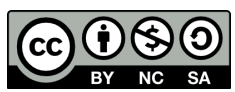
CDM

Finite Fields

Klaus Sutner

Carnegie Mellon University

Spring 2024



The first field one typically encounters early in life 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 some mysterious object

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

exists for each $0 \neq a \in \mathbb{Z}$, and that $a \cdot \frac{1}{a} = 1$.

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 old ring operations over $\mathbb{Z}$.

Here is a construction that builds fractions over an arbitrary integral domain R in a way that guarantees that the final result extends R and is a field (actually, the smallest field extending R).

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:

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

Equivalence classes are inevitable here; over $\mathbb{Z}$ we have

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

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 $1/1$.

The proof is not hard, but one needs to check all the relevant properties. E.g., one has to verify that addition is associative. Character-building exercise.

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.

A particularly interesting case of the quotient construction starts with an integer domain that is a polynomial ring $R[x]$. 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).

The scenario when one has nested fields

$$\mathbb{F}_1 \subseteq \mathbb{F}_2 \subseteq \mathbb{F}_3$$

and so on occurs with some frequency, one speaks of a tower of fields. Strictly speaking, $\mathbb{F}_i$ here should be a subfield of $\mathbb{F}_{i+1}$.

Usually one is very casual about isomorphisms, it is fine to have a field $\mathbb{F}'$ isomorphic to $\mathbb{F}_i$ such that $\mathbb{F}' \subseteq \mathbb{F}_{i+1}$. Pointing out the isomorphism gets to be really tedious, so one simply ignores this issue.

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, $\mathbb{Q} \cap \mathbb{R} = \emptyset$. Likewise for $\mathbb{R} \subseteq \mathbb{C}$.

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.

Algebraic numbers are computable, the associated polynomials provide a handle (though the details are quite messy).

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.

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 $\mathbb{Q}$.

The intersection-of-all-candidates definition is very elegant, but it leaves a number of questions wide open.

- So what exactly 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 $\mathbb{Q}(\sqrt{2})$ is none other than

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

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.

Again: Killing Denominators

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 $\mathbb{Q}(\sqrt{2})$, plus 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. □

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

Question:
Is there a nice taxonomy for finite fields?

The analogous question for rings is hopeless, and for infinite fields it is rather difficult. But for finite fields we can carry out a complete classification relatively easily.

Recall that the characteristic of a finite ring R is the least k such that

$$0 = \mathbf{1}_k = \underbrace{1 + \ldots + 1}_{k}$$

Lemma

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

$$P = \left\{ \frac{\pm \mathbf{1}_n}{\mathbf{1}_m} \,\middle|\, n \geq 0, m > 0, \mathbf{1}_m \neq 0 \right\}$$

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.

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 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.

The Mother of All Vector Spaces

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 always called n-dimensional vectors.

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

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

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

More Examples of Vector Spaces

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?

Solution: invoke the Axiom of Choice.

Using (AC) and transfinite induction one can construct a basis in any vector space whatsoever.

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: generate $\mathbb{F}^\times$ and then add 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 (and therefore 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

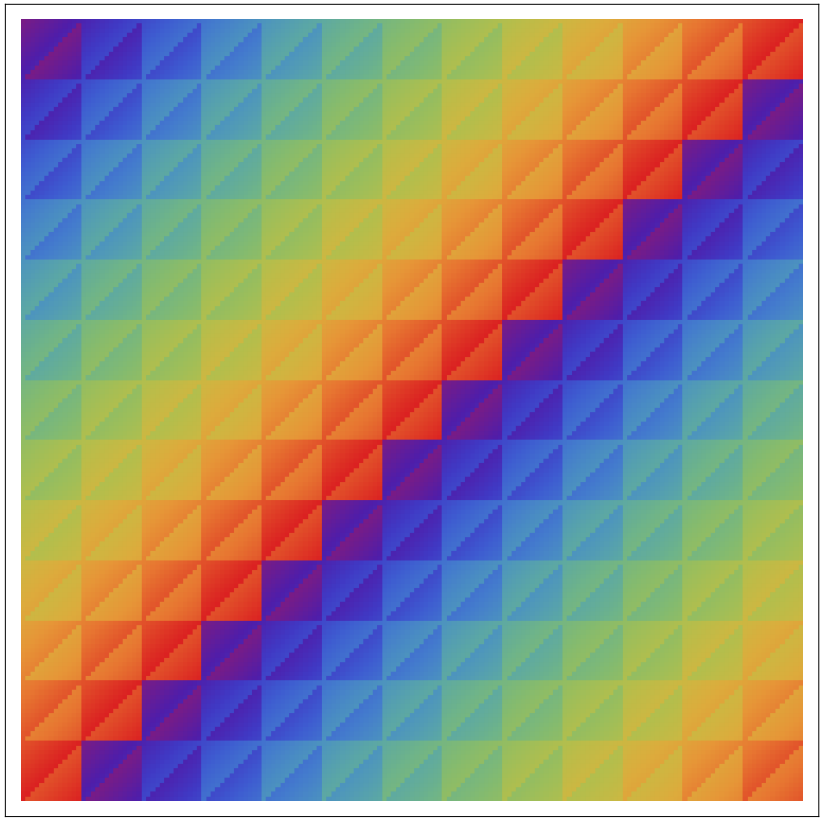
So either case comes down to plain modular arithmetic. Nice, but in typical applications we need to be able to freely mix both operations. Alas, everything breaks when we try to mix and match: who knows what

