MASTER'S THESIS

GROTHENDIECK'S GALOIS THEORY

TSANTILAS N. THEOPHILOS

HELLENIC REPUBLIC
National and Kapodistrian University of Athens
EST. 1837

Department of Mathematics

2022

A thesis submitted in partial fulfillment of the requirements for the degree of

M.Sc. in Theoretical Mathematics

at

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ATHENS, GREECE

THESIS COMMITTEE

I. Emmanouil (Advisor)
A. Kontogeorgis
M. Sykiotis

Αφιερωμένο στη μνήμη του πατέρα μου.
Στις καλύτερές μας αναμνήσεις.

ABSTRACT

There are many Galois Theories throughout Mathematics. The purpose of this dissertation is to intoduce the part of *Grothendieck's Galois Theory for Schemes* that is related to Artin's Galois Theory of finite field extensions; that is, *Galois Theory of Finite Étale Algebras*.

From early on, mathematicians had noticed the similarities the Fundamental Theorem of Galois Theory and the Classification Theorem of Covering Spaces share. Eventually, it was no other than Grothendieck who understood their bond on a deeper level and formulated a theory, Grothendieck's Galois Theory for Schemes, in which he succeeded to unify these two classification theorems. Grothendieck's theorem in its full generality classifies finite étale coverings of a connected scheme using sets on which its étale fundamental group acts continuously. When seen from a field theoretic point of view, the theorem classifies finite étale algebras of a field using sets on which its absolute Galois group acts continuously. This is the theorem we are pursuing.

The first chapter is dedicated to Artin's Galois Theory for finite Galois extensions as well as Krull's Galois Theory for arbitrary (not necessarily finite) Galois extensions. The understanding of Galois Theory for infinite extensions is crucial for the understanding of Grothendieck's approach. It reveals that Galois groups have a natural topology which makes them topological groups and this topology plays an important role when the extension is infinite.

The second chapter is dedicated to Galois Theory for Coverings. We study the important notions from Algebraic Topology that will allow us to prove the strikingly similar, in a manner of speaking, Classification Theorem for Covering Spaces. This similarity provides motivation for Grothendieck's Galois Theory for Schemes.

Finally, after realizing the similarities of the above classification theorems, we introduce Galois Theory of Étale Algebras in the third chapter as a generalization of Artin's Galois Theory for finite extensions. We also briefly treat Grothendieck's Galois Theory for

Schemes in its full generality and its relation to Galois Theory of étale algebras as a possible continouation of this thesis.

ΠΕΡΙΛΗΨΗ

Στα Μαθηματικά υπάρχουν πολλές Θεωρίες Γκαλουά. Σκοπός αυτής της εργασίας είναι να μελετήσει εκείνο το κομμάτι της Θεωρίας Γκαλουά του Γκρόθεντικ για Σχήματα το οποίο γενικεύει την κλασική Θεωρία Γκαλουά του Άρτιν για πεπερασμένες επεκτάσεις σωμάτων· αυτή είναι η Θεωρία Γκαλουά για Ετάλ Άλγεβρες.

Από νωρίς οι μαθηματικοί αντιλαμβανόντουσαν τις ομοιότητες του Θεμελιώδους Θεωρήματος της Θεωρίας Γκαλουά και του Θεωρήματος Ταξινόμησης Χώρων Επικάλυψης της Αλγεβρικής Τοπολογίας. Εν τέλει, ήταν ο Γκρόθεντικ εκείνος που κατανόησε το δεσμό που μοιράζονται σε βαθύτερο επίπεδο και διατύπωσε τη λεγόμενη Θεωρία Γκαλουά του Γκρόθεντικ για Σχήματα, ένα θεώρημα ταξινόμησης που ενοποιεί τα δύο προηγούμενα. Η Θεωρία Γκαλουά για Σχήματα, στη γενική της μορφή, ταξινομεί πεπερασμένες ετάλ επικαλύψεις συνεκτικών σχημάτων μέσω πεπερασμένων συνόλων στα οποία η ετάλ θεμελιώδης ομάδα του σχήματος δρα συνεχώς. Όταν προσεγγίσουμε το θεώρημα από τη σκοπιά των επεκτάσεων σωμάτων, το θεώρημα αυτό ταξινομεί ετάλ άλγεβρες ενός σώματος μέσω πεπεραμένων συνόλων στα οποία η απόλυτη ομάδα Γκαλουά δρα συνεχώς. Αυτό είναι το θεώρημα που θέλουμε να μελετήσουμε.

Το πρώτο κεφάλαιο ασχολείται με τη Θεωριά Γκαλουά του Άρτιν για πεπερασμένες επεκτάσεις Γκαλουά καθώς και με τη Θεωρία Γκαλουά του Κρουλ για τυχαίες (όχι απαραίτητα πεπερασμένες) επεκτάσεις Γκαλουά. Η κατανόηση της άπειρης Θεωρίας Γκαλουά είναι σημαντική όχι μόνο γιατί χρησιμοποιείται σε επόμενα απότελέσματα, αλλά και γιατί αναδεικνύει κάποια λεπτά μεν σημαντικά δε σημεία της Θεωρίας Γκαλουά που δεν έρχονται στην επιφάνεια στην πεπερασμένη περίπτωση. Φανερώνει ότι οι ομάδες Γκαλουά έρχονται φυσικά εφοδιασμένες με μία τοπολογία η οποία τις κάνει τοπολογικές ομάδες και αυτή τους η δομή είναι το κλειδί στην περίπτωση που η επέκταση είναι άπειρη.

Το δεύτερο κεφάλαιο ασχολείται με τη Θεωρία Γκαλουά για Χώρους Επικάλυψης. Μελετώνται εκείνες οι έννοιες της Αλγεβρικής Τοπολο-

γίας που μας επιτρέπουν να αποδείξουμε το Θεώρημα Ταξινόμησης των Χώρων Επικάλυψης, το οποίο είναι εξαιρετικά όμοιο κατά κάποιο τρόπο με αυτό της Θεωρίας Γκαλουά. Η ομοιότητα των δύο θεωρημάτων μας προοικονομεί την ύπαρξη μιας ενοποιημένης θεωρίας.

Τέλος, στο τρίτο κεφάλαιο διατυπώνεται και αποδεικνύεται το Θεώρημα Γκαλουά για ετάλ άλγεβρες και μελετάται η σχέση του με τη κλασική θεωρία Γκαλουά. Ως έναυσμα για περαιτέρω μελέτη, σκιαγραφούμε τη Θεωριά Γκαλουά για Σχήματα στη γενική της μορφή και στη σχέση της με τη Θεωρία Γκαλουά για Ετάλ Άλγεβρες.

AKNOWLEDGEMENTS

Θα ήθελα να χρησιμοποιήσω λίγο από το χώρο αυτής της εργασίας για να αποδώσω λοιπόν τα του Καίσαρος τω Καίσαρι και αφού κανείς δεν είναι εδώ για να με σταματήσει...

Θα ήθελα ειλικρινά να ευχαριστήσω τον επιβλέποντα καθηγητή κ. Εμμανουήλ για τη βοήθειά του, τη συνεχή του διαθεσημότητά του και πάνω απ όλα για την κατανόησή του κατά τη διάρκεια συγγραφής της διπλωματικής μου. Τον καθηγητή κ. Κοντογεώργη που δέχτηκε να είναι στη τριμελή επιτροπή, για τη βοήθεια που μου πρόσφερε κατά τη συγγραφή της διπλωματικής μου αλλά κυρίως γιατί όποτε και να τον ενόχλησα για οποιοδήποτε θέμα, ήταν πάντα πρόθυμος να με βοηθήσει. Και φυσικά τον καθηγητή κ. Συκιώτη που δέχτηκε να είναι μέλος της εξεταστικής επιτροπής. Ένα ευχαριστώ στον καθηγητή κ. Βάρσο ο οποίος από τα προπτυχιακά μου χρόνια αφιέρωνε πάντα χρόνο για να με ακούσει και να με συμβουλέψει.

Ένα τεράστιο ευχαριστώ σε όσους μου στάθηκαν και με βοήθησαν, οικογένεια, φίλοι και καθηγητές. Δεν θα τολμήσω να τους κατονομάσω όλους, μη τυχόν παραλείψω κάποιον.

Ωφείλω να αναφέρω ότι η ιδέα για το θέμα της διπλωματικής γεννήθηκε από τις συζητήσεις που είχα με τον Γ. Υφαντή όταν γράφαμε μαζί μια εργασία πάνω στις κατηγορίες.

Για τη συγγραφή της διπλωματικής σε LaTeXχρησιμοποιήθηκε το πανέμορφο πρότυπο του πακέτου ClassicThesis των André Miede και Ivo Pletikosić.

CONTENTS

Part I

A STRIKING SIMILARITY: THE CLASSIFICATION THEOREMS

Galois Theory and Algebraic Topology may appear as two unrelated disciplines but, as we shall see, there are aspects of both theories which are closely related. The following chapters contain a quick exposition of Galois Theory and Algebraic Topology that focuses on two highly important, and remarkably similar classification theorems: the Fundamental Theorem of Galois Theory and the Classification Theorem for Covering Spaces from Algebraic Topology.

Go to the roots, of these calculations! Group the operations. Classify them according to their complexities rather than their appearances! This, I believe, is the mission of future mathematicians. This is the road on which I am embarking in this work.

Évariste Galois (1811-1832).

1 ARTIN'S GALOIS THEORY

We are going to devote this chapter to the study of Galois Theory for field extensions. Our aim is not to provide a systematic or complete exposition of the subject but merely to set up the machinery needed in order to reach the *Fundamental Theorem of Galois Theory* for both *finite* and *infinite* field extensions. For this we assume a basic understanding of Abstract Algebra and General Topology, material usually covered in corresponding courses. The first twelve chapters of [10] and the first four of [29] are far more than enough.

Let us first introduce what Galois Theory is about with a few words.

For many centuries, one of the central problems in Mathematics was finding the solutions of polynomial equations

$$f(x) = a_n x^n + \ldots + a_1 x + a_0 = 0 \quad (n \in \mathbb{N})^1. \tag{1.1}$$

Formulas that gave the solutions of (1.1) when the *degree* of f, denoted ∂f, is 1 or 2 were found as early as the times of ancient Babylonians (around 2000-1500 B.C.!), while analogous formulas for the cases $\partial f = 3$ and 4 were discovered by the 16^{th} century.

All of these formulas had a common characteristic. They involved the four basic operations $+, -, \times, \div$ along with the extraction of roots applied on the coefficients of the polynomial; for example the roots of the general polynomial equation of degree $\partial f = 2$,

$$ax^2 + bx + c = 0,$$

are given by the well known formulas

$$\frac{-b + \sqrt{b^2 - 4ac}}{2a} \quad \text{and} \quad \frac{-b - \sqrt{b^2 - 4ac}}{2a}.$$

[1] Although not historically accurate, for simplicity we may assume at this point that $a_i \in \mathbb{C}\ \forall i = 1, \ldots, n$.

In other words, it was realized that *all* polynomial equations of degree $\partial f \leqslant 4$ can be *solved by radicals*.

It was therefore natural to search for similar formulas when $\partial f \geqslant 5$ as well. As Mathematics advanced, mathematicians realized that such formulas most likely did not exist and many attempted proving it. It was Niels Henrik Abel (1802-1829) who eventually gave a satisfactory proof of this fact for the general equation of degree $\partial f = 5$. But are there any special quintic equations that *can* be solved by radicals? What about equations of higher degrees? Immediately, the focus turned to find ways of deciding *when* an equation can be solved by radicals. The final answer was given by Évariste Galois.

Galois' idea was to derive information about the polynomial by studying the *permutations* of its roots. Although many mathematicians before Galois had linked the polynomials' behavior to such permutations, it was Galois who recognized that these permutations form a *group* under composition and used this extra structure to answer questions about the polynomial. He showed how information about a polynomial f could be derived from the group G of permutations of its roots and, as an application, gave an answer to one of the biggest problems of his time: *Eq.* (1.1) *can solved by radicals if and only if the group G is solvable* (using modern terminology).

His theory was proven to be most fruitful and over the years evolved. It required the work of many mathematicians, an adequate development of Group Theory and Field Theory (which was done in the 19^{th} century) and a world ready for such an abstract theory in order for Galois Theory to be formulated and established as we know it.

The ideas of modern Galois Theory, whose father is considered to be Emil Artin (1898-1962), trace back to Heinrich Martin Weber (1842-1913) and Julius Wilhelm Richard Dedekind (1831-1916). But it was Artin who managed to formulate the theory in the language of fields and field extensions independently of its main application, i.e. the solvability of polynomial equations by radicals.

Thus modern Galois Theory surpasses the scope of the original theory. Instead of studying polynomials using permutations of their roots, the focus is turned on how to use *automorphisms* to derive information about *field extensions*. In this setting, the original question about the solvability of (1.1) is translated into a question

about a *specific field extension* (the splitting field of the polynomial over the field where its coefficients lie).

1.1 FIELD EXTENSIONS

FIELD EXTENSIONS AND THEIR DEGREES An extension of a field k is just a bigger field E containing k or, more generally, containing an *isomorphic copy* of k.

Definition 1.1.1. Let k be a field. A **field extension** of k is a field E together with a field monomorphism $i : k \to E$. We refer to the field k as the **base field** and to E as the **extending field**.

Following the customary identification of k with its isomorphic image $i(k)$, we may think of k as a *subfield* of E and the monomorphism as the restriction $\mathrm{id}_E|_k$. This is why we will usually denote a field extension as $k \leqslant E$ or E/k, suppressing the monomorphism i from the notation.

When we are given a mathematical structure, e.g. sets, groups, fields, algebras, topological spaces, smooth manifolds etc., we do not confine ourselves to the study of the structure alone, but we are also interested in *substructures* and *structure preserving maps*.

In the theory of field extensions, a **subextension** or **intermediate field** of an extension $k \leqslant E$ is a field L such that $k \leqslant L$ and $L \leqslant E$. We use the notation $k \leqslant L \leqslant E$ to denote a subextension L of $k \leqslant E$.

More often than not we encounter multiple fields, one extending the other, i.e. multiple subextensions. For simplicity, we shall use the shorter notation

$$k \leqslant L_1 \leqslant L_2 \leqslant \ldots \leqslant L_n \leqslant E$$

and refer to such extensions as **towers of fields**.

Example 1.1.2. $\mathbb{Q} \leqslant \mathbb{R}$, $\mathbb{R} \leqslant \mathbb{C}$ and $\mathbb{Q} \leqslant \mathbb{C}$ are all field extensions. $\mathbb{Q} \leqslant \mathbb{R} \leqslant \mathbb{C}$ is a tower of fields.

Example 1.1.3. All fields can be considered extensions over their prime fields, i.e. over $\mathbb{Q}$ if their characteristic is 0 or over $\mathbb{F}_q = \mathbb{Z}/q\mathbb{Z}$ if their characteristic is $q > 0$.

Example 1.1.4. Let k be a field and $x_1, \ldots, x_n$ be indeterminates. We can form the ring of polynomial functions in n indeterminates $k[x_1, \ldots, x_n]$ and its fraction field

$$k(x_1, \ldots, x_n) = \left\{ \frac{f(x_1, \ldots, x_n)}{g(x_1, \ldots, x_n)} : f, g \in k[x_1, \ldots, x_n],\ g(x_1, \ldots, x_n) \neq 0 \right\}.$$

Then $k \leqslant k(x_1, \ldots, x_n)$ is a field extension since k can be viewed as *the subfield of constant polynomials* through the monomorphism

$$i : k \to k(x_1, \ldots, x_n) : a \mapsto \frac{a + 0x_1 + \ldots + 0x_n}{1 + 0x_1 + \ldots + 0x_n} = a \ \forall a \in k.$$

Remark 1.1.5. A natural question arises at this early stage: *Why are field extensions so important?* On the one hand, as it turned out, great mathematical problems of the antiquity such as the solvability of polynomial equations and the impossibility of geometric constructions by ruler and compass can be solved by translating them into questions about field extensions and using the tools of field extensions to solve the latter. On the other, field extensions are important on their own in the Theory of Fields.

Along with fields, we are also interested in field homomorphisms. Suppose that

$$\phi : k \to E$$

is a field homomorphism. Since $\ker \phi \lhd k$ is an ideal,[2]

$$\ker \phi = \{0\} \text{ or } k.$$

Therefore either $\phi = 0$ or ϕ is a monomorphism, i.e. $k \leqslant E$ is a field extension. That is, field extensions are the (non zero) *morphisms* in the *category of fields*.

Given a field extension $k \leqslant E$, the extending field E has the additional (rather trivial) structure of an k-vector space with external multiplication given by the multiplication of E restricted to k, i.e.

$$k \times E \to E.$$

This vector space structure is fundamental to Field Theory; it is one of the main tools we use to study the extension $k \leqslant E$.

[2] The only ideals of a field k are 0 and k because any non zero ideal I contains a unit, hence $I = k$.

Definition 1.1.6. The **degree** $[E : k]$ of the field extension $k \leqslant E$ is the dimension of E as an k-vector space, i.e. $[E : k] = \dim_k E$. The extension is called **finite** if $[E : k] < \infty$ and **infinite** otherwise.

Example 1.1.7. Consider the tower of fields $\mathbb{Q} \leqslant \mathbb{R} \leqslant \mathbb{C}$. It is immediate that both $[\mathbb{R} : \mathbb{Q}]$ and $[\mathbb{C} : \mathbb{Q}]$ are infinite because $\mathbb{Q}$ is countable while $\mathbb{R}$ and $\mathbb{C}$ are not.[3]

Example 1.1.8. On the other hand, $[\mathbb{C} : \mathbb{R}] = 2$; every complex number is an $\mathbb{R}$-linear combination of the set $\{1, i\}$.

Example 1.1.9. The extension $k \leqslant k(x)$ has infinite degree. Every polynomial is a finite k-linear combination of the linearly independent set $\{1, x, x^2, x^3, \ldots\}$.

Example 1.1.10. Lets see how we can use the vector space structure of an extension, and more specifically its degree, to derive information about the extension itself. We will show that $[E : k] = 1$ if and only if $E = k$.

If $E = k$ then $\{1_k\}$ is an k-basis for E and therefore $[E : k] = 1$. On the other hand, if $[E : k] = 1$ and, say, $\{u\}$ is an k-basis for E then $1_E = ru$ for some $r \in k$ and so $u = r^{-1}1_E = r^{-1} \in k$. Now $u \in k$ implies at once that $E = k$.

Proposition 1.1.11. *Let $k \leqslant L \leqslant E$ be a tower of fields. Then*

$$[E : k] = [E : L][L : k]$$

Proof. Let $\mathcal{B} = \{a_i : i \in I\}$ be an k-basis for L and let $\mathcal{B}' = \{b_j : j \in J\}$ be a L-basis for E; the set

$$\mathcal{B}'' = \{a_i b_j : i \in I, j \in J\}$$

is an k-basis for E. We can see that $\mathcal{B}''$ spans E using the distributive law and that any k-linear relation of its elements implies an L-linear relation among the elements of $\mathcal{B}'$, which is absurd. ◇

Using induction on $n \in \mathbb{N}$ we can prove

[3] Using a famous argument of Georg Ferdinand Ludwig Philipp Cantor (1845-1918), a vector space with a countable basis over a countable field is necessarily countable.

Corollary 1.1.12. *If $k \leqslant L_1 \leqslant L_2 \leqslant \ldots \leqslant L_n \leqslant E$ is a tower of fields then*

$$[E : k] = [E : L_n] \ldots [L_2 : L_1] [L_1 : k].$$

Corollary 1.1.13. *If $k \leqslant L \leqslant E$ is a tower of fields, then $k \leqslant E$ is finite if and only if both $k \leqslant L$ and $L \leqslant E$ are finite.*

CONSTRUCTING FIELD EXTENSIONS I: POLYNOMIALS We turn to more examples of field extensions and, in particular, a way of constructing field extensions using polynomials.

An important class of field extensions arise when trying to solve polynomial equations. For example, $x^2 + 1 = 0$ cannot be solved in the field $\mathbb{R}$ of real numbers but has two solutions in the extension $\mathbb{C}$.

More generally, given a field k and some *irreducible* polynomial $p(x) \in k[x]$, one may wonder whether there is some extension of k that contains a root of $p(x)$.

Example 1.1.14. Let k be a field and $p(x) \in k[x]$ an irreducible polynomial of degree $\partial p \geqslant 2$ (so that not all roots of p are in k). We can *construct* a field extension E of k that contains a root of $p(x)$. The motivating idea is simple: consider the polynomial ring in which $p(x)$ lives and force $p(x)$ to be 0 by taking the quotient.

Take the polynomial ring $k[x]$, its prime[4] ideal $I = \langle p(x) \rangle$ and form the quotient ring

$$E = k[x]/\langle p(x) \rangle = k[x]/I.$$

E is a field that extends k. Since $k[x]$ is a P.I.D., the ideal I is maximal[5] and therefore E is a field.[6] To see that E extends k, restrict the natural projection map $\pi : k[x] \to k[x]/I$ to k. The restriction $\pi|_k$ is a ring homomorphism with $\pi|_k(1) = 1$ and $\ker \pi|_k = \{0\}$ hence a field monomorphism.

