Math 533 Winter 2021, Lecture 14: Finite fields

website: `https://www.cip.ifi.lmu.de/~grinberg/t/21w/`

1. Finite fields (cont'd)

1.1. The char of a field (cont'd)

Remember how we defined the characteristic of a field:

Definition 1.1.1. Let F be a field. The **characteristic** of F is an integer called $\operatorname{char} F$, which is defined as follows:

- If there exists a positive integer n such that $n \cdot 1_F = 0$, then $\operatorname{char} F$ is defined to be the **smallest** such n.
- If such an n does not exist, then $\operatorname{char} F$ is defined to be 0.

Roughly speaking, $\operatorname{char} F$ is "how often you have to add 1_F to itself to obtain 0" (with the caveat that we define it to be 0 if you never obtain 0 by adding 1_F to itself). We refer to Lecture 13 for examples.

What does a characteristic satisfy in general?

Theorem 1.1.2 (Properties of characteristics)**.** Let F be a field. Let $p = \operatorname{char} F$. Then:

(a) The field F is a $\mathbb{Z}/p$-algebra. (Remember: $\mathbb{Z}/0 \cong \mathbb{Z}$.)

(b) We have $pa = 0$ for each $a \in F$.

(c) The number p is either prime or 0.

(d) If F is finite, then p is a prime.

(e) If F is finite, then $|F| = p^m$ for some positive integer m.

(f) If p is a prime, then F contains "a copy of $\mathbb{Z}/p$" (meaning: a subring isomorphic to $\mathbb{Z}/p$).

(g) If $p = 0$, then F contains "a copy of $\mathbb{Q}$" (meaning: a subring isomorphic to $\mathbb{Q}$): namely, the map

$$\begin{aligned} &\mathbb{Q} \to F, \\ &\frac{a}{b} \mapsto \frac{a \cdot 1_F}{b \cdot 1_F} \qquad (\text{for } a, b \in \mathbb{Z} \text{ with } b \neq 0) \end{aligned}$$

is an injective ring morphism.

Proof. We have $p \cdot 1_F = 0$. Indeed, if $p = 0$, then this is obvious; but otherwise it follows from the definition of $\operatorname{char} F$.

(b) Let $a \in F$. Then, $a = 1_F \cdot a$. Thus,

$$pa = p\left(1_F \cdot a\right) = \underbrace{\left(p \cdot 1_F\right)}_{=0} \cdot a = 0 \cdot a = 0.$$

This proves Theorem 1.1.2 **(b)**.

(a) We define an action of the ring $\mathbb{Z}/p$ on F by

$$\overline{k} \cdot a = ka \qquad \text{for all } k \in \mathbb{Z} \text{ and } a \in F.$$

Why is this well-defined? In other words, why is it true that if two integers k and ℓ satisfy $\overline{k} = \overline{\ell}$, then $ka = \ell a$ for all $a \in F$?

Let us check this directly: Let k and ℓ be two integers satisfying $\overline{k} = \overline{\ell}$ in $\mathbb{Z}/p$. This means $k \equiv \ell \bmod p$, so that $k - \ell$ is a multiple of p. That is, $k - \ell = pu$ for some $u \in \mathbb{Z}$. Consider this u. Now,

$$ka - \ell a = \underbrace{\left(k - \ell\right)}_{=pu} a = pua = 0$$

(by Theorem 1.1.2 **(b)**, applied to ua instead of a). Thus, $ka = \ell a$, which is precisely what we wanted to prove.

Thus, the action of $\mathbb{Z}/p$ on F is well-defined. Now, it remains to show that F is a $\mathbb{Z}/p$-module, and that the "scale-invariance" axiom is satisfied. All of this is easy and LTTR[1]. Thus, F becomes a $\mathbb{Z}/p$-algebra. This proves Theorem 1.1.2 **(a)**.

(c) Assume the contrary. Thus, p is neither a prime nor 0. Hence, p is either 1 or a composite[2] positive integer (since p is always a nonnegative integer).

Since F is a field, we have $1 \neq 0$ in F. In other words, $1_F \neq 0_F$. If we had $p = 1$, then we would thus have $\underbrace{p}_{=1} \cdot 1_F = 1 \cdot 1_F = 1_F \neq 0_F$, which would contradict $p \cdot 1_F = 0$. Thus, we cannot have $p = 1$. Hence, p must be composite (since p is either 1 or composite). In other words, $p = uv$ for some integers $u > 1$ and $v > 1$. Consider these integers u and v.

[1] For example, let us prove the associativity law, which says that $(rs)\, m = r\,(sm)$ for all $r, s \in \mathbb{Z}/p$ and $m \in F$. Indeed, let $r, s \in \mathbb{Z}/p$ and $m \in F$. Write r and s as $\overline{k}$ and $\overline{\ell}$ for some integers k and ℓ. Then, $rs = \overline{k} \cdot \overline{\ell} = \overline{k\ell}$, so that $(rs)\, m = \overline{k\ell} \cdot m = k\ell m$ (by our definition of the action of $\mathbb{Z}/p$ on F). Also, from $r = \overline{k}$ and $s = \overline{\ell}$, we obtain

$$\begin{aligned} r\,(sm) &= \overline{k} \cdot \left(\overline{\ell} \cdot m\right) = k\left(\overline{\ell} \cdot m\right) && \text{(by our definition of the action)} \\ &= k\,(\ell m) && \left(\text{since our definition of the action yields } \overline{\ell} \cdot m = \ell m\right) \\ &= k\ell m. \end{aligned}$$

Comparing this with $(rs)\, m = k\ell m$, we obtain $(rs)\, m = r\,(sm)$, qed.