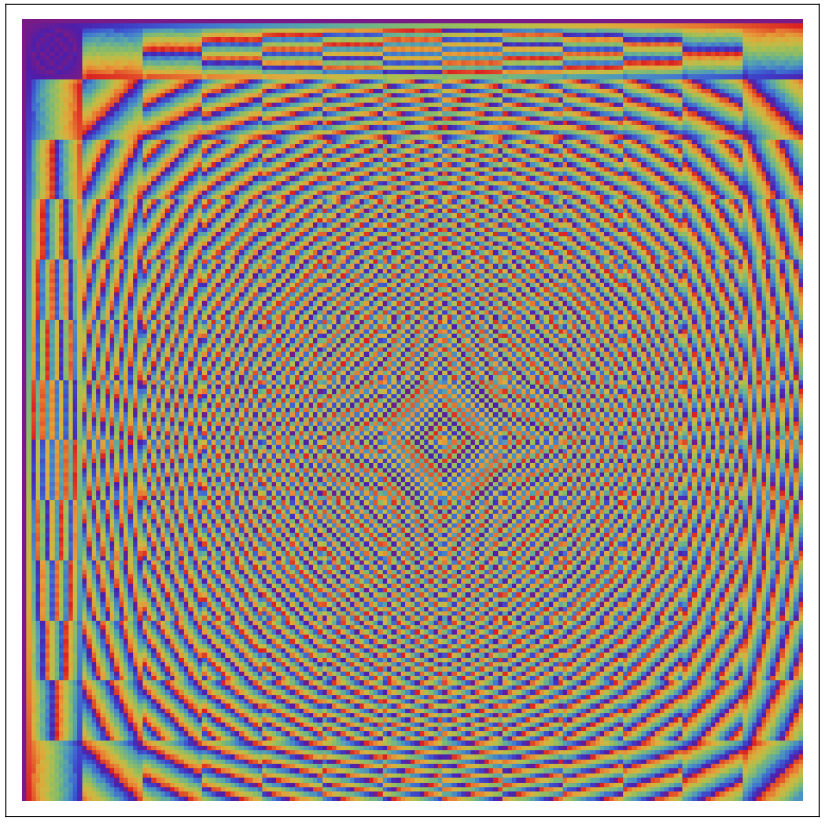
$$g^i + g^j \qquad \text{or} \qquad (a_1, \ldots, a_k) * (b_1, \ldots, b_k)$$

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.

Where Are We?

We know that every finite field carries two apparently separate structures: additive and multiplicative.

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

The problem is that we have absolutely no idea how to unify the two.

Time to get serious about building a finite field.

We would like to follow the construction of $\mathbb{Q}(\sqrt{2})$ from above, adjoining a root of $x^2 - 2 = 0$ to the rationals. But this time, we won't rely on intuition and prior knowledge of the reals. For example, consider the polynomial

$$f = x^2 + x + 1 \in \mathbb{F}_2[x]$$

We can easily check that f has no root over $\mathbb{F}_2$.

So how do we expand $\mathbb{F}_2$ to a field $\mathbb{F}$ where f has a root?

This time:

- We do not know a convenient big field like $\mathbb{R}$ that we can use as a safe sandbox, and
- we have no intuitive idea what a root of f looks like.

So, we can't just do

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

But: we can try to lift this construction to our new setting. To wit, we applied the simplification rule

$$x^2 \rightsquigarrow 2$$

to all polynomials over $\mathbb{Q}$. This produces expressions $a + bx$, $a, b \in \mathbb{Q}$, that turn out to form a field (the "unknown" x works just like the root we are after).

We want $x^2 + x + 1 = 0$, so we use the simplification rule

$$x^2 \rightsquigarrow x + 1$$

and apply it to all polynomials in $\mathbb{F}_2[x]$. We are in characteristic 2, so plus is minus.