To formally verify *there is a root of* $p(x)$ (or of $\pi|_k(p(x))$ to be more precise) *in* E, take the element $\tilde{x} = x + I \in k[x]/I$ and verify that

$$p(\tilde{x}) = p(x) + I = I = 0 \in E.$$

[4] A principal ideal of $k[x]$ that is generated by an irreducible element is prime.
[5] Prime ideals are maximal in P.I.D.s.
[6] If I is an ideal of a commutative ring R, then R/I *is a field iff* I *is maximal.*

It would be of benefit to us if we could have an explicit description of the elements of E. Using the extra structure E has as an k-vector space, we obtain a useful characterization of its elements as follows. The set

$$\mathcal{B} = \{1, \tilde{x}, \tilde{x}^2, \ldots, \tilde{x}^{\partial p - 1}\} = \{1, x + I, x^2 + I, \ldots, x^{\partial p - 1} + I\}$$

is a basis of E over k.

$\mathcal{B}$ spans E. Indeed, $k[x]$ is an Euclidean[7] domain with Euclidean function the usual degree function $\partial : k[x] \to \mathbb{N} : \ f \mapsto \partial f$. Using Euclid's algorithm, for every $g \in k[x]$ there are unique $b, r \in k[x]$ such that

$$g = bp + r, \quad \partial r < \partial p.$$

Therefore, $g + I = (bp + r) + I \overset{p=0}{=} r + I \in \langle \mathcal{B} \rangle$ since $\partial r < \partial p$.

Moreover, *$\mathcal{B}$ is k-linearly independent*. Any k-linear relation of the form

$$a_{\partial p - 1}(x^{\partial p - 1} + I) + \ldots + a_1(x + I) + a_0(1 + I) = 0 + I \in E$$

among the elements of $\mathcal{B}$ yields a polynomial

$$w(x) = a_{\partial p - 1} x^{\partial p - 1} + \ldots + a_1 x + a_0 \in k[x]$$

of degree $< \partial p$ which is equal to 0 in $k[x]/I$ or, equivalently, a polynomial of degree $< \partial p$ which is divided by p which is absurd.

Therefore,

$$[E : k] = |\mathcal{B}| = \partial p \tag{1.2}$$

and

$$\begin{aligned} E = k[x]/I &= \{b_0 + b_1 \tilde{x} + b_2 \tilde{x}^2 + \ldots + b_{\partial p - 1} \tilde{x}^{\partial p - 1} : \ b_i \in k\} \\ &= \{b_0 + b_1 x + b_2 x^2 + \ldots + b_{\partial p - 1} x^{\partial p - 1} + I : \ b_i \in k\}. \end{aligned}$$

Example 1.1.15. Take the field $\mathbb{R}$ and the irreducible polynomial $p(x) = x^2 + 1 \in \mathbb{R}[x]$ of degree $\partial p = 2$.[8]

By the example above, $E = \mathbb{R}[x]/\langle x^2 + 1 \rangle$ is a field extension of $\mathbb{R}$ which contains a root of $p(x)$. Moreover, E is spanned by

$$\mathcal{B} = \{1, x + \langle x^2 + 1 \rangle\}$$

[7] Euclid of Alexandria ($\sim$ 300 B.C.).

[8] The polynomial $x^2 + 1$ is irreducible since it has degree 2 and no real roots.

and, as a result, $[E : \mathbb{R}] = 2$ and

$$E = \{a + b(x + I) : a, b \in \mathbb{R}\}.$$

If we decide to just change the notation and set $i := x + I \in E$ (observe that $x + I$ is the root of $p(x)$ in E) then

$$E = \{a + b(x + I) : a, b \in \mathbb{R}\} \simeq \{a + bi : a, b \in \mathbb{R}\} = \mathbb{C}.$$

The formal way to do this is by constructing the field isomorphism

$$\sigma : E \to \mathbb{C} : a + bx + I \mapsto a + bi.$$

Thus we have constructed $\mathbb{C}$ in a very elegant, algebraic way.

ISOMORPHIC EXTENSIONS We now make a short pause from our examples to define a crucial notion.

Even in our first examples we came across two very different yet isomorphic fields that both extend $\mathbb{R}$, i.e. $\mathbb{C}$ and $E = \mathbb{R}[x]/\langle x^2 + 1\rangle$. Therefore it is only natural to consider both extensions $\mathbb{R} \leqslant E$ and $\mathbb{R} \leqslant \mathbb{C}$ to be the same from a field-theoretic point of view. This notion of isomorphic field extensions is the one we want to define.

Definition 1.1.16. A **morphism** between two abstract field extensions $i_1 : k_1 \to E_1$ and $i_2 : k_2 \to E_2$ is a pair of maps $\sigma : E_1 \to E_2$ and $\tau : k_1 \to k_2$ such that the following diagram commutes, i.e. $\sigma \circ i_1 = i_1 \circ \tau$.

$$\begin{array}{ccc} E_1 & \xrightarrow{\sigma} & E_2 \\ {\scriptstyle i_1}\uparrow & & \uparrow{\scriptstyle i_2} \\ k_1 & \xrightarrow{\tau} & k_2 \end{array}$$

The field extensions are called **isomorphic** if the maps σ and τ are *field isomorphisms*.

What this definition says is that an isomorphism of field extensions must *preserve the structure of the extensions* in question, i.e. *the base fields*, *the extending fields* and *the way they are related* (the monomorphisms).

With the identification of k_j with its isomorphic image $i_j(k_j) \leqslant E_j$ (in which case $i_j = \mathrm{id}_{E_j}|_{k_j}$), $j = 1, 2$, the commutativity of the above diagram gives

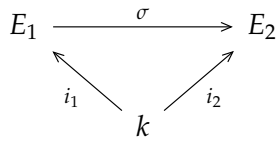
$$\sigma(i_1(x)) = i_2\big(\tau(x)\big) \iff \sigma(x) = \tau(x)\ \forall x \in k_1.$$

Hence we get the following *equivalent* definition.

Definition 1.1.17. Two field extensions $k_1 \leqslant E_1$ and $k_2 \leqslant E_2$ are **isomorphic** if there is a field isomorphism $\tau : k_1 \to k_2$ that can be *extended* to an isomorphism $\sigma : E_1 \to E_2$.

In many instances, such as Ex. 1.1.15, we will encounter the simpler case when the extensions are over the same base field.

Definition 1.1.18. Two abstract field extensions $i_1 : k \to E_1$ and $i_2 : k \to E_2$ *over the same field* are called **isomorphic** if there is a *field isomorphism* $\sigma : E_1 \to E_2$ so that the following diagram commutes, i.e. $\sigma \circ i_1 = i_2$.



Although we could simply use the first definition and take $\tau = \mathrm{id}_k$ when two extensions are over the same field, the importance of this special case dictates to formulate it separately.

In the case where we regard k as a subfield of both E_1 and E_2, the commutativity of the above diagram gives

$$\sigma(i_1(x)) = i_2(x) \iff \sigma(x) = x\ \forall x \in k.$$

That is, σ is a field isomorphism $E_1 \to E_2$ that *fixes k pointwise*. In this case we say that σ is an k-**isomorphism** from E_1 to E_2 (or in the case where $E_1 = E_2 = E$, an k-**automorphism** of E) and we can restate the previous definition as

Definition 1.1.19. Two field extensions $k \leqslant E_1$ and $k \leqslant E_2$ *over the same field* are called **isomorphic** if there is an *k-isomorphism* $\sigma : E_1 \to E_2$.

Lemma 1.1.20. *Field extension isomorphism is an equivalence relation.*

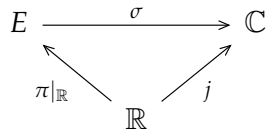
Example 1.1.21. The extensions $\mathbb{R} \leqslant \mathbb{C}$ and $\mathbb{R} \leqslant E = \mathbb{R}[x]/\langle x^2 + 1\rangle$ are isomorphic. If we define

$$\sigma : E \to \mathbb{C} : a + bx + I \mapsto a + bi$$

as before, and

$$j : \mathbb{R} \to \mathbb{C} : r \mapsto r + 0i,$$

to be the usual inclusion $\mathbb{R} \hookrightarrow \mathbb{C}$ then it is easily seen that the diagram



commutes. Indeed,

$$\sigma(\pi|_{\mathbb{R}}(r)) = \sigma(r + I) = \sigma(r + 0(x + I)) = r + 0i = j(r)\ \forall r \in \mathbb{R}.$$

Therefore, the extensions are isomorphic.

Field extension isomorphisms *over the same field* preserve all the information of a field extension including the structure of the extending fields as vector spaces over the base field.

Proposition 1.1.22. *If E_1, E_2 are fields extending a field k and $\sigma : E_1 \to E_2$ is an k-isomorphism then σ is a bijective k-linear transformation.*

Proof. The bijection is immediate; the linearity is clear from the corresponding commutative diagram. ◇

Corollary 1.1.23. *If $k \leqslant E_1$ and $k \leqslant E_2$ are isomorphic field extensions over the same field then $[E_1 : k] = [E_2 : k]$.*

Lastly, before we continue with our examples we introduce the concept of an automorphism of a field extension which will be needed later.

Definition 1.1.24. Let $k \leqslant E$ be a field extension. An **automorphism** of $k \leqslant E$ is an isomorphism from $k \leqslant E$ to itself, i.e. an k-automorphism of E. The set of all k-automorphisms of E is denoted by $\mathrm{Aut}(E/k)$.

Using elementary Group Theory, it is immediate that

Lemma 1.1.25. *If $k \leqslant E$ be a field extension then $\mathrm{Aut}(E/k)$ is a subgroup of $\mathrm{Aut}(E)$.*

CONSTRUCTING FIELD EXTENSIONS II: FINITELY GENERATED EXTENSIONS We continue with more examples and ways of constructing field extensions.

Given a fixed field extension $k \leqslant E$ and a subset X of E, the intersection of all subfields L of E that contain $X \cup k$ is non empty (E is such a field), and is a subfield of E.[9] It is the smallest extension of k with these properties and is denoted by $k(X)$;

$$k(X) = \bigcap_{X \cup k \subseteq L \leqslant E} L. \tag{1.3}$$

The elements of X are called the **generators** of the extension. A field extension $k \leqslant E$ is called **finitely generated** over k if there is some *finite* subset $X \subseteq E$ such that $E = k(X)$. If there exists a single element $a \in E$ such that $E = k(a)$, then the extension is called **simple**. Using (1.3) we can see that

$$k(a,b) = \bigcap_{\{a,b\} \cup k \subseteq L \leqslant E} L = \bigcap_{\{b\} \cup k(a) \subseteq L \leqslant E} L = \big[k(a)\big](b) \tag{1.4}$$

and by induction that $k(a_1, \ldots, a_n) = k(a_1, \ldots, a_{n-1})(a_n)$; this suggests that simple extensions act as building blocks for finitely generated extensions. Therefore, if we want to understand the structure of finitely generated extensions, we should first understand the structure of simple extensions.

Finitely generated extensions are the core of *finite* Galois Theory. As we shall see, every finite extension is finitely generated.

Example 1.1.26. We can take the extension $\mathbb{R} \leqslant \mathbb{C}$ and the element $i \in \mathbb{C}$. Then $\mathbb{R}(i)$ is a simple extension. It is the smallest field in $\mathbb{C}$ containing both $\mathbb{R}$ and i.

[9] The intersection of any family of subfields of E is again a subfield of E.

Example 1.1.27. To the same direction we can also take the extension $\mathbb{Q} \leqslant \mathbb{R}$ and the elements $\pi, \sqrt{2} \in \mathbb{R}$. Then $\mathbb{Q}(\sqrt{2}, \pi)$ is the smallest field $\subseteq \mathbb{R}$ that contains $\mathbb{Q}$ as well as π and $\sqrt{2}$.

Example 1.1.28. We will later see (Prop. 1.1.46) that every extension of the form $k \leqslant k[x]/\langle p(x)\rangle$ constructed as in Ex. 1.1.14 is isomorphic to a simple extension, generated by an element that satisfies some extra properties. So finitely generated extensions generalize the construction of Ex. 1.1.14.

So far we have no information about the structure of finitely generated extensions.

Proposition 1.1.29. *Let $k \leqslant E$ be a field extension and $a \in E$. Then*

$$k(a) = \left\{ \frac{f(a)}{g(a)} : f(x), g(x) \in k[x],\ g(a) \neq 0 \right\}.$$

Proof. Let L_0 be the right hand side of the above equality. L_0 is a subextension of $k \leqslant E$ such that $a \in L_0$. Therefore

$$k(a) = \bigcap_{\{a\} \cup k \subseteq L \leqslant E} L \subseteq L_0.$$

On the other hand, L_0 is obviously contained in every field L that contains both k and a. Hence

$$L_0 \subseteq \bigcap_{\{a\} \cup k \subseteq L \leqslant E} L = k(a)$$

The two inclusions imply the required equality. ◇

Corollary 1.1.30. *Let $k \leqslant E$ be a field extension and $a_1, \ldots, a_n \in E$. Then*

$$k(a_1, \ldots, a_n) = \left\{ \frac{f(a_1, \ldots, a_n)}{g(a_1, \ldots, a_n)} : f, g \in k[x_1, \ldots, x_n],\ g(a_1, \ldots, a_n) \neq 0 \right\}.$$

Example 1.1.31. Now we know that

$$\mathbb{R}(i) = \left\{ \frac{f(i)}{g(i)} : f(x), g(x) \in \mathbb{R}[x],\ g(i) \neq 0 \right\}.$$

and

$$\mathbb{Q}(\sqrt{2}, \pi) = \left\{ \frac{f(\sqrt{2}, \pi)}{g(\sqrt{2}, \pi)} : f, g \in \mathbb{Q}[x_1, x_2],\ g(\sqrt{2}, \pi) \neq 0 \right\}.$$

But these descriptions do not depict accurately the structure of the extensions. In particular, the form of the elements might not be as complex as described in the previous propositions as the generator may satisfy some polynomial equation. For example, the element

$$\frac{i^{15}+3i^6-2i+1}{i-1} \in \mathbb{R}(i)$$

can be simplified to

$$\begin{aligned}\frac{i^{15}+3i^6-2i+1}{i-1} &= (i^{15}+3i^6-2i+1)(i-1)^{-1} \\ &= (-3i-2)\left(-\frac{1}{2}i-\frac{1}{2}\right) = \frac{5}{2}i-\frac{1}{2} \in \mathbb{R}(i).\end{aligned}$$

After studying simple extensions in more depth, we will be able to get better descriptions of finitely generated extensions such as the above.

ALGEBRAIC EXTENSIONS I As we said, the understanding of finitely generated extensions requires a very good grasp of their building blocks, i.e. the simple extensions. In the general case, we can classify all simple extensions $k \leqslant k(a)$ up to isomorphism; their structure depends on whether the generator a is algebraic or not.

Definition 1.1.32. Let $k \leqslant E$ be a field extension and $a \in E$. The element $a \in E$ is said to be **algebraic over** k if there exists some $f(x) \in k[x]$ such that $f(a) = 0$. An extension whose elements are all algebraic is called an **algebraic extension**.

Example 1.1.33. The element $i \in \mathbb{C}$ is the root of $x^2+1 \in \mathbb{R}[x]$ hence algebraic over $\mathbb{R}$. In fact there is no point to commit ourselves to i. Any complex number $z = a+bi$ is the root of $x^2-2ax+(a^2+b^2) \in \mathbb{R}[x]$ hence algebraic over $\mathbb{R}$. So $\mathbb{R} \leqslant \mathbb{C}$ is an algebraic extension.

Definition 1.1.34. If the element $a \in E$ is not algebraic over k then it is said to be **transcendental over** k. An extension with transcendental elements is called **transcendental extension**.

Example 1.1.35. The extension $\mathbb{Q} \leqslant \mathbb{Q}(\sqrt{2}, \pi)$ is transcendental since π is transcendental over $\mathbb{Q}$.[10]

[10] For a proof of the transcendence of π see [30].

Example 1.1.36. An important example of a transcendental extension is $k \leqslant k(x)$, where x is an indeterminate. As we shall see, this is the only simple transcendental extension of k up to isomorphism.

CLASSIFYING SIMPLE EXTENSIONS Let $k \leqslant E = k(a)$ be a simple field extension ($a \in E$) and

$$e_a : k[x] \to k[a] : f(x) \mapsto f(a)$$

be the evaluation homomorphism. Note that e_a is clearly *onto*.

If $a \in E$ is transcendental, then there is no $g(x) \in k[x]$ such that $g(a) = 0$; in other words, $\ker e_a = \{0\}$. By the 1st Isomorphism Theorem we have a ring isomorphism

$$k[x] / \ker e_a = k[x] \simeq \operatorname{Im} e_a = k[a].$$

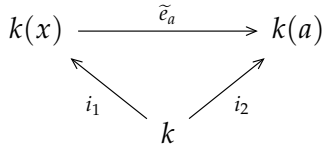
Since $k(x)$ and $k(a)$ are the fields of quotients (which are unique up to isomorphism) of $k[x]$ and $k[a]$ respectively, we get a *field isomorphism*

$$\widetilde{e}_a : k(x) \xrightarrow{\simeq} k(a) : \frac{f(x)}{g(x)} \mapsto \frac{f(a)}{g(a)}.$$

This field isomorphism together with the two inclusions

$$i_1 : k \to k(x) : z \mapsto \frac{f(x)}{g(x)} = \frac{z + 0x + 0x^2 + \dots}{1 + 0x + 0x^2 + \dots} = \frac{z}{1} \in k(x)$$
$$i_2 : k \to k(a) : z \mapsto \frac{f(a)}{g(a)} = \frac{z + 0a + 0a^2 + \dots}{1 + 0a + 0a^2 + \dots} = \frac{z}{1} \in k(a)$$

make the corresponding diagram commute.



Thus we have proven the following.

Proposition 1.1.37. *Let k be a field. $k \leqslant k(x)$, where x is an indeterminate, is the only simple transcendental extension of k up to isomorphism.*

If, on the other hand, $a \in E$ is algebraic over k then

$$\{0\} \neq \ker e_a \lhd k[x].$$

Hence $\ker e_a$ is generated by some non-zero element.[11]

Among all polynomials f that have a as a zero (i.e. $f \in \ker e_a$), there exist some with minimum degrees (by the well-ordering principle). Of all these we can choose a *monic* one.[12] This monic polynomial, denoted by $m(a,k)(x)$, such that $m(a,k)(a) = 0$ and whose degree $\partial m(a,k)$ is the smallest among all the polynomials that have a as root, is *unique*; if not, any other such polynomial $w(x)$ would result in a non-zero polynomial $m(a,k) - w$ with a as a root and degree less that ∂m (because both m and w are monic) - a contradiction.

Definition 1.1.38. This polynomial $m(a,k)$ is called the **minimal polynomial** of a over k and it obviously depends not only on a but on k as well.

Proposition 1.1.39. *With the above notation, the minimal polynomial* $m = m(a,k) \in k[x]$

(i) is irreducible

(ii) divides every polynomial $g(x) \in k[x]$ *such that* $g(a) = 0$.

Proof. (i) If not, then $m(x) = f_1(x)f_2(x)$ with $\partial f_1, \partial f_2 < \partial m$. But then $m(a) = f_1(a)f_2(a) = 0$ hence either $f_1(a) = 0$ or $f_2(a) = 0$ which contradicts the minimality of ∂m.

(ii) By Euclid's Algorithm (and the minimality of ∂m), there exist unique $q, r \in k[x]$ such that

$$g(x) = q(x)m(x) + r(x),\ \partial r < \partial m.$$

If $r \neq 0$ then $g(a) = q(a)m(a) + r(a) \Rightarrow r(a) = 0$ which again contradicts the minimality of ∂m. Therefore $r = 0$ and $m|g$.

◇

Corollary 1.1.40. *With the above notation,* $\ker e_a = \langle m(a,k) \rangle$.

[11] $k[x]$ is a P.I.D.

[12] Take one with minimum degree and divide it by its leading coefficient.

Corollary 1.1.41. *Let $k \leqslant E$ be a field extension, $a \in E$ and $m(x) \in k[x]$ a monic polynomial such that $m(a) = 0$. Then $m = m(a,k)$ if and only if m is irreducible.*

Corollary 1.1.42. *Let $k \leqslant L \leqslant E$ be a tower of fields and $a \in E$. Then $m(a,L)$ divides $m(a,k)$. In particular $\partial m(a,L) \leqslant \partial m(a,k)$.*

Example 1.1.43. $m(i,\mathbb{R}) = m(i,\mathbb{Q}) = x^2 + 1$.

Example 1.1.44. $m(\sqrt{2},\mathbb{Q}) = x^2 - 2 = m\big(\sqrt{2},\mathbb{Q}(\pi)\big)$.

Returning to the extension $k \leqslant k(a)$. The evaluation homomorphism e_a is a well defined ring epimorphism with kernel

$$\ker e'_a = \langle m(a,k)\rangle \equiv \langle m\rangle.$$

The 1st Isomorphism Theorem for Rings now gives a ring isomorphism

$$\widetilde{e}'_a : k[x]/\langle m\rangle \simeq k[a] : f(x) + \langle m\rangle \mapsto f(a). \tag{1.5}$$

However m is irreducible, so $\langle m\rangle$ is a maximal ideal of $k[x]$, hence $k[x]/\langle m\rangle$ is actually a field. That means that $k[a]$ is also a field, i.e. $k[a] = k(a)$, and therefore $k(a) \simeq k[x]/\langle m(a,k)\rangle$.

