Extension Fields

Throughout these notes, the letters F, E, K denote fields.

1 Introduction to extension fields

Let F, E be fields and suppose that $F \leq E$, i.e. that F is a subfield of E. We will often view F as the primary object of interest, and in this case refer to E as an *extension field* or simply *extension* of F. For example, $\mathbb{R}$ is an extension field of $\mathbb{Q}$ and $\mathbb{C}$ is an extension field of $\mathbb{R}$.

Now suppose that E is an extension field of F and that $\alpha \in E$. We have the evaluation homomorphism $\mathrm{ev}_\alpha \colon F[x] \to E$, whose value on a polynomial $f(x) \in F[x]$ is $f(\alpha)$. By definition, the image $\operatorname{Im} \mathrm{ev}_\alpha = F[\alpha]$ is a subring of E. It is the smallest subring of E containing both F and α, and it is an integral domain as it is a subring of a field. Note that, by definition,

$$F[\alpha] = \operatorname{Im} \mathrm{ev}_\alpha = \{f(\alpha) : f(x) \in F[x]\}.$$

There are now two cases:

Case I: $\operatorname{Ker} \mathrm{ev}_\alpha = \{0\}$. In other words, if $f(x) \in F[x]$ is a nonzero polynomial, then $f(\alpha) \neq 0$, i.e. α is not the root of any nonzero polynomial in $f(x)$. In this case, we say that α is *transcendental over* F. If α is transcendental over F, then $\mathrm{ev}_\alpha \colon F[x] \to E$ is injective, and hence ev_α is an isomorphism from $F[x]$ to $F[\alpha] \subseteq E$. In particular, $F[\alpha]$ is not a field, since $F[x]$ is not a field. By results on the field of quotients of an integral domain, ev_α extends to an injective homomorphism $\widehat{\mathrm{ev}}_\alpha \colon F(x) \to E$. Clearly, the image of $\widehat{\mathrm{ev}}_\alpha$ is the set of all quotients in E of the form $f(\alpha)/g(\alpha)$, where $f(x), g(x) \in F[x]$ and $g(x) \neq 0$. By general properties of fields of quotients, $f_1(\alpha)/g_1(\alpha) = f_2(\alpha)/g_2(\alpha) \iff f_1(\alpha)g_2(\alpha) = f_2(\alpha)g_1(\alpha) \iff f_1(x)g_2(x) = f_2(x)g_1(x)$. Defining

$$F(\alpha) = \operatorname{Im} \widehat{\mathrm{ev}}_\alpha = \{f(\alpha)/g(\alpha) : f(x), g(x) \in F[x], g(x) \neq 0\},$$

we see that $F(\alpha)$ is a field and it is the smallest subfield of E containing F and α.

For example, if $F = \mathbb{Q}$ and $E = \mathbb{R}$, "most" elements of $\mathbb{R}$ are transcendental over $\mathbb{Q}$. In fact, it is not hard to show that the set of elements of $\mathbb{R}$ which are not transcendental over $\mathbb{Q}$ is countable, and since $\mathbb{R}$ is uncountable there are an uncountable number of elements of $\mathbb{R}$ which are transcendental over $\mathbb{Q}$. It is much harder to show that a given element of $\mathbb{R}$ is transcendental over $\mathbb{Q}$. For example e and π are both transcendental over $\mathbb{Q}$. (The transcendence of π shows that it is impossible to "square the circle," in other words to construct a square with straightedge and compass whose area is π.) Hence, the subring $\mathbb{Q}[\pi]$ of $\mathbb{R}$ is isomorphic to the polynomial ring $\mathbb{Q}[x]$: every element of $\mathbb{Q}[\pi]$ can be uniquely written as a polynomial $\sum_{i=0}^{n} a_i\pi^i$ in π, where the $a_i \in \mathbb{Q}$. The field $\mathbb{Q}(\pi)$ is then the set of all quotients, $f(\pi)/g(\pi)$, where $f(x), g(x) \in \mathbb{Q}[x]$ and $g(x) \neq 0$. Finally, note that the property of transcendence is very much a relative property. Thus, $\pi \in \mathbb{R}$ is transcendental over $\mathbb{Q}$, but π is **not** transcendental over $\mathbb{R}$; in fact, π is a root of the nonzero polynomial $x - \pi \in \mathbb{R}[x]$.

For another example, let F be an arbitrary field and consider $F(x)$, the field of rational functions with coefficients in F. Thus $F(x)$ is the field of quotients of the polynomial ring $F[x]$, and the elements of $F(x)$ are quotients $f(x)/g(x)$, where $f(x), g(x) \in F[x]$ and $g(x) \neq 0$. However, when we think of $F(x)$ as a field in its own right, it is traditional to rename the variable x by some other letter such as t,which we still refer too as an "indeterminate," to avoid confusion with x which we reserve for the "variable" of a polynomial. With this convention, the field $F(t)$ (with t an indeterminate) is an extension field of F. Moreover, $t \in F(t)$ is transcendental over F, since, if $f(x) \in F[x]$ is a nonzero polynomial, then $\mathrm{ev}_t\, f(x) = f(t)$, which is a nonzero element of $F[t]$ and hence of $F(t)$.

Case II: $\operatorname{Ker} \mathrm{ev}_\alpha \neq \{0\}$. In other words, there exists a nonzero polynomial $f(x) \in F[x]$ $f(\alpha) = 0$. In this case, we say that α is *algebraic over F*. This will be the important case for us, so we state the main result as a proposition:

Proposition 1.1. *Suppose that E is an extension field of F and that $\alpha \in E$ is algebraic over F. Then $\operatorname{Ker} \mathrm{ev}_\alpha = (p(x))$, where $p(x) \in F[x]$ is an irreducible polynomial. Moreover, if $f(x) \in F[x]$ is any polynomial such that $f(\alpha) = 0$, then $p(x)\big|f(x)$. The homomorphism ev_α induces an isomorphism, denoted $\widetilde{\mathrm{ev}}_\alpha$, from $F[x]/(p(x))$ to $F[\alpha]$. Finally, $F[\alpha] = \operatorname{Im} \mathrm{ev}_\alpha$ is a field.*

Proof. By hypothesis, $\operatorname{Ker} \mathrm{ev}_\alpha$ is a nonzero ideal in $F[x]$. Moreover, the homomorphism ev_α induces an isomorphism, denoted $\widetilde{\mathrm{ev}}_\alpha$, from $F[x]/\operatorname{Ker} \mathrm{ev}_\alpha$

to $F[\alpha]$, and in particular $F[\alpha] \cong F[x]/\operatorname{Ker}\operatorname{ev}_\alpha$. Since $F[\alpha]$ is a subring of a field, it is an integral domain. Thus $F[x]/\operatorname{Ker}\operatorname{ev}_\alpha$ is also an integral domain, and hence $\operatorname{Ker}\operatorname{ev}_\alpha$ is a prime ideal. But we have seen that every nonzero prime ideal is maximal, hence $F[\alpha]$ is a subfield of E, and that the nonzero prime ideals are exactly those of the form $(p(x))$, where $p(x) \in F[x]$ is an irreducible polynomial. Thus $\operatorname{Ker}\operatorname{ev}_\alpha = (p(x))$ for some irreducible polynomial $p(x) \in F[x]$. By definition, $f(\alpha) = 0 \iff f(x) \in \operatorname{Ker}\operatorname{ev}_\alpha \iff p(x)\big|f(x)$. $\square$

Definition 1.2. Let E be an extension field of F and suppose that $\alpha \in E$ is algebraic over F. We set $F(\alpha) = F[\alpha]$. As in Case I, $F(\alpha)$ is a subfield of E and is the smallest subfield of E containing both F and α.