With luck, we might wind up with a finite field that has a root for f.

Here is one of the occasions where it is useful to think of a polynomial as an expression, a term in some formal language.

On that view, we can apply the rewrite rule $x^2 \rightsquigarrow x + 1$ to try to simplify the expression. More precisely, we use this rule plus all the standard simplifications we can apply to our terms (associativity, commutativity, cancellation, . . .).

For those concerned about the StringWorld approach to life, not to worry, we will unearth the actual algebraic meaning behind this rewrite process in a moment.

So what happens to an arbitrary polynomial $p(x) \in \mathbb{F}_2[x]$ if we apply this rule systematically? Essentially, we can smash all the higher powers of x. Here is an example.

$$\begin{aligned} x^6 + x^3 + x + 1 &\rightsquigarrow (x+1)^3 + x(x+1) + x + 1 \\ &\rightsquigarrow (x^3 + x^2 + x + 1) + (x^2 + x) + x + 1 \\ &\rightsquigarrow x(x+1) + (x+1) + 1 \\ &\rightsquigarrow x + 1 \end{aligned}$$

Proposition

x^k reduces to $1, x, x+1$, depending on $k \bmod 3$.

So $x^6 + x^3 + x + 1 \rightsquigarrow 1 + 1 + x + 1 = x + 1$.

The simplification process is highly nondeterministic, there are many choices along the way.

This might cause a huge headache: if we apply the rules in one particular way, we get a different result from when we apply the rules in another way.

One really needs to make sure the process is confluent: application order does not matter, the final result is always the same. More later.

In general, if we start with a polynomial $f \in \mathbb{F}[x]$ of degree d, we get a simplification rule

$$x^k \rightsquigarrow a_{k-1}x^{k-1} + a_{k-2}x^{k-2} + \ldots + a_1 x + a_0$$

But then we can reduce all polynomials down to polynomials of degree at most $d-1$. If the coefficient field has size q, the collection of polynomials of degree less than d, $\mathbb{F}_{<d}[x]$, has size q^d.

In particular if $\mathbb{F} = \mathbb{Z}_p$ for some prime p we get p^d reduced polynomials.

And Operations?

We want to use $\mathbb{F}_{<d}[x]$ as the carrier set for our extension field $\mathbb{F} \subseteq \mathbb{K}$. What are the operations?

- Addition is simply addition of polynomials in $\mathbb{F}[x]$.
- Multiplication is multiplication of polynomials in $\mathbb{F}[x]$ followed by a reduction: we have to apply the simplification rule until we get back to a polynomial of degree less than d.

We have an algorithm, but we need to work out the algebraic meaning of all of this.

Our simplification process induces an equivalence relation on $\mathbb{F}[x]$: two polynomials are equivalent if they reduce to the same polynomial in $\mathbb{F}_{<d}[x]$.

In fact, we get a congruence $\approx$: our simplification is compatible with the field operations.

So we can form a quotient ring, which turns out to be exactly the field we are looking for:

$$\mathbb{K} = \mathbb{F}[x]/\approx$$

Definition

Let R be a commutative ring. An ideal $I \subseteq R$ is a subset that is closed under addition and under multiplication by arbitrary ring elements: $a \in I, b \in R$ implies $ab \in I$.

So an ideal is much more constrained than a subring: it has to be closed by multiplication from the outside. Ideals are hugely important since they produce congruences and thus allow us to form a quotient structure:

$$a = b \pmod{I} \qquad \text{iff} \qquad a - b \in I.$$

As a consequence, arithmetic in this quotient structure is well-behaved: E.g.

$$a = a', b = b' \pmod{I} \quad \Rightarrow \quad a + b = a' + b', a\,b = a'b' \pmod{I}$$

Suppose $\mathbb{F}$ is a field and consider an irreducible polynomial $f(x)$ and the principal ideal $\big(f(x)\big) = f(x)\mathbb{F}[x]$ that it generates.

We identify two polynomials when their difference is divisible by f:

$$h(x) = g(x) \pmod{f(x)} \quad \Longleftrightarrow \quad f(x) \mid h(x) - g(x)$$

Let d be the degree of f. Then any polynomial h is equivalent to a polynomial g of degree less than d: write $h(x) = q(x)f(x) + g(x)$ by polynomial division.

What is the smallest ideal containing elements $a_1, \ldots, a_k \in R$?

All we need is linear combinations: the ideal generated by $a_1, \ldots, a_k$ is

$$(a_1, \ldots, a_k) = \{ r_1 a_1 + \ldots + r_k a_k \mid r_i \in R \}$$

In particular for $k = 1$ we have

$$(a) = \{ ra \mid x \in R \}$$

This is the principal ideal generated by a.