By the previous discussion, if $k \leqslant k(a)$ is a simple extension with a algebraic over k then $k(a)$ is isomorphic to an extension of the form $k[x]/\langle p(x)\rangle$ where $p(x)$ is a monic irreducible polynomial in $k[x]$, namely $p = m(a,k)$.

We can show that the converse also holds. Given a field k and an irreducible, monic polynomial $p(x) \in k[x]$, the quotient $E = k[x]/\langle p(x)\rangle$ is a field extension of k that contains a root of p, namely the element $\widetilde{x} = x + I \in E$ (Ex. 1.1.14). Since p is monic and irreducible, $m(\widetilde{x},k) = p$ by Cor. 1.1.41. Therefore

$$k[x]/\langle p(x)\rangle = k[x]/\langle m(\widetilde{x},k)\rangle \simeq k(\widetilde{x}).$$

The restriction that p be monic is actually superfluous. For if p is an irreducible polynomial in $k[x]$ with leading coefficient a then $q(x) = a^{-1}p(x) \in k[x]$ is *irreducible* and *monic* and

$$\langle q(x)\rangle = \langle p(x)\rangle.$$

Therefore

$$k[x]/\langle q(x)\rangle = k[x]/\langle p(x)\rangle.$$

Thus we have proven

Proposition 1.1.45. *Let k be a field. The simple field extensions of k which are generated by algebraic elements are exactly the fields $k[x]/\langle p(x)\rangle$ for $p(x) \in k[x]$ irreducible polynomials.*

Using now the data we have from Ex. 1.1.14 for the structure of extensions of the form $k[x]/\langle p(x)\rangle$, we can get a better description of the elements of simple extensions with algebraic generators than those provided by Prop. 1.1.29 and Cor. 1.1.30.

Corollary 1.1.46. *Let $k \leqslant E$ be a field extension and $a \in E$ an algebraic element over k. Then $[k(a) : k] = \partial m(a,k) = \partial m$ and*

$$k(a) = \{c_0 + c_1 a + c_2 a^2 + \ldots + c_{\partial m - 1} a^{\partial m - 1} : c_i \in k\}.$$

Proof. We already saw that $k(a) = k[a]$. Since $m(a) = 0$, we can replace every power of a in some $f(a) \in k[a]$ which is greater than ∂m by powers $\leqslant \partial m$.

Moreover, the set $\mathcal{B} = \{1, a, \ldots, a^{\partial m - 1}\}$ is k-linearly independent. Otherwise, any k-linear relation among the elements of $\mathcal{B}$ yields a polynomial $g \in k[x]$ such that $g(a) = 0$ and $\partial g < \partial m$ which is absurd. ◇

Example 1.1.47. Consider the extension $\mathbb{R} \leqslant \mathbb{R}(i)$. The element $i \in \mathbb{C}$ is algebraic over $\mathbb{R}$ with minimal polynomial $m(i, \mathbb{R}) = x^2 + 1$ of degree $\partial = 2$. Therefore $[\mathbb{R}(i) : \mathbb{R}] = 2$ and

$$\mathbb{R}(i) \simeq \{a + bi : \ a, b \in \mathbb{R}\} = \mathbb{C}.$$

Example 1.1.48. Now to the extension $\mathbb{Q} \leqslant \mathbb{Q}(\pi, \sqrt{2})$. Although

$$[\mathbb{Q}(\pi, \sqrt{2}) : \mathbb{Q}] = [\mathbb{Q}(\pi, \sqrt{2}) : \mathbb{Q}(\pi)]\,[\mathbb{Q}(\pi) : \mathbb{Q}] = \infty,$$

using the structure of simple extensions

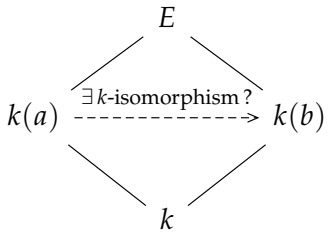
$$\mathbb{Q}(\pi, \sqrt{2}) = \big[\mathbb{Q}(\pi)\big](\sqrt{2}) \ \text{ and } \ \mathbb{Q}(\pi) = \{f(\pi) : f \in \mathbb{Q}(x)\}$$

(x being an indeterminate), we can again have a better description of the elements of the extension. Indeed, the element $\sqrt{2}$ is algebraic over $\mathbb{Q}(\pi)$ with minimal polynomial $m\big(\sqrt{2}, \mathbb{Q}(\pi)\big) = x^2 - 2$. Hence $[\mathbb{Q}(\pi, \sqrt{2}) : \mathbb{Q}(\pi)] = 2$ and

$$\mathbb{Q}(\pi, \sqrt{2}) = \big[\mathbb{Q}(\pi)\big](\sqrt{2}) = \{a + b\sqrt{2} : a, b \in \mathbb{Q}(\pi)\}.$$

INTRODUCING AN EXTENSION THEOREM Before continuing with the study of algebraic extensions, we take the opportunity the classification of simple extensions gives us to start discussing *extension theorems*. Our aim is to classify simple extensions *up to isomorphism*.

Suppose we begin with a field extension $k \leqslant E$ and two elements $a, b \in E$. We can then form the simple extensions $k(a)$ and $k(b)$. By definition, these are isomorphic if there is an k-isomorphism between them.



If one element is algebraic and the other is transcendental then the extensions cannot be isomorphic (the one is an infinite dimensional vector space over k while the other has finite dimension; see Cor. 1.1.23).

If both elements are transcendental then the extensions are isomorphic by Prop. 1.1.37.

So the interesting case is when both a and b are algebraic over k.

Example 1.1.49. The extensions $\mathbb{R} \leqslant \mathbb{R}(i)$ and $\mathbb{R} \leqslant \mathbb{R}(-i)$ are isomorphic through complex conjugation map $z \mapsto \overline{z}$ (which is an $\mathbb{R}$-isomorphism).

Counterexample 1.1.50. The extensions $\mathbb{Q} \leqslant \mathbb{Q}(\sqrt{2})$ and $\mathbb{Q} \leqslant \mathbb{Q}(\sqrt[3]{2})$ are *not* isomorphic; they have different degrees.

By Prop. 1.1.45,

$$k(a) \simeq k[x]/\langle m(a,k)\rangle \text{ and } k(b) \simeq k[x]/\langle m(b,k)\rangle.$$

Therefore a sufficient condition would be $\langle m(a,k)\rangle = \langle m(b,k)\rangle$. Since both $m(a,k)$ and $m(b,k)$ are irreducible and monic, this condition is equivalent to $m(a,k) = m(b,k)$. Therefore if a, b are roots of the same irreducible polynomial then the extensions are isomorphic.

Proposition 1.1.51. *If $k \leqslant E$ is a field extension and $a, b \in E$ algebraic elements over k such that $m(a,k) = m(b,k)$ then the extensions $k \leqslant k(a)$ and $k \leqslant k(b)$ are isomorphic.*

Proof. The required k-isomorphism is the composition

$$k(a) \xrightarrow{\simeq} k[x]/\langle m(a,k)\rangle = k[x]/\langle m(b,k)\rangle \xrightarrow{\simeq} k(b).$$

The details are easy to fill. ◇

Let us examine the general case where we have two different extensions $k_1 \leqslant E_1$ and $k_2 \leqslant E_2$ over *isomorphic* base fields with, say, $\tau : k_1 \to k_2$ being the base field isomorphism.

Given $\alpha \in E_1$ and $\beta \in E_2$ we can form the extensions $k_1 \leqslant k_1(\alpha)$ and $k_2 \leqslant k_2(\beta)$. We would like to know if these extensions are isomorphic. Again the non-trivial case is when α and β are algebraic over their respective base fields. In this case, the extensions are isomorphic if we can *extend* the given isomorphism τ to an isomorphism σ between $k_1(\alpha)$ and $k_2(\beta)$.

$$\begin{array}{ccc} k_1(\alpha) & \overset{\exists\sigma\,?}{\dashrightarrow} & k_2(\beta) \\ \Big| & & \Big| \\ k_1 & \xrightarrow{\quad\tau\quad} & k_2 \end{array}$$

Before proceeding, recall that if $\tau : k_1 \to k_2$ is a field homomorphism then τ induces a ring homomorphism

$$\widetilde{\tau} : k_1[x] \to k_2[x] : \sum_{i=0}^{m} a_i x^i \mapsto \sum_{i=0}^{m} \tau(a_i) x^i$$

and if τ is bijective then so is $\widetilde{\tau}$. Since

$$k_1(\alpha) \simeq k_1[x]/\langle m(\alpha, k_1)\rangle, \quad k_2(\beta) = k_2[x]/\langle m(\beta, k_2)\rangle$$

and $\widetilde{\tau} : k_1[x] \to k_2[x]$ is an isomorphism, a sufficient condition for the extensions to be isomorphic would be $\widetilde{\tau}\big(m(\alpha, k_1)\big) = m(\beta, k_2)$.

Indeed, if this is the case, then we can define

$$\sigma : k_1(\alpha) \to k_2(\beta) : f(\alpha) = \sum_{i=0}^{m} a_i \alpha^i \mapsto \big(\widetilde{\tau}(f)\big)(\beta) = \sum_{i=1}^{m} \tau(a_i)\beta^i.$$

This is a field isomorphism that makes the diagram

$$\begin{array}{ccc} k_1(\alpha) & \xrightarrow{\sigma} & k_2(\beta) \\ {\scriptstyle i}\uparrow & & \uparrow{\scriptstyle j} \\ k_1 & \xrightarrow{\tau} & k_2 \end{array}$$

commute (here i and j are the canonical inclusions). Moreover $\sigma(\alpha) = \beta$ and $\sigma|_{k_1} = \tau$, that is, σ *extends* τ. Thus we have proven

Proposition 1.1.52 (Extension Theorem for simple extensions)**.** *Suppose $k_1 \leqslant k_1(\alpha)$ and $k_2 \leqslant k_2(\beta)$ are algebraic simple extensions and $\tau : k_1 \to k_2$ is a field isomorphism such that $\widetilde{\tau}\big(m(\alpha, k_1)\big) = m(\beta, k_2)$. Then there exists an isomorphism $\sigma : k_1(\alpha) \to k_2(\beta)$ that extends τ. In other words, $k_1 \leqslant k_1(\alpha)$ and $k_2 \leqslant k_2(\beta)$ are isomorphic.*

ALGEBRAIC EXTENSIONS II Algebraic extensions are the core of classical Galois Theory and from now on we focus solely on them; no matter how interesting it is, the study of Galois Theory for arbitrary extensions is far beyond the scopes of this dissertation.

Convention. Henceforth all extensions are assumed to be algebraic unless explicitly stated otherwise.

So we need to put some extra effort into understanding algebraic extensions better.

Proposition 1.1.53. *A field extension is finite if and only if it is algebraic and finitely generated over the base field.*

Proof. ($\Rightarrow$) For a given finite extension $k \leqslant E$ with $[E : k] = n < \infty$, and any $z \in E$, the set

$$\{1, z, z, z^2, \ldots, z^n\} \subseteq E$$

is k-linearly dependent since it contains $n + 1 > [E : k]$ elements. That means we can find $a_0, \ldots, a_n \in k$, not all zero, such that

$$a_0 1 + a_1 z + a_2 z^2 + \ldots + a_n z^n = 0.$$

Therefore z is the root of $a_0 + a_1 x + \ldots + a_n x^n \in k[x]$ hence algebraic over k.

Moreover, since $[E : k] = n < \infty$, there exists an k-basis $\mathcal{B} = \{b_1, \ldots, b_n\} \subseteq E$ of E and thus

$$E = \{f_1 b_1 + \ldots + f_n b_n : f_i \in k\}.$$

Now on the one hand $k \leqslant k(\mathcal{B}) \leqslant E$ by definition of $k(\mathcal{B})$; so $k(\mathcal{B}) \subseteq E$. On the other hand,

$$\{f_1 b_1 + \ldots + f_n b_n : f_i \in k\} \subseteq L$$

for every extension L of k that contains $\mathcal{B}$ because

$$\underbrace{\underbrace{f_1}_{\in k \subseteq L}\underbrace{b_1}_{\in L}}_{\in L} + \ldots + \underbrace{\underbrace{f_n}_{\in k \subseteq L}\underbrace{b_n}_{\in L}}_{\in L} \in L \; \forall f_1, \ldots, f_n \in k.$$

But E *is* an extension of k that contains $\mathcal{B}$. Therefore $E \subseteq k(\mathcal{B})$ and as a result $E = k(\mathcal{B})$, i.e. E is finitely generated.

($\Leftarrow$) If $k \leqslant k(a_1, \ldots, a_n)$ is algebraic then in particular all a_i are algebraic over k. Applying 1.1.12 to the tower of fields

$$k \leqslant k(a_1) \leqslant k(a_1, a_2) \leqslant \ldots \leqslant k(a_1, \ldots, a_n),$$

we get

$$\begin{aligned}
[k(a_1, \ldots, a_n) : k] &= [k(a_1, \ldots, a_n) : k(a_1, \ldots, a_{n-1})] \ldots [k(a_1) : k] \\
&\overset{1.4}{=} [k(a_1, \ldots, a_{n-1})(a_n) : k(a_1, \ldots, a_{n-1})] \ldots [k(a_1) : k] \\
&\overset{1.1.46}{=} \partial m\big(a_n, k(a_1, \ldots, a_{n-1})\big) \ldots \partial m(a_1, k) \\
&\overset{1.1.42}{\leqslant} \partial m(a_n, k) \ldots \partial m(a_1, k) < \infty,
\end{aligned}$$

i.e. the extension is finite. ◇

Corollary 1.1.54. *Every finite field extension is algebraic.*

Corollary 1.1.55. *Suppose $k \leqslant E$ is a field extension and X a finite subset of E. Every $x \in X$ is algebraic if and only if the extension $k \leqslant k(X)$ is algebraic.*

Both hypotheses of 1.1.53 are essential for the converse. Example 1.1.35 gives a finitely generated but not finite, transcendental extension. Infinite algebraic extensions also exist.

Counterexample 1.1.56. The extension

$$\mathbb{Q} \leqslant \mathbb{A} = \{z \in \mathbb{C} : \ z \text{ is algebraic over } \mathbb{Q}\}$$

is a field extension which is by construction algebraic but not finite.[13]

Consider the n^{th} root of some number, say 2, and adjoin it to $\mathbb{Q}$. We get the tower of fields $\mathbb{Q} \leqslant \mathbb{Q}(\sqrt[n]{2}) \leqslant \mathbb{A}$. Since $\sqrt[n]{2}$ is algebraic over $\mathbb{Q}$ with minimal polynomial $m = x^n - 2$ (it is irreducible by Eisenstein's[14] criterion), the degree of the simple extension will be $[\mathbb{Q}(\sqrt[n]{2}) : \mathbb{Q}] = n$. Therefore, from 1.1.11, we get

$$[\mathbb{A} : \mathbb{Q}] = [\mathbb{A} : \mathbb{Q}(\sqrt[n]{2})][\mathbb{Q}(\sqrt[n]{2}) : \mathbb{Q}] \geqslant n.$$

Since $n \in \mathbb{N}$ was arbitrary, the degree of the extension is infinite.

Before closing this paragraph on algebraic extensions, we will see how they behave under subextensions.

Proposition 1.1.57 (Transitivity of algebraic extensions)**.** *Given a tower of fields $k \leqslant L \leqslant E$, the extension $k \leqslant E$ is algebraic if and only if both the extensions $k \leqslant L$ and $L \leqslant E$ are algebraic.*

Proof. ($\Rightarrow$) Immediate.

($\Leftarrow$) Let $a \in E$. Since $L \leqslant E$ is algebraic, we can find the minimal polynomial

$$m(a, L) = x^n + \ldots + a_1 x + a_0 \in L[x].$$

Consider the extension generated by the coefficients of $m(a, L)$ over k,

$$k \leqslant k(a_0, \ldots, a_{n-1}) \leqslant L.$$

Since $k \leqslant L$ is algebraic, so is $k \leqslant k(a_0, \ldots, a_{n-1})$ which is by construction finitely generated over k, hence, by 1.1.53, finite

$$[k(a_0, \ldots, a_{n-1}) : k] < \infty.$$

Moreover $m(a, k) \in k(a_0, \ldots, a_{n-1})[x]$, which means that a is algebraic over $k(a_0, \ldots, a_{n-1})$ and, again by 1.1.53,

$$[k(a_0, \ldots, a_{n-1}, a) : k(a_0, \ldots, a_{n-1})] < \infty.$$

[13] An accessible proof that $\mathbb{A}$ is indeed a field can be found in [15].
[14] Ferdinand Gotthold Max Eisenstein (1823–1852).

Therefore, by 1.1.11,

$$\begin{aligned}[k(a_0,\dots,a_{n-1},a):k] =& [k(a_0,\dots,a_{n-1},a):k(a_0,\dots,a_{n-1})]\\ &\cdot[k(a_0,\dots,a_{n-1}):k] < \infty.\end{aligned}$$

Thus the extension is finite and by 1.1.54, algebraic. In particular a is algebraic over k.

◇

CONSTRUCTING FIELD EXTENSIONS III: SPLITTING FIELDS For a given polynomial $f(x) \in k[x]$, we constructed in 1.1.14 a field extension that contains a zero of f. We can take one step further and construct a field that contains all roots of a given polynomial.

Definition 1.1.58. A polynomial $f(x) \in k[x]$ **splits over** k if all roots of f lie inside k. A **splitting field** for f is a minimal field over which f splits.

Example 1.1.59. The polynomial $x^2+1 \in \mathbb{Q}(x)$ does not split over $\mathbb{Q}$ or $\mathbb{R}$. It splits over $\mathbb{C}$ as well as over the smaller field $\mathbb{Q}(i)$.

Example 1.1.60. Suppose $k \leqslant E$ is a field extension and let $f(x) \in k[x]$ such that $\partial f = n$. If f splits over E and $a_1,\dots,a_n \in E$ are its roots, then $k(a_1,\dots,a_n)$ is by construction a splitting field of f.

Example 1.1.61. $\mathbb{Q}(i)$ is a splitting field of $x^2+1 \in \mathbb{Q}[x]$.

A famous theorem of Leopold Kronecker (1823-1891) states that any polynomial has a splitting field.

Proposition 1.1.62 (Kronecker)**.** *If $f(x) \in k[x]$ is a non-zero polynomial, then there exists a splitting field of f.*

Proof. Using induction on ∂f. The base case $\partial f = 1$ holds trivially. Assume that the theorem holds for all polynomials of degrees $\leqslant n$. Given a polynomial f with $\partial f = n+1$, the construction in 1.1.14 gives an extension E containing a root a of f. In E we can write $f(x) = (x-a)g(x)$, $\partial g \leqslant n$ and apply the induction hypothesis on $g(x)$. ◇

Example 1.1.63. Both $\mathbb{C} = \mathbb{R}(i)$ and $E = \mathbb{R}[x]/\langle x^2+1\rangle$ are splitting fields of $x^2+1 \in \mathbb{R}[x]$. To begin with, x^2+1 splits over both fields.

On the one hand, any splitting field $\mathbb{R} \leqslant L \leqslant \mathbb{R}(i)$ must contain both $\mathbb{R}$ and i. By the minimality of $\mathbb{R}(i)$, we deduce that $L = \mathbb{R}(i)$. On the other hand, for any splitting field $\mathbb{R} \leqslant L \leqslant E$, Prop. 1.1.11 gives

$$2 = [E : \mathbb{R}] = [E : L][L : \mathbb{R}].$$

But x^2+1 does not split over $\mathbb{R}$. Therefore, $[L : \mathbb{R}] > 1$ (so $[L : \mathbb{R}] = 2$) and $[E : L] = 1$ which means that $E = L$.

As we have already seen, the two fields are isomorphic. So in this case, these two splitting fields of $x^2+1 \in \mathbb{R}[x]$ are isomorphic.

Actually, any two splitting fields of a polynomial $f(x) \in k[x]$ are isomorphic. Inspired by the extension theorem 1.1.52, we will prove the result directly for the general case of two isomorphic base fields.

Proposition 1.1.64 (Extension Theorem for Splitting Fields)**.** *Let $\tau : k_1 \to k_2$ be a field isomorphism. If E_1 is the splitting field of some $f(x) \in k_1[x]$ and E_2 is the splitting field of $\widetilde{f} = \widetilde{\tau}(f) \in k_2[x]$ then there is a field isomorphism $\sigma : E_1 \to E_2$ that extends τ.*

Proof. With induction on ∂f.

If $\partial f = 1$ then k_1 is itself a splitting field of f and therefore, by the minimality of splitting fields, $k_1 = E_1$. Since $\widetilde{\tau}$ is an isomorphism, $\widetilde{f}$ also splits in k_2 and again $k_2 = E_2$. So the required extension of τ is itself.

Assume the theorem holds for all polynomials $f(x) \in k[x]$ of degree $\partial f < m$ for some $m \in \mathbb{N}$.