[2] A positive integer is said to be *composite* if it can be written as a product of two integers each larger than 1.

From $u > 1$ and $v > 1$ and $p = uv$, we see that both integers u and v are smaller than p. Hence, neither $u \cdot 1_F$ nor $v \cdot 1_F$ can be 0 (since $p = \operatorname{char} F$ was defined to be the **smallest** positive integer n such that $n \cdot 1_F = 0$). Since F is an integral domain (because F is a field), this yields that the product $(u \cdot 1_F) \cdot (v \cdot 1_F)$ is also nonzero.

Now, $p \cdot 1_F = 0$, so

$$0 = \underbrace{p}_{=uv} \cdot 1_F = uv \cdot 1_F = (u \cdot 1_F) \cdot (v \cdot 1_F).$$

This contradicts the fact that the product $(u \cdot 1_F) \cdot (v \cdot 1_F)$ is nonzero. This proves Theorem 1.1.2 **(c)**.

(d) Assume that F is finite. We must show that p is a prime.

According to Theorem 1.1.2 **(c)**, it suffices to show that $p \neq 0$. So let us show this. Assume the contrary. Then, $p = 0$. Hence, none of the elements $1 \cdot 1_F$, $2 \cdot 1_F$, $3 \cdot 1_F$, ... of F is 0 (by the definition of $\operatorname{char} F$). But F is finite, so two of these elements must be equal (by the Pigeonhole Principle). In other words, there exist positive integers $u < v$ such that $u \cdot 1_F = v \cdot 1_F$. Consider these u and v. Then, $v - u$ is a positive integer, and we have $(v - u) \cdot 1_F = v \cdot 1_F - u \cdot 1_F = 0$ (since $u \cdot 1_F = v \cdot 1_F$). But $(v - u) \cdot 1_F$ is one of the elements $1 \cdot 1_F, 2 \cdot 1_F, 3 \cdot 1_F, \ldots$ (since $u < v$), and we just said that none of these elements is 0. This contradicts $(v - u) \cdot 1_F = 0$. Thus, our assumption was false; hence, Theorem 1.1.2 **(d)** is proven.

(e) Assume that F is finite. Thus, by Theorem 1.1.2 **(d)**, we know that p is prime.

Since F is a field, we have $1 \neq 0$ in F. Hence, $|F| > 1$.

From part **(a)**, we know that F is a $\mathbb{Z}/p$-algebra. Thus, in particular, F is a $\mathbb{Z}/p$-module. But since p is prime, $\mathbb{Z}/p$ is a field.

Now, recall that a module over a field is nothing but a vector space. In particular, every module over a field is free (since any vector space has a basis[3]). Thus, in particular, the $\mathbb{Z}/p$-module F is free. In other words, the $\mathbb{Z}/p$-module F has a basis. This basis must be finite (since F itself is finite). Thus, $F \cong (\mathbb{Z}/p)^m$ as $\mathbb{Z}/p$-modules for some $m \in \mathbb{N}$. Consider this m. From $F \cong (\mathbb{Z}/p)^m$, we obtain $|F| = \left|(\mathbb{Z}/p)^m\right| = |\mathbb{Z}/p|^m = p^m$. It remains to prove that m is positive. But this is easy: If m was 0, then $|F| = p^m$ would imply $|F| = p^0 = 1$, which would contradict $|F| > 1$. Thus, the proof of Theorem 1.1.2 **(e)** is complete.

(f) We will be very brief, since we won't use Theorem 1.1.2 **(f)** in what follows.

[3]Once again, I haven't actually proved this fact in this course, but you can easily bridge this gap yourself or look it up in any text on linear algebra (or in Keith Conrad's `https://kconrad.math.uconn.edu/blurbs/linmultialg/dimension.pdf`). Our situation is simpler than the general case, since we know that F is finite, so it is clear that there is a finite list of vectors in F that span F (because you can just take a list of **all** elements of F). In order to obtain a basis from such a list, you only need to successively remove vectors that are linear combinations of other vectors; once no such vectors remain, the list will be a basis.

Assume that p is a prime. Then, F is a $\mathbb{Z}/p$-algebra (by Theorem 1.1.2 **(a)**), so we can define a map

$$\begin{aligned} \mathbb{Z}/p &\to F, \\ \alpha &\mapsto \alpha \cdot 1_F. \end{aligned}$$

It is straightforward to check that this map is a ring morphism; furthermore, it is easily seen to be injective[4]. Hence, its image is a subring of F that is isomorphic to $\mathbb{Z}/p$. This proves Theorem 1.1.2 **(f)**.

(g) Again, we will omit the details, since we won't use the fact.

