

1. Field Extensions

Throughout this chapter, F and E will be our regularly used notations for fields.

1.1. Basic questions

Throughout this section, let F be a field.

Let $f \in F[x]$ be a non-constant polynomial. We would like to be able to solve the equation

$$f(x) = 0.$$

There may or may not be a solution in F . If there is not, is there a larger field $E \supset F$ that contains a solution? How many such fields are there? Is there a field $E \supset F$ that contains the solutions to the polynomial equation $f(x) = 0$, for *every* $f \in F[x]$? How unique is E , if exists? Is there a field $E \supset F$ that contains the solutions to the polynomial equations $f(x) = 0$, for every $f \in E[x]$?

Definition 1.1. *Given $f \in F[x]$, an element $\alpha \in F$ is called a root of f in F , if $f(\alpha) = 0$.*

Definition 1.2. *We say that a field E is algebraically closed, if every non-constant polynomial $f \in E[x]$ has a root in E .*

Theorem 1.3. *Let E be an algebraically closed field. Then the only irreducible polynomials*

in $E[x]$ are the degree 1 polynomials.

Proof: By definition, every degree 1 polynomial is irreducible. Conversely, suppose $f \in E[x]$ is irreducible. Then it is non-constant and so it has a root $\alpha \in E$. By the Division Algorithm, $f = (x - \alpha)g + r$ for some $g, r \in E[x]$, with $\deg r < \deg(x - \alpha) = 1$. So r is a constant. But $0 = f(\alpha) = r$. Since f is irreducible, g must be constant, implying that f has degree 1. \square .

Corollary 1.4. *For any algebraically closed field E , every non-constant polynomial $f \in E[x]$ is the product of degree 1 polynomials. In particular, the roots of f are precisely those of its degree 1 divisors.*

Proof: This follows immediately from the preceding theorem and the Unique Factorization Theorem. \square

Thus, in an algebraically closed field, every polynomial factorizes into something that is as simple as possible. Therefore, given a field F , it is desirable to find a field $E \supset F$ in which this can be achieved. The oldest example of this situation is $F = \mathbf{R}$ and $E = \mathbf{C}$.

Theorem 1.5. *(The Fundamental Theorem of Algebra) The field of complex numbers \mathbf{C} is algebraically closed.*

Proof: We give an analytical proof. Let $f \in \mathbf{C}[x]$ be a non-constant polynomial. Then $\lim_{z \rightarrow \infty} |f(z)| = +\infty$. In particular, $1/|f(z)|$ is bounded for $|z|$ sufficiently large. Suppose f has no root in \mathbf{C} . Then $1/f(z)$ is an analytic function defined on \mathbf{C} . Since it is also bounded, $1/f(z)$ must be a constant, by Liouville's theorem in complex one-variable, contradicting that f is non-constant. \square

Thus some of the questions above are abstraction of the situation in this example. We now begin a systematic study of these questions, and along the way, raise more questions about polynomial equations and fields. We can use the following picture to organize the various players in these questions:

$$\begin{array}{ccc}
 F & \xrightarrow{\quad} & \alpha \in E \supset F, f(\alpha) = 0 \\
 & \searrow \quad \swarrow & \\
 & f \in F[x] &
 \end{array}$$

The questions raised will be concerned about, under various assumptions, existence and uniqueness of objects in one corner of this triangle, the given objects in the other two corners.

1.2. Basic construction

Definition 1.6. *Let E be a field. A subfield of E is an additive subgroup F of E which contains 1 and is closed under division, i.e. $a, b \in F, b \neq 0 \Rightarrow a/b \in F$. In this case, we also say that E is an extension field of F and we write $E \geq F$.*

Sub vs. homomorphism. Let F and E be fields. If $\phi : F \rightarrow E$ is a given field homomorphism (i.e. a ring homomorphism that preserves 1), then we can always think of F as a subfield of E , as follows. We think of F and $E - \phi(F)$ (the complement of $\phi(F)$ in E) as disjoint sets. Let K be their union. Then it is easy to check that K is a field, with $0, 1 \in F$, under the following addition rule on K

$$a - b := \begin{cases} a - b & \text{if } a, b \in F \\ \phi(a) - b & \text{if } a \in F, b \in E - \phi(F) \\ \phi^{-1}(a - b) & \text{if } a, b \in E - \phi(F), a - b \in \phi(F) \\ a - b & \text{if } a, b, a - b \in E - \phi(F) \end{cases}$$

and a similar division rule on K . Moreover F is a subfield of K such that $\phi : F \rightarrow E$ can be extended to a unique field isomorphism $K \rightarrow E$ that is the identity map on the subset $E - \phi(F) \subset K$. We can also think of F as a subset of E simply by identifying each $a \in F$ with its image $\phi(a) \in E$. Under this identification, the isomorphism $K \rightarrow E$ becomes the identity map.

We now return to our first basic question.

Theorem 1.7. *(Kronecker's Theorem) Given a non-constant $f \in F[x]$, there is extension field $E \geq F$ that contains a root of f .*

Proof: The idea is to find a suitable factor ring of $F[x]$ in which $f(x)$ is set to zero.

Consider the case when f is irreducible first. Then the principal ideal (f) is maximal, and hence $E = F[x]/(f)$ is field. Moreover, the map $F \rightarrow E, a \mapsto a + (f)$, is

clearly an injective ring homomorphism. Thus E can be regarded as an extension field of F . Put $\alpha = x + (f)$. Then for any given polynomial $g \in F[y]$, $g(\alpha) = g + (f) \in E$. In particular, for $g = f$ we have $f(\alpha) = 0$. Thus, E is an extension field of F that contains a root α of f , in this case.

For an arbitrary non-constant polynomial f , pick any irreducible factor p of f (which exists, by the Unique Factorization Theorem.) Then $f = pq$ for some $q \in F[x]$. The factor ring $E = F[x]/(p)$, again, becomes an extension field of F . For $\alpha = x + (p)$, we have $f(\alpha) = p(\alpha)q(\alpha) = 0$. Thus α is a root of f , as desired. \square

Let us can turn our basic question around and ask: given an extension field $E \geq F$ and an element $\alpha \in E$, when is α a root of a non-constant $f \in F[x]$?

Definition 1.8. *Given $\alpha \in E \geq F$, we say that α is algebraic over F if it is a root of a non-constant polynomial $f \in F[x]$. We say that α is transcendental over F , if it is not algebraic over F .*

The question above can be very hard even in the most familiar cases. For example, it is known that the real number $\pi = 3.141 \dots$ and the Euler number $e = 2.718 \dots$ are both transcendental over \mathbf{Q} . All known proofs uses analytic methods. (You can find proofs in Spivak's Calculus.) It is not known if the Euler-Macheroni constant, defined by

$$\gamma := \lim_{N \rightarrow \infty} \left(\sum_{n=1}^N 1/n - \ln(N) \right) = 0.577 \dots$$

is algebraic over \mathbf{Q} . It is not even known if it is irrational. Other settled but difficult cases include $e^\pi, \pi + e^\pi$, which are known to be transcendental. A famous theorem of Gelfond and Schneider (and Hilbert's 7th problem) says that if $a, b \in \mathbf{C}$ are algebraic over \mathbf{Q} , not equal to 0 or 1, and if b is not rational, then a^b is transcendental over \mathbf{Q} . But $e^{r\pi\sqrt{-1}} \in \mathbf{C}$ is algebraic, for any $r \in \mathbf{Q}$.

In each of these examples, a real or complex number can be specified by an (non-finitistic) algorithm that produces all its digits. It is usually difficult to use them effectively either to find a polynomial relation for the number or to exclude all polynomials relations for it. On the other hand, if a real number (such as $\sqrt{2}$) can specified by way of a purely algebraic characterization, then deciding its algebraicity is relatively easy.

We now motivate the rest of this chapter with the following more accessible algebraic question. Let $\alpha, \beta \in E \geq F$ be nonzero elements. If α, β are both algebraic over F , are $\alpha - \beta$ and α/β (if $\beta \neq 0$) also algebraic over F ? In other words, is algebraicity a property that is preserved under the usual arithmetical operations? The rest of this chapter will be about answering this, along with some of those questions raised in the introduction.