Let $f(x) \in k_1[x]$ be a polynomial of degree $\partial f = m$ and take $p(x)$ some *monic*, irreducible factor of f (it may be that $p = f$) of degree $\partial p \geqslant 2$; $p(x)$ has a root α in E_1 and $\widetilde{p} = \widetilde{\tau}(p)$ has a root β in E_2. By 1.1.52 there is an isomorphism σ_1 that extends τ to $k_1(\alpha)$.

$$\begin{array}{ccc} k_1(\alpha) & \xrightarrow{\sigma_1} & k_2(\beta) \\ \uparrow & & \uparrow \\ k_1 & \xrightarrow{\tau} & k_2 \end{array}$$

We now have that $f(x) = (x-\alpha)f_1(x) \in k(\alpha)[x]$ and $\widetilde{f}(x) = (x-\beta)\widetilde{f_1}(x) \in k(\beta)[x]$ where $\widetilde{f_1} = \widetilde{\tau}(f_1)$ and $\partial f_1 = \partial \widetilde{f_1} < m$. By the inductive hypothesis, E_1 is *the* splitting field of f_1 and E_2 is *the* splitting field of $\widetilde{f_1}$ (any other splitting field L of f_1 contains both α and the roots of f_1; this means that it is a splitting field of f inside E_1 and by minimality $L = E_1$; same for $\widetilde{f_1}$) and there is an isomorphism $\sigma : E_1 \to E_2$ extending σ_1.

$$\begin{array}{ccc} E_1 & \xrightarrow{\sigma} & E_2 \\ \uparrow & & \uparrow \\ k_1(\alpha) & \xrightarrow{\sigma_1} & k_2(\beta) \end{array}$$

Combining the two diagrams, we conclude that σ is an extension of τ. Therefore the extensions are isomorphic. ◇

Corollary 1.1.65. *Let k be a field, $f(x) \in k[x]$ some polynomial and E_1, E_2 two splitting fields of f. The extensions $k \leqslant E_1$ and $k \leqslant E_2$ are isomorphic. In particular, the splitting field of a polynomial $f(x)$ is unique up to isomorphism.*

ALGEBRAIC CLOSURES The results in the preceding paragraph can be generalized in the sense that we can construct the splitting field of any *finite* set of polynomials $\{f_i(x) \in F[x] : i = 1, 2, \ldots, n\}$ by carrying out the construction of Example 1.1.14 at most $\prod_{i=1}^{n} \partial f_i$ times.

Using *Zorn's Lemma*[15] we can take things one (huge) step further and consider an extension E of a field k that not only contains the roots of *every* polynomial $f(x) \in k[x]$, but also of *every polynomial* $g(x) \in E[x]$.

Lemma 1.1.66. *The following conditions are equivalent for a field k:*

i) There are no algebraic extensions $k \subsetneq E$.

ii) There are no finite extensions $k \subsetneq E$.

iii) Every $f(x) \in k[x]$ splits over k.

[15] Named after Max August Zorn (1906–1993).

iv) Every $f(x) \in k[x]$ has a root in k.

v) Every irreducible polynomial $p(x) \in k[x]$ has degree 1.

Proof. (i)⇒(ii): Immediate since every finite extension is algebraic.

(ii)⇒(iii): If some $f \in k[x]$ did not split, then we could construct an extension E of k that contains a root of f. By construction this would be a finite extension (of degree at most ∂f).

(iii)⇒(iv): Immediate.

(iv)⇒(v): If $p(x) \in k[x]$, then p has a root in k hence a linear factor $l(x) \in k[x]$. If p is irreducible then $p = l$ so $\partial p = 1$.

(v)⇒(i): Suppose we have an algebraic extension $k \leqslant E$ and let $a \in E$. By the hypothesis, $m(a,k)$ has degree 1 so $[k(a) : k] = \partial m(a,k) = 1$. Therefore $k(a) = k$ which means that $a \in k$ and thus $E = k$. ◇

Definition 1.1.67. A field that satisfies any of the above equivalent conditions is called **algebraically closed**.

Example 1.1.68. By the *Fundamental Theorem of Algebra*, $\mathbb{C}$ is algebraically closed.

Example 1.1.69. The field $\mathbb{A} = \{z \in \mathbb{C} : z \text{ algebraic over } \mathbb{Q}\}$ is algebraically closed. Indeed, suppose we have a finite (hence *finitely generated* and *algebraic*) extension $\mathbb{A} \leqslant \mathbb{A}(a_1, \ldots, a_n)$. The extension $\mathbb{Q} \leqslant \mathbb{A}$ is algebraic and, by Proposition 1.1.57, so is $\mathbb{Q} \leqslant \mathbb{A}(a_1, \ldots, a_n)$. Therefore $a_1, \ldots, a_n \in \mathbb{A}$ by the definition of $\mathbb{A}$ which implies that $\mathbb{A} = \mathbb{A}(a_1, \ldots, a_n)$.

Counterexample 1.1.70. Neither $\mathbb{Q}$ nor $\mathbb{R}$ are algebraically closed. In both cases, the irreducible polynomial $x^2 + 1$ has degree 2.

Definition 1.1.71. Let k be a field. An **algebraic closure** $\overline{k}$ of k is an *algebraic extension* of k that is *algebraically closed*.

Example 1.1.72. $\mathbb{C}$ is an algebraic closure of $\mathbb{R}$. It is a finite, hence algebraic, extension of $\mathbb{R}$ that is algebraically closed.

Counterexample 1.1.73. $\mathbb{C}$ is *not* an algebraic closure of $\mathbb{Q}$ because it is not an algebraic extension.

Example 1.1.74. $\mathbb{A}$ is an algebraic closure of $\mathbb{Q}$. It is by construction algebraic and, as we saw, algebraically closed.

Algebraic closures of arbitrary fields exist. As we mentioned, we are going to need Zorn's Lemma.

Lemma 1.1.75 (Kuratowski[16], Zorn)**.** *If $(P, \leqslant)$ is a partially ordered set so that every chain of P has an upper bound, then P has a maximal element.*

We take Zorn's Lemma as an axiom since it is equivalent to the Axiom of Choice. The interested reader can consult [28].

Proposition 1.1.76. *Suppose k is a field. Then an algebraic closure of k exists.*

Proof. For every non constant polynomial $f(x) \in k[x]$, we take an independent variable x_f and consider the polynomial ring R generated by all these variables over k. The ideal $I = \langle f(x_f) : f \in k[x] \rangle$ of R is proper. Otherwise we would find some $g_1, \ldots, g_k \in R$ and some $f_1, \ldots, f_m \in I$ such that

$$g_1 f_1(x_{f_1}) + \ldots + g_m f_m(x_{f_m}) = 1.$$

But each f_i has a root, say a_i. Evaluating at $(x_{f_1}, \ldots, x_{f_m}) = (a_1, \ldots, a_m)$, we get $0 = 1$ which is absurd.

Since the ideal I is proper, it is contained in a maximal ideal J, i.e. $I \subseteq J \subset R$. This is a standard result from Algebra that *requires Zorn's Lemma*. Take the set S of all proper ideals of R and show that every chain has an upper bound, the union of its elements. Zorn's Lemma ensures the existent of a maximal element of S. It is easy to see that this is the required ideal. The missing details are easy to fill.

Now J is maximal so R/J is a field. Using the restriction of the natural projection to F, i.e.

$$\pi|_k : k \to R/J$$

we can see that R/J is an extension of k. This extension contains a root of every $f(x) \in k[x]$, namely $x_f + I$, since

$$f(x_f) + I = I.$$

[16] Kazimierz Kuratowski (1896–1980) had indepedently proven this lemma a few years before Zorn.

We have constructed a field $k_1 = R/J$ that extends k and contains a root for every irreducible polynomial of $k[x]$. With the same arguments we can construct a tower of fields

$$k \leqslant k_1 \leqslant k_2 \leqslant k_3 \leqslant \ldots \leqslant k_s \leqslant \ldots$$

such that k_{j+1} contains a root for every irreducible polynomial in $k_j[x]$. Their union

$$E = \bigcup_{j=1}^{\infty} k_j$$

is clearly an extension of k. By construction, it contains the root of every polynomial $g \in E[x]$. In other words, it is algebraically closed.

We can now take

$$\overline{k} = \{z \in E : z \text{ algebraic over } k\}.$$

$\overline{k}$ is an algebraic extension of k that is algebraically closed. Thus we have constructed an algebraic closure of k. ◇

Using Zorn's Lemma again, it can be shown that

Proposition 1.1.77 (Extension Theorem for Algebraic Closures)**.** *Suppose $\tau : k_1 \to k_2$ is a field isomorphism, S_1 is a set of polynomials over k_1 and $S_2 = \widetilde{\tau}(S_1)$ where $\widetilde{\tau} : k_1[x] \to k_2[x]$ is the map induced by τ. If E_1 is a splitting field of S_1 and E_2 is a splitting field of S_2, then there is an isomorphism $\sigma : E_1 \to E_2$ extending τ. In particular, algebraic closures are unique up to isomorphism.*

Furthermore, if $a_1 \in E_1$ has minimal polynomial $m = m(a, k_1)$ and $a_2 \in E_2$ is any root of $\widetilde{\tau}(m)$, then σ can be chosen so that $\sigma(a_1) = a_2$.

Proof. Consider the set

$$S = \{(L, \vartheta) : L \leqslant E_1,\ \vartheta : L \to E_2 : \vartheta|_{k_1} = \tau\}.$$

Then $S \neq \emptyset$ since $(k_1, \tau) \in S$ and is partially ordered by defining

$$(L_1, \vartheta_1) \leqslant (L_2, \vartheta_2) \iff L_1 \leqslant L_2 \text{ and } \vartheta_2|_{L_1} = \vartheta_1.$$

If $c \equiv (L_i, \vartheta_i)$ is a chain in S and we define $L = \bigcup L_i$ and

$$\vartheta : L \to E_2 : \vartheta(x) = \vartheta_i(x) \text{ if } x \in L_i$$

then (L, ϑ) is an upper bound of c. By Zorn's Lemma, there is a maximal element (L_0, ϑ_0) in S. By definition, $L_0 \leqslant E_1$. If $L_0 \neq E_1$, then there is some $f_1 \in S_1$ that does not split over L_0. Take a root $a_1 \in E_1 \setminus L_0$ of f_1, its minimal polynomial $m_1 = m(a_1, k_1)$ and its image $m_2 = \widetilde{\tau}(m_1)$ and a root $a_2 \in E_2$ of m_2. From the Extension Theorem for simple extensions, τ can be extended to an isomorphism $\rho : L_0(a_1) \to \vartheta_0(L_0)(a_2)$. Then $(L_0(a_1), \rho)$ is an element of S which is bigger that (L_0, ϑ_0), a contradiction. Therefore, $L_0 = E_1$. From the Extension Theorem for splitting fields, $\vartheta_0(E_1) = E_2$. ◇

Corollary 1.1.78. *Every algebraic field extension E of a field k can be embedded in $\overline{k}$.*

Proof. The algebraic closure $\overline{E}$ of E is an algebraic closure of k as well since $k \leqslant E$ is algebraic. So there is an isomorphism $f : \overline{E} \to \overline{k}$ and E is then embedded in $\overline{k}$ as $E \simeq f(E)$. ◇

CONSTRUCTING FIELD EXTENSIONS IV: COMPOSITUMS Having defined algebraic closures, we obtain another way of constructing field extensions that will be proved useful.

Definition 1.1.79. Given any field k and any two *algebraic* field extensions L, M of k, we define their **compositum** LM to be the *smallest subfield of $\overline{k}$ that contains both L and M*, i.e.

$$LM = L(M) = M(L) \leqslant \overline{k}.$$

Similarly, we can define the compositum of any family $\{L_i\}_{i \in I}$ of algebraic extensions of k.

Obviously,

Lemma 1.1.80. *The compositum of any family of algebraic extensions of k is an algebraic extension of k.*

and

Lemma 1.1.81. *The compositum of any finite family of finite extensions of k is a finite extension of k.*

Proof. Suppose L and M are two finite extensions of k. From the previous lemma and Proposition 1.1.53, we need only show that their compositum is finitely generated. Since both L and M are finite, they are finitely generated; write $L = k(a_1, \ldots, a_s)$ and $M = k(b_1, \ldots, b_r)$. It is now immediate that $LM = k(a_1, \ldots, a_s, b_1, \ldots, b_r)$. We proceed with induction. ◇

1.2 THE GALOIS–ARTIN CORRESPONDENCE

The distinction, although artificial, between Field Theory and Galois Theory is in the tools we use to study field extensions. Until now, we have only used the theory of vector spaces. When Galois' ideas are introduced into the theory of fields, richer and deeper results are obtained.

In this section we define the Galois correspondence. We will see how we can associate each field extension to a suitably chosen group and what information can the latter give us about the extension. This group will be the group of automorphisms of the extension in question. So before we see how we can use it, lets state some important results.

MORE ON k-AUTOMORPHISMS Let us recall some definitions. Suppose $k \leqslant E$ be a field extension. An k-**automorphism** of E is a map $\sigma \in \mathrm{Aut}(E)$ such that $\sigma|_k = \mathrm{id}_k$. The k-automorphisms of E are exactly the field extension isomorphisms from $k \leqslant E$ to itself.

The set of all k-automorphisms of E, denoted by $\mathrm{Aut}(E/k)$, forms a group under the usual composition of maps; it is a *subgroup* of $\mathrm{Aut}(E)$ (Lemma 1.1.25).

For *finite* extensions, there is a rather straightforward way of computing $\mathrm{Aut}(E/k)$.

Example 1.2.1. Consider the *finite* (hence *finitely generated* and *algebraic*) extension $k \leqslant E$ where $E = k(X)$ for some finite subset $X = \{\alpha_1, \ldots, \alpha_n\}$ of E. Let $\sigma \in \mathrm{Aut}(E/k)$.

Since $\sigma|_k = \mathrm{id}_k$, the map σ is determined solely by its action on the elements of X. Indeed, using Cor. 1.1.30, for any $x \in E = k(\alpha_1, \ldots, \alpha_n)$, we have

$$\sigma(x) = \sigma\left(\frac{f(\alpha_1, \ldots, \alpha_n)}{g(\alpha_1, \ldots, \alpha_n)}\right) = \frac{f\big(\sigma(\alpha_1), \ldots, \sigma(\alpha_n)\big)}{g\big(\sigma(\alpha_1), \ldots, \sigma(\alpha_n)\big)}.$$

Since $k \leqslant E$ is *algebraic*, every $\alpha \in X$ is algebraic over k. If the minimal polynomial of an element $\alpha \in X$ over k is

$$m(x) \equiv m(\alpha, k)(x) = x^n + \ldots + a_1 x + a_0 \in k[x],$$

then $\sigma(\alpha)$ is just another root of m. Indeed,

$$\begin{aligned} m(\alpha) = 0 &\Rightarrow \sigma\big(m(\alpha)\big) = \sigma(0) \\ &\Rightarrow \sigma(\alpha^n + \ldots + a_1\alpha + a_0) = 0 \\ &\Rightarrow \sigma(\alpha^n) + \ldots + \sigma(a_1)\sigma(\alpha) + \sigma(a_0) = 0 \\ &\Rightarrow \sigma(\alpha)^n + \ldots + a_1\sigma(\alpha) + a_0 = 0 \\ &\Rightarrow m\big(\sigma(\alpha)\big) = 0. \end{aligned}$$

The roots of an irreducible polynomial are said to be **conjugate**. So the image $\sigma(\alpha)$ of some $\alpha \in X$ under $\sigma \in \mathrm{Aut}(E/k)$ *is a conjugate of* α.

Example 1.2.2. The only $\mathbb{R}$-automorphisms of $\mathbb{C} = \mathbb{R}(i)$ are the identical map $\mathrm{id}_{\mathbb{C}}$ and complex conjugation $z \mapsto \bar{z}$.

It is now apparent that

Proposition 1.2.3. *If $k \leqslant E$ is a finite extension, then* $\mathrm{Aut}(E/k)$ *is also finite.*

Proof. Immediate since (i) every $\sigma \in \mathrm{Aut}(E/k)$ is determined by its action on X, (ii) X is finite, (iii) $\sigma(\alpha)$ is a conjugate of α for every $\alpha \in X$ and (iv) $\partial m(\alpha, k) < \infty$ and therefore $\sigma(\alpha)$ can only take a finite number of values. ◇

THE CORRESPONDENCE It is time to describe the Galois correspondence. Let $k \leqslant E$ be a field extension. As we said, the group we are going to use to study the extension is the group $\mathrm{Aut}(E/k)$ of

k-automorphisms of E. The way we will pass from the extension to the group and vice versa is the following.

We associate every intermediate field $k \leqslant L \leqslant E$ with the group $\mathrm{Aut}(E/L)$ of L-automorphisms of E. Using elementary Group Theory it is easy to deduce that $\mathrm{Aut}(E/L)$ *is a subgroup of* $\mathrm{Aut}(E/F)$.

Lemma 1.2.4. *If L is a subextension of $k \leqslant E$ then $\mathrm{Aut}(E/L)$ is a subgroup of $\mathrm{Aut}(E/k)$.*

In the opposite direction, we associate with each subgroup $H \leqslant \mathrm{Aut}(E/k)$, the set

$$\mathrm{Fix}_E(H) = \{x \in E : \sigma(x) = x \ \forall \sigma \in H\}.$$

Using elementary properties of homomorphisms, it is easy to see that the above set is a *subextension* of $k \leqslant E$.

Lemma 1.2.5. *If $k \leqslant E$ is a field extension and H is a subgroup of $\mathrm{Aut}(E/k)$ then $\mathrm{Fix}_E(H)$ is a subextension of $k \leqslant E$.*

This establishes a correspondence between intermediate fields of a field extension and subgroups of its k-automorphism group.

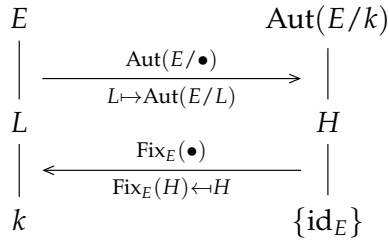


Figure 1.1: The Galois-Artin Correspondence

Example 1.2.6. Let $k \leqslant E$ be a field extension and $G = \mathrm{Aut}(E/k)$ its k-automorphism group. To the field k we assign the subgroup of G that fixes k which is, by definition, the whole group G. To the field E we assign the subgroup of G that fixes E, which is the trivial subgroup $\{\mathrm{id}_E\}$ since id_E is the only isomorphism $E \to E$ that fixes E. Observe how the *whole group* is associated to the *base field* while the *trivial subgroup* is associated to the *extending field*.

The above example reveals a crucial property of the correspondence.

Proposition 1.2.7. *The Galois correspondence of a field extension*

$$\{\text{subextensions } L:\ k \leqslant L \leqslant E\} \rightleftarrows \{\text{subgroups } H:\ H \leqslant \mathrm{Aut}(E/k)\}$$

as described above, is order reversing.

Proof. If $L_1 \leqslant L_2$ then given some automorphism $\sigma \in \mathrm{Aut}(E/L_2)$ that fixes L_2 pointwise, σ also fixes $L_1 \subseteq L_2$ pointwise.

On the other hand, if $H_1 \leqslant H_2$ and some element $a \in E$ is fixed by every automorphism $\tau \in H_2$, then it is also fixed by every automorphism $\tau' \in H_1 \subseteq H_2$. ◇

Another important property which is an immediate consequence of the definitions is

Proposition 1.2.8. *If $k \leqslant E$ is a field extension, then*

$$H \subseteq \mathrm{Aut}\big(E/\,\mathrm{Fix}_E(H)\big)\ \forall H \leqslant \mathrm{Aut}(E/k) \text{ and}$$
$$L \subseteq \mathrm{Fix}_E\big(\,\mathrm{Aut}(E/L)\big)\ \ \forall L : k \leqslant L \leqslant E.$$

BIJECTIVE GALOIS CORRESPONDENCES By its definition, the Galois correspondence is a correspondence between *the set of subextensions L of $k \leqslant E$ that arise as fixed fields*, i.e. $L = \mathrm{Fix}_E(H)$ for some $H \leqslant \mathrm{Aut}(E/k)$, and *the set of subgroups H of $\mathrm{Aut}(E/k)$ that arise as automorphism groups*, i.e. $H = \mathrm{Aut}(E/L)$ for some $k \leqslant L \leqslant E$.

$$\left\{\begin{array}{c}\text{subextensions } k \leqslant L \leqslant E:\\ \exists H \leqslant \mathrm{Aut}(E/k) : L = \mathrm{Fix}_E(H)\end{array}\right\} \rightleftarrows \left\{\begin{array}{c}\text{subgroups } H \leqslant \mathrm{Aut}(E/k):\\ \exists k \leqslant L \leqslant E : H = \mathrm{Aut}(E/L)\end{array}\right\}$$

The next result suggests that this is a bijective correspondence, i.e. the maps $\mathrm{Aut}(E/\bullet)$ and $\mathrm{Fix}_E(\bullet)$ are mutually inverse.

Proposition 1.2.9. *Let $k \leqslant E$ be a field extension. With the above notation,*

(i) If $H = \mathrm{Aut}(E/L)$ for some $k \leqslant L \leqslant E$ then

$$H = \mathrm{Aut}\big(E/\,\mathrm{Fix}_E(H)\big).$$

(ii) If $L = \mathrm{Fix}_E(H)$ for some $H \leqslant \mathrm{Aut}(E/k)$ then

$$L = \mathrm{Fix}\big(\,\mathrm{Aut}(E/L)\big).$$

Proof. (i) If $H = \mathrm{Aut}(E/L)$ for some subextension L of $k \leqslant E$ then

$$\begin{aligned} H = \mathrm{Aut}(E/L) &\Rightarrow L \subseteq \mathrm{Fix}_E(H) \\ &\Rightarrow \mathrm{Aut}(E/L) \supseteq \mathrm{Aut}\left(E/\mathrm{Fix}_E(H)\right) \\ &\Rightarrow H \supseteq \mathrm{Aut}\left(E/\mathrm{Fix}_E(H)\right) \end{aligned}$$

and we already know that $H \subseteq \mathrm{Aut}\left(E/\mathrm{Fix}_E(H)\right)$. The two inclusions give $H = \mathrm{Aut}\left(E/\mathrm{Fix}_E(H)\right)$.