Assume that $p = 0$. Then, for any nonzero integer b, the element $b \cdot 1_F$ of F is nonzero (why?) and therefore a unit of F (since F is a field). Hence, for any rational number $\frac{a}{b} \in \mathbb{Q}$ (written in such a way that $a, b \in \mathbb{Z}$ and $b \neq 0$), the element $\frac{a \cdot 1_F}{b \cdot 1_F} \in F$ is well-defined. Now, of course, the representation of a rational number as $\frac{a}{b}$ with $a, b \in \mathbb{Z}$ is not unique (for instance, $\frac{6}{4}$ and $\frac{3}{2}$ are the same rational number); however, it is not hard to show that $\frac{a \cdot 1_F}{b \cdot 1_F}$ is uniquely determined by $\frac{a}{b}$ (meaning that if $a, b, c, d \in \mathbb{Q}$ satisfy $\frac{a}{b} = \frac{c}{d}$, then we also have $\frac{a \cdot 1_F}{b \cdot 1_F} = \frac{c \cdot 1_F}{d \cdot 1_F}$). Thus, the map

$$\begin{aligned} \mathbb{Q} &\to F, \\ \frac{a}{b} &\mapsto \frac{a \cdot 1_F}{b \cdot 1_F} \qquad (\text{for } a, b \in \mathbb{Z} \text{ with } b \neq 0) \end{aligned}$$

is well-defined. Next, it can be shown that this map is a ring morphism and is injective[5]. Hence, its image is a subring of F that is isomorphic to $\mathbb{Z}/p$. This proves Theorem 1.1.2 **(g)**. □

Parts **(f)** and **(g)** of Theorem 1.1.2 show that any field F has at its "core" a "small" field: either (a copy of) $\mathbb{Z}/p$ (if its characteristic is a prime p) or (a copy of) $\mathbb{Q}$ (if its characteristic is 0).

Note that parts **(d)** and **(e)** of Theorem 1.1.2 (in combination) show that the size of any finite field is a power of a prime. Thus, there are no finite fields of size 6 or 10 or 12.

[4] This is actually best understood as a particular case of the following general fact: **Any** ring morphism from a field to a nontrivial ring is injective!

The proof of this general fact is actually pretty easy: If $f : K \to R$ is a ring morphism from a field K to a nontrivial ring R, then any $a \in \operatorname{Ker} f$ must be 0, because otherwise a would be a unit of K (since K is a field) and therefore $f(a)$ would be a unit of R (since ring morphisms send units to units); but $f(a) = 0$ cannot be a unit of R (since R is nontrivial). Thus, $\operatorname{Ker} f \subseteq \{0\}$, so that $\operatorname{Ker} f = \{0\}$ and therefore f is injective.

[5] The injectivity follows just as in part **(f)**.

Thus, we can limit our search for finite fields to those of size p^m for p prime and $m > 0$. We have already found such fields for $m = 1$ and for $m = 2$ (for all p), and briefly hinted at the cases $m = 3$ and $m = 4$, but we are still missing the case of general m.

1.2. Tools

We will approach the general case indirectly (no easy and direct proofs are known). We will need a bunch of tools. The first is the notion of a **splitting field**. We begin with a definition:

Definition 1.2.1. Let R be a commutative ring. Let $b \in R[x]$ be a polynomial over R. We say that b **splits** over R if there exist elements $r_1, r_2, \ldots, r_m$ of R such that

$$b = (x - r_1)(x - r_2) \cdots (x - r_m).$$

Note that in this definition, we must necessarily have $\deg b = m$ (unless R is trivial). Also, a polynomial cannot split unless it is monic. This might differ from how other authors define the notion of "splitting", but it is sufficient for what we will do with it.

Example 1.2.2. **(a)** The polynomial $x^2 - 1$ splits over $\mathbb{Q}$, since

$$x^2 - 1 = (x - 1)(x + 1) = (x - 1)(x - (-1)).$$

(b) The polynomial $x^2 + 1$ does not split over $\mathbb{R}$ (since it has no roots in $\mathbb{R}$), but it splits over $\mathbb{C}$, since

$$x^2 + 1 = (x - i)(x + i) = (x - i)(x - (-i)).$$

(c) The polynomial x^2 splits over $\mathbb{Q}$, since $x^2 = xx = (x - 0)(x - 0)$.

(d) The polynomial $x^4 - 9$ does not split over $\mathbb{R}$. Indeed, it has a factorization

$$x^4 - 9 = \left(x - \sqrt{3}\right)\left(x + \sqrt{3}\right)\left(x^2 + 3\right),$$

but the $x^2 + 3$ factor is still not of the form $x - r$. However, this polynomial does split over $\mathbb{C}$, since

$$x^4 - 9 = \left(x - \sqrt{3}\right)\left(x + \sqrt{3}\right)\left(x - \sqrt{3}i\right)\left(x + \sqrt{3}i\right).$$

(e) Any monic polynomial of degree 1 automatically splits over whatever field it is defined over. So does the constant polynomial 1 (since it is an empty product).

When a polynomial splits over a field, its roots can be read off directly from the splitting:

Proposition 1.2.3. Let F be a field. Let $r_1, r_2, \ldots, r_m \in F$. Then,

$$\begin{aligned}&\{\text{the roots of the polynomial } (x-r_1)(x-r_2)\cdots(x-r_m) \in F[x] \text{ in } F\}\\ &= \{r_1, r_2, \ldots, r_m\}.\end{aligned}$$

Proof. The ring F is a field, thus an integral domain. Thus, a product uv of two elements $u, v \in F$ is zero if and only if one of its factors is zero. Hence, a finite product $u_1 u_2 \cdots u_k$ of elements of F is zero if and only if one of its factors is zero[6].