1.3. Algebraicity criteria

Adjoining an element to a field. Let $\alpha \in E \geq F$. Let $F(\alpha)$ denotes the smallest subfield of E containing F and α . We say that $F(\alpha)$ is the field generated by α over F , or that we adjoint α to F . Here is a simple description of $F(\alpha)$. If K is a subfield of E containing F and α , then K necessarily contains $f(\alpha)$, for every $f \in F[x]$. Since K is also closed under division, it contains $f(\alpha)/g(\alpha)$, for every $f, g \in F[x]$ with $g(\alpha) \neq 0$. Conversely, the subset of E consisting of such ratios $f(\alpha)/g(\alpha)$ is clearly a subgroup of E which is closed under division. Therefore, these ratios form a subfield of E , and it is contained in every subfield of E that contains F and α . So, we have

$$F(\alpha) = \{f(\alpha)/g(\alpha) \mid f, g \in F[x], g(\alpha) \neq 0\}.$$

Put

$$F[\alpha] = \{f(\alpha) \mid f \in F[x]\} \subset F(\alpha).$$

This is clearly a subring (but not necessarily a subfield) of $F(\alpha)$.

Recall that we have ring homomorphism $E[x] \rightarrow E$, $f \mapsto f(\alpha)$, which we call the *evaluation* homomorphism at $\alpha \in E$. Since $F[x] \subset E[x]$, we can restrict this map to $F[x]$. Denote this restriction

$$\phi_\alpha : F[x] \rightarrow E.$$

This induces a ring isomorphism $F[x]/\text{Ker } \phi_\alpha \rightarrow \phi(F[x]) = F[\alpha] \subset E$. Since E is a field, hence a domain, it follows that $\text{Ker } \phi_\alpha$ is a prime ideal of $F[x]$.

Theorem 1.9. (*Algebraicity Criteria*) Let $\alpha \in E \geq F$. The following are equivalent.

- i. α is algebraic over F .

ii. *There exists an irreducible monic p in $F[x]$ such that $p(\alpha) = 0$.*

iii. *$\text{Ker } \phi_\alpha = (p)$ for some irreducible monic p in $F[x]$.*

iv. *$F[\alpha]$ is a field.*

v. *$F[\alpha] = F(\alpha)$.*

Proof: Assume i. Then there is a non-constant polynomial $f \in F[x]$ such that $f(\alpha) = 0$. By the Unique Factorization Theorem, we can write $f = cp_1 \cdots p_k$ where $c \in F$ and $p_1, \dots, p_k \in F[x]$ are irreducible monic. That $f(\alpha) = 0$ implies that $p_j(\alpha) = 0$ for some j , proving ii.

Assume ii. Then $p \in \text{Ker } \phi_\alpha$, proving iii.

Assume iii. Then $\text{Ker } \phi_\alpha$ is a maximal ideal of $F[x]$, since it is nonzero and prime. So, $F[x]/\text{Ker } \phi_\alpha$ is a field. But it is isomorphic to $F[\alpha]$, so this is a field, proving iv.

Assume iv. So, $F[\alpha]$ is a field containing F and α , implying that $F[\alpha]$ contains $F(\alpha)$. But the reverse containment is obvious. So v. is true.

Finally, assume v. If $\alpha = 0$ then i is true. If $\alpha \neq 0$ then it is invertible in the field $F[\alpha]$, and so we can find $g \in F[x]$ such that $\alpha g(\alpha) = 1$. In this case, α a root of the non-constant polynomial $xg(x) - 1 \in F[x]$, proving i.

This completes the proof. \square

Corollary 1.10. *α is transcendental over F iff ϕ_α is injective.*

If $\alpha \in E \geq F$ is algebraic, the theorem yields an irreducible monic p in the ideal $\text{Ker } \phi_\alpha$ of $F[x]$. Let q be the unique monic generator of this ideal. Then q divides p , implying that $q = p$, since p is irreducible and monic. So, if α is algebraic over F , we have shown that there is a unique irreducible monic having α as a root.

Definition 1.11. If $\alpha \in E \geq F$ is algebraic over F , the irreducible monic $p \in F[x]$ having α as a root is called the irreducible polynomial of α over F . We denote the degree of p by $\deg_F \alpha$, and call it the degree of α over F .

Warning. When speaking of the irreducible polynomial of $\alpha \in E$, it is important that we specify the field F over which α is said to be algebraic. Consider the example $\alpha = i\sqrt[4]{2} \in \mathbf{C}$. It is algebraic over \mathbf{R} . It satisfying no degree 1 relation over \mathbf{R} . But $\alpha^2 = -\sqrt{2}$. So, $p = x^2 + \sqrt{2}$ is the irreducible polynomial of α over \mathbf{R} . Now α is also algebraic over \mathbf{Q} , and it is easy to verify that α satisfies no polynomial relation of degree 2 or 3, and that it is a root of $x^4 - 2 \in \mathbf{Q}[x]$. So, this must be the irreducible polynomial of α over \mathbf{Q} .

Exercise. Find the degrees of $\alpha = \sqrt{2} + i \in \mathbf{C}$ over \mathbf{Q} , and over $\mathbf{Q}[\sqrt{2}]$.

Exercise. Suppose that $\alpha \in \mathbf{C}$ satisfies the relation

$$\alpha^3 + \alpha + 2 = 0.$$

Express $(\alpha^2 + 1)(\alpha + 2)$ and $(\alpha - 1)^{-1}$ in the form

$$a\alpha^2 + b\alpha + c$$

with $a, b, c \in \mathbf{Q}$.

Example. We know that π is transcendental over \mathbf{Q} . But π is algebraic over the field $F = \mathbf{Q}(\pi^3)$ of degree ≤ 3 , since π is a root of $q = x^3 - \pi^3 \in F[x]$. We will show that q is the irreducible polynomial of π over F , hence π has degree 3 over F . Let p be the irreducible polynomial of π over F . Then $q \in (p) = \text{Ker } \phi_\pi$. So, p divides q in $F[x]$. Since π is transcendental over \mathbf{Q} , $\pi \notin \mathbf{Q}(\pi^3)$. Since the only degree 1 monic having π as a root is $x - \pi \notin F[x]$, p can only have degree 2 or 3. Suppose p has degree 2. Then its roots in \mathbf{C} must also be roots of p , which are π and $-\frac{1}{2}\pi \pm \frac{1}{2}i\sqrt{3}\pi$ because $q = (x - \pi)(x^2 + \pi x + \pi^2)$. Since p is real, the only possibility left for p is that $p = x^2 + \pi x + \pi^2 \in F[x]$. So, we have $\pi \in F$, a contradiction. Thus p must have degree 3, and so $p = q$.

Exercise. Let F be a finite field with q elements, and let $\alpha \in E \geq F$ be algebraic over F of degree $n + 1$. Show that the map

$$F^{n+1} \rightarrow F[\alpha], \quad (a_0, \dots, a_n) \mapsto \sum_{i=0}^n a_i \alpha^i$$

is a group isomorphism. In particular, the field $F[\alpha]$ has exactly q^{n+1} elements.

1.4. Brief review of linear algebra

The concepts and basic results reviewed in this section are straightforward generalizations of some of those found in Chapters 6-7 of my book, *Linear Algebra*. The generalizations here involves no more than replacing \mathbf{R} or \mathbf{C} there, by an abstract field F . The proofs carry over verbatim. Therefore, details of the proofs will not be repeated here, but main ideas of some of the theorems will be reviewed.

Fix a field F .

Definition 1.12. *A vector space over F is an additive group V , equipped with one additional operation $F \times V \rightarrow V$, $(a, v) \mapsto av$, which we call scaling, satisfying the following axioms. For $a, b \in F$ and $u, v \in V$,*

$$V1. \quad a(u + v) = au + av.$$

$$V2. \quad (a + b)v = av + bv.$$

$$V3. \quad a(bv) = (ab)v.$$

$$V4. \quad 1v = v.$$

Let V be a vector space over F . Elements of V are referred to as vectors, and elements of F as scalars. A subgroup $W \subset V$ which is closed under scaling is called a vector (or linear) subspace of V over F .

Example. *The primary example.* Let $E \geq F$ be any given extension field. Then the multiplication $F \times E \rightarrow E$, $(a, \alpha) \mapsto a\alpha$, in the field E , defines a vector space structure on E over the subfield F . The identities V1-V4 in this case follows immediately from the ring axioms.

Example. $\mathbf{C} \geq \mathbf{R}$ is a vector space over \mathbf{R} . In this case, every vector in \mathbf{C} can be uniquely expressed as $a + bi$, for some $a, b \in \mathbf{R}$. We will see that this means that the dimension of \mathbf{C} over \mathbf{R} is 2.

Definition 1.13. Let V be a vector space over F , and S a subset of V . We introduce the following terminology.