(ii) Similarly, if $L = \mathrm{Fix}_E(H)$ for some subgroup $H \leqslant \mathrm{Aut}(E/k)$ then

$$\begin{aligned} L = \mathrm{Fix}_E(H) &\Rightarrow H \subseteq \mathrm{Aut}(E/L) \\ &\Rightarrow \mathrm{Fix}_E(H) \supseteq \mathrm{Fix}_E\left(\mathrm{Aut}(E/L)\right) \\ &\Rightarrow L \supseteq \mathrm{Fix}_E\left(\mathrm{Aut}(E/L)\right) \end{aligned}$$

and we already know that $L \subseteq \mathrm{Fix}_E\left(\mathrm{Aut}(E/L)\right)$. The two inclusions give $L = \mathrm{Fix}\left(\mathrm{Gal}(E/L)\right)$.

◇

Our main objective now is to examine to which extend the Galois correspondence for an arbitrary field extension $k \leqslant E$ is a bijective correspondence between the set of *all* subextensions of $k \leqslant E$ and the set of *all* subgroups of $\mathrm{Aut}(E/k)$. In other words, to what extend the maps $\mathrm{Aut}(E/\bullet)$ and $\mathrm{Fix}_E(\bullet)$ are onto or, equivalently by Prop. 1.2.9, mutually inverse, i.e.

$$H = \mathrm{Aut}\left(E/\mathrm{Fix}_E(H)\right) \ \forall H \leqslant \mathrm{Aut}(E/k) \tag{1.6}$$

and

$$L = \mathrm{Fix}_E\left(\mathrm{Aut}(E/L)\right) \ \forall L : k \leqslant L \leqslant E. \tag{1.7}$$

In general these maps are *not* mutualy inverse. As we saw (Prop. 1.2.8), they satisfy the weaker conditions

$$H \subseteq \mathrm{Aut}\left(E/\mathrm{Fix}_E(H)\right) \ \forall H \leqslant \mathrm{Aut}(E/k) \tag{1.8}$$

and

$$L \subseteq \mathrm{Fix}_E\left(\mathrm{Aut}(E/L)\right) \ \forall L : k \leqslant L \leqslant E. \tag{1.9}$$

There are examples where these inclusions can be equalities as shown below.

Example 1.2.10. Consider the extension $\mathbb{R} \leqslant \mathbb{R}(i)$ inside $\mathbb{C}$. On the one hand, $G = \mathrm{Aut}(\mathbb{R}(i)/\mathbb{R})$ is a group of order 2 (Ex. 1.2.2). On the other, by Ex. 1.1.10 and Prop. 1.1.11, the only intermediate fields $\mathbb{R} \leqslant L \leqslant \mathbb{R}(i)$ are $L = \mathbb{R}$ and $L = \mathbb{R}(i)$. Thus the Galois correspondence associates

$$\mathbb{R} \rightleftarrows G \quad \text{and} \quad \mathbb{R}(i) \rightleftarrows \{\mathrm{id}_{\mathbb{R}(i)}\}$$

and the two inclusions are (rather trivially) equalities.

But there are also examples where the inclusions are strict.

Counterexample 1.2.11. A case where (1.8) might be strict is when a given extension $k \leqslant E$ has infinite degree. In this case we will see in due time that the corresponding group $\mathrm{Aut}(E/k)$ is also infinite, hence too big to be handled properly. In particular, $\mathrm{Aut}(E/k)$ has too many subgroups! Indeed, as we shall see in Ex. 1.5.3, in this case not every subgroup H of $\mathrm{Aut}(E/k)$ arises as an automorphism group, i.e. there might not exist L such that $k \leqslant L \leqslant E$ and $H = \mathrm{Aut}(E/L)$.

So if H is a subgroup that cannot arise as an automorphism group of some intermediate field, then $H \subsetneq \mathrm{Aut}\left(E/\mathrm{Fix}_E(H)\right)$.

Counterexample 1.2.12. Consider the extension $\mathbb{Q} \leqslant \mathbb{Q}(\sqrt[3]{2})$. If $\sigma \in \mathrm{Aut}(\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q})$ then σ is determined by its image $\sigma(\sqrt[3]{2})$ which is a conjugate of $\sqrt[3]{2}$. But the conjugates of $\sqrt[3]{2}$ are all complex numbers that are not in $\mathbb{Q}(\sqrt[3]{2})$. Therefore, the group of automorphisms $\mathrm{Aut}(\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q})$ is trivial, hence its fixed field is

$$\mathrm{Fix}_{\mathbb{Q}(\sqrt[3]{2})}\left(\mathrm{Aut}(\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q})\right) = \mathbb{Q}(\sqrt[3]{2}) \supsetneq \mathbb{Q}$$

and that shows that the inclusion (1.9) may also be strict.

Counterexample 1.2.13. There is another instance where (1.9) may be strict, but for completely different reasons this time. Consider the finite field $\mathbb{F}_2$ and an indeterminate t. We can then form the extension $\mathbb{F}_2 \leqslant \mathbb{F}_2(t)$. If we take the element $t^2 \in \mathbb{F}_2(t)$, we can further form the tower

$$\mathbb{F}_2 \leqslant \mathbb{F}_2(t^2) \leqslant \mathbb{F}_2(t).$$

We will focus on the extension $\mathbb{F}_2(t^2) \leqslant \mathbb{F}_2(t)$.

First of all, it is intuitively clear (but requires a long, technical proof which we will avoid) that $\mathbb{F}_2(t) = [\mathbb{F}_2(t^2)](t)$ and therefore the extension is simple (every polynomial in t with coefficients in $\mathbb{F}_2$ can be viwed as a polynomial in t with coefficients from $\mathbb{F}_2(t^2)$ since $\mathbb{F}_2 \leqslant \mathbb{F}_2(t^2)$; on the other hand, a polynomial in t with coefficients in $\mathbb{F}_2(t^2)$ can be considered as a polynomial in t using the distributive law and grouping together all the t's in every monomial).

The element t does not belong in $\mathbb{F}_2(t^2)$ but is algebraic over $\mathbb{F}_2(t^2)$: the polynomial

$$m(x) = x^2 - t^2 \in [\mathbb{F}_2(t^2)][x]$$

is monic, of minimal degree such that $m(t) = 0 \in \mathbb{F}_2(t^2)$ hence by definition is the minimal polynomial of t over $\mathbb{F}_2(t^2)$, i.e.

$$m(x) = m\big(t, \mathbb{F}_2(t^2)\big)(x).$$

Observe at this point that since we are inside $\mathbb{F}_2$, we have

$$m(x) = x^2 - t^2 = (x - t)^2$$

so the only root of m is t. We can now imagine where this leads us.

If $\sigma \in \operatorname{Aut}\big(\mathbb{F}_2(t)/\mathbb{F}_2(t^2)\big)$ then σ is determined by its action on t and $\sigma(t)$ is a conjugate of t. But the only conjugate of t is t itself. Therefore $\sigma(t) = t$ and the automorphism group is again the trivial, which means

$$\operatorname{Fix}_{\mathbb{F}_2(t)}\left[\operatorname{Aut}\big(\mathbb{F}_2(t)/\mathbb{F}_2(t^2)\big)\right] = \mathbb{F}_2(t) \supsetneq \mathbb{F}_2(t^2).$$

By the above examples it is apparent that the Galois correspondence is *not* bijective in general. Our next plan is to understand why it failed to be bijective in the above counterexamples and define for which extensions the correspondence is actually bijective.

THE CORRESPONDENCE FOR FINITE EXTENSIONS If we try to understand why

$$H = \operatorname{Aut}\big(E/\operatorname{Fix}_E(H)\big)\ \forall H \leqslant \operatorname{Aut}(E/k) \tag{1.6}$$

fails to hold for infinite extensions (1.2.11), we end up thinking if it is due to the automorphism group being too big. So it is natural to

wonder whether finite extensions suffice for such an equality. We will see that the answer is *yes*.

If $k \leqslant E$ is a *finite* field extension then, by Prop. 1.2.3, $\mathrm{Aut}(E/k)$ is finite and so are its subgroups H and $\mathrm{Aut}\left(E/\operatorname{Fix}_E(H)\right)$.

So we have two *finite* subgroups of $\mathrm{Aut}(E/k)$ for which we already know that

$$H \subseteq \mathrm{Aut}\left(E/\operatorname{Fix}_E(H)\right).$$

So it suffices to show that

$$|H| \geqslant |\mathrm{Aut}\left(E/\operatorname{Fix}_E(H)\right)|.$$

We will proceed by finding an upper bound b for the cardinality $|\mathrm{Aut}\left(E/\operatorname{Fix}_E(H)\right)|$ and we will then show that $|H| = b$.

Lemma 1.2.14 (Dedekind). *Let $\vartheta_1, \ldots, \vartheta_n : k \to E$ be distinct field monomorphisms. Then the ϑ_i's are linearly independent over E.*

Proof. Suppose, for the contrary, that there are $c_1, \ldots, c_n \in E$ not all zero such that

$$c_1\vartheta_1(x) + \ldots + c_n\vartheta_n(x) = 0 \ \forall x \in k.$$

Omitting the terms whose $c_i = 0$ and rearranging the rest if necessary, we get a minimal $m \leqslant n$ such that

$$c_1\vartheta_1(x) + \ldots + c_k\vartheta_m(x) = 0 \ \forall x \in k \tag{1.10}$$

and $c_i \neq 0$ for all $i = 1, \ldots, m$. Since the monomorphisms are distinct, there is some $x_0 \in k$ such that $\vartheta_1(x_0) \neq \vartheta_2(x_0)$. If we multiply both sides of (1.10) by $\vartheta_1(x_0)$ we get

$$c_1\vartheta_1(x)\vartheta_1(x_0) + \ldots + c_m\vartheta_m(x)\vartheta_1(x_0) = 0 \ \forall x \in k. \tag{1.11}$$

Taking $x = x_0x$ in (1.10) we have

$$c_1\vartheta_1(x_0x) + \ldots + c_m\vartheta_m(x_0x) = 0 \ \forall x \in k$$

or, equivalently,

$$c_1\vartheta_1(x_0)\vartheta_1(x) + \ldots + c_m\vartheta_m(x_0)\vartheta_m(x) = 0 \ \forall x \in k. \tag{1.12}$$

Substracting (1.12) from (1.11) gives us

$$c_1\vartheta_1(x)\big(\vartheta_1(x_0)-\vartheta_1(x_0)\big)+\ldots+c_m\vartheta_m(x)\big(\vartheta_1(x_0)-\vartheta_m(x_0)\big)=0$$

or, equivalently,

$$c_2\vartheta_2(x)\big(\vartheta_1(x_0)-\vartheta_2(x_0)\big)+\ldots+c_m\vartheta_m(x)\big(\vartheta_1(x_0)-\vartheta_m(x_0)\big)=0$$

which contradicts the minimality of m. ◇

Proposition 1.2.15. *If $k \leqslant E$ is a finite field extension then $|\operatorname{Aut}(E/k)|$ is at most $[E:k]$.*

Proof. Since $k \leqslant E$ is finite, so is $\operatorname{Aut}(E/k)$. Suppose that $\operatorname{Aut}(E/k) = \{\vartheta_1,\ldots,\vartheta_n\}$ and that $[E:k] = m < n$. If $\mathcal{B} = \{a_1,\ldots,a_m\}$ is an k-basis of E, the matrix

$$A = \begin{pmatrix} \vartheta_1(a_1) & \vartheta_1(a_2) & \ldots & \vartheta_1(a_m) \\ \vartheta_2(a_1) & \vartheta_2(a_2) & \ldots & \vartheta_2(a_m) \\ \vdots & \vdots & \ddots & \vdots \\ \vartheta_n(a_1) & \vartheta_n(a_2) & \ldots & \vartheta_n(a_m) \end{pmatrix}$$

has rank $\operatorname{rank}(A) \leqslant m < n$. Therefore, its rows are linearly dependent over k; so there are $c_{ji} \in E$, $i = 1,\ldots,n$, not all zero, such that $\sum_{i=1}^{n} c_{ji}\vartheta_i(a_j) = 0$ for all $j = 1,\ldots,m$. As a result, for every $x = \sum_{j=1}^{m} x_j a_j \in E$ ($x_j \in F$) we have

$$\begin{aligned}\sum_{i=1}^{n} c_{ji}\vartheta_i(x) &= \sum_{i=1}^{n} c_{ji}\vartheta_i\left(\sum_{j=1}^{m} x_j a_j\right) = \sum_{i=1}^{n} c_{ji}\left(x_j \sum_{j=1}^{m} \vartheta_i(a_j)\right) \\ &= \sum_{j=1}^{m} x_j\left(\sum_{i=1}^{n} c_{ji}\vartheta_i(a_j)\right) = 0\end{aligned}$$

so the ϑ_i's are E-lineraly dependent which contradicts Dedekind's Lemma since the ϑ_i's are distinct monomorphisms $E \to E$. Therefore, $|\operatorname{Aut}(E/k)| \leqslant [E:k]$ ◇

The above, together with Cor 1.1.13, give us an upper bound for $|\operatorname{Aut}\big(E/\operatorname{Fix}_E(H)\big)|$.

Corollary 1.2.16. *Suppose $k \leqslant E$ is a finite field extension and H is a subgroup of $\operatorname{Aut}(E/k)$. Then $|\operatorname{Aut}\big(E/\operatorname{Fix}_E(H)\big)| \leqslant [E:\operatorname{Fix}_E(H)]$.*

Proposition 1.2.17. *If E is a field and H is a finite subgroup of* $\mathrm{Aut}(E)$ *then*

$$|H| = [E : \mathrm{Fix}_E(H)].$$

Proof. Let

$$H = \{\vartheta_1 = 1, \vartheta_2, \ldots, \vartheta_n\}$$

and $[E : \mathrm{Fix}_E(H)] = m$. We will show that $m < n$ and $m > n$ cannot happen.

Suppose $m < n$ and $\{a_1, \ldots, a_m\}$ is a $\mathrm{Fix}_E(H)$-basis of E. The homogenous system

$$\begin{cases} \vartheta_1(a_1)x_1 + \ldots + \vartheta_n(a_1)x_n = 0 \\ \quad \vdots \\ \vartheta_1(a_m)x_1 + \ldots + \vartheta_n(a_m)x_n = 0 \end{cases}$$

has m linear equations and $n < m$ unknowns. Hence it has a non-zero solution $(y_1, \ldots, y_n) \in E^n$, i.e.

$$\vartheta_1(a_i)y_1 + \ldots + \vartheta_n(a_i)y_n = 0 \quad \forall i = 1, \ldots, m.$$

For an arbitrary $x = \sum_{j=1}^{m} c_j a_j \in E$, $c_j \in \mathrm{Fix}_E(H)$, we get

$$\begin{aligned} \vartheta_1(x)y_1 + \ldots + \vartheta_n(x)y_n &= \vartheta_1\left(\sum_{j=1}^{m} c_j a_j\right) y_1 + \ldots + \vartheta_n\left(\sum_{j=1}^{m} c_j a_j\right) y_n \\ &= \sum_{j=1}^{m} c_j[\vartheta_1(a_j)y_1 + \ldots + \vartheta_n(a_j)y_n] = 0. \end{aligned}$$

That is, the monomorphisms $\vartheta_1, \ldots, \vartheta_n$ are linearly dependent. But this contradicts Dedekind's Lemma. Therefore, $m \geqslant n$.

Suppose now that $m > n$ and take again some $\mathrm{Fix}_E(H)$-lineraly independent set of $n+1$ elements, say $\{a_1, \ldots, a_{n+1}\}$. Once more, we have a homogenous linear system

$$\begin{cases} \vartheta_1(a_1)x_1 + \ldots + \vartheta_1(a_{n+1})x_{n+1} = 0 \\ \quad \vdots \\ \vartheta_n(a_1)x_1 + \ldots + \vartheta_n(a_{n+1})x_{n+1} = 0 \end{cases}$$

with n equations and $n+1 > n$ unknowns. Hence it has a non-zero solution. We can choose a solution $(y_1, \ldots, y_{n+1}) \in E^n$ that has the

fewest possible non-zero coordinates. Without loss of generality we may assume that

$$y_1, \dots, y_r \neq 0, \;\; y_{r+1}, \dots, y_{n+1} = 0$$

for some $1 \leqslant r \leqslant n+1$ so that we have

$$\vartheta_i(a_1)y_1 + \dots + \vartheta_i(a_r)y_r = 0 \;\; \forall i = 1, \dots, n. \tag{$*$}$$

For all $\vartheta \in H$ we have

$$\vartheta\vartheta_i(a_1)\vartheta(y_1) + \dots + \vartheta\vartheta_i(a_r)\vartheta(y_r) = 0 \;\; \forall i = 1, \dots, n.$$

or equivalently, since the map $H \to H : \vartheta_i \mapsto \vartheta\vartheta_i$ is a bijection,

$$\vartheta_i(a_1)\vartheta(y_1) + \dots + \vartheta_i(a_r)\vartheta(y_r) = 0 \;\; \forall i = 1, \dots, n. \tag{$**$}$$

Multiplying $(*)$ by $\vartheta(a_1)$ and $(**)$ by a_1 and substracting, we get

$$[y_2\vartheta(y_1) - \vartheta(y_2)y_1]\vartheta_i(a_2) + \dots + [y_r\vartheta(y_1) - \vartheta(y_r)y_1)]\vartheta_i(a_r) = 0$$

for all $i = 1, \dots, n$ which contradicts the minimality of r. Therefore

$$y_j\vartheta(y_1) - y_1\vartheta(y_j) = 0 \iff y_j y_1^{-1} = \vartheta(y_j y_1^{-1}) \;\; \forall j = 1, \dots, r$$

for all $\vartheta \in H$. That means we can find some $z_1, \dots, z_r \in \mathrm{Fix}_E(H)$ and some $k \in E$ so that $y_j = kz_j$ for all $j = 1, \dots, r$. Then, for $i = 1$, $(*)$ becomes

$$kz_1\, a_1 + \dots + kz_r\, a_r = 0 \overset{k\neq 0}{\iff} z_1a_1 + \dots + z_ra_r = 0$$

which contradicts the linear independence of $\{a_1, \dots, a_{n+1}\}$. Therefore $m \leqslant n$ and we conclude that $m = n$. ◇

Corollary 1.2.18. *For any finite field extension $k \leqslant E$ and any subgroup H of* $\mathrm{Aut}(E/k)$*,*

$$H = \mathrm{Aut}\big(E/\,\mathrm{Fix}_E(H)\big).$$

Proof. We have $H \subseteq \mathrm{Aut}\big(E/\,\mathrm{Fix}_E(H)\big)$ and

$$|\,\mathrm{Aut}\big(E/\,\mathrm{Fix}_E(H)\big)| \overset{1.2.16}{\leqslant} [E : \mathrm{Fix}_E(H)] \overset{1.2.17}{=} |H| \overset{1.2.3}{<} \infty.$$

Therefore $H = \mathrm{Aut}\big(E/\,\mathrm{Fix}_E(H)\big)$. ◇

GALOIS EXTENSIONS I It remains to examine the extensions for which (1.7) holds. These extensions are called *Galois extensions* because they naturally generalize the setting Galois used to work in, to abstract fields.

Definition 1.2.19 (1st definition of Galois extensions)**.** A field extension $k \leqslant E$ is called **Galois** if

$$L = \mathrm{Fix}_E\left(\mathrm{Aut}(E/L)\right)$$

for every subextension L of $k \leqslant E$. In this case we write $\mathrm{Gal}(E/k)$ for $\mathrm{Aut}(E/k)$ and call it the **Galois group** of the extension.

Example 1.2.20. The extension $\mathbb{R} \leqslant \mathbb{R}(i)$ of Ex. 1.2.10 is Galois as we already saw.

Counterexample 1.2.21. The extensions $\mathbb{Q} \leqslant \mathbb{Q}(\sqrt[3]{2})$ and $\mathbb{F}_2(t^2) \leqslant \mathbb{F}_2(t)$ of Ex. 1.2.12 and 1.2.13 respectively are *not* Galois.

This definition gives us no information on the structure of a Galois extension and is in general hard to work with especially since the defining condition (1.7) has to be checked for *every* intermediate field L. We want to understand the structure of Galois extensions better and see if we can simplify their definition. This is done in the next section where we will see a number conditions equivalent to (1.7) that will help us understand Galois extensions better.

