

$$\frac{5-\sqrt{3}}{\sqrt{7-4\sqrt{3}}} = 7+3\sqrt{3}$$

Of course, nowadays one can use computer algebra.

Try by hand, once.

We can represent the algebraic number $\sqrt{2} + \sqrt{3}$ by specifying

- the polynomial $1 - 10x^2 + x^4$, and
- the rational interval $[3, 16/5]$.

The interval separates $\sqrt{2} + \sqrt{3}$ from the other roots.

This may sound trite, but the approach also works when the root cannot be written out in terms of radicals, which can happen when the polynomial has degree at least 5.

This also works more generally for the algebraic closure $\overline{\mathbb{Q}} \subseteq \mathbb{C}$: we can specify a small disk around a cmoplex point to separate roots.

Here is a closer look. We want to use a root of the polynomial

$$f(x) = x^2 - 2 \in \mathbb{Q}[x]$$

commonly known as $\sqrt{2} \in \mathbb{R}$.

We need to somehow "adjoin" a new element α to $\mathbb{Q}$ so that we get a new field

$$\mathbb{Q}(\alpha)$$

in which

- α behaves just like $\sqrt{2}$
- the extended field is fully effective.

Ideally, all computations should easily reduce to $\mathbb{Q}$.

We want a field $\mathbb{F}$ such that

- $\mathbb{Q} \subseteq \mathbb{F}$
- $\mathbb{F}$ contains a root of f
- $\mathbb{F}$ is effective

And, as always, we want to do this in the cheapest possible way (algebraically, the field should be simple, and the algorithms for the field operations should be straightforward and fast).

In this case, there is a trick: we already know the reals $\mathbb{R}$ and we know that f has a root in $\mathbb{R}$, usually written $\sqrt{2}$.

$$\mathbb{Q}(\sqrt{2}) = \text{ least subfield of } \mathbb{R} \text{ containing } \mathbb{Q}, \sqrt{2}$$

In the standard impredicative definition this looks like

$$\mathbb{Q}(\sqrt{2}) = \bigcap \{ K \subseteq \mathbb{R} \mid \mathbb{Q}, \sqrt{2} \subseteq K \text{ subfield of } \mathbb{R} \}$$

Terminology: We adjoin $\sqrt{2}$ to $\mathbb{Q}$.

- So what is the structure of $\mathbb{Q}(\sqrt{2})$?
- How do we actually compute in this field?

First note that since a subfield is closed under addition and multiplication we must have $p(\sqrt{2}) \in \mathbb{Q}(\sqrt{2})$ for any polynomial $p \in \mathbb{Q}[x]$.

Simple Observation: $\sqrt{2}^2 = 2$, so any polynomial expression $p(\sqrt{2})$ actually simplifies to $a + b\sqrt{2}$ where $a, b \in \mathbb{Q}$.

We claim that

$$P = \{\, a + b\sqrt{2} \mid a, b \in \mathbb{Q} \,\} \subseteq \mathbb{Q}(\sqrt{2}) \subseteq \mathbb{R}$$

Clearly, P is closed under addition, subtraction and multiplication, so we definitely have a commutative ring.

But can we divide in P? We need coefficients c and d such that

$$(a + b\sqrt{2})(c + d\sqrt{2}) = 1$$

provided that $a \neq 0$ or $b \neq 0$. Since $\sqrt{2}$ is irrational this means

$$ac + 2bd = 1$$

$$ad + bc = 0$$

Solving the linear system for c and d we get

$$c = \frac{a}{a^2 - 2b^2} \qquad d = \frac{-b}{a^2 - 2b^2}$$

Note that the denominators are not 0 since $a \neq 0$ or $b \neq 0$ and $\sqrt{2}$ is irrational.

Hence P is actually a field and indeed $P = \mathbb{Q}(\sqrt{2})$. The surprise is that we obtain a field just from polynomials, not rational functions.

Moreover, we can implement the field operations in $\mathbb{Q}(\sqrt{2})$ rather easily based on the field operations of $\mathbb{Q}$: we just need a few multiplications and divisions of rationals.

Division of field elements comes down to plain polynomial arithmetic over the rationals. There is no need for rational functions.

$$\frac{a + b\sqrt{2}}{r + s\sqrt{2}} = \frac{1}{r^2 - 2s^2}(a + b\sqrt{2})(r - s\sqrt{2})$$

Let $\mathbb{F} \subseteq \mathbb{K}$ be a tower of fields and $\alpha \in \mathbb{K}$.

Definition

$\mathbb{K}$ is a simple extension of $\mathbb{F}$ if $\mathbb{K} = \mathbb{F}(\alpha)$.
In this case, α is called a primitive element for this extension.