- i. A vector of the form $\sum_{j=1}^k a_j v_j = a_1 v_1 + \cdots + a_k v_k$ is called a linear combination of $v_1, \dots, v_k \in V$ with coefficients $a_1, \dots, a_k \in F$.
- ii. We say that S is linearly dependent if there are distinct vectors $v_1, \dots, v_k \in S$, and scalars $a_1, \dots, a_k \in F$, not all zero, such that $\sum_{j=1}^k a_j v_j = 0$. Otherwise, we say that S is linearly independent.
- iii. We call the set $\{\sum_{j=1}^k a_j v_j \mid v_1, \dots, v_k \in S, a_1, \dots, a_k \in F\}$, the linear span of S , and denote it by $\text{Span}_F S$. We say that S spans V , if $V = \text{Span}_F S$.
- iv. We call S a basis of V , if S spans V and if S is linearly independent.
- v. If V' is a vector space over F , we call a group homomorphism $\phi : V \rightarrow V'$, a linear homomorphism if ϕ preserves scaling, i.e. $\phi(av) = a\phi(v)$, for all $a \in F$ and $v \in V$. A linear isomorphism is a bijective linear homomorphism.

Let V be a vector space over F , and S a subset of V . It is easy to check that $\text{Span}_F S$ is a linear subspace of V . Similarly, a linear subspace W of V is always closed under taking linear combinations of vectors in W . By definition, the empty set is linearly independent, and its linear span is the subspace $\{0\}$.

Remark 1.14. To be brief, we often suppress the use of the word linear, when no confusion arises. Since a field E is a vector space over any given subfield F of E , it is important that we are specific about which subfield is under consideration, especially when more than one subfields are in play in a context. To avoid confusion, we will use the phrase “over F ” for emphasis, when considering independence, bases, spans, homomorphisms, etc, with respect to the subfield F .

Exercise. Let $\phi : V \rightarrow V'$ be a linear homomorphism over F . Show the following results (cf. their familiar analogues in group theory.)

- i. The subgroup $\text{Ker } \phi$ of V is a subspace of V .
- ii. If W is a subspace of V , then $\phi(W)$ is a subspace of V' .
- iii. If W' a subspace of V' , then $\phi^{-1}(W')$ is a subspace of V .
- iv. If ϕ is an isomorphism, then so is ϕ^{-1} .

Exercise. *More Algebraicity Criteria.* Show that $\alpha \in E \geq F$ is algebraic over F iff $1, \alpha, \dots, \alpha^n$ are dependent over F , for some positive integer n .

Example. *The primary example.* Given $\alpha \in E \geq F$, the evaluation map is the ring homomorphism

$$\phi_\alpha : F[x] \rightarrow E, f \mapsto f(\alpha).$$

It is linear over F , since $af \mapsto af(\alpha)$, for $a \in F$ and $f \in F[x]$.

Theorem 1.15. (*Finite Basis Theorem*) Let S be a finite subset of a vector space V that spans V . Then there is a subset B of S which is a basis of V .

Proof: Let v_1, \dots, v_k be the distinct vectors in S . If S is independent, then S is a basis of V . Suppose S is dependent. Then at least one vector, say v_k , can be expressed as a combination of the rest v_1, \dots, v_{k-1} . It follows that $V = \text{Span}_F\{v_1, \dots, v_k\} = \text{Span}_F\{v_1, \dots, v_{k-1}\}$. Repeat this process with $\{v_1, \dots, v_{k-1}\}$. It terminates at some independent set $B \subset S$ such that $V = \text{Span}_F B$. \square

Theorem 1.16. (*Uniqueness of Coefficients*) Let S be a basis of V . Then every vector in V can be expressed as a combination of vectors in B in exactly one way.

Proof: Let C be the set consisting of all functions $\lambda : B \rightarrow F$ such that $\lambda(u) = 0$ for all but finitely many $u \in B$. Specifying the coefficients for a linear combination of vectors in B is equivalent to specifying a function $\lambda \in C$. Let $v \in V$. To specify one way of expressing v as a combination of vectors in B is to specify a function $\lambda \in C$, such that $v = \sum_{u \in B} \lambda(u)u$. Suppose $\lambda' \in C$ is another such function, i.e.

$$v = \sum_{u \in B} \lambda(u)u = \sum_{u \in B} \lambda'(u)u.$$

It follows that $\sum_{u \in B} (\lambda(u) - \lambda'(u))u = 0$. By independence of B , it follows that $\lambda(u) - \lambda'(u) = 0$ for all $u \in B$, implying that $\lambda = \lambda'$. Thus the two ways of expressing v as a combination of vectors in B , are in fact the same. \square

The next theorem gives meaning to the notion of finite dimensions.

Theorem 1.17. (*Dimension Theorem*) *Suppose V has a finite basis with exactly k vectors.*

- i. Any set of $> k$ vectors in V is dependent.*
- ii. Any set of k independent vectors in V is a basis of V .*
- iii. Any set of $< k$ vectors in V does not span V .*
- iv. Any set of k vectors that spans V is a basis of V .*

Sketch of proof: Part i uses the fact that if a homogeneous system of linear equations, with coefficients in F , has more variables than equations then we can find a nontrivial solution by row reduction.

If S is a set of k independent vectors in V that fails to be a basis of V , then S fails to span V . In this case, we can find $v \in V - \text{Span}_F S$, and so $S \cup \{v\}$ is a set of $k + 1$ vectors in V which is independent, contradicting i.

If S is a set of $< k$ vectors in V that spans V , then we can find a subset B of S which is a basis of V , by the Finite Basis Theorem. In this case, the basis of k vectors we began with would be dependent, by i. This is a contradiction.

If S is a set of k vectors that spans V , then again we can find a subset B of S which is a basis of V . By iii, B can't have $< k$ vectors, and so $S = B$ is a basis of V . \square

Corollary 1.18. *If a vector space has a finite basis, then any two of its bases have the same cardinality.*

Definition 1.19. *Let V be a vector space over F . If V has a finite basis, the cardinality of the basis is called the dimension of V over F , and is denoted by $\dim_F V$. If V has no*

finite basis, we say that V has infinite dimension. If V has a (infinite) countable basis, we say that V has countable dimension over F . If V has neither a finite nor a countable basis, we say that V has uncountable dimension over F .

Corollary 1.20. *Let W be a subspace of a finite dimensional vector space V over F . Then $\dim_F W \leq \dim_F V$, and equality holds iff $W = V$.*

Proof: Exercise.

Theorem 1.21. *(Invariance of Dimension) Let $\phi : V \rightarrow V'$ be an isomorphism over F . Then ϕ maps each basis of V to a basis of V' . In particular, V and V' have the same dimension over F .*

Proof: That ϕ is injective implies that it preserves independence. For if B is an independent set in V , then for distinct $v_1, \dots, v_k \in V$ and for $a_1, \dots, a_k \in F$ such that

$$\sum_i a_i \phi(v_i) = 0,$$

we have $\phi(\sum_i a_i v_i) = 0$ by linearity, and we have $\sum_i a_i v_i = 0$ by injectivity of ϕ , implying that $a_1 = \dots = a_k = 0$.

That ϕ is surjective implies that it preserves the spanning property. For if B is a set that spans V , then by linearity, we have $\phi(V) = \phi(\text{Span}_F B) = \text{Span}_F \phi(B)$. But $\phi(V) = V'$, by surjectivity of ϕ , implying that $\phi(B)$ spans V' .

Therefore, if B is a basis of V , then $\phi(B)$ is independent in V' and spans V' , and so it is a basis of V' . \square

Example. (1) The n -fold Cartesian product $F^n = F \times \dots \times F$ is a vector space over F . The vectors $e_i = (0, \dots, 0, 1, 0, \dots, 0)$, where 1 is the i th entry, with $i = 1, \dots, n$, form a basis of F^n . So, $\dim_F F^n = n$.

(2) If F is a finite field with q elements and if $\alpha \in E \rightarrow F$ has degree n , recall that we have a linear isomorphism $F^n \rightarrow F(\alpha) = F[\alpha]$, $(a_0, \dots, a_{n-1}) \mapsto \sum_i a_i \alpha^i$. We will generalize this to an arbitrary field F later.

(3) \mathbf{C} is a vector space over \mathbf{R} with basis $\{1, \sqrt{-1}\}$. So, $\dim_{\mathbf{R}}\mathbf{C} = 2$.

(4) Likewise $\mathbf{Q}(\sqrt{2})$ is a vector space over \mathbf{Q} with basis $\{1, \sqrt{2}\}$.

(5) The polynomial ring $F[x]$ is a vector space over F with a countable basis $\{1, x, x^2, \dots\}$.

(6) In a homework problem, you will prove that if $\alpha \in E \geq F$ is transcendental over F , then the extension field $F(\alpha) \geq F$ has uncountable dimension over F .

Exercise. Show that \mathbf{R} , as a vector space over \mathbf{Q} , has uncountable dimension.

Proposition 1.22. (*Factor Space*) Let V be a vector space and U a subspace over F . Then the factor group V/U has a vector space structure over F , with scaling defined by

$$a(v + U) = av + U$$

for $a \in F$ and $v \in V$. We call V/U the quotient or factor space of V by U .