With E and α as above, suppose that $p_1(x), p_2(x) \in F[x]$ are two polynomials such that $\operatorname{Ker}\operatorname{ev}_\alpha = (p_1(x)) = (p_2(x))$. Then $p_1(x)\big|p_2(x)$ and $p_2(x)\big|p_1(x)$. It is then easy to see that there exists a $c \in F^*$ such that $p_2(x) = cp_1(x)$. In particular, there is a unique monic polynomial $p(x) \in F[x]$ such that $\operatorname{Ker}\operatorname{ev}_\alpha = (p(x))$.

Definition 1.3. Let E be an extension field of F and suppose that $\alpha \in E$ is algebraic over F. The unique monic irreducible polynomial which is a generator of $\operatorname{Ker}\operatorname{ev}_\alpha$ will be denoted $\operatorname{irr}(\alpha, F, x)$.

Thus, if E is an extension field of F and $\alpha \in E$ is algebraic over F, then $\operatorname{irr}(\alpha, F, x)$ is the unique monic irreducible polynomial in $F[x]$ for which α is a root. One way to find $\operatorname{irr}(\alpha, F, x)$ is as follows: suppose that $p(x) \in F[x]$ is an irreducible monic polynomial such that $p(\alpha) = 0$. Then $\operatorname{irr}(\alpha, F, x)$ divides $p(x)$, but since $p(x)$ is irreducible, there exists a $c \in F^*$ such that $p(x) = c\operatorname{irr}(\alpha, F, x)$. Finally, since both $p(x)$ and $\operatorname{irr}(\alpha, F, x)$ are monic, $c = 1$, i.e. $p(x) = \operatorname{irr}(\alpha, F, x)$.

For example, $x^2 - 2 = \operatorname{irr}(\sqrt{2}, \mathbb{Q}, x)$, since $p(x) = x^2 - 2$ is monic and (as we have seen) irreducible and $p(\sqrt{2}) = 0$. As in Case I, the definition of $\operatorname{irr}(\alpha, F, x)$ is relative to the field F. For example, $\operatorname{irr}(\sqrt{2}, \mathbb{Q}(\sqrt{2}), x) = x - \sqrt{2}$. Note that $x - \sqrt{2}$ is a factor of $x^2 - 2$ in $\mathbb{Q}(\sqrt{2})[x]$, but that $x - \sqrt{2}$ is not an element of $\mathbb{Q}[x]$.

One problem with finding $\operatorname{irr}(\alpha, F, x)$ is that we don't have many ways of showing that a polynomial is irreducible. So far, we just know that a polynomial of degree 2 or 3 is irreducible $\iff$ it does not have a root. Here are a few more examples that we can handle by this method:

Example 1.4. (1) $\operatorname{irr}(\sqrt[3]{2}, \mathbb{Q}, x) = x^3 - 2$ since $\sqrt[3]{2}$ is irrational.

(2) There is no element $\alpha \in \mathbb{Q}(\sqrt{2})$ such that $\alpha^2 = 3$ (by a homework problem). In other words, $\sqrt{3} \notin \mathbb{Q}(\sqrt{2})$. Thus $\mathrm{irr}(\sqrt{3}, \mathbb{Q}(\sqrt{2}), x) = x^2 - 3$.

(3) If $\alpha = \sqrt{2} + \sqrt{3}$, then it is easy to check (homework) that α is a root of the polynomial $x^4 - 10x^2 + 1$. Thus $\mathrm{irr}(\alpha, \mathbb{Q}, x)$ divides $x^4 - 10x^2 + 1$, but awe cannot conclude that they are equal unless we can show that $x^4 - 10x^2 + 1$ is irreducible, or by some other method. We will describe one such method in the next section.

Definition 1.5. Let E be an extension field of F. Then we say that E is a *simple extension of* F if there exists an $\alpha \in E$ such that $E = F(\alpha)$. Note that this definition makes sense both in case α is algebraic over F and in case it is transcendental over F. However, we shall mainly be interested in the case where α is algebraic over F.

In many cases, we want to consider extension fields which are not necessarily simple extensions.

Definition 1.6. Let E be an extension field of F and let $\alpha_1, \dots, \alpha_n \in E$. We define $F(\alpha_1, \dots, \alpha_n)$ to be the smallest subfield of E containing F and $\alpha_1, \dots, \alpha_n$. If $E = F(\alpha_1, \dots, \alpha_n)$, we say that E is *generated over* F *by* $\alpha_1, \dots, \alpha_n$. It is easy to see that $F(\alpha_1, \dots, \alpha_n) = F(\alpha_1, \dots, \alpha_{n-1})(\alpha_n)$. In fact, by definition, both sides of this equality are the smallest subfield of E containing F, $\alpha_1, \dots, \alpha_{n-1}$, and α_n. More generally, for every k, $1 \leq k \leq n$, $F(\alpha_1, \dots, \alpha_n) = F(\alpha_1, \dots, \alpha_k)(\alpha_{k+1}, \dots, \alpha_n)$.

Example 1.7. Consider the field $\mathbb{Q}(\sqrt{2}, \sqrt{3})$. Then $\alpha = \sqrt{2} + \sqrt{3} \in \mathbb{Q}(\sqrt{2}, \sqrt{3})$, and hence $\mathbb{Q}(\alpha) \leq \mathbb{Q}(\sqrt{2}, \sqrt{3})$. However, it is another homework problem to show that $\sqrt{2} \in \mathbb{Q}(\alpha)$ and that $\sqrt{3} \in \mathbb{Q}(\alpha)$. Thus $\mathbb{Q}(\sqrt{2}, \sqrt{3}) \leq \mathbb{Q}(\alpha)$ and hence $\mathbb{Q}(\sqrt{2}, \sqrt{3}) = \mathbb{Q}(\alpha)$. In conclusion, a field such as $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ which is not obviously a simple extension may turn out to be a simple extension. We shall analyze this in much greater detail later.

2 Finite and algebraic extensions

Let E be an extension field of F. Then E is an F-vector space.

Definition 2.1. Let E be an extension field of F. Then E is a *finite extension of* F if E is a finite dimensional F-vector space. If E is a finite extension of F, then the positive integer $\dim_F E$ is called the *degree of* E *over* F, and is denoted $[E : F]$. Note that $[E : F] = 1 \iff E = F$.

Proposition 2.2. *Suppose that $E = F(\alpha)$ is a simple extension of F. Then E is a finite extension of $F \iff \alpha$ is algebraic over F. In this case*

$$[E : F] = \deg_F \alpha,$$

where by definition $\deg_F \alpha$ is the degree of $\mathrm{irr}(\alpha, F, x)$. Finally, if α is algebraic over F and $\deg_F \alpha = \mathrm{irr}(\alpha, F, x) = d$, then $1, \alpha, \ldots, \alpha^{d-1}$ is a basis for $F(\alpha)$ as an F-vector space.

Proof. First suppose that α is transcendental over F. Then we have seen that $F \leq F[\alpha] \leq F(\alpha)$, and that $\mathrm{ev}_\alpha \colon F[x] \to F[\alpha]$ is an isomorphism, which is clearly F-linear. Since $F[x]$ is not a finite dimensional F-vector space, $F[\alpha]$ is also not a finite dimensional F-vector space. But then $F(\alpha)$ is also not a finite dimensional F-vector space, since every vector subspace of a finite dimensional F-vector space is also finite dimensional. Hence $F(\alpha)$ is not a finite extension of F.