1.3 GALOIS EXTENSIONS

GALOIS EXTENSIONS II Turning our attention to

$$L = \mathrm{Fix}_E\left(\mathrm{Aut}(E/L)\right)\ \forall L : k \leqslant L \leqslant E \tag{1.7}$$

and why that failed to hold in 1.2.12 and 1.2.13 for $L = k$, we find that, contrary to the previous discussion, the automorphism group in both cases is too small and thus contains little information about the field extension in the sense that it does not match its upper bound set in Prop. 1.2.15:

$$|\mathrm{Aut}(\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q})| = 1 < 3 = \partial m(\sqrt[3]{2}, \mathbb{Q}) = [\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}]$$

and

$$|\operatorname{Aut}\left(\mathbb{F}_2(t)/\mathbb{F}(t^2)\right)| = 1 < 2 = \partial m\big(t, \mathbb{F}_2(t^2)\big) = [\mathbb{F}_2(t) : \mathbb{F}_2(t^2)].$$

Therefore it would be reasonable to suspect that if $|\operatorname{Aut}(E/L)|$ equals $[E : L]$ then (1.7) holds. In fact, the conditions

$$L = \operatorname{Fix}_E\big(\operatorname{Aut}(E/L)\big) \ \forall L : k \leqslant L \leqslant E$$

and

$$|\operatorname{Aut}(E/L)| = [E : L] \ \forall L : k \leqslant L \leqslant E$$

are equivalent for *finite* extensions.

Proposition 1.3.1. *Suppose $k \leqslant E$ is a finite extension and L is a subextension, i.e. $k \leqslant L \leqslant E$. Then*

$$L = \operatorname{Fix}_E\big(\operatorname{Aut}(E/L)\big) \quad \textit{iff} \quad |\operatorname{Aut}(E/L)| = [E : L].$$

Proof. ($\Rightarrow$) Suppose $L = \operatorname{Fix}_E\big(\operatorname{Aut}(E/L)\big)$. Since $\operatorname{Aut}(E/L)$ is a subgroup of the finite group $\operatorname{Aut}(E/k)$ (and therefore a finite subgroup of $\operatorname{Aut}(E)$), 1.2.17 implies that

$$|\operatorname{Aut}(E/L)| = [E : \operatorname{Fix}_E\big(\operatorname{Aut}(E/L)\big)] = [E : L].$$

($\Leftarrow$) For the contrary we assume that $|\operatorname{Aut}(E/L)| = [E : L]$. By 1.2.17 again we have that

$$[E : L] = |\operatorname{Aut}(E/L)| = [E : \operatorname{Fix}_E\big(\operatorname{Aut}(E/L)\big)].$$

Since, by (1.9), $L \subseteq \operatorname{Fix}_E\big(\operatorname{Aut}(E/L)\big)$, 1.1.11 gives us that

$$[E : L] = \underbrace{[E : \operatorname{Fix}_E\big(\operatorname{Aut}(E/L)\big)]}_{=[E:L]}[\operatorname{Fix}_E\big(\operatorname{Aut}(E/L)\big) : L] < \infty$$

and therefore $[\operatorname{Fix}_E\big(\operatorname{Aut}(E/L)\big) : L] = 1$ which means that $\operatorname{Fix}_E\big(\operatorname{Aut}(E/L)\big) = L$.

◇

Definition 1.3.2 (2nd definition of Galois extensions - *finite* case)**.** A *finite* field extension $k \leqslant E$ is called **Galois** if

$$|\operatorname{Aut}(E/L)| = [E : L] \tag{1.13}$$

for every subextension L of $k \leqslant E$. In this case we write $\operatorname{Gal}(E/k)$ for $\operatorname{Aut}(E/k)$ and call it the **Galois group** of the extension.

Example 1.3.3. The extension $\mathbb{R} \leqslant \mathbb{R}(i)$ of Ex. 1.2.10 is Galois because, as we saw, $\mathrm{Aut}(\mathbb{R}(i)/\mathbb{R})$ has only two elements (the identity function $z \mapsto z$ and the complex conjugation $z \mapsto \bar{z}$) and therefore

$$|\mathrm{Aut}(\mathbb{R}(i)/\mathbb{R})| = 2 = \partial m(i, \mathbb{R}) = [\mathbb{R}(i) : \mathbb{R}].$$

Counterexample 1.3.4. In the beginning of this section we saw that the condition (1.13) does not hold for the extensions $\mathbb{Q} \leqslant \mathbb{Q}(\sqrt[3]{2})$ and $\mathbb{F}_2(t^2) \leqslant \mathbb{F}_2(t)$ so these are *not* Galois.

But there is also a different path we can take to define Galois extensions, one that reveals more about these extensions and how their study arise naturally in Galois' work.

To give some motivation for the work that follows, lets see what it means for a *finite* and *simple* extension to be Galois using the definitions we have formulated so far. Since simple algebraic extensions are the building blocks of finite extensions, the next example will be useful to generalize the ideas to arbitrary extensions.

Example 1.3.5. Suppose $k \leqslant k(a)$ is a simple algebraic extension. We saw in 1.2.15 that

$$|\mathrm{Aut}(k(a)/k)| \leqslant [k(a) : k] = \partial m(a, k).$$

If $k \leqslant k(a)$ is Galois then $|\mathrm{Gal}(k(a)/L)| = [k(a) : L]$ for every subextension L and in particular

$$|\mathrm{Gal}(k(a)/k)| = [k(a) : k] = \partial m(a, k). \tag{1.14}$$

But we know that every $\sigma \in \mathrm{Gal}(k(a)/k)$ is determined by its action on a and $\sigma(a)$ is a root of $m(a, k)$. So by (1.14), there must be $\partial m(a, k)$ *distinct* elements in $\mathrm{Gal}(k(a)/k) \iff \sigma(a)$ takes *exactly* $\partial m(a, k)$ *distinct values*. In other words:

- *every* root of $m(a, k)$ *must lie in* $k(a)$ and
- there aren't any repetitions, i.e. *there are no multiple roots.*

It is not hard to see that the converse also holds. If every root of $m(a, k)$ lies in $k(a)$ and there are no multiple roots then

$$|\mathrm{Aut}(k(a)/k)| = \partial m(a, k) = [k(a) : k]$$

and the extension $k \leqslant F(k)$ is Galois.

NORMAL EXTENSIONS In the last example, we saw that if a finite simple extension $k \leqslant k(a)$ is Galois then all the roots of $m(a,k)$ are inside $k(a)$. We give extensions with the suitably generalized property a name.

Definition 1.3.6. A field extension $k \leqslant E$ is **normal** if it is algebraic and $m(a,k)$ splits in E for every $a \in E$.

Example 1.3.7. If $k \leqslant E$ is an arbitrary field extension such that $[E:k] = 2$, then the extension is normal. First of all, the extension is finite hence algebraic. Let $a \in E \setminus k$.[17] Since $[E:k] = 2$, from

$$[E:k] = [E:k(a)][k(a):k]$$

we get $E = k(a)$ and consequently $\partial m(a,k) = 2$. The minimal polynomial $m = m(a,k)$ has a root in E and has degree 2 so it splits over E. Since a was arbitrary, the extension is normal.

Example 1.3.8. The extension $\mathbb{R} \leqslant \mathbb{R}(i)$ of Ex. 1.2.10 is normal since $[\mathbb{R}(i):\mathbb{R}] = 2$.

Counterexample 1.3.9. The extension $\mathbb{Q} \leqslant \mathbb{Q}(\sqrt[3]{2})$ of Ex. 1.2.12 is *not* normal. The two complex roots of $m(\sqrt[3]{2}, \mathbb{Q}) = x^3 - 2$ are not in $\mathbb{Q}$.

Example 1.3.10. The extension $k \leqslant k_{sep}$ is normal. Take some $a \in k_{sep}$ and $m = m(a,k)$. If $a_1, \ldots, a_r$ are the roots of m in the algebraic closure $\overline{k}$ of k, then every a_i is in k_{sep}. For $a_i \notin k_{sep}$ implies that a_i is not separable over k, i.e. $m(a_i,k)$ is not separable which is absurd since $m(a_i,k) = m$.

Normal extensions have nice equivalent formulations that are sometimes more appropriate to work with, depending on the context.

Lemma 1.3.11. *Let $k \leqslant E$ be an algebraic extension. The following are equivalent*

i) Every irreducible polynomial $p(x) \in k[x]$ that has a root in E splits.

ii) E is normal over k.

[17] The case $a \in k$ is trivial.

iii) *E is the splitting field of a set of polynomials over k.*

iv) *If $\tau : E \to \overline{E}$ is a k-homomorphism then $\tau(E) = E$.*

v) *If $k \leqslant L_1 \leqslant E \leqslant L_2$ is a tower of fields and $\sigma : L_1 \to L_2$ is a k-homomorphism, then $\sigma(L_1) \subseteq E$ and we can find some $\tau \in \mathrm{Aut}(E/k)$ that extends σ.*

Proof. (i) $\Longrightarrow$ (ii) $\Longrightarrow$ (iii): It is immediate. Every minimal polynomial is irreducible and E being normal over k implies that E is the splitting field of $\{m(a,k) : a \in E\}$.

(iii) $\Longrightarrow$ (iv): If E is the splitting field of some $S \subseteq k[x]$ then so is $\tau(E)$ in $\overline{E}$ by the Extension Theorem for splitting fields. But then both E and $\tau(E)$ are generated over k by the same roots. Hence $E = \tau(E)$.

(iv) $\Longrightarrow$ (v): Suppose $k \leqslant L_1 \leqslant E \leqslant L_2$ is a tower of fields and $\sigma : L_1 \to L_2$ is a k-homomorphism. Since $k \leqslant E$ is algebraic, so is $k \leqslant L_1$ and, therefore, so is $k \leqslant \sigma(L_1)$ (any $a = \sigma(x) \in \sigma(L_1)$ is the root of $\widetilde{\sigma}[m(x,k)]$). If we take

$$k' = \{x \in L_2 : x \text{ is algebraic over } k\}$$

and its algebraic closure $\overline{k'}$ then E, being a subextension of $k \leqslant k'$, can be embedded in $\overline{k'}$ so $\overline{k'} = \overline{E}$ by the uniqueness of algebraic closures. By the Extension Theorem for algebraic closures, there is some $\rho : \overline{k'} \to \overline{k'}$ that extends σ, i.e. $\rho|_{L_1} = \sigma$. Take

$$\tau = \rho|_E : E \to \overline{k'} = \overline{E}.$$

By our hypothesis, $\tau(E) = E$ so $\sigma(L_1) = \rho|_{L_1}(L_1) = \rho|_E(L_1) = \tau(L_1) \subseteq \tau(E) = E$, τ extends σ, and $\tau|k = \sigma|k = \mathrm{id}_k$; that is, $\tau \in \mathrm{Aut}(E/k)$.

(v) $\Longrightarrow$ (i): Take an irreducible polynomial $p(x) \in k[x]$ and a root $a \in E$ of p. Then we have the tower

$$k \leqslant k(a) \leqslant E \leqslant \overline{E}$$

and we can find another root $b \in \overline{E}$ of $p(x)$ and a k-homomorphism $\sigma : k(a) \to \overline{E}$ such that $\sigma(a) = b$ (define $f(a) \overset{\sigma}{\mapsto} f(b)$ for all $f(a) \in k(a)$). By our hypothesis, $\sigma\big(k(a)\big) \subseteq E$ so $b \in E$ and p splits in E. $\diamond$

As a property of great interest, following 1.1.57, we also want to know how normality behaves under subextensions.

Proposition 1.3.12. *Let $k \leqslant L \leqslant E$ be a tower of fields. If $k \leqslant E$ is normal, then so is $L \leqslant E$.*

Proof. From 1.1.57, $L \leqslant E$ is algebraic. And if $m(a,k)$ splits in E then so does $m(a,L)$ since $m(a,L)|m(a,k)$ and $k[x]$ is a U.F.D. ◇

SEPARABLE EXTENSIONS We also saw that if $k \leqslant k(a)$ is a Galois extension then $m(a,k)$ has no multiple roots in $k(a)$. We will, as before, generalize this property to arbitrary extensions.

Definition 1.3.13. An irreducible polynomial $p(x) \in k[x]$ is called **separable over** k if it has no repeating roots in its splitting field. A polynomial $f(x) \in k[x]$ is **separable over** k if every one of its irreducible factors is separable. Otherwise, f is called **inseparable**.

Definition 1.3.14. An element $a \in E$ of an algebraic field extension $k \leqslant E$ is a **separable element over** k if $m(a,k)$ is separable over k. A **separable extension** is an *algebraic* field extension whose elements are all separable. Again an element that is not separable is called **inseparable** and an extension with inseparable elements is called **inseparable extension**.

Remark 1.3.15. An easy way to see if a polynomial is separable or not is by using the derivative criterion. For any polynomial

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0$$

we define its **formal derivative** the usual way, namely

$$f'(x) = na_n x^{n-1} + (n-1)a_{n-1}x^{n-2} + \ldots + a_1.$$

It is easy to see now that $f(x)$ *is separable if and only if* $(f, f') = 1$.

Example 1.3.16. Every extension over a field of characteristic zero is separable. Suppose $k \leqslant E$ is an extension with $\mathrm{char}(k) = 0$ and let $a \in E$. If $m = m(a,k)$ had a multiple root b then $m(b) = 0$ and $m'(b) = 0$. But m is also the minimal polynomial of b since it is irreducible; the relation $m'(b) = 0$ contradicts the minimality of ∂m. Therefore m is separable.

Example 1.3.17. The extension $\mathbb{R} \leqslant \mathbb{R}(i)$ of Ex. 1.2.10 is separable since $\mathrm{ch}(\mathbb{R}) = 0$.

Counterexample 1.3.18. The field extension $\mathbb{F}_2(t^2) \leqslant \mathbb{F}_2(t)$ of Ex. 1.2.13 is *inseparable*. As we saw, the minimal polynomial

$$m\big(t, \mathbb{F}_2(t^2)\big)(x) = x^2 - t^2 = (x-t)^2$$

has a double root. Observe that in this example, the characteristic is positive.

Example 1.3.19. Suppose k is a field and $\overline{k}$ is the algebraic closure of k. We define the **separable closure** k_{sep} of k to be the compositum of all simple separable extensions of k in $\overline{k}$. Obviously, k_{sep} is a separable extension of k.

It should come as no surprise that we want to study separability under subextensions.

Proposition 1.3.20. *For a tower of fields $k \leqslant L \leqslant E$, if $k \leqslant E$ is separable then so is $L \leqslant E$.*

Proof. From 1.1.57, $L \leqslant E$ is algebraic; if the roots of $m(a,k)$ are all simple then so are the roots of $m(a,L)$ since $m(a,L) | m(a,k)$. ◇

GALOIS EXTENSIONS III We are ready to give the last and most important definition of Galois extensions.

Using the new terminology we now have, we restate the result of Ex. 1.3.5: *a simple algebraic extension is Galois if and only if it is normal and separable*. It is not a surprise that this result also holds for arbitrary extensions.

First let's see how the transitivity results for algebraic, normal and separable extensions allow us to simplify the first definition of Galois extensions we gave.

Proposition 1.3.21. *Let $k \leqslant E$ be an algebraic extension. The following are equivalent*

(i) $k = \mathrm{Fix}_E\big(\mathrm{Aut}(E/k)\big)$.

(ii) The extension $k \leqslant E$ is both normal and separable.

Proof. (i) $\implies$ (ii): We will show that an arbitrary $a \in E$ is normal and separable. Take its minimal polynomial $m = m(a,k)$. If $\{a_1, \ldots, a_r\}$ are the roots of m in E, the set $S = \{\sigma a_i : \sigma \in \mathrm{Aut}(E/k)\}$ is finite and the polynomial

$$f(x) = \prod \big(x - \sigma(a_i)\big) \in E[x],$$

where the product is taken over the distinct elements of S, is fixed by any element of $\mathrm{Aut}(E/k)$. So the coefficients of f lie in $\mathrm{Fix}_E\big(\mathrm{Aut}(E/k)\big) = k$. We thus deduce that $m|f$ so m splits over k and has no multiple roots.

(ii) $\implies$ (i): If $k \leqslant E$ is separable, then E is contained in k_{sep}. In that case every $\sigma \in \mathrm{Aut}(k_{sep}/k)$ satisfies $\sigma(E) \subseteq E$. Indeed, the extension is normal hence every $a \in E$ is normal and $\sigma(a)$ is a root of $m(a,k)$ and thus $\sigma(E) \subseteq E$.

We will show that every element not in k is moved by some automorphism. Given any $a \in E \setminus k$ we can find an element $\sigma \in \mathrm{Aut}(k_{sep}/k)$ with $\sigma(a) \neq a$ since k_{sep} is normal. But if σ preserves E, the restriction $\sigma|_E \in \mathrm{Aut}(E/k)$ and $\sigma|_E(a) \neq a$. So $\mathrm{Fix}_E\big(\mathrm{Aut}(E/k)\big) = k$ ◇

Now the above proposition enables us to restate the *first* definition of a Galois extension without the universal quantifier.

Corollary 1.3.22. *An algebraic extension $k \leqslant E$ is Galois if and only if $k = \mathrm{Fix}_E\big(\mathrm{Aut}(E/k)\big)$.*

Proof. ($\Rightarrow$) Immediate from the definition. Take $L = k$.

($\Leftarrow$) If $k = \mathrm{Fix}_E\big(\mathrm{Aut}(E/k)\big)$ then the proposition tells us that $k \leqslant E$ is normal and separable. By the transitivity results, so is $L \leqslant E$ for every intermediate field L of $k \leqslant E$ and therefore, by the proposition again, $L = \mathrm{Fix}_E\big(\mathrm{Aut}(E/L)\big)$ for every intermediate field L.

◇

Definition 1.3.23 (3rd definition of Galois extensions)**.** A field extension $k \leqslant E$ is called **Galois** if

$$k = \mathrm{Fix}_E\big(\mathrm{Aut}(E/k)\big).$$

In this case we write $\mathrm{Gal}(E/k)$ for $\mathrm{Aut}(E/k)$ and call it the **Galois group** of the extension.

For finite extensions we can still use the second equivalent definition and again, as a consequence of the previous proposition, we can drop the quantifier as well.

Once more, using the proposition and the transitivity results we have the most important result of this section.

Corollary 1.3.24. *An algebraic extension $k \leqslant E$ is Galois if and only if it is normal and separable.*

Proof. Immediate from Prop. 1.3.21 and Cor. 1.3.22. ◇

Definition 1.3.25 (4th definition of Galois extensions)**.** A field extension $k \leqslant E$ is called **Galois** if it is both normal and separable. In this case we write $\mathrm{Gal}(E/k)$ for $\mathrm{Aut}(E/k)$ and call it the **Galois group** of the extension.

Example 1.3.26. The extension $\mathbb{R} \leqslant \mathbb{R}(i)$ of Ex. 1.2.10 is Galois.

Example 1.3.27. The separable closure k_{sep} of k is a Galois extension of k. By its definition, it is the maximal Galois extension of k in the sense that any other Galois extension of k is contained in k_{sep}.

Counterexample 1.3.28. The extension $\mathbb{Q} \leqslant \mathbb{Q}(\sqrt[3]{2})$ of Ex. 1.2.12 is *not* normal hence *not* Galois.

Counterexample 1.3.29. The extension $\mathbb{F}_2(t^2) \leqslant \mathbb{F}_2(t)$ of Ex. 1.2.13 is *not* separable hence *not* Galois.

In view of the last definition, we finish with the last transitivity result (essentially a rephrasing of 1.3.12 and 1.3.20).

Corollary 1.3.30. *If $k \leqslant L \leqslant E$ is a tower of fields and $k \leqslant E$ is Galois, then so is $L \leqslant E$.*

Convention. From now on, when we refer to an automorphism group as Galois group we will always mean that the extension is Galois.

1.4 THE FUNDAMENTAL THEOREM

We have reached the first important classification theorem, the Fundamental Theorem of Galois Theory for *finite* Galois extensions. We

will focus only on the part of the Fundamental Theorem that concerns the Galois correspondence since it is the *correspondence* we are mainly interested in.

Theorem 1.4.1 (Fundamental Theorem of Galois Theory for finite extensions)**.** *For a finite Galois extension $k \leqslant E$, the Galois-Artin correspondence*

$$\{L : \ k \leqslant L \leqslant E\} \rightleftarrows \{H : \ H \leqslant \mathrm{Gal}(E/k)\}$$

is bijective, i.e. the maps $\mathrm{Aut}(E/\bullet)$ *and* $\mathrm{Fix}_E(\bullet)$ *are mutual inverses, and order reversing.*

Proof. Immediate from Proposition 1.2.7, Corollary 1.2.18 and Corollary 1.3.24. ◇

Having established a bijective correspondence, the remaining part illustrates of the theorem how we can derive information about the extension using its Galois group and the Galois correspondence.

Theorem 1.4.2 (Fundamental Theorem of Galois Theory for finite extensions; part 2)**.** *If L is a subextension of a finite Galois extension $k \leqslant E$, then $L \leqslant E$ is Galois and*

$$[E : L] = |\mathrm{Gal}(E/L)| \quad \textit{and} \quad [L : k] = [\mathrm{Gal}(E/k) : \mathrm{Gal}(E/L)].$$

Moreover, the extension $k \leqslant L$ is normal if and only if $\mathrm{Gal}(E/L)$ *is a normal subgroup of* $\mathrm{Gal}(E/k)$*. In that case*

$$\mathrm{Gal}(L/k) \simeq \mathrm{Gal}(E/k)/\mathrm{Gal}(E/L).$$

Proof. $L \leqslant E$ is Galois from 1.3.30. Therefore, from the definition of finite Galois extensions, we have

$$[E : L] = |\mathrm{Gal}(E/L)|$$

and, using Prop. 1.1.11, we get

$$[L : F] = \frac{[E : F]}{[E : L]} = \frac{|\mathrm{Gal}(E/F)|}{|\mathrm{Gal}(E/L)|} = [\mathrm{Gal}(E/F) : \mathrm{Gal}(E/L)]$$

where the last equality is Lagrange's[18] theorem for finite groups.