Proof: We first show that the scaling operation is well-defined. Let $a \in F$, and $v, v' \in V$ such that $v + U = v' + U$. Then $v - v' \in U$. Since U is closed under scaling in V , $a(v - v') = av - av' \in U$, implying that $av + U = av' + U$. So $av + U$ does not depend on the choice of representative v of the coset $v + U$.

Now checking properties V1-V4 for V/U is straightforward and is left to the reader.

□

Exercise. Verify that the projection homomorphism $V \rightarrow V/U$, $v \mapsto v + U$ is linear.

Here comes the most important example in field theory. Let N be a given ideal of $F[x]$. Then N is a subspace of $F[x]$ over F . So, $F[x]/N$ is a vector space. Let's compute its dimension over F . If $N = (0)$, then $F[x]/N = F[x]$ has countable dimension, as before. If $N = F[x]$, then $F[x]/N$ is the zero space and has dimension 0 in this case. Finally, we consider $N = (f)$, $f \in F[x]$ a non-constant polynomial.

Theorem 1.23. (*Basis Theorem for $F[\bar{x}]$.*) Let $f \in F[x]$ be a polynomial of degree $n \geq 1$, and put $\bar{x} = x + (f)$. Then the set $S = \{1, \bar{x}, \dots, \bar{x}^{n-1}\}$ is a basis of the factor space $F[x]/(f)$ over F .

Proof: Recall that the ring homomorphism $\gamma : F[x] \rightarrow F[x]/(f)$, $g \mapsto g + (f)$, is the same as the “evaluation” map $g \mapsto g(\bar{x})$. Put $f = \sum_{i=0}^n a_i x^i$. Since f has degree n , $a_n \neq 0$. Since $f(\bar{x}) = f + (f) = 0$, it follows that

$$\bar{x}^n = -\frac{1}{a_n} \sum_{i=0}^{n-1} a_i \bar{x}^i \in \text{Span}_F S.$$

Obviously, we have $1, \bar{x}, \dots, \bar{x}^{n-1} \in \text{Span}_F S$ as well. We do induction. Take $k \geq n$, and assume that $\bar{x}^k \in \text{Span}_F S$. Then

$$\bar{x}^{k+1} = \sum_{i=0}^{n-1} b_i \bar{x}^{i+1}$$

for some $b_0, \dots, b_{n-1} \in F$. Since $\bar{x}, \dots, \bar{x}^n \in \text{Span}_F S$, by the inductive assumption, it follows that $\bar{x}^{k+1} \in \text{Span}_F S$. We have shown that all powers of \bar{x} lies in $\text{Span}_F S$. It follows that $F[x]/(f) = \gamma(F[x]) = F[\bar{x}] = \text{Span}_F S$, i.e. S spans $F[x]/(f)$ over F .

To show linear independence of S , let $\sum_{i=0}^{n-1} b_i \bar{x}^i = 0$ for some $b_i \in F$. We can write this as $\gamma(g) = 0$, where $g = \sum_{i=0}^{n-1} b_i x^i \in F[x]$. Thus $g \in \text{Ker } \gamma = (f)$, and so f divides g . Since $\deg g < n = \deg f$, it follows that $g = 0$, i.e. $b_0 = \dots = b_{n-1} = 0$. \square

Corollary 1.24. Suppose $\alpha \in E \geq F$ is algebraic over F , of degree n . Then

$$\dim_F F(\alpha) = n.$$

Proof: Let p be the irreducible polynomial of α . By the Algebraicity Criteria, the evaluation map $\phi_\alpha : F[x] \rightarrow E$, $g \mapsto g(\alpha)$, induces a linear isomorphism $F[x]/(p) \rightarrow \phi(F[x]) = F[\alpha] = F(\alpha)$. By the Invariance of Dimension, all four vector spaces have the same dimension over F , which is $\deg p = n$, by the preceding theorem. \square

1.5. Algebraic extensions

In this section, let F be a given field.

Definition 1.25. Let $E \geq F$ be an extension field. We denote by $[E : F]$, the dimension of E over F , and also call it the degree of E over F . We say that E is an algebraic extension of F or that E is algebraic over F , if every element of E is algebraic over F , i.e. $\deg_F \alpha = [F(\alpha) : F]$ is finite for every $\alpha \in E$. We say that E is finite extension of F or that E is finite over F , if $[E : F]$ is finite. Warning: E being a finite extension of F does not mean that E is a finite set.

Remark 1.26. The reader might be wondering why we use two different names – dimension and degree – for the same number. One name comes from the linear algebra viewpoint, while the other comes from the polynomial algebra viewpoint, and field theory is one place where the two viewpoints converge and provide complementary tools for questions in arithmetic. It is therefore helpful to keep both in mind.

Exercise. Show that a finite extension E of F has degree 1 iff $E = F$.

Proposition 1.27. Let $E \geq F$ be an extension field. If E is finite over F , then it is algebraic over F . In fact, for any $\alpha \in E$, we have $[F(\alpha) : F] \leq [E : F]$.

Proof: Let $\alpha \in E$. Then $F(\alpha)$ is a subspace of the finite dimensional vector space E over F . By a corollary to the Dimension Theorem,

$$[F(\alpha) : F] \leq [E : F]. \quad \square$$

We shall see later that an algebraic extension over F need not be a finite extension.

The next theorem is the field theory analogue of Lagrange's theorem in group theory.

Theorem 1.28. (*Degree Factorization Theorem*) Let E be a finite extension over F , and K a finite extension over E . Then K is a finite extension over F , and

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

Proof: Let $\{\alpha_1, \dots, \alpha_m\}$ be a basis of E over F , and $\{\beta_1, \dots, \beta_n\}$ a basis of K over E . We will show that the set $S = \{\alpha_i \beta_j \mid i = 1, \dots, m, j = 1, \dots, n\}$ is a basis of K over F , with

mn elements. Let $\gamma \in K$. Then $\gamma = \sum_{j=1}^n b_j \beta_j$ for some $b_j \in E$. Likewise, for each j , $b_j = \sum_{i=1}^m a_{ij} \alpha_i$ for some $a_{ij} \in F$. So,

$$\gamma = \sum_{i=1}^m \sum_{j=1}^n a_{ij} \alpha_i \beta_j.$$

We have shown that S spans K over F .

To show that the list of mn elements in S are independent over F , let

$$0 = \sum_{i=1}^m \sum_{j=1}^n a_{ij} \alpha_i \beta_j$$

for some $a_{ij} \in F$. Since $\sum_{i=1}^m a_{ij} \alpha_i \in E$ for each j , and since $\beta_1, \dots, \beta_n \in K$ are independent over E , it follows that $\sum_{i=1}^m a_{ij} \alpha_i = 0$ for each j . Since $\alpha_1, \dots, \alpha_m \in E$ are independent over F , it follows that $a_{ij} = 0$ for all i, j . \square

Corollary 1.29. *Let $F_r \geq F_{r-1} \geq \dots \geq F_1$ be a list of finite extensions. Then*

$$[F_r : F_1] = [F_r : F_{r-1}] \cdots [F_2 : F_1].$$

Proof: Applying the theorem repeatedly, we see that each F_j is a finite extension of F_i , for $j > i$. Moreover,

$$[F_r : F_1] = [F_r : F_{r-1}][F_{r-1} : F_1] = \cdots = [F_r : F_{r-1}][F_{r-1} : F_{r-2}] \cdots [F_2 : F_1]. \quad \square$$

Corollary 1.30. *Let $\alpha \in E \geq F$ be algebraic over F , and $\beta \in F(\alpha)$. Then $\deg_F \beta$ divides $\deg_F \alpha$.*

Proof: We have $F(\alpha) \geq F(\beta) \geq F$. By the theorem, $[F(\alpha) : F] = [F(\alpha) : F(\beta)][F(\beta) : F]$. \square

Exercise. Let E be a finite extension of F , of degree p which is a prime number. Show that E and F are the only extension fields of F that are contained in E .

Exercise. Let $\alpha = \sqrt[3]{2}$. Show that there is no $\beta \in \mathbf{Q}(\alpha)$ which is algebraic over \mathbf{Q} , of degree 2.

Exercise. Let $\alpha \in E \geq F$ be algebraic over F , of odd degree. Show that $F(\alpha^2) = F(\alpha)$.