Now suppose that α is algebraic over F. Then ev_α induces an isomorphism $\widetilde{\mathrm{ev}}_\alpha \colon F[x]/(\mathrm{irr}(\alpha, F, x)) \to F[\alpha] = F(\alpha)$. Concretely, given $g(x) \in F[x]$, $\widetilde{\mathrm{ev}}_\alpha(g(x) + (\mathrm{irr}(\alpha, F, x)) = g(\alpha)$. Moreover, every coset in the quotient ring $F[x]/(\mathrm{irr}(\alpha, F, x))$ can be uniquely written as $\sum_{i=0}^{d-1} c_i x^i + (\mathrm{irr}(\alpha, F, x))$, where $d = \deg \mathrm{irr}(\alpha, F, x)$. It follows that every element of $F[\alpha] = F(\alpha)$ can be uniquely written as $\sum_{i=0}^{d-1} c_i \alpha^i$. Thus, $1, \alpha, \ldots, \alpha^{d-1}$ is a basis for $F(\alpha)$ as an F-vector space. It then follows that $\dim_F F(\alpha) = d$. $\square$

To be able to calculate the degree $[E : F]$ and use it to extract more information about field extensions, we shall need to consider a sequence of extension fields:

Proposition 2.3. *Suppose that F, E, and K are fields such that $F \leq E \leq K$, i.e. that E is an extension field of F and that K is an extension field of E. Then K is a finite extension field of $F \iff K$ is a finite extension field of E and E is a finite extension field of F. Moreover, in this case*

$$[K : F] = [K : E][E : F].$$

Proof. First suppose that K is a finite extension field of F. Then E is an F-vector subspace of the finite dimensional F-vector space K, hence E is finite dimensional and thus is a finite extension of F. Also, there exists an F-basis $\alpha_1, \ldots, \alpha_n$ of K. Thus every element of K is a linear combination of the α_i with coefficients in F and hence with coefficients in E. Thus $\alpha_1, \ldots, \alpha_n$ span K as an E-vector space, so that K is a finite dimensional E-vector space. Thus K is a finite extension field of E.

Conversely, suppose that K is a finite extension field of E and E is a finite extension field of F. The proof then follows from the following more general lemma (taking $V = K$): $\square$

Lemma 2.4. *Let E be a finite extension field of F and let V be an E-vector space. Then, viewing V as an F-vector space, V is a finite-dimensional F-vector space $\iff$ V is a finite-dimensional E-vector space, and in this case*

$$\dim_F V = [E : F] \dim_E V.$$

Proof. $\implies$: As in the proof above, an F-basis of V clearly spans V over E, hence if V is a finite-dimensional F-vector space, then it is a finite-dimensional E-vector space.

$\impliedby$: Let $v_1, \dots, v_n$ be an E-basis for V and let $\alpha_1, \dots, \alpha_m$ be an F-basis for E. We claim that $\alpha_i v_j$ is an F-basis for V. First, the $\alpha_i v_j$ span V: if $v \in V$, since the v_j are an E-basis for V, there exist $a_j \in E$ such that $\sum_{j=1}^n a_j v_j = v$. Since the α_i are an F-basis of E, there exist $b_{ij} \in F$ such that $a_j = \sum_{i=1}^m b_{ij}\alpha_i$. Hence

$$v = \sum_{j=1}^n a_j v_j = \sum_{i,j} b_{ij}\alpha_i v_j.$$

Thus the $\alpha_i v_j$ span V.

Finally, to see that the $\alpha_i v_j$ are linearly independent, suppose that there exist $b_{ij} \in F$ such that $\sum_{i,j} b_{ij}\alpha_i v_j = 0$. We must show that all of the b_{ij} are 0. Regrouping this sum as

$$0 = \sum_{i,j} b_{ij}\alpha_i v_j = \sum_{j=1}^n \left(\sum_{i=1}^m b_{ij}\alpha_i \right) v_j,$$

and using the fact that the v_j are linearly independent over E, it follows that, for every j, the sum $\sum_{i=1}^m b_{ij}\alpha_i$ is 0. But since the α_i are linearly independent over F, we must have $b_{ij} = 0$ for all i and j. Hence the $\alpha_i v_j$ are linearly independent, and therefore a basis. $\square$

Corollary 2.5. *If $F \le E \le K$ and K is a finite extension of F, then $[K : E]$ and $[E : F]$ both divide $[K : F]$.*

Proof. This is immediate from the formula above. $\square$

The proof also shows the following:

Corollary 2.6. *If K is a finite extension field of E with basis $\beta_1, \ldots, \beta_n$ and E is a finite extension field of F with basis $\alpha_1, \ldots, \alpha_m$, then $\alpha_i\beta_j$, $1 \le i \le m$ and $1 \le j \le n$, is an F-basis of K.* $\square$

Example 2.7. By a homework problem, $\sqrt{3} \notin \mathbb{Q}(\sqrt{2})$. Thus

$$[\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}] = [\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}(\sqrt{2})][\mathbb{Q}(\sqrt{2}) : \mathbb{Q})] = 2 \cdot 2 = 4.$$

A $\mathbb{Q}$-basis for $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ is $1, \sqrt{2}, \sqrt{3}, \sqrt{6}$. Furthermore, with $\alpha = \sqrt{2} + \sqrt{3}$, $\mathbb{Q}(\alpha) = \mathbb{Q}(\sqrt{2}, \sqrt{3})$. Hence $[\mathbb{Q}(\alpha) : \mathbb{Q}] = 4$ and so $\deg_{\mathbb{Q}} \alpha = 4$.

Example 2.8. The real number $\sqrt{2} \notin \mathbb{Q}(\sqrt[3]{2})$, because if it were, then $\mathbb{Q}(\sqrt{2})$ would be a subfield of $\mathbb{Q}(\sqrt[3]{2})$, hence $2 = [\mathbb{Q}(\sqrt{2} : \mathbb{Q}]$ would divide $3 = [\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}]$.

Example 2.9. The above corollary is the main point in showing that various geometric constructions with straightedge and compass such as trisecting every angle or doubling the cube are impossible.

Returning to a general extension of fields, we have the following basic definition:

Definition 2.10. Let E be an extension field of F. Then E is an *algebraic extension* of F if, for every $\alpha \in E$, α is algebraic over F.

The following two lemmas are then easy corollaries of Proposition **??**:

Lemma 2.11. *Let E be a finite extension of F. Then E is an algebraic extension of F.*

Proof. If $\alpha \in E$, then we have a sequence of extensions

$$F \le F(\alpha) \le E.$$

Since E is a finite extension of F, $F(\alpha)$ is a finite extension of F as well, by Proposition **??**. Thus α is algebraic over F. $\square$

Lemma 2.12. *Let E be an extension field of F and let $\alpha, \beta \in E$ be algebraic over F. Then $\alpha \pm \beta$, $\alpha \cdot \beta$, and (if $\beta \ne 0$) α/β are all algebraic over F.*

Proof. Consider the sequence of extensions

$$F \le F(\alpha) \le F(\alpha)(\beta) = F(\alpha, \beta).$$

Then $F(\alpha)$ is a finite extension of F since α is algebraic over F. Moreover, β is the root of a nonzero polynomial with coefficients in F, and hence with coefficients in $F(\alpha)$. Thus β is algebraic over $F(\alpha)$, so that $F(\alpha)(\beta)$ is a finite extension of $F(\alpha)$. By Proposition **??**, $F(\alpha, \beta)$ is then a finite extension of F, and by the previous lemma it is an algebraic extension of F. Thus every element of $F(\alpha, \beta)$ is algebraic over F, in particular $\alpha \pm \beta$, $\alpha \cdot \beta$, and α/β if $\beta \neq 0$. $\square$

Definition 2.13. Let E be an extension field of F. We define the *algebraic closure of F in E* to be

$$\{\alpha \in E : \alpha \text{ is algebraic over } F\ \}.$$

Thus the algebraic closure of F in E is the set of all elements of E which are algebraic over F.