Now, we have

$$\begin{aligned}&\{\text{the roots of the polynomial } (x-r_1)(x-r_2)\cdots(x-r_m) \in F[x] \text{ in } F\}\\ &= \{a \in F \mid ((x-r_1)(x-r_2)\cdots(x-r_m))[a] = 0\}\\ &\qquad \text{(by the definition of a "root")}\\ &= \{a \in F \mid (a-r_1)(a-r_2)\cdots(a-r_m) = 0\}\\ &\qquad \left(\begin{array}{c}\text{since the evaluation } ((x-r_1)(x-r_2)\cdots(x-r_m))[a]\\ \text{equals } (a-r_1)(a-r_2)\cdots(a-r_m)\end{array}\right)\\ &= \{a \in F \mid \text{one of } a-r_1, a-r_2, \ldots, a-r_m \text{ is zero}\}\\ &\qquad \left(\begin{array}{c}\text{since a finite product } u_1u_2\cdots u_k \text{ of elements of } F \text{ is zero}\\ \text{if and only if one of its factors is zero}\end{array}\right)\\ &= \{a \in F \mid a = r_1 \text{ or } a = r_2 \text{ or } \cdots \text{ or } a = r_m\}\\ &= \{r_1, r_2, \ldots, r_m\}.\end{aligned}$$

This proves Proposition 1.2.3. □

Remark 1.2.4. It is worth noting that Proposition 1.2.3 still holds if we replace "field" by "integral domain" (and the same proof applies); but it does not hold when F is just a general commutative ring. For example, if $F = \mathbb{Z}/4$, then the polynomial $(x-0)(x-0)(x-1)(x-3) \in F[x]$ has roots $0, 1, 2, 3$, rather than just $0, 1, 3$ as Proposition 1.2.3 would predict.

The Fundamental Theorem of Algebra says that each monic univariate polynomial over $\mathbb{C}$ splits over $\mathbb{C}$. This is not actually a theorem of algebra, since it relies on the definition of $\mathbb{C}$ (which is analytic); however, it explains some of the significance of $\mathbb{C}$. In general, a field F is said to be **algebraically closed** if each monic univariate polynomial over F splits over F. The field $\mathbb{C}$ is not the only algebraically closed field, but it is perhaps the best-known.

We won't need algebraically closed fields in this course; we will need a more "local" notion: that of a splitting field. To introduce it, we make a simple

[6] Indeed, this follows easily by induction on k, using the preceding sentence in the induction step.

observation, which we have already (tacitly) used in Example 1.2.2 (as we have been treating the same polynomial x^2+1 first as a polynomial in $\mathbb{R}[x]$ and then as a polynomial in $\mathbb{C}[x]$):

Proposition 1.2.5. Let S be a commutative ring. Let R be a subring of S. Then, any polynomial over R automatically is a polynomial over S as well (since its coefficients lie in R and therefore also lie in S), and thus the polynomial ring $R[x]$ becomes a subring of $S[x]$.

For example, $\mathbb{R}[x]$ is a subring of $\mathbb{C}[x]$. Polynomials like x^2+1 might not split over $\mathbb{R}$, but they split over $\mathbb{C}$. This suggests that if a monic polynomial does not split over a ring, we might fix this by making the ring larger ("extending" the ring, possibly by "adjoining" some roots), just as $\mathbb{C}$ was constructed from $\mathbb{R}$ in order to make x^2+1 split. Thus we make the following definition:

Definition 1.2.6. Let F be a field. Let $b \in F[x]$ be a monic polynomial over F. Then, a **splitting field** of b (over F) means a field S such that

- F is a subring of S;
- the polynomial b (regarded as a polynomial in $S[x]$) splits over S.

Examples:

- $\mathbb{C}$ is a splitting field of x^2+1 over $\mathbb{R}$.
- $\mathbb{C}$ is a splitting field of x^2-2 over $\mathbb{Q}$, but so is $\mathbb{R}$ (since x^2-2 already splits over $\mathbb{R}$) or even the smaller field $\mathbb{Q}\left[\sqrt{2}\right] = \left\{a+b\sqrt{2} \mid a,b \in \mathbb{Q}\right\}$.
- $\mathbb{Q}$ itself is a splitting field of x^2-1 over $\mathbb{Q}$.

(Be careful with the literature: Many authors have a more restrictive concept of a "splitting field", which requires not only that the polynomial split over it, but also that the field – in some reasonable way – is minimal with this property. For example, these authors do not accept $\mathbb{R}$ as a splitting field of x^2-2 over $\mathbb{Q}$, since the much smaller field $\mathbb{Q}\left[\sqrt{2}\right]$ suffices to split the polynomial. But our definition suffices for our purposes.)

The most important fact about splitting fields is that they always exist:

Theorem 1.2.7. Let F be a field. Let $b \in F[x]$ be a monic polynomial over F. Then:

(a) We can write b as a product $b = c_1 c_2 \cdots c_k$ of monic irreducible polynomials $c_1, c_2, \ldots, c_k \in F[x]$.

(b) If $\deg b > 0$, then there is a field that contains F as a subring and that contains a root of b.

(c) There exists a splitting field of b over F.