Definition 1.31. (Notations) Let $\alpha_1, \dots, \alpha_n \in E \geq F$. We denote by $F(\alpha_1, \dots, \alpha_n)$ the smallest subfield of E containing F and all $\alpha_1, \dots, \alpha_n$. Thus, $F(\alpha_1, \dots, \alpha_n)$ consists of all ratios $f(\alpha_1, \dots, \alpha_n)/g(\alpha_1, \dots, \alpha_n)$, where $f, g \in F[x_1, \dots, x_n]$ are polynomials in n variables such that $g(\alpha_1, \dots, \alpha_n) \neq 0$. We call $F(\alpha_1, \dots, \alpha_n)$ the field generated by $\alpha_1, \dots, \alpha_n$ (in E) over F . We also say that we generate the field by adjoining $\alpha_1, \dots, \alpha_n$ to F .

Lemma 1.32. In the notation above, for $1 \leq k \leq n$, we have

$$F(\alpha_1, \dots, \alpha_n) = F(\alpha_1, \dots, \alpha_k)(\alpha_{k+1}, \dots, \alpha_n).$$

Proof: Both sides of the equation are fields that contain F and all $\alpha_1, \dots, \alpha_n$. The left side is the *smallest* of such fields. So, the left side is contained in the right side. Now, both sides are also fields that contain $F(\alpha_1, \dots, \alpha_k)$ and all $\alpha_{k+1}, \dots, \alpha_n$, and the right side is the smallest of such fields. So, the right side is contained in the left side. \square

Corollary 1.33. (Finite Generation of Finite Extensions) Let $E \geq F$ be a finite extension. Then E is finitely generated over F . That is, there exist finitely many $\alpha_1, \dots, \alpha_n \in E$ such that

$$E = F(\alpha_1, \dots, \alpha_n).$$

Proof: We do strong induction on the degree $[E : F]$. If $[E : F] = 1$, then $E = F = F(1)$. Assume that our assertion is true for any finite extensions of degree $\leq k$, over any field. Let $E \geq F$ be a finite extension, of degree $k + 1$. Pick any $\alpha_1 \in E - F$. Then

$$E \geq F(\alpha_1) > F$$

and $[E : F] = [E : F(\alpha_1)][F(\alpha_1) : F]$. Since $\alpha_1 \notin F$, we have $[F(\alpha_1) : F] > 1$, hence $[E : F(\alpha_1)] < [E : F]$. Applying the inductive assumption to the finite extension $E \geq F(\alpha_1)$, we can find $\alpha_2, \dots, \alpha_n \in E$ such that

$$E = F(\alpha_1)(\alpha_2, \dots, \alpha_n) = F(\alpha_1, \dots, \alpha_n). \quad \square$$

We now discuss the converse of the preceding corollary.

Lemma 1.34. *Let $\alpha \in E \geq F$ be algebraic over F . Then α is algebraic over any field $K \geq F$ contained in E .*

Proof: Since α is algebraic over F , for some $n > 0$, the list $1, \alpha, \dots, \alpha^n$ is dependent over F . The same list is dependent over K , since $F \subset K$. So, α is algebraic over K . \square

Corollary 1.35. *Let $\alpha_1, \dots, \alpha_n \in E \geq F$ be algebraic over F . Then the field $F(\alpha_1, \dots, \alpha_n)$ is finite over F .*

Proof: Let $1 \leq i \leq n$. Since α_i is algebraic over F , it is algebraic over $F(\alpha_1, \dots, \alpha_{i-1})$, by the lemma. It follows that $F(\alpha_1, \dots, \alpha_i) = F(\alpha_1, \dots, \alpha_{i-1})(\alpha_i) \geq F(\alpha_1, \dots, \alpha_{i-1})$ is a finite extension. By a corollary to the Degree Factorization Theorem, $F(\alpha_1, \dots, \alpha_n) \geq F$ is a finite extension. \square

Finally, this yields the answer to one of our basic questions.

Corollary 1.36. *Let $\alpha, \beta \in E \geq F$ be algebraic over F . Then $\alpha - \beta$ and α/β (if $\beta \neq 0$) are algebraic over F . Thus, the set E^a consisting of all $\alpha \in E$ that are algebraic over F , is a field.*

Proof: Since $F(\alpha, \beta)$ has finite dimension over F , so does the subspace $F(\alpha - \beta) \subset F(\alpha, \beta)$, implying that $\alpha - \beta$ is algebraic over F . Likewise, if $\beta \neq 0$ then $F(\alpha/\beta)$ has finite dimension over F . \square

Exercise. Let $\alpha, \beta \in E \geq F$ be algebraic over F , of respective degrees m, n . Show that $[F(\alpha, \beta) : F] \leq mn$. Given an example where the equality fails.

Example. *Algebraic numbers.* An algebraic number is a complex number that is algebraic over \mathbf{Q} . In a homework problem, you will show that the field \mathbf{C}^a of all algebraic numbers, is algebraically closed. It is called *the algebraic closure* of \mathbf{Q} , and is usually denoted by $\bar{\mathbf{Q}}$.

1.6. Algebraic closure

In this section, let F be a given field.

Definition 1.37. *An extension field $E \geq F$ is called an algebraic closure of F , if E is an algebraic extension of F and if E is algebraically closed.*

Example. \mathbf{C} is an algebraic closure of \mathbf{R} , since $[\mathbf{C} : \mathbf{R}] = 2$ and since \mathbf{C} is algebraically closed (Fundamental Theorem of Algebra.)

Example. $\bar{\mathbf{Q}}$ is an algebraic closure of \mathbf{Q} , as you will show in a homework problem.

We now move on to the existence and uniqueness of algebraic closure of F . The construction (due to Artin.) has two main steps.

- i. For every field K , construct an extension field $L \geq K$ such that every non-constant polynomial $f \in K[x]$ has a root in L . Put $L^a = \{\alpha \in L \mid \alpha \text{ algebraic over } K\}$. Then L^a is an algebraic extension of K such that every non-constant polynomial $f \in K[x]$ has a root in L^a .
- ii. Using i, construct a chain of algebraic extensions $F = E_0 \leq E_1 \leq E_2 \leq \dots$, such that every non-constant $f \in E_i[x]$ has a root in E_{i+1} . Show that $E = \cup_{i \geq 1} E_i$ is an algebraic closure of F .

Steps i is carried out in your homework in full generality. We will do step i in the case of a finite or countable field K without using Zorn's lemma, and we will do step ii as well. Finally, we will also prove that any two algebraic closures of a given field F are isomorphic.

Theorem 1.38. *Let $E \geq F$ be an algebraic extension of F , and $K \geq E$ an algebraic extension of E . Then $K \geq F$ is an algebraic extension of F .*

Proof: Let $\alpha \in K$. Since K is algebraic over E , α is a root of a non-constant polynomial $f \in E[x]$, say $f = \sum_{i=0}^n a_i x^i$. Since E is algebraic over F , all $a_0, \dots, a_n \in E$ are algebraic over F . By a corollary to the Degree Factorization Theorem, the field $L = F(a_0, \dots, a_n)$ is

finite over F . Since α is a root of $f \in L[x] \supset E[x]$, $L(\alpha)$ is finite over L . By the Degree Factorization Theorem, $L(\alpha)$ is finite over F , and so every $\beta \in L(\alpha)$ is algebraic over F . In particular, α is algebraic over F . \square

Lemma 1.39. *Let $E_0 \leq E_1 \leq E_2 \leq \dots$ be a chain of field extensions. Then the set $E = \cup_{i \geq 0} E_i$ is a field. If, in addition, for each i , E_{i+1} is algebraic over E_i , then E is algebraic over E_0 .*

Proof: Let $\alpha, \beta \in E$. Then $\alpha, \beta \in E_i$ for some i . Since E_i is a field, $\alpha - \beta$ and α/β (if $\beta \neq 0$) lies in E_i . So, E has well-defined addition, negation, multiplication and inversion operations. Likewise, given any $\alpha, \beta, \gamma \in E$, all three lie in E_i for some i , and they satisfy all field axioms, because E_i is a field.

Assume that for each i , E_{i+1} is algebraic over E_i . Then by the theorem and induction, each E_i is algebraic over E_0 . It follows that every $\alpha \in E$ is algebraic over E_0 . \square

Lemma 1.40. *(Step ii.) Let $E_0 \leq E_1 \leq E_2 \leq \dots$ be a chain of field extensions, such that every non-constant polynomial $f \in E_i[x]$ has a root in E_{i+1} . Then the field $E = \cup_{i \geq 0} E_i$ is algebraically closed. If, in addition, for each i , E_{i+1} is algebraic over E_i , then E is an algebraic closure of E_0 .*

Proof: Let $f \in E[x]$ be a non-constant polynomial, say $f = \sum_{j=0}^n a_j x^j$. Again, $a_0, \dots, a_n \in E_i$ for some i , so $f \in E_i[x]$. By hypothesis, f has a root in $E_{i+1} \subset E$. This shows that E is algebraically closed.