Corollary 2.14. *The algebraic closure of F in E is a subfield of E containing F. Moreover, it is an algebraic extension of F.*

Proof. It clearly contains F, since every $a \in F$ is algebraic over F, and it is a subfield of E by Lemma **??**. By definition, the algebraic closure of F in E is an algebraic extension of F. $\square$

Example 2.15. There are many fields which are algebraic over $\mathbb{Q}$ but not finite over $\mathbb{Q}$. For example, it is not hard to see that the smallest subfield $\mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{5}, \dots)$ of $\mathbb{R}$ which contains the square roots of all of the prime numbers, and hence of every positive integer, is not a finite extension of $\mathbb{Q}$.

For another important example, let $\mathbb{Q}^{\text{alg}}$, the *field of algebraic numbers*, be the algebraic closure of $\mathbb{Q}$ in $\mathbb{C}$. Thus

$$\mathbb{Q}^{\text{alg}} = \{\alpha \in \mathbb{C} : \alpha \text{ is algebraic over } \mathbb{Q}\ \}.$$

Then $\mathbb{Q}^{\text{alg}}$ is a subfield of $\mathbb{C}$, and by definition it is the largest subfield of $\mathbb{C}$ which is algebraic over $\mathbb{Q}$. The extension field $\mathbb{Q}^{\text{alg}}$ is not a finite extension of $\mathbb{Q}$, since for example it contains $\mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{5}, \dots)$.

Finally, let F be an arbitrary field and consider the extension $F(t)$ of F, where t is an indeterminate. As we have seen $F(t)$ is not an algebraic extension of F. In fact, one can show that the algebraic closure of F in $F(t)$ is F, in other words that if a rational function $f(t)/g(t)$ is the root of a nonzero polynomial with coefficients in F, then $f(t)/g(t)$ is constant, i.e. lies in the subfield F of $F(t)$.

We now give another characterization of finite extensions:

Lemma 2.16. *Let E be an extension of F. Then E is a finite extension of F $\iff$ there exist $\alpha_1, \dots, \alpha_n \in E$, all algebraic over F, such that $E = F(\alpha_1, \dots, \alpha_n)$.*

Proof. $\Longleftarrow$: By induction on n. In case $n = 1$, this is just the statement that, if α_1 is algebraic over F, then the simple extension $F(\alpha_1)$ is a finite extension of F. For the inductive step, suppose that we have showed that $F(\alpha_1, \dots, \alpha_i)$ is a finite extension of F. Then α_{i+1} is algebraic over F, hence over $F(\alpha_1, \dots, \alpha_i)$ as in the proof of Lemma **??**. Thus $F(\alpha_1, \dots, \alpha_{i+1}) = F(\alpha_1, \dots, \alpha_i)(\alpha_{i+1})$ is a finite extension of $F(\alpha_1, \dots, \alpha_i)$. Now consider the sequence of extensions

$$F \leq F(\alpha_1, \dots, \alpha_i) \leq F(\alpha_1, \dots, \alpha_{i+1}).$$

Since $F(\alpha_1, \dots, \alpha_{i+1})$ is a finite extension of $F(\alpha_1, \dots, \alpha_i)$ and $F(\alpha_1, \dots, \alpha_i)$ is a finite extension of F, it follows from Proposition **??** that $F(\alpha_1, \dots, \alpha_{i+1})$ is a finite extension of F. This completes the inductive step, and hence the proof that $E = F(\alpha_1, \dots, \alpha_n)$ is a finite extension of F.

$\Longrightarrow$: Let $N = [E : F]$. The proof is by complete induction on N, and the case $N = 1$ is clear since then $E = F$ and we can just take $\alpha_1 = 1$. Now suppose that we have showed that, for every finite extension $F_1 \leq E_1$ with degree $[E_1 : F_1] < N$, there exist $\beta_1, \dots, \beta_k \in E_1$ such that $E_1 = F_1(\beta_1, \dots, \beta_k)$. Let E be a finite extension of F with $[E : F] = N > 1$. Since $E \neq F$, there exists an $\alpha_1 \in E$ with $\alpha_1 \notin F$. Hence $[F(\alpha_1) : F] > 1$. Since $N = [E : F] = [E : F(\alpha_1)][F(\alpha_1) : F]$, it follows that $[E : F(\alpha_1)] < N$. By the inductive hypothesis, there exist $\alpha_2, \dots, \alpha_n \in E$ such that $E = F(\alpha_1)(\alpha_2, \dots, \alpha_n) = F(\alpha_1, \dots, \alpha_n)$. Finally, the α_i are automatically algebraic over F since E is a finite extension of F. This completes the proof of the inductive step and hence of the lemma. $\square$

Lemma 2.17. *Let $F \leq E \leq K$ be a sequence of field extensions, with E an algebraic extension of F, and let $\alpha \in K$. Then α is algebraic over F $\iff$ α is algebraic over E.*

Proof. $\Longrightarrow$: This is clear since, if α is a root of a nonzero polynomial $f(x) \in F[x]$, then since $F[x] \subseteq E[x]$, α is also a root of the nonzero polynomial $f(x)$ viewed as an element of $E[x]$.

$\Longleftarrow$: Write $\mathrm{irr}(\alpha, E, x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$, where the $a_i \in E$ and hence the a_i are algebraic over F. By the previous lemma, $F(a_0, \dots, a_{n-1})$ is a finite extension of F, and clearly α is algebraic over $F(a_0, \dots, a_{n-1})$ since it is the root of a nonzero polynomial with coefficients in $F(a_0, \dots, a_{n-1})$. Thus

$$F(a_0, \dots, a_{n-1})(\alpha) = F(a_0, \dots, a_{n-1}, \alpha)$$

is a finite extension of $F(a_0, \dots, a_{n-1})$. It follows from Proposition **??** that $F(a_0, \dots, a_{n-1}, \alpha)$ is a finite extension of F, hence an algebraic extension of F. Hence α is algebraic over F. $\square$

Corollary 2.18. *Let $F \le E \le K$ be a sequence of field extensions. Then K is an algebraic extension of F $\iff$ K is an algebraic extension of E and E is an algebraic extension of F.*

Proof. $\implies$: If K is an algebraic extension of F, then clearly E is an algebraic extension of F. Moreover, every element α of K is the root of a nonzero polynomial with coefficients in F and hence in E, hence α is algebraic over E. Thus K is an algebraic extension of E.

$\impliedby$: Follows immediately from the preceding lemma. $\square$

Definition 2.19. A field K is *algebraically closed* if every nonconstant polynomial $f(x) \in K[x]$ has a root in K.

Lemma 2.20. *Let K be a field. Then the following are equivalent:*

(i) *K is algebraically closed.*

(ii) *If $f(x) \in K[x]$ is a nonconstant polynomial, then $f(x)$ is a product of linear factors. In other words, the irreducible polynomials in $K[x]$ are linear.*

(iii) *The only algebraic extension of K is K.*

Proof. (i) $\implies$ (ii): Let $f(x) \in K[x]$ be a nonconstant polynomial. Then $f(x)$ factors into a product of irreducible polynomials, so it suffices to show that every irreducible polynomial is linear. Let $p(x)$ be irreducible. Then, since K is algebraically closed, there exists a root α of $p(x)$ in K, and hence a linear factor $x-\alpha \in K[x]$ of $p(x)$. Since $p(x)$ is irreducible, $p(x) = c(x-\alpha)$ for some $c \in K^*$, and hence $p(x)$ is linear.