Proof. **(a)** Any nonzero polynomial in $F[x]$ can be made monic by scaling it with a nonzero scalar (namely, if $g \in F[x]$ is a nonzero polynomial, and if c is its leading coefficient, then $c^{-1}g$ is a monic polynomial). This scaling does not interfere with its divisibility properties; thus, if g is irreducible, then it remains so after the scaling.

Hence, it suffices to show that we can write b as a product $b = c_1 c_2 \cdots c_k$ of irreducible polynomials $c_1, c_2, \ldots, c_k \in F[x]$.

Abstractly, this follows easily from the fact that $F[x]$ is a UFD. In a more down-to-earth manner, this can be shown just like the classical fact that each positive integer can be written as a product of primes. The proof proceeds by strong induction on $\deg b$; the main idea is "either b is itself irreducible, in which case we are done; or b can be written as a product of two polynomials of smaller degree, in which case the induction hypothesis applies".

(Note that this proof is constructive when F is finite, since we can actually try out all polynomials of degree smaller than $\deg b$ and check which of them divide b.)

Theorem 1.2.7 **(a)** is thus proved.

(b) Assume that $\deg b > 0$. We must find a field that contains F as a subring and that contains a root of b.

It is tempting to take $F[x]/b$, but this might fail to be a field (since b might fail to be irreducible).

Instead, we use Theorem 1.2.7 **(a)** to write b as a product $b = c_1 c_2 \cdots c_k$ of monic irreducible polynomials $c_1, c_2, \ldots, c_k \in F[x]$, and then we take the field $F[x]/c_1$ (which is indeed a field, because c_1 is irreducible[7]). This field will contain a root of c_1, and thus also contain a root of b (since a root of c_1 is always a root of b). So Theorem 1.2.7 **(b)** is proved.

(Where did I use the assumption $\deg b > 0$ in this proof? Hint: Why is there a c_1 ?)

(c) Here is a proof by example: Assume that $\deg b = 3$.

Theorem 1.2.7 **(b)** says that there is a field F' that contains F as a subring and that contains a root of b. Consider this F', and let r_1 be the root of b that it contains. Thus, $x - r_1 \mid b$ in $F'[x]$ (since r_1 is a root of b). Hence, the polynomial $\dfrac{b}{x - r_1} \in F'[x]$ is well-defined. Moreover, this polynomial $\dfrac{b}{x - r_1}$ has degree $3 - 1 = 2$ and is monic[8].

Now, we apply Theorem 1.2.7 **(b)** again, but this time to the field F' and the monic polynomial $\dfrac{b}{x - r_1}$ over it. Thus we conclude that there is a field F'' that contains F' as a subring and that contains a root of $\dfrac{b}{x - r_1}$. Consider this F'',

[7] We are using a result from Lecture 13 here.

[8] Here, we are using the fact that when we divide a monic polynomial by a monic polynomial, the quotient will again be monic. (The proof is LTTR. Note that this holds even if there is a remainder!)

and let r_2 be the root of $\dfrac{b}{x-r_1}$ that it contains. Thus, $x - r_2 \mid \dfrac{b}{x-r_1}$ in $F''[x]$ (since r_2 is a root of $\dfrac{b}{x-r_1}$). Hence, the polynomial $\dfrac{b}{x-r_1} / (x-r_2) \in F''[x]$ is well-defined. In other words, the polynomial $\dfrac{b}{(x-r_1)(x-r_2)} \in F''[x]$ is well-defined. Moreover, this polynomial $\dfrac{b}{(x-r_1)(x-r_2)}$ has degree $3-2=1$ and is monic.

Now, we apply Theorem 1.2.7 **(b)** again, but this time to the field F'' and the monic polynomial $\dfrac{b}{(x-r_1)(x-r_2)}$ over it. Thus we conclude that there is a field F''' that contains F'' as a subring and that contains a root of $\dfrac{b}{(x-r_1)(x-r_2)}$. Consider this F''', and let r_3 be the root of $\dfrac{b}{(x-r_1)(x-r_2)}$ that it contains. Thus, $x - r_3 \mid \dfrac{b}{(x-r_1)(x-r_2)}$ in $F'''[x]$. Hence, the polynomial $\dfrac{b}{(x-r_1)(x-r_2)(x-r_3)} \in F'''[x]$ is well-defined. Furthermore, this polynomial has degree $3-3=0$ and is monic. In other words, this polynomial equals 1. In other words, $b = (x-r_1)(x-r_2)(x-r_3)$ in $F'''[x]$. This shows that b splits over F'''. Moreover, by construction, F''' is a field that contains F as a subring (since $F \subseteq F' \subseteq F'' \subseteq F'''$, and each of these "$\subseteq$" signs is not just a subset but actually a subring).

Thus, we have proved Theorem 1.2.7 **(c)** in our example. Proving it in the general case is just a matter of formalizing what we did as an induction on $\deg b$. □

Next, to something different. The following is a rather surprising property of fields of positive characteristic:

Theorem 1.2.8 (Idiot's Binomial Formula, aka Freshman's Dream)**.** Let p be a prime number. Let F be a field of characteristic p, or, more generally, any commutative $\mathbb{Z}/p$-algebra. Then:

(a) We have $(a+b)^p = a^p + b^p$ for any $a, b \in F$.

(b) We have $(a+b)^{p^m} = a^{p^m} + b^{p^m}$ for any $a, b \in F$ and $m \in \mathbb{N}$.