Assume that for each i , E_{i+1} is algebraic over E_i . By the preceding lemma, E is algebraic over E_0 . Since E is also algebraically closed, it is an algebraic closure of E_0 . \square

Lemma 1.41. *(Step i.) Let $f_1, f_2, \dots \in F[x]$ be a list of non-constant polynomials. Then there is an algebraic extension E_1 of F such that every f_i has a root in E_1 . If, in addition, F is countable, then E_1 can be chosen to be countable.*

Proof: By Kronecker's theorem, there is a finite extension F_1 of F that contains a root of f_1 . Likewise, since $f_2 \in F_1[x] \supset F[x]$, there is a finite extension F_2 of F_1 that contains a

root of f_2 , and so on. So, we have a chain of finite (hence algebraic) extensions

$$F = F_0 \leq F_1 \leq F_2 \leq \cdots$$

such that for each i , F_i contains a root of f_i . It follows that the field $E_1 = \cup_{j \geq 0} F_j$ is algebraic over $F = F_0$, by Lemma 1.39, and it contains a root of every f_i . Note that if F is countable, then each F_i is also countable, and so being the union of a countable family of countable sets, E_1 is also countable. \square

Theorem 1.42. *Let F be a countable field. Then F has an algebraic closure which is countable.*

Proof: For $n \geq 0$, the set $F[x]_n$ of polynomials of degree $\leq n$ in $F[x]$ is a vector space over F , and is isomorphic to F^{n+1} , which is countable. It follows that $F[x] = \cup_{n \geq 0} F[x]_n$ is countable. So, we can list the non-constant polynomials in $F[x]$, say f_1, f_2, \dots . By the Step i Lemma, there is a countable algebraic extension E_1 of F every f_i has a root in E_1 .

Repeating this process with E_1 , we find a countable algebraic extension E_2 of E_1 such that every non-constant polynomial in $E_1[x]$ has a root in E_2 . Continuing this way, we have a chain of countable algebraic extensions

$$F = E_0 \leq E_1 \leq E_2 \leq \cdots$$

such that for each $j \geq 0$, every non-constant polynomial in $E_j[x]$ has a root in E_{j+1} . By the Step ii Lemma, the field $E = \cup_{j \geq 0} E_j$ is an algebraic closure of $F = E_0$. Since E is the union of a countable family of countable sets, E is also countable. \square

We now move on to the uniqueness of algebraic closure.

Lemma 1.43. *(Homomorphism Extension) Let $\alpha \in E \geq F$ be algebraic over F , and $\sigma : F \rightarrow L$ a field homomorphism into an algebraically closed field L . Then we can extend σ to a field homomorphism $\sigma' : F(\alpha) \rightarrow L$, i.e. $\sigma'(a) = \sigma(a)$ for all $a \in F$.*

Proof: Let $p \in F[x]$ be the irreducible polynomial of α , say $p = \sum_{i=0}^n a_i x^i$. For each $f = \sum_{i=0}^m b_i x^i \in F[x]$, we put $f^\sigma = \sum_{i=0}^m \sigma(b_i) x^i \in L[x]$. Since L is algebraically closed, f^σ has a root in L , say β . By the Algebraicity Criteria, we have the isomorphism (induced

by the evaluation map) $F[x]/(p) \rightarrow F[\alpha] = F(\alpha)$, $f + (p) \mapsto f(\alpha)$, which we will denote by ϕ .

Define the unital ring homomorphism $F[x] \rightarrow L$, $f \mapsto f^\sigma(\beta)$. Then p is in its kernel, since $p^\sigma(\beta) = 0$. Thus, we have an induced homomorphism $F[x]/(p) \rightarrow L$. It is injective, since $F[x]/(p)$ is a field. Compose it with $\phi^{-1} : F[\alpha] \rightarrow F[x]/(p)$, we get a field homomorphism

$$\sigma' : F[\alpha] \rightarrow L, \quad f(\alpha) \mapsto f^\sigma(\beta).$$

For $a \in F$, which is a constant polynomial, we have $\sigma'(a) = \sigma(a)$, as desired.

Theorem 1.44. (*Uniqueness of Algebraic Closure*) *Let K and L be both algebraic closures of F . Then we can extend the identity map $F \rightarrow F$ to an isomorphism $K \rightarrow L$.*

Proof: The main idea is quite generic: we consider all possible ways of extending the identity map of F to some larger fields contained in K , partially order all those different ways by the inclusion relation, and then use Zorn's lemma to find the maximal way of extending.

Let S be the set of all pairs (E, τ) , such that $F \leq E \leq K$ and $\tau : E \rightarrow L$ is a homomorphism, with $\tau(a) = a$ for all $a \in F$. We have the homomorphism $\iota : F \rightarrow L$, $\tau(a) = a$, so $(F, \iota) \in S$. Let $(E, \tau) \geq (E', \tau')$ iff $E \subset E'$ and $\tau'(b) = \tau(b)$ for all $b \in E$. This makes S a nonempty partially ordered set. Let $T = \{(E_i, \tau_i)\}$ be any linearly ordered subset of S . That means that for any i, j , either $(E_i, \tau_i) \geq (E_j, \tau_j)$ or $(E_j, \tau_j) \leq (E_i, \tau_i)$ holds. Put $E = \cup_i E_i$. As in Lemma 1.39, we can check that E is field such that $F \leq E \leq K$. Define $\tau : E \rightarrow L$, by setting $\tau(b) = \tau_i(b)$ for $b \in E_i$. Then it is easy to check (using the fact that T is linearly ordered) that τ is a well-defined homomorphism, with $\tau(a) = a$ for all $a \in F$. It follows that $(E, \tau) \in S$. Clearly, we have $(E, \tau) \geq (E_i, \tau_i)$ for all i , implying that (E, τ) is an upper bound of T . Thus, we have shown that every linearly ordered subset of S has an upper bound. So, we can apply Zorn's lemma to the partially ordered set S , and conclude that S has a maximal element (E, τ) . We will show that $E = K$.

Suppose the contrary, and let $\alpha \in K - E$. Since K is an algebraic extension of F , α is algebraic over F , hence over $E \supset F$. By the preceding lemma, we can extend $\tau : E \rightarrow L$

to a field homomorphism $\tau' : E(\alpha) \rightarrow L$. Since $K \geq E(\alpha) > E$ and $\tau'(a) = \tau(a) = a$ for all $a \in F$, we have $(E(\alpha), \tau') \in \mathcal{S}$ and $(E(\alpha), \tau') > (E, \tau)$, contradicting the maximality of (E, τ) . So, $E = K$.

It remains to show that $\tau : K \rightarrow L$ is an isomorphism. We have $L \geq \tau(K) \geq F$, and that L is algebraic over F . By Lemma 1.34, L is also algebraic over $\tau(K)$. But $\tau(K)$ is algebraically closed, since it is isomorphic to K . So, $\tau(K)$ has no proper algebraic extension, implying that $\tau(K) = L$. Since $\tau : K \rightarrow L$ is also injective, it is an isomorphism.

□

2. Euclidean Constructions

We will discuss one of the historically earliest applications of field theory. This application is primarily about proving the impossibility of a variety of geometric constructions on the Euclidean plane. One notable example is the proof that it is impossible to trisect certain angles using a compass and an unmarked straightedge.

2.1. Constructible objects

Roughly speaking, a constructible object is one that we *can* draw on the plane in a finite number of steps, using no more than a compass and a straightedge, starting from some simple initial data, such as a finite set of points. To make precise this notion, we begin with some notations.

We will think of the Euclidean plane \mathbf{R}^2 as \mathbf{C} , the field of complex numbers, under the usual identification that the point $(x, y) \in \mathbf{R}^2$ corresponds to $x + iy \in \mathbf{C}$. We refer to x, y as the coordinates of this point. Lines and circles in \mathbf{C} will have their usual elementary geometric meanings, and the rules of Euclidean geometry apply. For example, given two distinct points in \mathbf{C} , we can draw exactly one line (or rather, one arbitrarily long line segment) through the points, with a straightedge. The subset $\mathbf{R} \subset \mathbf{C}$ will be thought of as the real line in the complex plane. Let \mathcal{F} denote the set consisting of all finite subsets of \mathbf{C} .

Definition 2.1. Let $S \in \mathcal{F}$, i.e. S is a finite subset of \mathbf{C} . An S -constructible curve is either a line through two distinct points $a, b \in S$, or a circle centered at a point $c \in S$ with radius $|d - e|$, for some distinct $d, e \in S$.

Definition 2.2. Define a function $\varepsilon : \mathcal{F} \rightarrow \mathcal{F}$, which we call the Euclidean construction, as follows. For $S \in \mathcal{F}$, let

$$\varepsilon(S) = S \cup (\cup_{C \neq C'} C \cap C')$$

where the union is taken over all possible pairs of distinct S -constructible curves C, C' .