The ideals $\{0\}$ and R are called trivial, all others are proper.

Note that a field is a commutative ring that has no proper ideals.

Definition

A principal ideal domain (PID) is an integral domain, all of whose ideals are principal.

Important examples of PIDs are

- the integers $\mathbb{Z}$ (think GCD)
- the Gaussian integers $\mathbb{Z}[\mathbf{i}]$
- a polynomial ring $\mathbb{F}[x]$ where $\mathbb{F}$ is a field

Counterexamples: $\mathbb{Z}[x]$ and $\mathbb{F}[x, y]$ both fail to be PIDs.

Suppose we have an extension $\mathbb{F} \subseteq \mathbb{K}$ with $\alpha \in \mathbb{K}$ algebraic over $\mathbb{F}$. Let

$$I = \{ f \in \mathbb{F}[x] \mid f(\alpha) = 0 \}$$

Then I is an ideal and we must have $I = (g)$.

The polynomial g has minimal degree among all the annihilators of α, and we may safely assume that g is monic.

Definition

This polynomial g is the minimal polynomial of α over $\mathbb{F}$.

In algebra it is important to come up with the right notion of substructure: just picking a subset that is closed under the algebraic operations is often not very interesting.

- For groups, normal subgroups are arguably more important than plain subgroups.
- For rings, ideals are arguably more important than subrings.
- But for vector spaces, sub-vector-spaces are just the right notion.

Ideals provide the right type of equivalence relation for the construction of a finite field from a polynomial ring. Alas, the ideals cannot be chosen arbitrarily, we need to start from special polynomials, in analogy to the modulus being prime in the integer case.

Definition

A polynomial is irreducible if it is not the product of polynomials of smaller degree.

Irreducibility is necessary when we try to construct a field $\mathbb{F}[x]/(f)$: otherwise we do not even get an integral domain.

For suppose $f(x) = f_1(x)f_2(x)$ where both f_1 and f_2 have degree at least 1. Then $1 \leq \deg(f_i) < \deg(f)$, so neither f_1 or f_2 can be simplified in $\mathbb{F}[x]/(f)$.

In particular both elements in $\mathbb{F}[x]/(f)$ are non-zero, but their product is zero.

Fix some prime p.

Question:
How many irreducible polynomials of degree m are there in $\mathbb{F}_p[x]$?

Let's write I_m^p for this number, so trivially $I_m^p \leq p^m$.

Lemma (Gauss)

$$I_m^p = \frac{1}{m} \sum_{d|m} \mu(m/d)\, p^d$$

Recall the Möbius function μ:

$$\mu(n) = \begin{cases} +1 & \text{if } n \text{ square-free, even number of prime factors} \\ -1 & \text{if } n \text{ square-free, odd number of prime factors} \\ 0 & \text{otherwise.} \end{cases}$$

One can show that

$$(p^m - 2p^{m/2})/m \leq I_m^p \leq p^m/m$$

E.g., $I_{50}^2 = 22517997465744$, about 2 percent.

Here are some numerical values for characteristic 2.

1–5	2	1	2	3	6
6–10	9	18	30	56	99
11–15	186	335	630	1161	2182
16–20	4080	7710	14532	27594	52377

Example

$$x, 1+x$$

$$1+x+x^2$$

$$1+x^2+x^3, 1+x+x^3$$

$$1+x^3+x^4, 1+x+x^4, 1+x+x^2+x^3+x^4$$

$$1+x^3+x^5, 1+x^2+x^5, 1+x^2+x^3+x^4+x^5, 1+x+x^3+x^4+x^5$$

$$1+x+x^2+x^4+x^5, 1+x+x^2+x^3+x^5$$

All irreducibles in $\mathbb{F}_2[x]$ up to degree 5.

Lemma

$$x^{p^k} - x = \prod \big(f \mid f \text{ monic, irreducible, } \deg(f) \mid k \big)$$

There is a fairly good test for irreducibility that assumes we have access to the prime factors of m (a reasonable assumption).

Theorem (Rabin)

Suppose $f \in \mathbb{F}_p[x]$ is a monic polynomial of degree m. Then f is irreducible iff f divides $x^{p^d} - x$ but f and $x^{p^{d/q}} - x$ are coprime for all prime divisors q of m.

Another Example

Over $\mathbb{F}_2$, the polynomial

$$f(x) = x^3 + x + 1$$

is irreducible. Let $I = \big(f(x)\big)$ be the ideal generated by f.