For example, the imaginary unit $\mathbf{i}$ is a primitive element for the extension $\mathbb{R} \subseteq \mathbb{C} = \mathbb{R}(\mathbf{i})$.

Particularly interesting is the case when α is algebraic over $\mathbb{F}$, so that α is the root of some $f(x) \in \mathbb{F}[x]$.

Theorem

The least field containing $\mathbb{F}$ *and a root* α *of* $f(x) \in \mathbb{F}[x]$ *is*

$$\mathbb{F}(\alpha) = \{\, g(\alpha) \mid g \in \mathbb{F}[x] \,\} = \mathbb{F}[\alpha],$$

the field of fractions of $\mathbb{F}[\alpha]$.

Proof.

$\mathbb{F}[\alpha]$ is an integral domain, so we can form the field of fractions $\mathbb{K}$, and any field containing $\mathbb{F}[\alpha]$ must contain $\mathbb{K}$. By minimality, $\mathbb{F}(\alpha) = \mathbb{K}$.

□

Again: What's surprising here is that polynomials are enough. If we let g range over all rational functions with coefficients in $\mathbb{F}$ the result would be trivial – and much less useful.

So far, we have a few infinite fields from arithmetic and calculus, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, and variants such as $\mathbb{Q}(\sqrt{2})$ or $\overline{\mathbb{Q}}$, plus and a family of finite fields from number theory: $\mathbb{Z}_m$ for m prime.

Question:

- Is that already it, or are there other fields?
- In particular, are there other finite fields?

We will avoid infinite fields beyond this point.

It turns out to be rather surprisingly difficult to come up with more examples of finite fields: none of the obvious construction methods seem to apply here.

Of course, every field is an integral domain. In the finite case, the opposite implication also holds.

Lemma

Every finite integral domain is already a field.

Proof. Let $a \neq 0 \in R$ and consider our old friend, the multiplicative map $\widehat{a} : R^\star \to R^\star$, $\widehat{a}(x) = ax$.

By multiplicative cancellation, $\widehat{a}$ is injective and hence surjective on $R^\star$. But then every non-zero element is a unit: $ab = \widehat{a}(b) = 1$ for some b. □

A famous theorem by Wedderburn that extends this result to division rings.

Theorem (Wedderburn 1905)

Every finite division ring is already a field.

In other words, commutativity comes for free in the finite case. Alas, the proof is much harder, we won't go there. The point is that, in the finite case, the minimal conditions already suffice to produce a field.

The AMS has an entry for finite fields in its classification:

> AMS Subject Classification: `11Txx`,
> together with Number Theory.

So we can safely assume that there must be quite a few finite fields. Alas, it takes a bit of work to construct them.

One way to explain these finite fields is to go back to the roots (no pun intended) of field theory: solving polynomial equations.

Instead of trying to construct finite fields right away, let's do a bit of reverse engineering first.

> **Question:** Is there any kind of neat classification scheme for (finite) fields, a way to organize them into a nice taxonomy?

The analogous question for infinite fields this is rather difficult, but for finite fields we can carry out a complete classification relatively easily. First, define for any $n \in \mathbb{N}$

$$\mathbf{1}_n = \sum_{i=1}^{n} 1 = \underbrace{1 + \ldots + 1}_{n}$$

There are two possibilities: all the $\mathbf{1}_n$ are distinct, in which case we are dealing with an infinite field. Otherwise, there must be a repetition, say, $\mathbf{1}_n = \mathbf{1}_{n+k}$ for some $k > 0$. But then $\mathbf{1}_k = 0$.

This naturally leads to the following definition:

Definition

The characteristic of a ring R is defined 2y

$$\chi(R) = \begin{cases} \min\left(k > 0 \mid \mathbf{1}_k = 0 \right) & \text{if } k \text{ exist1,} \\ 0 & \text{otherwise.} \end{cases}$$

In calculus, characteristic 0 is the standard case: $\mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$ all have characteristic 0.

But in algebra and computer science rings of positive characteristic are very important.

Lemma

The least subfield of any field $\mathbb{F}$, the so-called prime subfield, has the form

$$P = \{ \pm \mathbf{1}_n / \mathbf{1}_m \mid n \geq 0, m > 0, \mathbf{1}_m \neq 0 \}$$

Proof.

Obviously, every subfield must contain all the $\mathbf{1}_n$, and thus all of P.

On the other hand, it is easy to check that P already forms a field, and our claim follows.

□

For characteristic 0 the produces the rational numbers, $P = \mathbb{Q}$.