(ii) $\implies$ (iii): Let E be an algebraic extension of K and let $\alpha \in E$. Then $p(x) = \mathrm{irr}(\alpha, K, x)$ is a monic irreducible polynomial, hence necessarily of the form $x - \alpha$. Since $p(x) \in K[x]$, it follows that $\alpha \in K$.

(iii) $\implies$ (i): If $f(x) \in K[x]$ is a nonconstant polynomial, then there exists an extension field E of K and an $\alpha \in E$ which is a root of $f(x)$. Clearly, α is algebraic over K and hence the extension field $F(\alpha)$ is an algebraic extension of K. By assumption, $K(\alpha) = K$, i.e. $\alpha \in K$. Hence there exists a root of $f(x)$ in K. $\square$

The most important example of an algebraically closed field comes from the following theorem, essentially due to Gauss (1799):

Theorem 2.21 (The Fundamental Theorem of Algebra)**.** *The field $\mathbb{C}$ of complex numbers is algebraically closed.* $\square$

Despite its name, the Fundamental Theorem of Algebra cannot be a theorem strictly about algebra, since the real numbers and hence the complex numbers are not defined algebraically. There are many proofs of the Fundamental Theorem of Algebra. A number of proofs use some basic complex analysis, or some topological properties of the plane. We will give a (mostly) algebraic proof at the end of the course.

Definition 2.22. Let F be a field. Then an extension field K of F is an *algebraic closure* of F if the following hold:

1. K is an algebraic extension of F, and
2. K is algebraically closed.

With this definition, $\mathbb{C}$ is **not** an algebraic closure of $\mathbb{Q}$, because $\mathbb{C}$ is not an algebraic extension of $\mathbb{Q}$.

So far, we have defined three confusingly similar sounding concepts: **the** algebraic closure of the field F in an extension field E, when a field K **is** algebraically closed (with no reference to any subfield), and when an extension field K is **an** algebraic closure of the field F. One way these concepts are related is as follows:

Proposition 2.23. *Let F be a field, let K be an extension field of F, and suppose that K is algebraically closed. Then the algebraic closure of F in K is an algebraic closure of F.*

Proof. Let E be the algebraic closure of F in K. Then E is an algebraic extension of F, and we must prove that E is algebraically closed. Let $f(x) \in E[x]$ be a nonconstant polynomial. Then, since $E[x] \subseteq K[x]$, there exists a root $\alpha \in K$ of $f(x)$. Clearly α is algebraic over E. By Lemma **??**, α is algebraic over F, hence $\alpha \in E$. Thus E is algebraically closed. $\square$

Corollary 2.24. *The field $\mathbb{Q}^{\text{alg}}$ of algebraic numbers is an algebraic closure of $\mathbb{Q}$.* $\square$

The following theorem, which we shall not prove, guarantees the existence of an algebraic closure for every field:

Theorem 2.25. *Let F be a field. Then there exists an algebraic closure of F. Moreover, every two algebraic closures of F are isomorphic. More precisely, if $F \leq K_1$ and $F \leq K_2$, then there exists an isomorphism $\rho\colon K_1 \to K_2$ such that $\rho(a) = a$ for all $a \in F$, viewing F as a subfield both of K_1 and K_2.* □

The isomorphism ρ in the previous theorem is far from unique. In fact, understanding the possible isomorphisms is, in a very vague sense, the central problem in Galois theory.

3 Derivatives and multiple roots

We begin by recalling the definition of a repeated root.

Definition 3.1. Let F be a field and let $\alpha \in F$. Then there is a unique integer $m \geq 0$ such that $(x-\alpha)^m$ divides $f(x)$ but $(x-\alpha)^{m+1}$ does not divide $f(x)$. We define this integer m to be the *multiplicity* of the root α in $f(x)$. Note that, by the correspondence between roots of a polynomial and its linear factors, α has multiplicity 0 in $f(x)$, i.e. $m = 0$ above, $\iff$ $f(\alpha) \neq 0$. More generally, if α has multiplicity m in $f(x)$, then $f(x) = (x-\alpha)^m g(x)$ with $g(\alpha) \neq 0$, and conversely.

If α has multiplicity 1 in $f(x)$, we call α a *simple root* of $f(x)$. If α has multiplicity $m \geq 2$ in $f(x)$, then we call α a *multiple root* or *repeated root* of $f(x)$.

We would like to find conditions when a nonconstant polynomial does, or does not have a multiple root in F or in some extension field E of F. To do so, we introduce the *formal derivative*:

Definition 3.2. Let F be a field. Define the function $D\colon F[x] \to F[x]$ by the formula

$$D(\sum_{i=0}^{n} a_i x^i) = \sum_{i=1}^{n} i a_i x^{i-1}.$$

Here the notation ia_i means the ring element $i \cdot a_i = \underbrace{a_i + \cdots + a_i}_{i \text{ times}}$. We usually write $D(f(x))$ as $Df(x)$. Note that either $Df(x) = 0$ or $\deg Df(x) \leq \deg f(x) - 1$.

Clearly, the function D is compatible with field extension, in the sense that, if $F \le E$, then we have $D\colon F[x] \to F[x]$ and $D\colon E[x] \to E[x]$, then, given $f(x) \in F[x]$, $Df(x)$ is the same whether we view $f(x)$ as an element of $F[x]$ or of $E[x]$. Also, an easy calculation shows that:

Proposition 3.3. *$D\colon F[x] \to F[x]$ is F-linear.* $\square$

This result is equivalent to the *sum rule*: for all $f(x), g(x) \in F[x]$, $D(f + g) = Df + Dg$ as well as the *constant multiple rule*: for all $f(x) \in F[x]$ and $c \in F$, $D(cf) = cDf$. Once we know that D is F-linear, it is specified by the fact $D(1) = 0$ and, that, for all $i > 0$, $Dx^i = ix^{i-1}$. Also, viewing D as a homomorphism of abelian groups, we can try to compute

$$\operatorname{Ker} D = \{f(x) \in F[x] : Df(x) = 0\}.$$

Our expectation from calculus is that a function whose derivative is 0 is a constant. But if $\operatorname{char} F = p > 0$, something strange happens:

Proposition 3.4. *If $\operatorname{Ker} D = \{f(x) \in F[x] : Df(x) = 0\}$, then*

$$\operatorname{Ker} D = \begin{cases} F, & \text{if } \operatorname{char} F = 0; \\ F[x^p], & \text{if } \operatorname{char} F = p > 0. \end{cases}$$

Here $F[x^p] = \{\sum_{i=0}^n a_i x^{ip} : a_i \in F\}$ is the subring of all polynomials in x^p.

Proof. Clearly, $f(x) = \sum_{i=0}^n a_i x^i$ is in $\operatorname{Ker} D \iff$ for every i such that the coefficient a_i is nonzero, the monomial $ix^{i-1} = 0$. In case $\operatorname{char} F = 0$, this is only possible if $i = 0$, in other words $f(x) \in F$ is a constant polynomial. In case $\operatorname{char} F = p > 0$, this happens exactly when $p|i$ for every i such that $a_i \neq 0$. This is equivalent to saying that $f(x)$ is a polynomial in x^p. $\square$

As is well-known in calculus, D is **not** a ring homomorphism. In other words, the derivative of a product of two polynomials is **not** in general the product of the derivatives. Instead we have:

Proposition 3.5 (The product rule)**.** *For all $f(x), g(x) \in F[x]$,*

$$D(f \cdot g)(x) = Df(x) \cdot g(x) + f(x) \cdot Dg(x).$$

Proof. If $f(x) = x^a$ and $g(x) = x^b$, then we can verify this directly:

$$D(x^a x^b) = D(x^{a+b}) = (a+b)x^{a+b-1};$$
$$(Dx^a)x^b + x^a(Dx^b) = ax^{a-1}x^b + bx^a x^{b-1} = (a+b)x^{a+b-1}.$$

The general case follows from this by writing $f(x)$ and $g(x)$ as sums of monomials and expanding (but is a little messy to write down). Another approach using formal difference quotients is in the HW. $\square$

If R is a ring, a function $d\colon R \to R$ which is an additive homomorphism (i.e. $d(r+s) = d(r) + d(s)$ for all $r, s \in R$) satisfying $d(rs) = d(r)s + rd(s)$ for all $r, s \in R$ is called a *derivation* of R. Thus, D is a derivation of $F[x]$.