(c) We have $(a-b)^p = a^p - b^p$ for any $a, b \in F$.

(d) We have $(a-b)^{p^m} = a^{p^m} - b^{p^m}$ for any $a, b \in F$ and $m \in \mathbb{N}$.

For example, for $p = 3$, Theorem 1.2.8 **(a)** says that $(a+b)^3 = a^3 + b^3$. And indeed, we can show this directly: Theorem 1.1.2 **(b)** shows that $3u = 0$ for any $u \in F$ (for $p = 3$), and the binomial formula yields

$$(a+b)^3 = a^3 + \underbrace{3a^2b}_{\substack{=0 \\ \text{(since } 3u=0 \\ \text{for any } u \in F\text{)}}} + \underbrace{3ab^2}_{\substack{=0 \\ \text{(since } 3u=0 \\ \text{for any } u \in F\text{)}}} + b^3 = a^3 + b^3.$$

To prove Theorem 1.2.8 **(a)** in general, we will argue in the same way; we will just need to know that all but the leftmost and rightmost addends in the binomial formula vanish. This is a consequence of the following lemma:

Lemma 1.2.9. Let p be a prime number. Let $k \in \{1, 2, \ldots, p-1\}$. Then, $p \mid \dbinom{p}{k}$.

Note that this does indeed depend on p being a prime. For example, the number 4 is not prime, and we **do not** have $4 \mid \underbrace{\dbinom{4}{2}}_{=6}$.

Proof of Lemma 1.2.9. There is an easy-to-prove formula saying that

$$\binom{p}{k} = \frac{p}{k} \cdot \binom{p-1}{k-1}.$$

Hence,

$$k\binom{p}{k} = p\binom{p-1}{k-1}.$$

Hence, $k\dbinom{p}{k}$ is divisible by p. But k is coprime to p (since p is prime), so we can cancel k from this divisibility, and conclude that $\dbinom{p}{k}$ is divisible by p. Lemma 1.2.9 is proved. □

Proof of Theorem 1.2.8. **(a)** Let $a, b \in F$. Then, $ab = ba$ (since F is commutative); thus, the Binomial Formula yields

$$(a+b)^p = \sum_{k=0}^{p} \binom{p}{k} a^k b^{p-k} = a^p + \sum_{k=1}^{p-1} \binom{p}{k} a^k b^{p-k} + b^p. \tag{1}$$

Now, we claim that all the addends in the sum $\sum\limits_{k=1}^{p-1} \dbinom{p}{k} a^k b^{p-k}$ vanish. Indeed, let $k \in \{1, 2, \ldots, p-1\}$. Then, Lemma 1.2.9 tells us that $\dbinom{p}{k} = mp$ for some $m \in \mathbb{Z}$. Consider this m. Then, each $u \in F$ satisfies $\dbinom{p}{k} u = m \underbrace{pu}_{=0} = m \cdot 0 = 0$. Hence, in particular, we have

(by Theorem 1.1.2 **(b)**)

$$\binom{p}{k} a^k b^{p-k} = 0. \tag{2}$$

Now, forget that we fixed k. We thus have shown that (2) holds for each $k \in \{1, 2, \ldots, p-1\}$. Hence, (1) becomes

$$(a+b)^p = a^p + \sum_{k=1}^{p-1} \underbrace{\binom{p}{k} a^k b^{p-k}}_{\substack{=0 \\ \text{(by (2))}}} + b^p = a^p + b^p.$$

This proves Theorem 1.2.8 **(a)**.

(b) This follows by induction on m using Theorem 1.2.8 **(a)**, since any $u \in F$ satisfies $u^{p^m} = \left(u^{p^{m-1}}\right)^p$.

(c) Let $a, b \in F$. Applying Theorem 1.2.8 **(a)** to $a - b$ instead of a, we get

$$((a-b)+b)^p = (a-b)^p + b^p.$$

Solving this for $(a-b)^p$, we get

$$(a-b)^p = \left(\underbrace{(a-b)+b}_{=a}\right)^p - b^p = a^p - b^p.$$

This proves Theorem 1.2.8 **(c)**.

(d) This follows by induction on m using Theorem 1.2.8 **(c)**. □

Corollary 1.2.10. Let p be a prime number. Let F be a field of characteristic p, or, more generally, any commutative $\mathbb{Z}/p$-algebra. Then, the map

$$\begin{aligned} F &\to F, \\ a &\mapsto a^p \end{aligned}$$

is a ring morphism.

Proof. Theorem 1.2.8 **(a)** says that this map respects addition. But it is also clear that this map respects multiplication (since $(ab)^p = a^p b^p$ for any $a, b \in F$) and respects zero and unity (since $0^p = 0$ and $1^p = 1$). Thus, it is a ring morphism. □

The ring morphism in Corollary 1.2.10 is known as the **Frobenius endomorphism** of F. It exists for arbitrary commutative $\mathbb{Z}/p$-algebras, but it is particularly well-behaved for finite fields. (In particular, it is bijective when F is a finite field; this will be on homework set #4.)