[18] Joseph-Louis Lagrange (1736–1813)

For the second assertion it is not hard to show that

$$\sigma \operatorname{Gal}(E/L)\sigma^{-1} = \operatorname{Gal}\big(E/\sigma(L)\big) \tag{1.15}$$

for all $\sigma \in \operatorname{Gal}(E/F)$. Indeed, given $\tau \in \operatorname{Gal}(E/L)$ and $x \in L$,

$$(\sigma\tau\sigma^{-1})\big(\sigma(x)\big) = (\sigma\tau)(x) = \sigma(x) \quad \forall \sigma \in \operatorname{Gal}(E/F)$$

so $\sigma \operatorname{Gal}(E/L)\sigma^{-1} \subseteq \operatorname{Gal}\big(E/\sigma(L)\big)$. Similarly,

$$\sigma^{-1} \operatorname{Gal}\big(E/\sigma(L)\big)\sigma \subseteq \operatorname{Gal}\big(E/\sigma^{-1}\sigma(L)\big) = \operatorname{Gal}(E/L)$$

which proves (1.15).

Suppose now that the extension $F \leqslant L$ is normal. To show that $\operatorname{Gal}(E/L)$ is a normal subgroup of $\operatorname{Gal}(E/F)$, it suffices to show that $\sigma(L) = L$ for all $\sigma \in \operatorname{Gal}(E/F)$ since we would then have

$$\sigma \operatorname{Gal}(E/L)\sigma^{-1} \overset{(1.15)}{=} \operatorname{Gal}\big(E/\sigma(L)\big) = \operatorname{Gal}(E/L).$$

Given $\sigma \in \operatorname{Gal}(E/F)$ and $a \in L$, $\sigma(a)$ is a root of $m(a, F)$ which is in L since $F \leqslant L$ is normal. So $\sigma(L) \subseteq L$. But σ is injective and the extensions are all finite. Hence $\sigma(L) = L$.

For the contrary, suppose the extension $F \leqslant L$ is normal. In that case, the restriction map

$$\bullet|_L : \operatorname{Gal}(E/F) \to \operatorname{Gal}(L/F) : \sigma \mapsto \sigma|_L$$

is a well defined group homomorphism. The kernel of this homomorphism is

$$\ker = \{\sigma \in \operatorname{Gal}(E/F) : \sigma|_L = \mathrm{id}_L\} = \operatorname{Gal}(E/L).$$

So $\operatorname{Gal}(E/L) \lhd \operatorname{Gal}(E/F)$. Furthermore, the map $\bullet|_L$ is surjective by the Isomorphism Extension Theorem. Therefore, by the First Isomorphism Theorem for Groups,

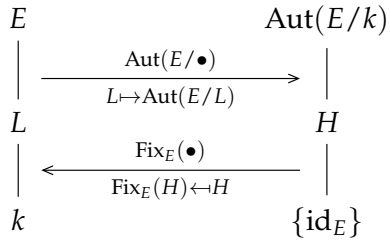
$$\operatorname{Gal}(L/F) \simeq \operatorname{Gal}(E/F)/\operatorname{Gal}(E/L).$$

$\diamond$

1.5 KRULL'S GALOIS THEORY

GALOIS CORESPONDENCE FOR INFINITE EXTENSIONS Let us now drop the finiteness assumption. Suppose $k \leqslant E$ is a possibly *infinite* but still *algebraic* field extension and consider the Galois correspondence

$$\{\text{subextensions } k \leqslant L \leqslant E\} \rightleftarrows \{\text{subgroups } H \leqslant \mathrm{Aut}(E/k)\}.$$



Proposition 1.2.7 did not assume any finiteness, so the correspondence is *order reversing* even for infinite extensions. Furthermore, the inclusions

$$H \subseteq \mathrm{Aut}\left(E/\,\mathrm{Fix}_E(H)\right)\ \forall H \leqslant \mathrm{Aut}(E/k) \tag{1.8}$$

and

$$L \subseteq \mathrm{Fix}_E\left(\,\mathrm{Aut}(E/L)\right)\ \forall L : k \leqslant L \leqslant E \tag{1.9}$$

are still valid since they are independent of the degree of the extension too.

Our aim once more is to examine for which extensions $k \leqslant E$, $\mathrm{Aut}(E/\bullet)$ and $\mathrm{Fix}_E(\bullet)$ are mutually inverse, i.e. for which extensions the relations

$$H = \mathrm{Aut}\left(E/\,\mathrm{Fix}_E(H)\right)\ \forall H \leqslant \mathrm{Aut}(E/k) \tag{1.6}$$

and

$$L = \mathrm{Fix}_E\left(\,\mathrm{Aut}(E/L)\right)\ \forall L : k \leqslant L \leqslant E \tag{1.7}$$

hold. We have already seen that (1.7) holds if and only if the extension is normal and separable and both normality and separability are defined independently of the degree of an extension. So we turn our focus to (1.6).

THE GALOIS GROUP OF AN INFINITE EXTENSION We saw in Cor. 1.2.18 that when $k \leqslant E$ is a finite extension, $\mathrm{Aut}(E/k)$ is also finite and (1.6) holds.

But when we drop the finiteness assumption, the theory breaks. First of all, $\mathrm{Aut}(E/k)$ is no longer finite. To prove this, we first need a variant of the *Primitive Element Theorem*. The proof we present is taken from [5].

Proposition 1.5.1 (The Primitive Element Theorem)**.** *If $k \leqslant E$ is a finite separable extension, then there is some element $\gamma \in E$ such that $E = k(\gamma)$.*

Such an element γ is called a **primitive element**, hence the name of the theorem.

Proof. Since $k \leqslant E$ is finite, there are $m \in \mathbb{N}$ and $a_1, \ldots, a_m \in E$ such that $E = k(a_1, \ldots, a_m)$.

If k is finite then so is E and it is easy to check that any generator γ of the cyclic group E^* will do.

If k is infinite we proceed with induction on m. Suppose $E = k(a, b)$ and that $m(a, k)$ and $m(b, k)$ have roots $a = a_1, \ldots, a_r$ and $b = b_1, \ldots, b_s$ in some extension of E (such extension can be constructed using Prop. 1.1.62 twice in a row). Using the separability of a and b we can prove that $\gamma = a + \lambda b$ is a primitive element for all $\lambda \in k$ except when

$$\lambda = \frac{a_i - a}{b - b_j}, \quad i = 1, \ldots, r, \ j = 1, \ldots, s$$

which are finitely many exceptions in an infinite field. The inductive step is now immediate. ◇

Proposition 1.5.2. *If $k \leqslant E$ is an infinite Galois extension then $\mathrm{Gal}(E/k)$ is also infinite.*

Proof. If $G = \mathrm{Gal}(E/k)$ were finite, say $|G| = n$ for some $n \in \mathbb{N}$, then $\sigma^n = \mathrm{id}_E$ for all $\sigma \in G$, which implies that

$$\sigma(a^n) = \sigma^n(a) = \mathrm{id}_E(a) = a \ \forall a \in E. \tag{1.16}$$

Therefore, if $a \in E$ then (using the fact that a is algebraic over k since $k \leqslant E$ is Galois)

$$[k(a) : k] = \partial m(a, k) \leqslant n$$

because if $\partial m(a,k) > n$, any $\sigma \in G$ acting on $m(a,k)(a) = 0$ yields a monic polynomial which is zero at a and whose degree is smaller than $\partial m(a,k)$ (any power of a greater than n becomes less than n from (1.16)) - a contradiction. But if $[k(a) : k] \leqslant n$ for all $a \in E$, then we can choose an element $a_0 \in E$ whose degree over k, $[k(a_0) : k] = m_0 \leqslant n$ is maximal among the degrees of elements of E. In that case we can prove that $E = k(a_0)$. Indeed, if not, there would be some $b_0 \in E \setminus k(a_0)$. Looking at the tower of fields

$$k \leqslant k(a_0) \leqslant k(a_0, b_0) \leqslant E$$

we conclude that $k \leqslant k(a_0, b_0)$ is separable (since $k \leqslant E$ is Galois, hence separable and using 1.3.20). Moreover, $[k(a_0) : k] = m_0$ maximal and finite, and since b_0 is algebraic over k, it is also algebraic over $k(a_0)$. That means, using 1.1.53, that $k(a_0) \leqslant k(a_0, b_0)$ is also of finite degree. From 1.1.11, we conclude that

$$[k(a_0, b_0) : k] = [k(a_0, b_0) : k(a_0)][k(a_0) : k] > m_0$$

But from the primitive element theorem, as $k \leqslant k(a_0, b_0)$ is separable and finite, there exists some $\gamma \in E$ such that $k(a_0, b_0) = k(\gamma)$. The above arguments imply that

$$[k(\gamma) : k] > m_0$$

which contradicts the maximality of m_0 among the degrees of elements of E. Therefore $E = k(a_0)$; under these circumstances, E is a finitely generrated, algebraic extension of k, hence finite over k - a contradiction. ◇

Our informal discussion in 1.2.11 should have prepared us to understand why infinite Galois groups do not behave well; they have too many subgroups so not every subgroup can arise as an automorphism group of some intermediate field.

Counterexample 1.5.3. Consider the extension

$$\mathbb{Q} \leqslant \mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{5}, \sqrt{7}, \sqrt{11}, \ldots) = E$$

constructed by adjoining the roots of all equations of the form $x^2 - p = 0$ where p is a prime number. It is easy to see that $\mathbb{Q} \leqslant E$

is normal (by definition) and separable (we are in characteristic 0), hence Galois. So (1.7) holds. It is in (1.6) where the correspondence breaks because $\mathrm{Aut}(E/\bullet)$ is *not* onto, i.e. not every subgroup can arise as an automorphism group. Let's see why.

We will focus on extensions of $\mathbb{Q}$ of degree 2 inside E, i.e. **quadratic number fields** inside E. It is not hard to see what a quadratic number field looks like in general. If $[L:\mathbb{Q}]=2$ then $L\neq\mathbb{Q}$ and we can find some $a\in L\setminus\mathbb{Q}$. From

$$[L:\mathbb{Q}(a)][\mathbb{Q}(a):\mathbb{Q}]=[L:\mathbb{Q}]=2,$$

we can deduce that $L=\mathbb{Q}(a)$. Therefore

$$2=[L:\mathbb{Q}]=[\mathbb{Q}(a):\mathbb{Q}]=\partial m(a,\mathbb{Q})$$

so $m(a,\mathbb{Q})=x^2+d$ or, equivalently, $a=\sqrt{d}$ for some square-free $d\in\mathbb{Q}$. The converse also holds; namely, any field of the form $\mathbb{Q}(\sqrt{d})$ with $d\in\mathbb{Q}$ square-free, is an algebraic number field. So there are countably many quadratic number fields, hence *countably many quadratic number fields inside E*.

By Theorem 1.4.2, any quadratic field L is mapped through $\mathrm{Aut}(E/\bullet)$ to a subgroup $\mathrm{Gal}(E/L)$ of index

$$[\mathrm{Gal}(E/\mathbb{Q}):\mathrm{Gal}(E/L)]=[L:\mathbb{Q}]=2$$

in $\mathrm{Gal}(E/\mathbb{Q})$ and only quadratic number fields can be mapped to subgroups of $\mathrm{Gal}(E/\mathbb{Q})$ of index 2.

But, any $\sigma\in G=\mathrm{Gal}(E/\mathbb{Q})$ is determined by its action on the square roots and, since $\sigma(\sqrt{p})$ is also a root of $m(\sqrt{p},\mathbb{Q})=x^2-p$, we have

$$\sigma(\sqrt{p})=\pm\sqrt{p}\ \ \forall \text{ prime } p.$$

We can therefore deduce that, as groups, $G\simeq\prod_{i=1}^{\infty}\mathbb{Z}_2$. Indeed the map

$$\begin{aligned}\Phi:\prod_{i=1}^{\infty}\mathbb{Z}_2&\to\mathrm{Gal}(E/\mathbb{Q})\\ (a_i)_{i=1}^{\infty}&\mapsto\sigma:\sigma(\sqrt{p_i})=(-1)^{a_i}\sqrt{p_i},\end{aligned}$$

where p_i is the i-th prime number, is easily seen to be a group isomorphism. So G is uncountable.[19] Moreover, G is an infinite dimensional vector space over $\mathbb{Z}_2$, which implies that its dual space $\mathrm{Hom}_{\mathbb{Z}_2}(G, \mathbb{Z}_2)$ is uncountable.[20] The kernels of all these (uncountably many) linear functionals are subgroups of G of index 2. So we have *uncountably many subgroups of G of index* 2.

Since there only countably many extensions of $\mathbb{Q}$ of degree 2 inside E but uncountably many subgroups of G of index 2, $\mathrm{Aut}(E/\bullet)$ cannot be onto and (1.6) cannot hold.

So when the extension is infinite, its Galois group is also infinite and the Galois correspondence may not be bijective as the previous example suggests. A natural question to ask is whether there are infinite Galois extensions for which the correspondence *is* bijective. We shall see that answer is *no*.

We therefore need to find a way to distinguish among the subgroups that *can* arise as Galois groups and the subgroups that *cannot*.

This is done by defining a topology on $\mathrm{Gal}(E/k)$, the *Krull topology*,[21] which distinguishes between "good" and "bad" subgroups. In particular, the subgroups that are *closed* with respect to the Krull topology will be exactly those which arise as Galois groups.

Convention. For the rest of this section, fix $k \leqslant E$ a (possibly infinite) Galois extension and $G = \mathrm{Gal}(E/k)$ its Galois group. We define

$$\mathcal{L} = \{L : \ k \leqslant L \leqslant E, \ [L : k] < \infty \text{ and } k \leqslant L \text{ Galois}\}$$

to be the set of subextensions L of $k \leqslant E$ for which $k \leqslant L$ is a finite Galois extension and

$$\mathcal{N} = \{N \leqslant G : \ N = \mathrm{Gal}(E/L), \ L \in \mathcal{L}\}$$

to be the set of subgroups of G that can arise as Galois groups of subextensions $L \in \mathcal{L}$.

[19] Another famous argument of Cantor says that the set of all binary sequences is uncountable (its cardinality equals the continuum); see [28].

[20] If V is an infinite dimensional vector space over a field F and V^* is its dual space, then $\dim_F V^* > \dim_F V$; see [2].

[21] Named after Wolfgang Krull (1899-1971) who was the first to develop the theory for infinite Galois extensions in [20].

Proposition 1.5.4. *The set*

$$\mathcal{B} = \{\sigma N : \ \sigma \in G, \ N \in \mathcal{N}\}$$

constitutes a basis for a topology $\mathcal{T}$ *on* $\mathrm{Gal}(E/k)$.

Proof. In other words, we need to show that the set

$$\mathcal{T} = \left\{ \bigcup_{i \in I} \sigma_i N_i : \ \sigma_i \in G, \ N \in \mathcal{N} \right\}$$

constitutes a topology on G. Obviously $\emptyset, G \in \mathcal{T}$. For the latter consider any arbitrary extension $L \in \mathcal{L}$ with Galois group $N = \mathrm{Gal}(E/L) \in \mathcal{N}$ and write

$$G = \bigcup_{\sigma \in G} \sigma N.$$

Moreover, it is apparent that the union of an arbitrary family of elements of $\mathcal{T}$ is again in $\mathcal{T}$. It remains to show that given a finite number of elements of $\mathcal{T}$, their intersection also lies in $\mathcal{T}$. It suffices to prove our claim for only two elements of $\mathcal{T}$. From the set-theoretic equality

$$\left(\bigcup_{i \in I} \sigma_i N_i \right) \cap \left(\bigcup_{j \in J} \tau_j N_j \right) = \bigcup_{i,j} \left(\sigma_i N_i \cap \tau_j N_j \right)$$

and the fact that $\mathcal{T}$ is closed under unions, it suffices to prove that $\sigma_1 N_1 \cap \sigma_2 N_2 \in \mathcal{T}$ for any two elements of $\mathcal{B}$. In fact we will prove that this intersection lies in $\mathcal{B} \subseteq \mathcal{T}$. Observe that if $\tau \in \sigma_1 N_1 \cap \sigma_2 N_2$ then

$$\sigma_1 N_1 \cap \sigma_2 N_2 = \tau N_1 \cap \tau N_2 = \tau (N_1 \cap N_2) \in \mathcal{B}$$

since $N_1 \cap N_2 \in \mathcal{N}$. Indeed, if

$$N_1 = \mathrm{Gal}(E/L_1) \text{ and } N_2 = \mathrm{Gal}(E/L_2), \ L_1, L_2 \in \mathcal{L}$$

then $\mathrm{Gal}(E/L_1 L_2) = N_1 \cap N_2$ since

$$\begin{aligned} \rho \in N_1 \cap N_2 &\iff \rho|_{L_1} = \mathrm{id} \text{ and } \rho|_{L_2} = \mathrm{id} \\ &\iff L_1 \subseteq \mathrm{Fix}_E(\rho) \text{ and } L_2 \subseteq \mathrm{Fix}_E(\rho) \\ &\iff L_1 L_2 \subseteq \mathrm{Fix}_E(\rho) \iff \rho \in \mathrm{Gal}(E/L_1 L_2) \end{aligned}$$

which completes the proof ($F \leqslant L_1 L_2$ is again a finite, Galois extension). ◇

Definition 1.5.5. The topology $\mathcal{T}$ defined in the previous proposition is called the **Krull topology**.

Example 1.5.6. The Krull topology of a finite Galois extension is the discrete topology; that is, every subgroup of its Galois groups is both open and closed.[22] Indeed, if $k \leqslant E$ is finite with Galois group G then $E \in \mathcal{L}$ and $\mathrm{Gal}(E/E) = \{\mathrm{id}_E\} \in \mathcal{N}$. Therefore

$$\{\sigma\} = \sigma \,\mathrm{Gal}(E/E) \in \mathcal{B} \quad \forall \sigma \in G.$$

That is, every singleton is open (and in particular basic); hence every subset of G is clopen.

The converse of the above claim also holds.

Lemma 1.5.7. *The Krull topology is discrete if and only if the extension is finite.*

Proof. See [3], Proposition 3.12.8. ◇

We will now prove our claim that the closed subsets of G in the Krull topology are exactly those that can arise as Galois groups of subextensions.

This result and the above lemma imply that *there are no infinite Galois extensions for which* (1.6) *holds*.

Proposition 1.5.8. *If $H \subseteq G$ then $\mathrm{Gal}(E/\operatorname{Fix}_E(H)) = \overline{H}$, the closure of H in the Krull topology.*

Proof. To simplify the notation, set $H' = \mathrm{Gal}(E/\operatorname{Fix}_E(H))$. In order to prove that $\overline{H} = H'$ we need to show two inclusions. For the inclusion $\overline{H} \subseteq H'$, we need only prove that H' is closed. Indeed, we have already established that $H \subseteq H'$ which implies that $\overline{H} \subseteq \overline{H'}$; therefore, if we prove that H' is closed then $H' = \overline{H'}$ and the inclusion $\overline{H} \subseteq \overline{H'}$ becomes $\overline{H} \subseteq H'$ which is exactly what we want. To prove that H' is closed take some $\sigma \in G \setminus H'$. Now

$$\sigma \in G \setminus H' \Rightarrow \exists a \in \operatorname{Fix}_E(H') : \ \sigma(a) \neq a$$

[22] A set that is both open and closed with respect to some topolgy is called **clopen**.

Take some $L \in \mathcal{L}$ such that $a \in L$ and set $N = \mathrm{Gal}(E/L) \in \mathcal{N}$. The set σN is a basic open set containing σ and is disjoint from H' since

$$\tau(a) = a \ \forall \tau \in N \text{ and } \sigma\tau(a) = \sigma(a) \neq a$$

Therefore, $G \setminus H'$ is open, hence H' is closed. For the other inclusion, $H' \subseteq \overline{H}$, set $L = \mathrm{Fix}_E(H)$ and let $\sigma \in H'$ and $N \in \mathcal{N}$. If $K = \mathrm{Fix}_E(N) \in \mathcal{L}$ and $H_0 = \{\tau|_K : \tau \in H\} \leqslant \mathrm{Gal}(K/k)$, then, since

$$\mathrm{Fix}_K(H_0) = \mathrm{Fix}_K(H) \cap K = L \cap K,$$

the Fundamental Theorem for the finite Galois extensions implies that $H_0 = \mathrm{Gal}(K/K \cap L)$. Now $\sigma \in H'$, so $\sigma|_L = \mathrm{id}_L$ and $\sigma|_K \in H_0$. Therefore, there is some $\tau \in H$ such that $\tau|_K = \sigma|_K$. Thus $\sigma^{-1}\tau \in \mathrm{Gal}(E/K) = N$, so $\tau \in \sigma N \cap H$. In other words, every basic open neighborhood σN of $\sigma \in H'$ meets H, so $\sigma \in \overline{H}$. ◇

From the above proposition, the Galois correspondence between

$$\{\text{subextensions } k \leqslant L \leqslant E\} \rightleftarrows \{\text{closed subgroups } H \leqslant \mathrm{Aut}(E/k)\}$$

is bijective. Hence we have proved the analogue of Theorem 1.4.1 for infinite extensions.