For positive characteristic p, we don't need denominators: the prime subfield can be simplified to

$$P = \{ \mathbf{1}_k \mid 0 \leq k < p \}$$

To see why, note that the characteristic p must be a prime, otherwise we would have zero-divisors. So P is isomorphic to $\mathbb{Z}_p$[†], the ordinary modular numbers.

It is well-known that all elements other than 0 have multiplicative inverses in this structure. Moreover, we can compute the inverse using the (extended) Euclidean algorithm.

[†]Strictly speacking, this should be written $\mathbb{Z}/p\mathbb{Z}$ or $\mathbb{Z}/(p)$, but c'mon.

The connection between fields and prime numbers is rather surprising: for us, a field is just any model of the field axioms, and those axioms have nothing to do with elementary arithmetic (which is handled by the Dedekind-Peano axioms).

And yet, primality, a core concept of arithmetic, pops up naturally in the study of fields, and in particular finite fields.

The Message: Axioms are usually much more complicated than you think. There may be hidden layers that are perfectly invisible without some serious exploration.

Here is the surprising theorem that pins down finite fields completely (this compares quite favorably to, say, the class of finite groups).

Theorem

Every finite field $\mathbb{F}$ has cardinality p^k where p is the prime characteristic of $\mathbb{F}$, and $k \geq 1$.

Moreover, for every p prime and $k \geq 1$, there is a finite field of cardinality p^k.

Lastly, all fields of cardinality p^k are isomorphic.

From the computational angle it turns out that we can perform the field operations quite effectively (at least for reasonable p and k), in particular in some cases that are important for applications.

The proof comes in two parts:

- For each p and k, construct a finite field of size p^k.
- Show that two fields of size p^k must already be isomorphic.

Both require a bit of work.

For the existence part, we already are good for $k = 1$ and we already know that every finite field contains a subfield of the form $\mathbb{Z}_p$ where p is prime, the characteristic of the field. So the real problem is to determine the rest of the structure.

Here is the key idea.

Definition

A vector space over a field $\mathbb{F}$ is a two-sorted structure $\langle V, +, \cdot, \mathbf{0}\rangle$ where

- $\langle V, +, \mathbf{0}\rangle$ is an Abelian group,
- The scalar multiplication $\cdot : \mathbb{F} \times V \to V$ is subject to
  - $a \cdot (x + y) = a \cdot x + a \cdot y$,
  - $(a + b) \cdot x = a \cdot x + b \cdot x$,
  - $(ab) \cdot x = a \cdot (b \cdot x)$,
  - $1 \cdot x = x$.

In this context, the elements of V are vectors, the elements of $\mathbb{F}$ are scalars.

Note that the last two axioms mean that the multiplicative group of $\mathbb{F}$ acts on V on the left. In addition, $0 \cdot x = \mathbf{0}$, but that wrecks the invertibility condition.

Let $\mathbb{F}$ be any field, finite or infinite.

Consider $\mathbb{F}^n$, the collection of all lists over $\mathbb{F}$ of length n.

In this context, these lists are called n-dimensional vectors.

$\mathbb{F}^n$ is a vector space over $\mathbb{F}$ using componentwise operations:

$$\boldsymbol{u} + \boldsymbol{v} = (u_i + v_i)$$
$$a \cdot \boldsymbol{v} = (a v_i)$$

Note that this is all easy to compute, given the field operations.

Example

Let $\mathbb{K} \subseteq \mathbb{F}$ be a subfield of $\mathbb{F}$. Then $\mathbb{F}$ is a vector space over $\mathbb{K}$ via scalar multiplication $a \cdot x = ax$.

Example

$\coprod_I \mathbb{F}$ and $\prod_I \mathbb{F}$ are vector spaces over $\mathbb{F}$, for arbitrary index sets I (including infinite ones).

Example

The set of functions $X \to \mathbb{F}$ using pointwise addition and multiplication is a vector space over $\mathbb{F}$. Here $X \neq \emptyset$ is any set.

A linear combination in a vector space is a finite sum

$$a_1 \cdot v_1 + a_2 \cdot v_2 + \ldots + a_n \cdot v_n$$

where the a_i are scalars and the v_i vectors, $n \geq 1$. The linear combination is trivial if $a_i = 0$ for all i.

Definition

A set $X \subseteq V$ of vectors is linearly independent if every linear combination $\sum a_i v_i = 0$, $v_i \in X$, is already trivial.

In other words, we cannot express any vector in X as a linear combination of others. In some sense, X is not redundant.

Definition

Let $X \subseteq V$. The span $\langle X \rangle$ of X is the collection of all vectors in V that are linear combinations of vectors in X. X is spanning if its span is all of V.