The first few powers of x modulo I are:

$$1, x, x^2, x+1, x^2+x, x^2+x+1, x^2+1$$

These are actually all polynomials of degree less-than 3, except 0.

So $\mathbb{F}_{<3}[x]$ forms an integral domain, and hence a field, if multiplication is understood modulo I.

OK, but where is the root of f?

We write α for (the equivalence class of) x for emphasis, $\alpha = x \bmod f(x)$.

Then $\alpha \in \mathbb{K}$ is a root of f in the extension field $\mathbb{K}$.

Why? We have by brute force

$$f(\alpha) = x^3 + x + 1 = 0 \pmod{I}$$

Yes, this is a bit lame. One would have hoped for some kind of fireworks, some clever way of writing down the root in terms of some fancy polynomial.

But, it's really no different from the $\sqrt{2}$ example, just less familiar.

Again, algebraically, it is best to think of the extension field $\mathbb{F}_2 \subseteq \mathbb{K}$ as a quotient structure, as the polynomials modulo f:

$$\mathbb{K} = \mathbb{F}_2[x]/\big(f(x)\big)$$

With a view towards algorithms, we can make things more combinatorial by keeping track of coefficient vectors, in this case

$$c_2x^2 + c_1x + c_0 \rightsquigarrow (c_2, c_1, c_0)$$

where $c_i \in \mathbb{F}_2$ is just a single bit.

In this setting the additive structure is trivial: it's just componentwise addition of these triples mod 2.

$$(c_2, c_1, c_0) + (c_2', c_1', c_0') = (c_2 + c_2', c_1 + c_1', c_0 + c_0')$$

As observed before, the additive group of these fields is just a Boolean group. Note that this operation is trivial to implement (xor on bit-vectors, can even be done in 32 or 64 bit blocks).

For other characteristics, though, we have to use modular numbers.

How about multiplication? Since multiplication increases the degree, we can't just multiply out, but we have to simplify using our rule $x^3 \to x + 1$ afterwards.

The product

$$(c_2, c_1, c_0) \cdot (c'_2, c'_1, c'_0) = (d_2, d_1, d_0)$$

is given by the coefficient triple

$$\begin{aligned} d_2 &= c_2\, c'_0 + c_1\, c'_1 + c_0\, c'_2 + c_2\, c'_2 \\ d_1 &= c_1\, c'_0 + c_0\, c'_1 + c_2\, c'_1 + c_1\, c'_2 + c_2\, c'_2 \\ d_0 &= c_0\, c'_0 + c_2\, c'_1 + c_1\, c'_2 \end{aligned}$$

This is a bit messy, and it gets more messy when we deal with larger degree polynomials. Still, we could hard-wire a circuit.

Recall that α is the equivalence class of x. We have already checked that α is the generator of $\mathbb{F}^\times$. Here are the corresponding vector representations.

$$\begin{aligned}
\alpha^0 &= 1 && = (0,0,1)\\
\alpha^1 &= \alpha && = (0,1,0)\\
\alpha^2 &= \alpha^2 && = (1,0,0)\\
\alpha^3 &= \alpha + 1 && = (0,1,1)\\
\alpha^4 &= \alpha^2 + \alpha && = (1,1,0)\\
\alpha^5 &= \alpha^2 + \alpha + 1 && = (1,1,1)\\
\alpha^6 &= \alpha^2 + 1 && = (1,0,1)
\end{aligned}$$

Careful, though, it is in general **not** the case that α generates the whole multiplicative group.

For this to work, we need to choose particular irreducible polynomials in our construction, so-called primitive polynomials.

For example, there are 9 monic irreducibles of degree 6 in $\mathbb{F}_2[x]$:

$$1+x^5+x^6, 1+x^3+x^6, 1+x^2+x^4+x^5+x^6, 1+x^2+x^3+x^5+x^6, 1+x+x^6,$$
$$1+x+x^4+x^5+x^6, 1+x+x^3+x^4+x^6, 1+x+x^2+x^5+x^6, 1+x+x^2+x^4+x^6$$

But 3 of them fail to be primitive:

$$1+x^3+x^6, 1+x+x^2+x^4+x^6, 1+x^2+x^4+x^5+x^6$$

We really obtain a field this way, not just some ring.

	h	h^{-1}
1	1	1
2	α	$1+\alpha^2$
3	α^2	$1+\alpha+\alpha^2$
4	$1+\alpha$	$\alpha+\alpha^2$
5	$1+\alpha^2$	α
6	$\alpha+\alpha^2$	$1+\alpha$
7	$1+\alpha+\alpha^2$	α^2

This table duly defines an involution: $(h^{-1})^{-1} = h$.