As a corollary of the product rule, we obtain:

Corollary 3.6 (The power rule)**.** *For all $f(x) \in F[x]$ and $n \in \mathbb{N}$,*

$$D(f(x))^n = n(f(x))^{n-1} Df(x).$$

Proof. This is an easy induction using the product rule and starting with the case $n = 1$. □

The connection between derivatives and multiple roots is as follows:

Lemma 3.7. *Let $f(x) \in F[x]$ be a nonconstant polynomial. Then $\alpha \in F$ is a multiple root of $f(x) \iff f(\alpha) = Df(\alpha) = 0$.*

Proof. Write $f(x) = (x-\alpha)^m g(x)$ with m equal to the multiplicity of α in $f(x)$ and $g(x) \in F[x]$ a polynomial such that $g(\alpha) \neq 0$. If $m = 0$, then $f(\alpha) = g(\alpha) \neq 0$. Otherwise,

$$Df(x) = m(x-\alpha)^{m-1} g(x) + (x-\alpha)^m Dg(x).$$

If $m = 1$, then $Df(\alpha) = g(\alpha) \neq 0$. If $m \geq 2$, then $f(\alpha) = Df(\alpha) = 0$. Thus we see that $\alpha \in F$ is a multiple root of $f(x) \iff m \geq 2 \iff f(\alpha) = Df(\alpha) = 0$. □

In practice, an (unknown) root of $f(x)$ will only exist in some (unknown) extension field E of F. We would like to have a criterion for when a polynomial $f(x)$ has **some** multiple root α in **some** extension field E of F, without having to know what E and α are explicitly. In order to find such a criterion, we begin with the following lemma, which says essentially that divisibility, greatest common divisors, and relative primality are unchanged after passing to extension fields.

Lemma 3.8. *Let E be an extension field of a field F, and let $f(x), g(x) \in F[x]$, not both 0.*

(i) *$f(x)\big|g(x)$ in $F[x] \iff f(x)\big|g(x)$ in $E[x]$.*

(ii) *The polynomial $d(x) \in F[x]$ is a* gcd *of $f(x), g(x)$ in $F[x] \iff d(x)$ is a* gcd *of $f(x), g(x)$ in $E[x]$.*

(iii) *The polynomials $f(x), g(x)$ are relatively prime in $F[x] \iff f(x), g(x)$ are relatively prime in $E[x]$.*

Proof. (i): $\Longrightarrow$: obvious. $\Longleftarrow$: We can assume that $f(x) \neq 0$, since otherwise $f(x)\big|g(x)$ (in either $F[x]$ or $E[x]$) $\iff$ $g(x) = 0$. Suppose that $f(x)\big|g(x)$ in $E[x]$, i.e. that $g(x) = f(x)h(x)$ for some $h(x) \in E[x]$. We must show that $h(x) \in F[x]$. By long division with remainder in $F[x]$, there exist $q(x), r(x) \in F[x]$ with either $r(x) = 0$ or $\deg r(x) < \deg f(x)$, such that $g(x) = f(x)q(x) + r(x)$. Now, in $E[x]$, we have both $g(x) = f(x)h(x)$ and $g(x) = f(x)q(x) + r(x)$. By uniqueness of long division with remainder in $E[x]$, we must have $h(x) = q(x)$ (and $r(x) = 0$). In particular, $h(x) = q(x) \in F[x]$, as claimed.

(ii): $\Longrightarrow$: Let $d(x) \in F[x]$ be a gcd of $f(x), g(x)$ in $F[x]$. Then, by (i), since $d(x)\big|f(x)$, $d(x)\big|g(x)$ in $F[x]$, $d(x)\big|f(x)$, $d(x)\big|g(x)$ in $E[x]$. Moreover, there exist $a(x), b(x) \in F[x]$ such that $d(x) = a(x)f(x) + b(x)g(x)$. Now suppose that $e(x) \in E[x]$ and that $e(x)\big|f(x)$, $e(x)\big|g(x)$ in $E[x]$. Then $e(x)\big|a(x)f(x)+b(x)g(x) = d(x)$. It follows that $d(x)$ satisfies the properties of being a gcd in $E[x]$. $\Longleftarrow$: Let $d(x) \in F[x]$ be a gcd of $f(x), g(x)$ in $E[x]$. Then $d(x)\big|f(x)$, $d(x)\big|g(x)$ in $E[x]$, hence by (i) $d(x)\big|f(x)$, $d(x)\big|g(x)$ in $F[x]$. Suppose that $e(x) \in F[x]$ and that $e(x)\big|f(x)$, $e(x)\big|g(x)$ in $F[x]$. Then $e(x)\big|f(x)$, $e(x)\big|g(x)$ in $E[x]$. Hence $e(x)\big|d(x)$ in $E[x]$. Since both $e(x), d(x) \in F[x]$, it again follows by (i) that $e(x)\big|d(x)$ in $F[x]$. Thus $d(x)$ is a gcd of $f(x), g(x)$ in $F[x]$.

(iii): The polynomials $f(x), g(x)$ are relatively prime in $F[x]$ $\iff$ $1 \in F[x]$ is a gcd of $f(x)$ and $g(x)$ in $F[x]$ $\iff$ $1 \in F[x]$ is a gcd of $f(x)$ and $g(x)$ in $F[x]$, by (ii), $\iff$ $f(x), g(x)$ are relatively prime in $E[x]$. $\square$

Corollary 3.9. *Let $f(x) \in F[x]$ be a nonconstant polynomial. Then there exists an extension field E of F and a multiple root of $f(x)$ in E $\iff$ $f(x)$ and $Df(x)$ are not relatively prime in $F[x]$.*

Proof. $\Longrightarrow$: If E and α exist, then, by Lemma **??**, $f(x)$ and $Df(x)$ have a common factor $x - \alpha$ in $E[x]$ and hence are not relatively prime. Thus by Lemma **??** $f(x)$ and $Df(x)$ are not relatively prime in $F[x]$.

$\Longleftarrow$: Suppose that $f(x)$ and $Df(x)$ are not relatively prime in $F[x]$, and let $g(x)$ be a common nonconstant factor of $f(x)$ and $Df(x)$. There exists an extension field E of F and an $\alpha \in E$ which is a root of $g(x)$. Then α is a common root of $f(x)$ and $Df(x)$, and hence a multiple root of $f(x)$. $\square$

We now apply the above to an **irreducible** polynomial $f(x) \in F[x]$.

Corollary 3.10. *Let $f(x) \in F[x]$ be an irreducible polynomial. Then there exists an extension field E of F and a multiple root of $f(x)$ in E $\iff$ $Df(x) = 0$.*

Proof. $\Longrightarrow$: By the previous corollary, if there exists an extension field E of F and a multiple root of $f(x)$ in E, then $f(x)$ and $Df(x)$ are not relatively prime in $F[x]$. In this case, since $f(x)$ is irreducible, it must be that $f(x)$ divides $Df(x)$. Hence, if $Df(x) \neq 0$, then $\deg Df(x) \geq \deg f(x)$. But we have seen that either $\deg Df(x) < \deg f(x)$ or $Df(x) = 0$. Thus, we must have $Df(x) = 0$.