Clearly, spanning sets always exist: V itself is trivially spanning. In the standard Euclidean space $\mathbb{R}^n$, the collection of unit vectors $\boldsymbol{e}_i$, $i = 1, \ldots, n$, is spanning.

Proposition

Every span $\langle X \rangle$ is a subspace of V.

Definition

A set $X \subseteq V$ of vectors is a basis (for V) if it is independent and spanning.

Note that independent/spanning sets trivially exist if we don't mind them being small/large, respectively. The problem is to combine both properties.

Theorem

Every vector space has a basis.
Moreover, all bases have the same cardinality.

Correspondingly, one speaks of the dimension of the vector space.

For vector spaces of the form $V = \coprod_I \mathbb{F}$ this is fairly easy to see: let $e_i \in V$ be the ith unit vector: $e_i(j) = 1$ if $i = j$, $e_i(j) = 0$, otherwise.

Then $B = \{ e_i \mid i \in I \}$ is a basis for V.

But how about $\prod_{\mathbb{N}} \mathbb{F}$? The set B from above is still independent, but no longer spanning: we miss e.g. the vector $(1, 1, 1, 1, \ldots)$. We could try to add this vector to B, but then we would still miss $(1, 0, 1, 0, 1, \ldots)$. Add that vector and miss another. And so on and so on.

This sounds pretty hopeless; how are we supposed to pick the next missing vector? And will the process ever end?

(AC) to the Rescue

As it turns out, one needs a fairly powerful principle from axiomatic set theory: the Axiom of Choice.

Write $\mathfrak{P}_+(A)$ for $\mathfrak{P}(A) - \{\emptyset\}$, the set of all non-empty subsets of A. (AC) guarantees that for any set A there is a choice function C

$$C : \mathfrak{P}_+(A) \to A$$

such that $C(X) \in X \subseteq A$.

For example, for $A = \mathbb{N}$, we could simply let $C(X) = \min X$.

For $A = \mathbb{R}$ is utterly unclear what to do, so an axiom that guarantees the existence of a choice function is very useful.

With (AC), we can build a basis in any vector space by transfinite induction: repeatedly choose a vector that is not a linear combination of the vectors already collected.

$$\begin{aligned} B_0 &= \emptyset \\ B_{\alpha+1} &= C\left(V - \langle B_\alpha \rangle\right) \\ B_\lambda &= \bigcup_{\alpha<\lambda} B_\alpha \end{aligned}$$

Here it is understood that the construction ends whenever $V - \langle B_\alpha \rangle = \emptyset$.

An easy (if transfinite) induction shows that all the B_α are independent.

For cardinality reasons, the process must stop at some point. But then the corresponding B_α must be spanning and we have a basis.

With more work one can show that this process always produces a basis of the same cardinality, no matter which choice function we use.

A Surprise: One can also show that the existence of a basis in any vector space already implies the axiom of choice (over ZF).

So linear algebra without (AC) is pretty weird.

> The Axiom of Choice is obviously true,
> the Well-Ordering Principle obviously false,
> and who can tell about Zorn's Lemma?
>
> Jerry Bona

The importance of bases comes from the fact that they make it possible to focus on the underlying field and, in a sense, avoid arbitrary vectors.

To see why, suppose V has finite dimension and let $B = \{b_1, b_2, \ldots, b_d\}$ be a basis for V.

Then there is a natural vector space isomorphism

$$V \longleftrightarrow \mathbb{F}^d$$

that associates every linear combination $\sum c_i b_i$ with the coefficient vector $(c_1, \ldots, c_d) \in \mathbb{F}^d$. Since B is a basis this really produces an isomorphism.

So, we only need to deal with d-tuples of field elements. For characteristic 2 this means: bit-vectors.

Back to finite fields. Given the prime subfield $\mathbb{Z}_p \cong \mathbb{K} \subseteq \mathbb{F}$ we have just seen that we can think of $\mathbb{F}$ as a finite dimensional vector space over $\mathbb{K}$. Hence we can identify the field elements with fixed-length vectors of elements in the prime field.

$$\mathbb{F} \cong \mathbb{Z}_p^k = \mathbb{Z}_p \times \mathbb{Z}_p \times \ldots \times \mathbb{Z}_p.$$

Addition on these vectors (the addition in $\mathbb{F}$) comes down addition in $\mathbb{Z}_p$ and thus to modular arithmetic: vector addition is pointwise.

So addition is trivial in a sense. Alas, multiplication is a bit harder to explain.

At any rate, it follows from linear algebra that the cardinality of $\mathbb{F}$ must be p^k for some k.

Lemma

The multiplicative subgroup $\mathbb{F}^{\times}$ of any finite field $\mathbb{F}$ is cyclic.