Since the intersection of two *distinct* curves, each being either a line or a circle, is always a finite set, it follows that $\varepsilon(S)$ above is finite. Geometrically, $z \in \varepsilon(S)$ means precisely that we can construct the point z by using compass and straightedge to draw circles and lines, starting from points in S .

Example. We have $\varepsilon(\emptyset) = \emptyset$, and $\varepsilon(\{0\}) = \{0\}$. For $S = \{0, 1\}$, there are exactly three S -constructible curves: the real axis, and the two unit circles centered respectively at 0, 1. Their pairwise intersections yield the set $\varepsilon(S)$ with exactly 6 points: $-1, 0, 1, 2, \frac{1}{2} \pm i\frac{\sqrt{3}}{2}$

Definition 2.3. We say that $z \in \mathbf{C}$ is constructible or that z is a constructible point, if

$$z \in \varepsilon^n(\{0, 1\})$$

for some $n \in \mathbf{N}$. We say that a line or a circle is a constructible curve, if it is S -constructible for some finite set S of constructible points.

Geometrically a constructible point z is a point we can construct, using a straight-edge and a compass, in a finite number of steps.

Proposition 2.4. Let $z \in \mathbf{C}$. Then z is constructible if and only if $z \in C \cap C'$, for some distinct constructible curves C, C' .

Proof: Assume z is constructible, say $z \in \varepsilon^n(\{0, 1\})$ for some $n \in \mathbf{N}$. Since $S = \varepsilon^{n-1}(\{0, 1\})$ is a set of constructible points (including the case $n = 1$), and $z \in \varepsilon(S)$ is an intersection point of two distinct S -constructible curves.

Conversely, assume that $z \in C \cap C'$, where C and C' are distinct constructible curves. Each of C, C' is a line or circle we can construct using compass and edge, and a finite number of constructible points in $\varepsilon^n(\{0, 1\})$, for some $n \in \mathbf{N}$. This implies that $z \in \varepsilon^{n+1}(\{0, 1\})$, so z is constructible. \square

Exercise. Show that every integer $n \in \mathbf{Z}$ is constructible. Show that every half integer is constructible. Show that the point i is constructible. (Hint: i is in the intersection of the line of imaginary numbers, and the unit circle centered at 0.)

Exercise. Show that every Gaussian integer $m + in \in \mathbf{C}$ ($m, n \in \mathbf{Z}$) is constructible.

Exercise. *Bisector.* Let $a, b \in \alpha$ be distinct constructible. Show that the bisector line of the segment \overline{ab} is constructible. In particular, $\frac{1}{2}a$ is also constructible.

Exercise. *Absolute value.* Let $a \in \mathbf{C}$ be constructible. Then $|a|$ is constructible.

Proposition 2.5. (*Parallelogram Trick*) Let $a, b, c, d \in \mathbf{C}$ be vertices of a parallelogram such that at least three of them are constructible and not collinear. Then the fourth vertex is also constructible.

Proof: We can arrange the points so that a, b, c, d are in counterclockwise order, and that a, b, c are constructible. Then d is an intersection point of the circles with respective centers a, c and radii $|b - c|, |a - b|$. Thus, d is constructible. \square

Exercise. *Parallel translation.* Show that if $a, b, c \in \mathbf{C}$ are constructible and not collinear, then the line parallel to segment \overline{ab} and passing through c is constructible.

Exercise. *Additivity.* Let $a, b \in \mathbf{C}$ be constructible. Show that $a - b$ is constructible. Conclude that the set of constructible points is an additive subgroup of \mathbf{C} . (Hint: Apply the parallelogram trick to $0, a, b, a - b$ in the non collinear case.)

Exercise. *Conjugation.* Let $a \in \mathbf{C}$ be constructible. Show that its complex conjugate \bar{a} is also constructible. (Hint: If $a \notin \mathbf{R}$, the circle centered at a of radius $|a|$ intersects \mathbf{R} at two points $0, b$. Then consider the parallelogram with vertices $0, a, b, \bar{a}$.)

Exercise. *Reflection.* More generally, let $a \in \mathbf{C}$ be constructible, and L be a constructible line. Then the reflection image a' of a along L is also constructible.

Exercise. Squaring. Let $a \in \mathbf{R}$ be constructible. Show that a^2 is constructible. (Hint: Construct two *similar* triangles with vertices $0, 1, b$ and $0, a, c$, where $0, b, c$ are collinear. Then $|a|/|c| = 1/|a|$.)

Exercise. Inverting. Let $a \in \mathbf{R}$ be nonzero constructible. Show that $1/a$ is constructible. (Hint: Construct two similar triangles to get $|b|/1 = 1/a$.)

Theorem 2.6. *Let $\alpha, \beta \in \mathbf{C}$ be constructible. Then $\alpha - \beta$, $\alpha\beta$ and $1/\beta$ (if $\beta \neq 0$) are constructible. Therefore, the set of constructible complex numbers form a subfield K of \mathbf{C} .*

Proof: The second assertion follows from the first assertion. For the first assertion, the case when one of α, β is zero, is trivial. So, consider the case when neither is zero. In an exercise above (on additivity), we saw that $\alpha - \beta$ is constructible.

Next, we show that $\alpha\beta$ is constructible. The line L through $0, \alpha$ is constructible. The reflection image γ of 1 along L is constructible. So, the line L' through 0 and γ is constructible. Since the absolute value $|\alpha|$ is constructible, its square $|\alpha|^2$ is also constructible. Now α^2 is an intersection point of the circle centered at 0 of radii $|\alpha|^2$, with the line L' . So, α^2 is constructible. Likewise β^2 and $(\alpha + \beta)^2$ are also constructible. By additivity and bisecting properties, it follows that

$$\alpha\beta = \frac{1}{2}[(\alpha + \beta)^2 - \alpha^2 - \beta^2]$$

is constructible.

Finally, since $\bar{\beta}$ and $1/|\beta|^2$ are also constructible, it follows that $1/\beta = \bar{\beta}/|\beta|^2$ is also constructible. \square

Corollary 2.7. *The set of constructible real numbers form a subfield of \mathbf{R} .*

Proof: This set is $K \cap \mathbf{R}$, which is the intersection of two subfields of \mathbf{C} , and is therefore a field. \square

We now give another useful description of K . Consider a field with the following property:

(*) *it is a subfield of \mathbf{R} which is closed under taking the square root of any of its positive element.*

Note that \mathbf{R} has this property. Let k be the intersection of all fields with this property. Then k is the *smallest* field with this property. Note that $\mathbf{Q} \subset k$.

Lemma 2.8. *The field $K \cap \mathbf{R}$ has property (*). In particular $k \subset K \cap \mathbf{R}$.*

Proof: Let $a \in K \cap \mathbf{R}$ with $a > 0$. Then the point $\frac{1}{2}(a - 1)$ in between a and -1 is constructible, and so the circle centered at $\frac{1}{2}(a - 1)$ with radius $\frac{1}{2}(a + 1)$ is constructible. This circle intersects the line \mathbf{R} at a and -1 . The circle also intersects the line $i\mathbf{R}$ at some point ib with $b > 0$. The triangle with vertices $0, a, ib$, and the triangle with vertices $0, ib, -1$, are two similar right triangles. Comparing their sides, we find that

$$b/a = 1/b.$$

It follows that $ib = i\sqrt{a}$, hence also b , is constructible, i.e. $\sqrt{a} \in K \cap \mathbf{R}$. This shows that the subfield $K \cap \mathbf{R}$ of \mathbf{R} has property (*). \square

Lemma 2.9. *Let $S \subset \mathbf{C}$ be any nonempty finite set, and F be the subfield of \mathbf{R} , generated over \mathbf{Q} by all the real and imaginary parts of the elements of S . Let $u + iv \in \varepsilon(S)$ with $u, v \in \mathbf{R}$. Then u and v each is a root of a quadratic polynomial over F .*

Proof: Either $u + iv \in S$, in which case $u, v \in F$ and the assertion is true, or $u + iv$ is an intersection point of two distinct S -constructible curves. It remains to consider this second case.

Let C be an S -constructible curve. If C is the line, then it passes through two distinct points $\alpha, \beta \in S$, and so C can be defined by the linear equation

$$(1) \operatorname{Im}[(\alpha - \beta)(\bar{z} - \bar{\beta})] = 0$$

in two real variables x, y with $z = x + iy$. Expressing this equation in the real form $ax + by + c = 0$, and noting that $\alpha, \beta \in S$, we see that the coefficients a, b, c lie in F . If C is a circle, then it is centered at some $\alpha \in S$, of radius $|\beta|$ with $\beta \in S$, and so C can be defined by the quadratic equation

$$(2) |z - \alpha|^2 = |\beta|^2.$$

Again, expressing this in real form $(x - a)^2 + (y - b)^2 = c^2$, we see that $a, b, c \in F$ as well. Let C, C' be distinct S -constructible curves, and $u + iv \in C \cap C'$. Then each of C, C' can be defined by an equation of the form (1) or (2), with all coefficients in F , and $(u, v) \equiv u + iv$ is a solution to two such equations.