Our last tool is essentially a criterion for a polynomial to have distinct roots. The criterion is in terms of its derivative, which is defined as follows:

Definition 1.2.11. Let R be a commutative ring. Let $f \in R[x]$ be a polynomial. The **derivative** f' of f is a polynomial in $R[x]$ defined as follows: Writing f in the form $f = \sum_{k \in \mathbb{N}} f_k x^k$ for some $f_0, f_1, f_2, \ldots \in R$, we set $f' = \sum_{k>0} f_k k x^{k-1}$.

For example, if $f = 7x^4 + 2x + 3$, then $f' = 7 \cdot 4x^3 + 2 \cdot 1x^0 = 28x^3 + 2$ (where we have, of course, ignored zero coefficients).

Definition 1.2.11 is obviously inspired by the formula for the derivative of a polynomial function in calculus. Unlike in calculus, we are not wasting our time with little εs and convergence issues; instead, we are just defining f' using the explicit formula that probably took you some time to prove back in calculus. There is no free lunch here – with this definition, you cannot re-use anything you have learned about derivatives in your analysis classes (already because you are working in a much more general setting now, with a commutative ring R instead of the real numbers); thus, a host of basic properties of derivatives need to be proven before the notion becomes useful. In particular, the following needs to be shown:

Proposition 1.2.12. Let R be a commutative ring. Let $f, g \in R[x]$. Then:
(a) We have $(f+g)' = f' + g'$.
(b) We have $(fg)' = f'g + fg'$. (This is called the **Leibniz rule**.)

Proof. This is part of homework set #3 exercise 7. □

The following corollary is an algebraic analogue of the well-known fact "a double root of a polynomial is a root of its derivative":

Corollary 1.2.13. Let R be a commutative ring. Let $f \in R[x]$ and $r \in R$. If $(x-r)^2 \mid f$ in $R[x]$, then $x - r \mid f'$ in $R[x]$.

Proof. Assume that $(x-r)^2 \mid f$. Thus, we can write f as $f = (x-r)^2 g$ for some $g \in R[x]$. Consider this g. From $f = (x-r)^2 g$, we obtain

$$f' = \left((x-r)^2 g\right)' = \underbrace{\left((x-r)^2\right)'}_{\substack{=2(x-r) \\ \text{(this is easy to} \\ \text{check directly)}}} g + (x-r)^2 g' \qquad \text{(by the Leibniz rule)}$$

$$= 2(x-r) g + (x-r)^2 g' = (x-r)\left(2g + (x-r) g'\right).$$

Thus, $x - r \mid f'$, so that Corollary 1.2.13 is proven.

(Note that $\left((x-r)^2\right)' = 2(x-r)$ could also be obtained from the chain rule for polynomials, which says – just like the chain rule in calculus – that $(f[g])' = f'[g] \cdot g'$ for any two polynomials $f, g \in R[x]$. But then you would have to prove this chain rule – which is a nice exercise in fact.) □

1.3. Existence of finite fields

Now we are in walking distance of the existence of fields of size p^m:

Theorem 1.3.1. Let p be a prime number. Let m be a positive integer. Then, there exists a finite field of size p^m.

Proof. From $p > 1$ and $m > 0$, we obtain $p^m > 1$. Hence, the polynomial $x^{p^m} - x$ is monic. Thus, by Theorem 1.2.7 **(c)**, there exists a splitting field of this polynomial over $\mathbb{Z}/p$. Let S be such a splitting field. Thus, the polynomial $x^{p^m} - x$ splits over S. In other words, there exist elements $r_1, r_2, \ldots, r_{p^m}$ of S such that

$$x^{p^m} - x = (x - r_1)(x - r_2) \cdots \left(x - r_{p^m}\right). \tag{3}$$

Consider these $r_1, r_2, \ldots, r_{p^m}$.

Let

$$L = \left\{r_1, r_2, \ldots, r_{p^m}\right\}.$$

Our goal will be to show that L is a finite field of size p^m.

Everything in this statement needs proof![9] Even the size is not obvious, let alone that L is a field. Let us start with the size:

Claim 1: We have $|L| = p^m$.

[*Proof:* This amounts to showing that $r_1, r_2, \ldots, r_{p^m}$ are distinct (since this will immediately yield that $L = \left\{r_1, r_2, \ldots, r_{p^m}\right\}$ is a p^m-element set). Let us thus do this. Indeed, assume the contrary. Then, $r_i = r_j$ for some $i < j$. Hence, the $x - r_i$ and $x - r_j$ factors on the right hand side of (3) are identical. Thus, $x - r_i$ appears twice as a factor on this right hand side; consequently, (3) entails that $(x - r_i)^2 \mid x^{p^m} - x$. Hence, Corollary 1.2.13 (applied to $R = S$ and $f = x^{p^m} - x$ and $r = r_i$) yields $x - r_i \mid \left(x^{p^m} - x\right)'$. But Definition 1.2.11 yields

$$\left(x^{p^m} - x\right)' = \underbrace{p^m x^{p^m - 1}}_{\substack{=0 \\ \text{(since } pu=0 \\ \text{for any } u \in S\text{)}}} - 1 = -1.$$

Thus, $x - r_i \mid \left(x^{p^m} - x\right)' = -1 \mid 1$. But it is impossible for the degree-1 polynomial $x - r_i$ to divide the degree-0 polynomial 1 (for degree reasons). So we have found a contradiction.]