$\Longleftarrow$: Clearly, if $Df(x) = 0$, then $f(x)$ is a gcd of $f(x)$ and $Df(x)$, hence $f(x)$ and $Df(x)$ are not relatively prime in $F[x]$. $\square$

Corollary 3.11. *Let F be a field of characteristic 0 and let $f(x) \in F[x]$ be an irreducible polynomial. Then there does not exist an extension field E of F and a multiple root of $f(x)$ in E. In particular, if E is an extension field of F such that $f(x)$ factors into linear factors in E, say*

$$f(x) = c(x - \alpha_1) \cdots (x - \alpha_n),$$

then the α_i are distinct, i.e. for $i \neq j$, $\alpha_i \neq \alpha_j$. $\square$

If $\operatorname{char} F = p > 0$, then it is possible for an irreducible polynomial $f(x) \in F[x]$ to have a multiple root in some extension field, but it takes a little effort to produce such examples. The basic example arises as follows: consider the field $\mathbb{F}_p(t)$, where t is an indeterminate (here we could replace $\mathbb{F}_p$ by any field of characteristic p). Then t is not a p^{th} power in $\mathbb{F}_p(t)$, and in fact one can show that the polynomial $x^p - t$ is irreducible in $\mathbb{F}_p(t)[x]$. Let E be an extension field of $\mathbb{F}_p(t)$ which contains a root α of $x^p - t$, so that by definition $\alpha^p = t$. Then

$$x^p - t = x^p - \alpha^p = (x - \alpha)^p,$$

because we are in characteristic p. Thus α is a multiple root of $x^p - t$, of multiplicity p.

The key property of the field $\mathbb{F}_p(t)$ which made the above example work was that t was not a p^{th} power in $\mathbb{F}_p(t)$. More generally, define a field F of characteristic p to be *perfect* if every element of F is a p^{th} power, or equivalently if the Frobenius homomorphism $F \to F$ is surjective. For example, we shall show below that a finite field is perfect. An algebraically closed field is also perfect. We also declare every field of characteristic zero to be perfect. By a problem on HW, if F is a perfect field and $f(x) \in F[x]$ is an irreducible polynomial, then there does not exist an extension field E of F and a multiple root of $f(x)$ in E.

4 Finite fields

Our goal in this section is to classify finite fields up to isomorphism and, given two finite fields, to describe when one of them is isomorphic to a subfield of the other. We begin with some general remarks about finite fields.

Let $\mathbb{F}$ be a finite field. As the additive group $(\mathbb{F}, +)$ is finite, $\operatorname{char} \mathbb{F} = p > 0$ for some prime p. Thus $\mathbb{F}$ contains a subfield isomorphic to the prime field $\mathbb{F}_p$, which we will identify with $\mathbb{F}_p$. Since $\mathbb{F}$ is finite, it is clearly a finite-dimensional vector space over $\mathbb{F}_p$. Let $n = \dim_{\mathbb{F}_p} \mathbb{F}$. Then $\#(\mathbb{F}) = p^n$. It is traditional to use the letter q to demote a prime power p^n in this context.

We note that the multiplicative group $(\mathbb{F}^*, \cdot)$ is cyclic. If γ is a generator, then every nonzero element of $\mathbb{F}$ is a power of γ. In particular, $\mathbb{F} = \mathbb{F}_p(\gamma)$ is a simple extension of $\mathbb{F}_p$.

With $\#(\mathbb{F}) = p^n = q$ as above, by Lagrange's theorem, since $\mathbb{F}^*$ is a finite group of order $q - 1$, for every $\alpha \in \mathbb{F}^*$, $\alpha^{q-1} = 1$. Hence $\alpha^q = \alpha$ for all $\alpha \in \mathbb{F}$, since clearly $0^q = 0$. Thus every element of $\mathbb{F}$ is a root of the polynomial $x^q - x$.

Since $\operatorname{char} \mathbb{F} = p$, the function $\sigma_p \colon \mathbb{F} \to \mathbb{F}$ is a homomorphism, the *Frobenius homomorphism*. Clearly $\operatorname{Ker} \sigma_p = \{0\}$ since $\alpha^p = 0 \iff \alpha = 0$, and hence σ_p is injective. (In fact, this is always true for homomorphisms from a field to a nonzero ring.) As $\mathbb{F}$ is **finite**, since σ_p is injective, it is also surjective and hence an isomorphism (by the pigeonhole principle). Thus, every element of $\mathbb{F}$ is a p^{th} power, so that $\mathbb{F}$ is perfect as defined above. Note that every power σ_p^k is also an isomorphism. We have

$$\sigma_p^2(\alpha) = \sigma_p(\sigma_p(\alpha)) = \sigma_p(\alpha^p) = (\alpha^p)^p = \alpha^{p^2},$$

and so $\sigma_p^2 = \sigma_{p^2}$, where by definition $\sigma_{p^2}(\alpha) = \alpha^{p^2}$. More generally, an easy induction shows that $\sigma_p^k = \sigma_{p^k}$, where by definition $\sigma_{p^k}(\alpha) = \alpha^{p^k}$. In particular, taking $k = n$, where $\#(\mathbb{F}) = q = p^n$, we see that $\sigma_q(\alpha) = \alpha^q = \alpha$. Thus $\sigma_q = \mathrm{Id}$. (**Warning:** although $\alpha^q = \alpha$ for every $\alpha \in \mathbb{F}$, it is **not** true that $x^q - x \in \mathbb{F}[x]$ is the zero polynomial.)

With this said, we can now state the classification theorem for finite fields:

Theorem 4.1 (Classification of finite fields)**.** *Let p be a prime number.*

(i) *For every $n \in \mathbb{N}$, there exists a field $\mathbb{F}$ with $q = p^n$ elements.*

(ii) *If $\mathbb{F}_1$ and $\mathbb{F}_2$ are two finite fields with $\#(\mathbb{F}_1) = \#(\mathbb{F}_2)$, then $\mathbb{F}_1$ and $\mathbb{F}_2$ are isomorphic.*

(iii) *Let* $\mathbb{F}$ *and* $\mathbb{F}'$ *be two finite fields, with* $\#(\mathbb{F}) = q = p^n$ *and* $\#(\mathbb{F}') = q' = p^m$. *Then* $\mathbb{F}'$ *is isomorphic to a subfield of* $\mathbb{F}$ $\iff$ m *divides* n $\iff$ $q = (q')^d$ *for some positive integer* d.

Proof. First, we prove (i). Viewing the polynomial $x^q - x$ as a polynomial in $\mathbb{F}_p[x]$, we know that there exists an extension field E of $\mathbb{F}_p$ such that $x^q - x$ is a product of linear factors in $E[x]$, say

$$x^q - x = (x - \alpha_1) \cdots (x - \alpha_q)$$

where the $\alpha_i \in E$. We claim that the α_i are all distinct: $\alpha_i = \alpha_j$ for some $i \neq j \iff x^q - x$ has a multiple root in $E \iff x^q - x$ and $D(x^q - x)$ are not relatively prime in $\mathbb{F}_p[x]$. But $D(x^q - x) = qx^{q-1} - 1 = -1$, since q is a power of p and hence divisible by p. Thus the gcd of $x^q - x$ and $D(x^q - x)$ divides -1 and hence is a unit, so that $x^q - x$ and $D(x^q - x)$ are relatively prime. It follows that $x^q - x$ does not have any multiple roots in E.