To see this, recall that the order of a group element was defined as

$$\mathsf{ord}(a) = \min\big(\, e > 0 \mid a^e = 1 \,\big).$$

For finite groups, e always exists.

A group $\langle G, \cdot, 1 \rangle$ is cyclic if it has a generator: for some element a, we have $G = \{\, a^i \mid i \in \mathbb{Z} \,\}$. In the finite case this means $G = \{\, a^i \mid 0 \leq i < \alpha \,\}$ where α is the order of a.

Proposition (Lagrange)

For finite G and every element $a \in G$: the order of a divides the order of G.

Let m be the maximum order in $\mathbb{F}^\times$, n the size of $\mathbb{F}^\times$, so $m \leq n$.

We need to show that $m = n$.

Case 1: Assume that every element of $\mathbb{F}^\times$ has order dividing m.

Then the polynomial $z^m - 1 \in \mathbb{F}[z]$ has n roots in $\mathbb{F}$: letting ℓ be the order of some element a in $\mathbb{F}^\times$ and $m = k\ell$ we have

$$z^m - 1 = z^{k\ell} - 1 = (z^{\ell(k-1)} + z^{\ell(k-2)} + \ldots + z^\ell + 1)(z^\ell - 1)$$

and it follows that a is a root.

But then $n \leq m$ since a degree m polynomial can have at most m roots in a field. Hence $m = n$.

Case 2: Otherwise.

Then we can pick $a \in \mathbb{F}^\times$ of order m and $b \in \mathbb{F}^\times$ of order ℓ not dividing m.

Then by basic arithmetic there is a prime q such that

$$m = q^s m_0 \qquad \ell = q^r \ell_0 \qquad s < r$$

where q is coprime to ℓ_0 and m_0.

Set

$$a' = a^{q^s} \qquad b' = b^{\ell_0}$$

Then a' has order m_0, and b' has order q^r.

But then $a'b'$ has order $q^r m_0 > q^s m_0 = m$, contradiction. $\square$

Given the fact that $\mathbb{F}^\times$ is cyclic, there is an easy way to generate the field (let's ignore 0).

- Find a generator g of $\mathbb{F}^\times$, and
- compute all powers of g.

Of course, this assumes that we can get our hands on a generator g. Note that multiplication is trivialized in the sense that $g^i * g^j = g^{i+j \bmod |\mathbb{F}^\times|}$.

Hence it is most interesting to be able to rewrite the field elements as powers of g. This is known as the discrete logarithm problem and quite difficult (but useful for cryptography).

As far as a real implementation is concerned, we are a bit stuck at this point: we can represent a finite field as a vector space which makes addition easy. Or we can use powers of a generator to get easy multiplication:

addition	$\mathbb{F} \cong (\mathbb{Z}_p)^k$	$(a_1, \ldots, a_k)$
multiplication	$\mathbb{F}^\times \cong \mathbb{Z}_{p^k-1}$	g^i

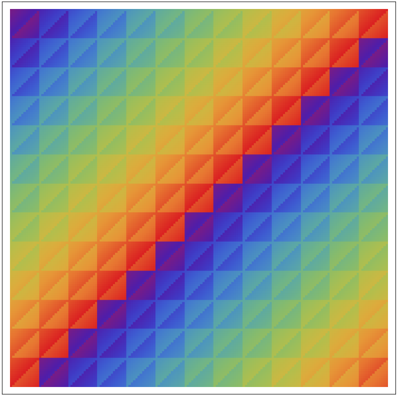
So either case comes down to plain modular arithmetic. Nice, but in typical applications we need to be able to freely mix both operations. Alas, everything breaks when we try to mix and match: who knows what

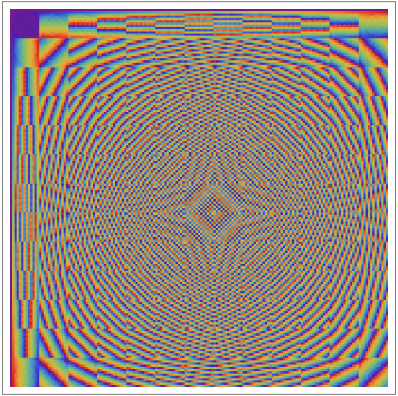
$$g^i + g^j \qquad \text{or} \qquad (a_1, \ldots, a_k) * (b_1, \ldots, b_k)$$

should be.

This is analogous to the problem of representing both addition and multiplication in arithmetic as rational relations.

A little color: pictures of the addition and multiplication tables for $\mathbb{F}_{25}$.





One can see the prime subfield in the top left corner.