Next, let us characterize L somewhat differently:

Claim 2: We have

$$L = \left\{\text{the roots of } x^{p^m} - x \text{ in } S\right\} = \left\{a \in S \mid a^{p^m} - a = 0\right\}$$
$$= \left\{a \in S \mid a^{p^m} = a\right\}.$$

[9] Except for the "finite" part, which is obvious but not overly helpful by itself.

[*Proof of Claim 2:* The equation (3) yields that

$$\begin{aligned}
&\left\{\text{the roots of } x^{p^m} - x \text{ in } S\right\} \\
&= \left\{\text{the roots of } (x - r_1)(x - r_2)\cdots\left(x - r_{p^m}\right) \text{ in } S\right\} \\
&= \left\{r_1, r_2, \ldots, r_{p^m}\right\} \qquad \text{(by Proposition 1.2.3)} \\
&= L.
\end{aligned}$$

Hence,

$$\begin{aligned}
L &= \left\{\text{the roots of } x^{p^m} - x \text{ in } S\right\} = \left\{a \in S \mid a^{p^m} - a = 0\right\} \\
&= \left\{a \in S \mid a^{p^m} = a\right\}.
\end{aligned}$$

This proves Claim 2.]

Now, why is L a field? First, let us check that L is a ring:

Claim 3: The set L is a subring of S.

[*Proof of Claim 3:* Claim 2 yields that

$$L = \left\{a \in S \mid a^{p^m} = a\right\}. \tag{4}$$

Hence, $0 \in L$ (since $0^{p^m} = 0$) and $1 \in L$ (since $1^{p^m} = 1$). Furthermore, I claim that L is closed under addition. Indeed, if $a, b \in L$, then (4) yields $a^{p^m} = a$ and $b^{p^m} = b$, so that

$$\begin{aligned}
(a+b)^{p^m} &= \underbrace{a^{p^m}}_{=a} + \underbrace{b^{p^m}}_{=b} \qquad \text{(by Theorem 1.2.8 } \textbf{(b)}\text{)} \\
&= a + b,
\end{aligned}$$

and this means $a + b \in L$ (again because of (4)). This shows that L is closed under addition For a similar reason, L is closed under subtraction[10], so that L is closed under negation. Finally, L is closed under multiplication, since $(ab)^{p^m} = a^{p^m} b^{p^m}$ for any $a, b \in L$. Hence, L is a subring of S.]

Thus, in particular, L is a commutative ring (since S is a field, thus a commutative ring). Now, let us see that L is a field:

Claim 4: The ring L is a field.

[*Proof of Claim 4:* We know that S is a field, so that $0 \neq 1$ in S, and this of course means that $0 \neq 1$ in L. It thus remains to show that every nonzero element of L is a unit.

[10] Use Theorem 1.2.8 **(d)** instead of Theorem 1.2.8 **(b)** here.

Let $a \in L$ be nonzero. Then, a has an inverse in S, since S is a field. This inverse a^{-1} satisfies $\left(a^{-1}\right)^{p^m} = \left(a^{p^m}\right)^{-1}$ (indeed, this is a particular case of the identity $\left(g^{-1}\right)^k = \left(g^k\right)^{-1}$, which holds whenever g is an element of a group and k is an integer). But $a \in L$ and thus $a^{p^m} = a$ (by (4)). Hence,

$$\left(a^{-1}\right)^{p^m} = \left(\underbrace{a^{p^m}}_{=a}\right)^{-1} = a^{-1},$$

so that $a^{-1} \in L$ (by (4) again). Thus, a has an inverse in L; in other words, a is a unit of L.

Thus, we have shown that every nonzero element of L is a unit. As we said, this finishes the proof of Claim 4.]

Combining Claims 1 and 4, we conclude that L is a field of size p^m. Thus, such a field exists. This proves Theorem 1.3.1. □

So we are done with the first deep result of this course! There is much more to say about finite fields:

- We have obtained L rather indirectly: First, we took a splitting field S of the huge polynomial $x^{p^m} - x$; then we carved L out of it by taking the roots of this polynomial. Could we get L more directly? For example, if there is an irreducible polynomial f of degree m over $\mathbb{Z}/p$, then we can just take the field $(\mathbb{Z}/p)[x]/f$. Is there such an f ?

- Can there be several non-isomorphic fields of size p^m (for fixed p and m)? For example, can there be two non-isomorphic fields of size p^2 ? It is not hard to see that any field of size p^2 can be obtained (up to isomorphism) by adjoining a root of an irreducible quadratic polynomial to $\mathbb{Z}/p$; thus, the question is whether different such polynomials can lead to different fields.

  If we were working with infinite fields, examples of such behavior would be easy to find. For example, adjoining a root of $x^2 - 2$ to $\mathbb{Q}$ yields the field $\mathbb{Q}\left[\sqrt{2}\right]$, whereas adjoining a root of $x^2 - 3$ to $\mathbb{Q}$ yields the field $\mathbb{Q}\left[\sqrt{3}\right]$. It is not hard to see that $\mathbb{Q}\left[\sqrt{2}\right]$ is not isomorphic to $\mathbb{Q}\left[\sqrt{3}\right]$ (for example, you can show that 2 is a square in $\mathbb{Q}\left[\sqrt{2}\right]$ but not in $\mathbb{Q}\left[\sqrt{3}\right]$). Can this happen with $\mathbb{Z}/p$ instead of $\mathbb{Q}$?

You will see some answers on homework set #4.