Now define the subset $\mathbb{F}$ of E by

$$\mathbb{F} = \{\alpha_1, \ldots, \alpha_q\} = \{\alpha \in E : \alpha^q - \alpha = 0\} = \{\alpha \in E : \alpha^q = \alpha\}.$$

By what we have seen above, $\#(\mathbb{F}) = q$. Moreover, we claim that $\mathbb{F}$ is a subfield of E, and hence is a field with q elements. It suffices to show that $\mathbb{F}$ is closed under addition, subtraction, multiplication, and division. This follows since σ_q is a homomorphism. Hence, if $\alpha, \beta \in \mathbb{F}$, i.e. if $\alpha^q = \alpha$ and $\beta^q = \beta$, then $(\alpha \pm \beta)^q = \alpha^q \pm \beta^q = \alpha \pm \beta$, $(\alpha\beta)^q = \alpha^q\beta^q = \alpha\beta$, and, if $\beta \neq 0$, then $(\alpha/\beta)^q = \alpha^q/\beta^q = \alpha/\beta$. In other words, then $\alpha \pm \beta$, $\alpha\beta$, and (for $\beta \neq 0$) α/β are all in $\mathbb{F}$. Hence $\mathbb{F}$ is a subfield of E, and in particular it is a field with q elements. (Remark: $\mathbb{F}$ is the *fixed field* of σ_q, i.e. $\mathbb{F} = \{\alpha \in E : \sigma_q(\alpha) = \alpha\}$.)

Next we prove (iii). Let $\mathbb{F}$ and $\mathbb{F}'$ be two finite fields with $\#(\mathbb{F}) = q = p^n$ and $\#(\mathbb{F}') = q' = p^m$. Clearly, if $\mathbb{F}'$ is isomorphic to a subfield of $\mathbb{F}$, which we can identify with $\mathbb{F}'$, then $\mathbb{F}$ is an $\mathbb{F}'$-vector space. Since $\mathbb{F}$ is finite, it is finite-dimensional as an $\mathbb{F}'$-vector space. Let $d = \dim_{\mathbb{F}'} \mathbb{F} = [\mathbb{F} : \mathbb{F}']$. Then $p^n = q = \#(\mathbb{F}) = (q')^d = p^{md}$, proving that m divides n and that q is a power of q'. Conversely, suppose that $\mathbb{F}$ is the finite field with $q = p^n$ elements constructed in the proof of (i), so that $x^q - x$ factors into linear factors in $\mathbb{F}[x]$. Let $\mathbb{F}'$ be a finite field with $\#(\mathbb{F}') = q' = p^m$ and suppose that $q = p^n = (q')^d$, or equivalently $n = md$. We shall show first that $\mathbb{F}$ contains a subfield isomorphic to $\mathbb{F}'$ and then that every field with q elements is isomorphic to $\mathbb{F}$, proving the converse part of (iii) as well as (ii).

As we saw in the remarks before the statement of Theorem **??**, there exists a $\beta \in \mathbb{F}'$ such that $\mathbb{F}' = \mathbb{F}_p(\beta)$. Since $\beta \in \mathbb{F}'$, $\sigma_{q'}(\beta) = (\beta)^{q'} = \beta$, and hence $\sigma_q(\beta) = \sigma_{q'}^d(\beta) = \beta$. Thus β is a root of $x^q - x$. Hence $\mathrm{irr}(\beta, \mathbb{F}_p, x)$ divides $x^q - x$ in $\mathbb{F}_p[x]$. On the other hand, $x^q - x$ factors into linear factors in $\mathbb{F}[x]$, so that one of these linear factors must divide $\mathrm{irr}(\beta, \mathbb{F}_p, x)$ in $\mathbb{F}[x]$. It follows that there exists a root γ of $\mathrm{irr}(\beta, \mathbb{F}_p, x)$ in $\mathbb{F}$. Since $\mathrm{irr}(\beta, \mathbb{F}_p, x)$ is a monic irreducible polynomial and γ is a root of $\mathrm{irr}(\beta, \mathbb{F}_p, x)$, we must have $\mathrm{irr}(\gamma, \mathbb{F}_p, x) = \mathrm{irr}(\beta, \mathbb{F}_p, x)$. Let $p(x) = \mathrm{irr}(\gamma, \mathbb{F}_p, x) = \mathrm{irr}(\beta, \mathbb{F}_p, x)$. Then since $\mathbb{F}' = \mathbb{F}_p(\beta)$, ev_β induces an isomorphism $\widetilde{\mathrm{ev}}_\beta \colon \mathbb{F}_p[x]/(p(x)) \cong \mathbb{F}'$. On the other hand, we have $\mathrm{ev}_\gamma \colon \mathbb{F}_p[x] \to \mathbb{F}$, with $\operatorname{Ker} \mathrm{ev}_\gamma = (p(x))$ as well, so there is an induced injective homomorphism $\widetilde{\mathrm{ev}}_\gamma \colon \mathbb{F}_p[x]/(p(x)) \to \mathbb{F}$. The situation is summarized in the following diagram:

$$\begin{tikzcd}
\mathbb{F}_p[x]/(p(x)) \arrow[r, "\widetilde{\mathrm{ev}}_\gamma"] \arrow[d, "\widetilde{\mathrm{ev}}_\beta"', "\cong"] & \mathbb{F} \\
\mathbb{F}' &
\end{tikzcd}$$

The homomorphism $\widetilde{\mathrm{ev}}_\gamma \circ (\widetilde{\mathrm{ev}}_\beta)^{-1}$ is then an injective homomorphism from $\mathbb{F}'$ to $\mathbb{F}$ and thus identifies $\mathbb{F}'$ with a subfield of $\mathbb{F}$. This proves the converse direction of (iii), for the specific field $\mathbb{F}$ constructed in (i), and hence for any field which is isomorphic to $\mathbb{F}$.

Now suppose that $\mathbb{F}$ is the specific field with q elements constructed in the proof of (i) and that $\mathbb{F}_1$ is another finite field with q elements. By what we have proved so far above, $\mathbb{F}_1$ is isomorphic to a subfield of $\mathbb{F}$, i.e. there is an injective homomorphism $\rho \colon \mathbb{F}_1 \to \mathbb{F}$. But since $\mathbb{F}_1$ and $\mathbb{F}$ have the same number of elements, ρ is necessarily an isomorphism, i.e. $\mathbb{F}_1 \cong \mathbb{F}$. Hence, if $\mathbb{F}_2$ is yet another field with q elements, then also $\mathbb{F}_2 \cong \mathbb{F}$ and hence $\mathbb{F}_1 \cong \mathbb{F}_2$, proving (ii). Finally, the converse direction of (iii) now holds for every field with q elements, since every such field is isomorphic to $\mathbb{F}$. $\square$

If $q = p^n$, we often write $\mathbb{F}_q$ to denote any field with q elements. Since any two such fields are isomorphic, we often speak of **the** field with q elements.

Remark 4.2. Let $q = p^n$. The polynomial $x^q - x$ is reducible in $\mathbb{F}_p[x]$. For example, for every $a \in \mathbb{F}_p$, $x - a$ is a factor of $x^q - x$. Using Theorem **??**, one can show that the irreducible monic factors of $x^q - x$ are exactly the irreducible monic polynomials in $\mathbb{F}_p[x]$ of degree m, where m divides n. From this, one can show the following beautiful formula: let $N_p(m)$ be the number of irreducible monic polynomials in $\mathbb{F}_p[x]$ of degree m. Then

$$\sum_{d|n} dN_p(d) = p^n.$$