We consider two possibilities. Suppose one curve, say C , is a line. Then its equation is linear in x, y . Using it to eliminate one variable (say y), the remaining equation becomes $f(x) = 0$, for some $f \in F[x]$ of degree 1 or 2. Since $u + iv \in C \cap C'$, we have $f(u) = 0$. Likewise, v also satisfies a quadratic equation. Finally, suppose both C, C' are circles, and consider the real form of their equations:

$$(x - a)^2 + (y - b)^2 = c^2, \quad (x - a')^2 + (y - b')^2 = c'^2$$

where $(a, b) \neq (a', b')$ (i.e. different centers.) Subtract one from another, we get

$$2(a - a')x + 2(b - b')y = c'^2 - a'^2 - b'^2 - c^2 + a^2 + b^2.$$

which is a linear equation with coefficients in F . Using it to eliminate one variable, again, we see that u, v each satisfies a quadratic equation with coefficients in F . \square

Theorem 2.10. $K = k(i)$.

Proof: By the first lemma, $k \subset K \cap \mathbf{R} \subset K$. Since $i \in K$, it follows that $k(i) \subset K$.

To show $K \subset k(i)$, we will show $\varepsilon^n(\{0, 1\}) \subset k(i)$ by induction. We saw in an example above that the six points in $\varepsilon(\{0, 1\})$ are all in $k(i)$, so the assertion holds for $n = 1$. Assume it holds for some n , i.e. $S := \varepsilon^n(\{0, 1\}) \subset k(i)$. Let $\gamma \in \varepsilon^{n+1}(\{0, 1\}) = \varepsilon(S)$. Then either $\gamma \in S$, in which case $\gamma \in k(i)$ by the inductive assumption, or γ is an intersection point of two distinct S -constructible curves. Consider the second case, and let F be the subfield of \mathbf{R} , generated over \mathbf{Q} by all the real and imaginary parts of the elements of S , as in the second lemma. Let u, v be the real and imaginary parts of γ . We have $F \subset k$, by the inductive assumption. By the second lemma, u, v each lies in either F or a quadratic extension of F . In either case, $u, v \in k$, since k has property (*). It follows that $\gamma = u + iv \in k(i)$. \square

Corollary 2.11. *The elements of k can be listed in such a way, $\alpha_1, \alpha_2, \alpha_3, \dots$, that*

$$[\mathbf{Q}(\alpha_1, \dots, \alpha_j) : \mathbf{Q}(\alpha_1, \dots, \alpha_{j-1})] \leq 2, \quad \forall j \geq 1.$$

Proof: For $\alpha \in k$, we define $length(\alpha)$ to be the smallest integer $n \geq 0$, such that α is the real or imaginary part of some element in $\varepsilon^n(\{0, 1\})$. Since $\varepsilon^n(\{0, 1\})$ is finite for each n , the number of elements $\alpha \in k$ having length up to a given m , is finite. So, we can list all elements of k of length 1 first, followed by those of length 2, and so on. Denote the resulting list by $\alpha_1, \alpha_2, \alpha_3, \dots$. We now verify the inequality.

The length 0 elements of k are 0, 1, and α_1, α_2 are them. So the inequality holds for $j = 1, 2$. Suppose the inequality holds up to some j , and put $n = length(\alpha_{j+1}) > 0$. So, α_{j+1} is the real or the imaginary part of some $\gamma \in \varepsilon^n(\{0, 1\})$. By the second lemma above, α_{j+1} is a root of a polynomial $f \in F[x]$ of degree 1 or 2, where F is the field generated over \mathbf{Q} by the real and imaginary parts of the elements of $S = \varepsilon^{n-1}(\{0, 1\})$, say $\alpha_1, \dots, \alpha_p$, for some $p \leq j$. But we can also view $f \in \mathbf{Q}(\alpha_1, \dots, \alpha_j)[x]$, so that α_{j+1} still has degree 1 or 2 over the field $\mathbf{Q}(\alpha_1, \dots, \alpha_j)$. That is to say,

$$[\mathbf{Q}(\alpha_1, \dots, \alpha_{j+1}) : \mathbf{Q}(\alpha_1, \dots, \alpha_j)] \leq 2.$$

This completes the proof. \square

Corollary 2.12. $[\mathbf{Q}(\gamma) : \mathbf{Q}]$ is a power of 2, for any $\gamma \in k$.

Proof: Let $\gamma \in k$, say $\gamma = \alpha_p$ in the list above. By the Degree Factorization Theorem,

$$[\mathbf{Q}(\alpha_1, \dots, \alpha_p) : \mathbf{Q}] = \prod_{j=1}^p [\mathbf{Q}(\alpha_1, \dots, \alpha_j) : \mathbf{Q}(\alpha_1, \dots, \alpha_{j-1})].$$

By the preceding corollary, the right side is a power of 2. Since $\mathbf{Q}(\alpha_1, \dots, \alpha_p) \geq \mathbf{Q}(\gamma) \geq \mathbf{Q}$, it follows that $[\mathbf{Q}(\gamma) : \mathbf{Q}]$ divides that power of 2, by the Degree Factorization Theorem.

\square

Exercise. Let $\gamma \in \mathbf{R}$. Prove that $\gamma \in k$ iff there exist finitely many real numbers $\alpha_1, \dots, \alpha_n = \gamma$, such that for $j = 1, \dots, n$,

$$[\mathbf{Q}(\alpha_1, \dots, \alpha_j) : \mathbf{Q}(\alpha_1, \dots, \alpha_{j-1})] = 2.$$

(Hint: Use a corollary above in one direction, and use property (*) of k to show that $\alpha_1, \dots, \alpha_n \in k$ in the opposite direction.)

2.2. Impossibility of Some Geometric Constructions

Theorem 2.13. *Doubling the cube is impossible. In other words, given a cube, it is not always possible to construct, using a straightedge and a compass, the side of a cube whose volume is double the volume of the original cube.*

Proof: Take a cube of side length 1. Its volume is 1. A cube with double this volume has side length $\sqrt[3]{2}$. Since its irreducible polynomial over \mathbf{Q} is $x^3 - 2$, we have $[\mathbf{Q}(\sqrt[3]{2}) : \mathbf{Q}] = 3$, which is not a power of 2. So, $\sqrt[3]{2}$ is not constructible. \square

Theorem 2.14. *Squaring the circle is impossible. In other words, given a circle, it is not always possible to construct, using a straightedge and a compass, a square whose area is equal to the area enclosed by the circle.*

Proof: Take a unit circle. Its enclosed area is π . A square with area π has side length $\sqrt{\pi}$. But this number is transcendental over \mathbf{Q} . So, $\sqrt{\pi}$ is not constructible. \square

Theorem 2.15. *Trisecting the angle is impossible. In other words, given an angle, it is not always possible to construct, using a straightedge and a compass, an angle which is equal to $1/3$ the original angle.*

Proof: We can construct an angle θ iff we can construct two lines intersecting at that angle. We can do so iff we can construct a right triangle with hypotenuse 1 and base $\cos \theta$. From trigonometric identities,

$$\cos 3\theta = \cos(2\theta + \theta) = (2\cos^2\theta - 1)\cos \theta - (2\sin \theta \cos \theta)\sin \theta = 4\cos^3\theta - 3\cos \theta.$$

Take the angle $\theta = 20^\circ$, and put $\alpha = \cos 20^\circ$. Then the identity gives

$$4\alpha^3 - 3\alpha^2 = \frac{1}{2}.$$

So, α is a root of $f = 8x^3 - 6x - 1$. This polynomial is irreducible in $\mathbf{Q}[x]$, because it cannot be factor in $\mathbf{Z}[x]$. For factoring f in $\mathbf{Z}[x]$ would imply that f has a linear factor of the form

$$8x \pm 1, 4x \pm 1, 2x \pm 1, \text{ or } x \pm 1$$

which in turn would imply that f has a root $\pm\frac{1}{8}, \pm\frac{1}{4}, \pm\frac{1}{2}$, or ± 1 . It is easily checked that this is not the case, hence f must be irreducible in $\mathbf{Q}[x]$. So, $[\mathbf{Q}(\alpha) : \mathbf{Q}] = 3$. This implies that α is not constructible. So, we cannot construct a right triangle with hypotenuse 1 and base α , which means that we cannot construct the angle 20° . So, we cannot trisect 60° . \square