Theorem 1.5.9 (Fundamental Theorem of Galois Theory for infinite extensions; Krull)**.** *If $k \leqslant E$ is a (possibly infite) Galois extension, the correspondence*

$$\{L : \ k \leqslant L \leqslant E\} \rightleftarrows \{H : \ H \leqslant \mathrm{Gal}(E/k), \ H \text{ closed}\}$$

is bijective, i.e., the maps $\mathrm{Aut}(E/\bullet)$ *and* $\mathrm{Fix}_E(\bullet)$ *are mutual inverses, and order reversing.*

Proof. Immediate from Proposition 1.2.7, Corollary 1.2.18 and Proposition 1.5.8. ◇

We can also establish an analogue of Theorem 1.4.2.

Theorem 1.5.10 (Fundamental Theorem of Galois Theory for infinite extensions, Krull; part 2)**.** *If L is a subextension of $k \leqslant E$, then $L \leqslant E$ is Galois and if $H = \mathrm{Gal}(E/L)$ then*

$$|G : H| < \infty \iff H \text{ is open} \iff [L : k] < \infty$$

and in that case, $|G : H| = [L : k]$. *Moreover, the extension* $k \leqslant L$ *is normal (hence Galois) if and only if* $\mathrm{Gal}(L/k)$ *is a normal subgroup of* $\mathrm{Gal}(E/k)$. *In that case*

$$\mathrm{Gal}(L/k) \simeq \mathrm{Gal}(E/k)/\mathrm{Gal}(E/L)$$

If we endow $\mathrm{Gal}(E/k)/\mathrm{Gal}(E/L)$ *with the quotient topology then the above isomorphism is a homeomorphism.*

We must note however, that in order to prove the second part of the Fundamental Theorem we need more information on the Krull topology. In particular we need

Proposition 1.5.11. *The set* G *endowed with the Krull topology is Hausdorff, compact and totally disconnected.*

The proofs of these propositions are given in the next paragraph. A last remark is in order before we proceed.

Remark 1.5.12. Krull's Galois Theory is a generalization of the Galois Theory for finite extensions. If $k \leqslant E$ is a finite Galois extension then the Krull topology on $\mathrm{Gal}(E/k)$ is the discrete topology; hence every subgroup is clopen and we retrieve Theorems 1.4.1 and 1.4.2.

PROFINITE TOPOLOGICAL GROUPS The previous discussion of infinite Galois Theory is somewhat elementary and probably raises more questions than it answers. One might wonder for example how did we come up with the Krull topology or even why did we use topology to begin with.

To see how our work in infinite Galois Theory comes naturally, we need the notion of a *profinite topological group*. For a more elaborate study of profinite groups and how they are related to Galois Theory we refer the reader to [16].

Definition 1.5.13. A **topological group** is a group G endowed with a topology such that the maps

$$\cdot : G \times G \to G : (g,h) \mapsto gh \quad \text{and} \quad {}^{-1} : G \to G : g \mapsto g^{-1}$$

are continuous. A **homomorphism of topological groups** is a *group homomorphism* that is also *continuous*.

Thus a topological group is a group that is also a topological space for which the group structure and the topology are compatible in the sense described above and a homomorphism of topological groups is a map that respects both the group structure and the topology.

Example 1.5.14. If $k \leqslant E$ is a Galois extension, $\mathrm{Gal}(E/k)$ endowed with the Krull topology is a topological group.

If the extension is finite, so is $\mathrm{Gal}(E/k)$ and therefore its Krull topology is the discrete topology. Hence the maps $(g,h) \mapsto gh$ and $g \mapsto g^{-1}$ are (trivially) continuous.

The interesting case is when $k \leqslant E$ is infinite. Suppose $(g,h) \in G \times G$ and take a basic open neighborhood $gh\,\mathrm{Gal}(E/L)$ of gh. Then, the basic open neighborhood $g\,\mathrm{Gal}(E/L) \times h\,\mathrm{Gal}(E/L)$ of (g,h) is contained in $gh\,\mathrm{Gal}(E/L)$ through the map $(g,h) \mapsto gh$. The element (g,h) was arbitrary, therefore $(g,h) \mapsto gh$ is continuous.

Similarly, if $g \in G$ and $g^{-1}\,\mathrm{Gal}(E/L)$ is a basic open neighborhood of g^{-1}, then the open neighborhood $g\,\mathrm{Gal}(E/L)$ of g is contained in $g^{-1}\,\mathrm{Gal}(E/L)$ through the map $g \mapsto g^{-1}$. The element g was arbitrary, therefore $g \mapsto g^{-1}$ is also continuous.

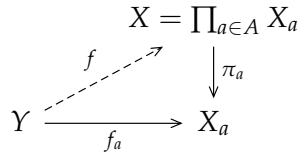
Proposition 1.5.15. *A subgroup of topological group is a topological group.*

Proof. Let G be a topological group and $H \leqslant G$. Recall from General Topology that *the restriction of a continuous map is continuous*. Therefore the restrictions $\cdot|_H : H \times H \to G$ and ${}^{-1}|_H : H \to G$ are continuous. But since H is a subgroup of G, their range is H, i.e. $\cdot|_H : H \times H \to H$ and ${}^{-1}|_H : H \to H$ so H is a topological group. $\diamond$

Proposition 1.5.16. *The product $G = \prod_{a\in A} G_a$ of a family $\{G_a\}_{a\in A}$ of topological groups is a topological group when endowed with the product topology.*

Before proving that, recall from General Topology that if $\{X_a\}_{a\in A}$ is a family of topological spaces, then $X = \prod_{a\in A} X_a$ endowed with the product topology is a topological space. X (with projections $\pi_a : X \to X_a$) has the universal property that for every other topological space Y and every family $f_a : Y \to X_a$ of continuous maps, there is

a unique continuous map $f : Y \to X$ so that the following diagram commutes.



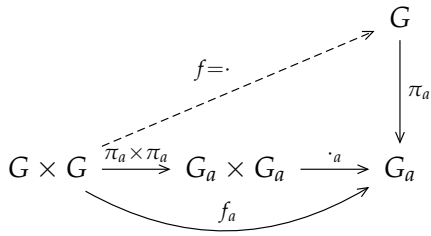
Proof. First of all, recall from Group Theory that $G = \prod_{a \in A} G_a$ is a group with multiplication defined componentwise, i.e.

$$(g_a)_{a \in A} \cdot (h_a)_{a \in A} = (g_a \cdot h_a)_{a \in A}.$$

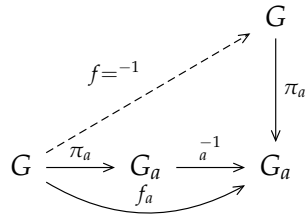
To show that $\cdot : G \times G \to G$ is continuous take $Y = G \times G$ and f_a to be the composition

$$G \times G \xrightarrow{\pi_a \times \pi_a} G_a \times G_a \xrightarrow{\cdot_a} G_a$$

in the universal property of the product topology. Then $\cdot$ is the unique map f, hence continuous.



Similarly, to prove that $^{-1} : G \to G$ is continuous take $Y = G$ and f_a to be the composition $G \xrightarrow{\pi_a} G_a \xrightarrow{^{-1}_a} G_a$.



◇

We are not interested in topological groups in general but rather in a spacial kind of topological groups, profinite groups.

Definition 1.5.17. An **inverse system of topological groups** is a pair of families $(\{G_a\}_{a\in A}, \{\phi_b^a\}_{a\leqslant b\in A})$ indexed by some *directed set*[23] $(A, \leqslant)$, where $\{G_a\}_{a\in A}$ is a family of topological groups and for every $a \leqslant b \in A$, ϕ_b^a is a continuous group homomorphism $G_b \to G_a$ such that

i) $\phi_a^a = \mathrm{id}_{G_a}$

ii) $\phi_c^a = \phi_b^a \circ \phi_c^b$ for all $a \leqslant b \leqslant c \in A$.

We shall write (G_a, ϕ_b^a) for an inverse system of topological groups.

Definition 1.5.18. Suppose (G_a, ϕ_b^a) is an inverse system of topological groups. The **inverse limit** of the system is the subset of the product $\prod_{a\in A} G_a$ consisting of sequences (g_a) such that $\phi_b^a(g_b) = g_a$ for all $a \leqslant b$. The inverse limit is usually denoted by $\varprojlim G_a$.

Lemma 1.5.19. *If (G_a, ϕ_b^a) is an inverse system of topological groups, then the inverse limit $\varprojlim G_a$ is a subgroup of $\prod_{a\in A} G_a$.*

Proof. First of all, $\varprojlim G_a \neq \emptyset$. Indeed, $(e_a)_{a\in A} \in \varprojlim G_a$ since $\phi_b^a : G_b \to G_a$ is a group homomorphism and therefore sends $e_b \mapsto e_a$ for every $a \leqslant b$.

Furthermore, for every $(g_a), (h_a) \in \varprojlim G_a$,

$$(g_a) \cdot (h_a)^{-1} = (g_a) \cdot (h_a^{-1}) = (g_a \cdot h_a^{-1}) \in \varprojlim G_a$$

because for every $a \leqslant b$,

$$\phi_b^a(g_b \cdot h_b^{-1}) = \phi_b^a(g_b) \cdot \phi_b^a(h_b^{-1}) = \phi_b^a(g_b) \cdot \phi_b^a(h_b)^{-1} = g_a h_a^{-1}.$$

Therefore $\varprojlim G_a \leqslant \prod_{a\in A} G_a$. ◇

Corollary 1.5.20. *The inverse limit $\varprojlim G_a$ is a topological group.*

Proof. Immediate from Propositions 1.5.15 and 1.5.16, and the previous lemma. ◇

[23] Recall that a **directed set** is a partially ordered set $(A, \leqslant)$ with the property that for every $a, b \in A$ there is some $c \in A$ such that $a \leqslant c$ and $b \leqslant c$.

Profinite topological groups are a special kind of inverse limit.

Definition 1.5.21. The inverse limit of a system of *finite* topological groups (endowed with the *discrete* topology) is called a **profinite** (**topological**) **group**.

Corollary 1.5.22. *Profinite groups are topological groups.*

Example 1.5.23. Every finite topological group G is profinite. Just take (G_a, ϕ_b^a) to be (G, id_G).

Example 1.5.24. The Galois group $\mathrm{Gal}(E/k)$ of a Galois extension $k \leqslant E$ is a profinite group.

If the extension is finite then so is $\mathrm{Gal}(E/k)$ and the assertion follows from the previous example.

Suppose the extension is infinite. The pair $(\mathcal{L}, \subseteq)$ is a directed set. Indeed, for every $L_1, L_2 \in \mathcal{L}$, their compositum $L_3 = L_1 L_2 \in \mathcal{L}$ (for $i = 1, 2$, L_i is a finite Galois extension of k, so it is generated by a finite set S_i of elements whose minimal polynomials are separable and split over k; L_3 is then generated by $S_1 \cup S_2$ and therefore $L_3 \in \mathcal{L}$) and $L_1 \subseteq L_3$, $L_2 \subseteq L_3$.

For every $L \in \mathcal{L}$ take the finite group $\mathrm{Gal}(L/k)$ endowed with the discrete topology and for every $L_1 \subseteq L_2 \in \mathcal{L}$ consider the restriction homomorphism

$$\phi_{L_2}^{L_1} : \mathrm{Gal}(L_2/k) \to \mathrm{Gal}(L_1/k) : \sigma \mapsto \sigma|_{L_1} \quad \forall \sigma \in \mathrm{Gal}(L_2/k).$$

Thus we have an inverse system $(\mathrm{Gal}(L/k), \phi_{L_2}^{L_1})$ of topological groups. We will show that

$$\mathrm{Gal}(E/k) \cong \varprojlim \mathrm{Gal}(L/k),$$

that is, they are isomorphic as groups and homeomorphic as topological spaces.

Define

$$\begin{aligned} \vartheta : \mathrm{Gal}(E/k) &\to \varprojlim \mathrm{Gal}(L/k) \\ \sigma &\mapsto (\sigma|_L)_{L \in \mathcal{L}}. \end{aligned}$$

The map ϑ is obviously well defined.

ϑ is also a group homomorphism since

$$\sigma|_L \tau|_L = (\sigma\tau)|_L.$$

We will show that the kernel of ϑ is trivial so ϑ is injective. Suppose $\sigma \in \mathrm{Gal}(E/k)$ such that $\sigma|_L = \mathrm{id}_L$ for all $L \in \mathcal{L}$. Then $\sigma = \mathrm{id}_E$. Indeed, for every $x \in E$ take the splitting field L_x of $m(x,k)$. It is immediate that $L_x \in \mathcal{L}$. The hypothesis $\sigma|_{L_x} = \mathrm{id}_{L_x}$ implies that $\sigma(x) = \sigma|_{L_x}(x) = x$. The element $x \in E$ was arbitrary, therefore $\sigma = \mathrm{id}_E$.

To show that ϑ is surjective take some $(\tau_L) \in \varprojlim \mathrm{Gal}(L/k)$ and define

$$\sigma : E \to E : \sigma(x) = \tau_{L_x}(x) \ \ \forall x \in E.$$

It is a trivial procedure to show that $\sigma \in \mathrm{Gal}(E/k)$ and, by its definition, $\vartheta(\sigma) = (\sigma|_L) = (\tau_L)$.

ϑ is continuous and open. The basic open sets of $\mathrm{Gal}(E/k)$ are of the form σN where $N = \mathrm{Gal}(E/L) \in \mathcal{N}$ for some $L \in \mathcal{L}$ and $\sigma \in \mathrm{Gal}(E/k)$. The topology of $\varprojlim \mathrm{Gal}(L/k)$ is, by definition of the product and the subspace topology, the smallest topology that contains the sets $\pi_L^{-1}(\{\tau\})$ where $L \in \mathcal{L}$, $\pi_L : \mathrm{Gal}(E/k) \to \mathrm{Gal}(L/k)$ is the usual projection (restriction) and $\tau \in \mathrm{Gal}(L/k)$. We compute

$$\begin{aligned}
\vartheta^{-1}\big(\pi_L^{-1}(\{\tau\})\big) &= \{\sigma \in \mathrm{Gal}(E/k) : \sigma|_L = \tau\} \\
&= \{\sigma \in \mathrm{Gal}(E/k) : \sigma \text{ extends } \tau \text{ to } E\} \\
&= \bigcup_{\sigma \in \mathrm{Gal}(E/k)} \sigma\,\mathrm{Gal}(E/L)
\end{aligned}$$

which is a union of open sets in $\mathrm{Gal}(E/k)$, hence open. So ϑ is continuous. Moreover,

$$\begin{aligned}
\vartheta(\sigma N) &= \{(\sigma\tau_H)_{H\in\mathcal{L}} : \tau_H|_L = \mathrm{id}_{H\cap L}\} \\
&= \{(\tau_H)_{H\in\mathcal{L}} : \sigma^{-1}\tau_H|_L = \mathrm{id}_{H\cap L}\} \\
&= \{(\tau_H)_{H\in\mathcal{L}} : \tau_H|_L = \sigma|_{H\cap E}\} \\
&= \pi_H^{-1}(\{\sigma|_E\})
\end{aligned}$$

so the image of a basic open set is open in $\varprojlim \mathrm{Gal}(L/k)$. Therefore ϑ is open.

The Krull topology has some nice properties which can now be almost immediately derived.

Recall from General Topology that a space X is called **compact** if every open cover of X has a finite subcover, **Hausdorff**[24] if for every $x, y \in X$ there are open sets V_x, V_y such that $x \in V_x$, $y \in V_y$ and $V_x \cap V_y = \emptyset$, and **totally disconnected** if its only connected components are the singletons. It is immediate that *every finite space equipped with the discrete topology is compact, Hausdorff and totally disconnected*. The not (at all) obvious results we will need are

i) *Any product of compact spaces is compact.*[25]

ii) *Any product of Hausdorff spaces is Hausdorff.*

iii) *Any product of totally disconnected spaces is totally disconnected.*

iv) The subspace of a compact space need not be compact; for example $[0,1]$ is compact while $(0,1)$ is not. However, *a closed subspace of a compact space is compact.*

v) *The subspace of a Hausdorff space is Hausdorff.*

vi) *The subspace of a totally disconnected space is totally disconnected.*

vii) *If $f, g : X \to Y$ are continuous maps and Y is Hausdorff, then the set*

$$\{x \in X : f(x) = g(x)\}$$

is a closed subset of X.

Lemma 1.5.25. *$\varprojlim G_a$ is a closed subset of $\prod_{a \in A} G_a$.*

Proof. Write $\varprojlim G_a$ as

$$\begin{aligned}\varprojlim G_a &= \{(g_a) \in \prod_{a \in A} G_a : \phi_b^a(g_b) = g_a \ \forall b > a\} \\ &= \bigcap_{b>a} \{(g_a) \in \prod_{a \in A} G_a : \phi_b^a \circ \pi_b(g_b) = \pi_a(g_a)\}\end{aligned}$$

and apply (vii) for $f = \phi_b^a \circ \pi_b$ and $g = \pi_a$ as continuous functions $\prod_{a \in A} G_a \to G_a$. $\varprojlim G_a$ is then closed as the intersection of closed subsets of $\prod_{a \in A} G_a$. ◇

[24] Named after Felix Hausdorff (1868–1942).

[25] This is Tychonoff's theorem, named after Andrey Nikolayevich Tikhonov (1906-1993).

Corollary 1.5.26. *The Galois group of an extension $k \leqslant E$ endowed with the Krull topology is a compact, Hausdorff and totally disconnected space.*

We can now prove the second part of the Fundamental Theorem of infinite Galois Theory.

Theorem 1.5.27 (Fundamental Theorem of Galois Theory for infinite extensions, Krull; part 2)**.** *If L is a subextension of $k \leqslant E$, then $L \leqslant E$ is Galois and if $H = \mathrm{Gal}(E/L)$ then*

$$|G : H| < \infty \iff H \text{ is open} \iff [L : k] < \infty$$

and in that case, $|G : H| = [L : k]$. Moreover, the extension $k \leqslant L$ is normal (hence Galois) if and only if $\mathrm{Gal}(L/k)$ is a normal subgroup of $\mathrm{Gal}(E/k)$. In that case

$$\mathrm{Gal}(L/k) \simeq \mathrm{Gal}(E/k)/\mathrm{Gal}(E/L)$$

If we endow $\mathrm{Gal}(E/k)/\mathrm{Gal}(E/L)$ with the quotient topology then the above isomorphism is a homeomorphism.

Proof. Suppose $[G : H] = m < \infty$. If the left cosets of H in G are $\{H, g_1H, \dots, g_mH\}$ then

$$G \setminus H = \bigsqcup_{i=1}^{m} g_iH$$

which is a finite union of closed subsets of G. Indeed, $H = \mathrm{Gal}(E/L)$ is closed as the Galois group of some subextension and the map

$$\cdot|_{\{g_i\}\times H} : \{g_i\} \times H \to gH \subseteq G : h \mapsto g_ih$$

is bijective and continuous as the restriction of a continuous map. Therefore g_iH is closed for every $i = 1, \dots, m$. Thus $G \setminus H$ is closed which implies that H is open.

Suppose now that H is an open subgroup of G. Then $\mathrm{id}_E \in H$ so there is some basic open neighborhood $N = \mathrm{Gal}(E/S) \in \mathcal{N}$, $S \in \mathcal{L}$, so that $\mathrm{id}_E \in N \leqslant H$. In that case,

$$\begin{aligned} N \leqslant H &\implies \mathrm{Fix}_E(H) \leqslant \mathrm{Fix}_E(N) \\ &\implies \underbrace{\mathrm{Fix}_E\left(\mathrm{Gal}(E/L)\right)}_{=L \text{ since } E/L \text{ is Galois}} \leqslant \underbrace{\mathrm{Fix}_E\left(\mathrm{Gal}(E/S)\right)}_{=S \text{ since } E/S \text{ is Galois}} \\ &\implies L \leqslant S \end{aligned}$$

and since $S \in \mathcal{N}$, that is, S is finite over k, so is L.

Lastly, suppose that $[L : k] < \infty$. Take E_f to be the splitting field of all the minimal polynomials of the generators of L over k. Then $E_f \in \mathcal{L}$ and $k \leqslant L \leqslant E_f \leqslant E$. Set $N = \mathrm{Gal}(E/E_f) \in \mathcal{N}$. Similarly to the proof of the fundamental theorem for finite extensions, the map

$$\vartheta_{E_f} : \mathrm{Gal}(E/k) \to \mathrm{Gal}(E_f/k) : \sigma \mapsto \sigma|_{E_f}$$

is a surjective (from Lemma 1.3.11) group homomorphism with $\ker \vartheta_{E_f} = N$. From the 1st Isomorphism Theorem we have

$$\mathrm{Gal}(E_f/k) \simeq \mathrm{Gal}(E/k)/\mathrm{Gal}(E/E_f) = G/N.$$

The inclusion $L \leqslant E_f$ now implies that $N = \mathrm{Gal}(E/E_f) \leqslant \mathrm{Gal}(E/L) = H$ and therefore

$$[G : H] \leqslant [G : N] = |G/N| = |\mathrm{Gal}(E_f/k)|$$

and $|\mathrm{Gal}(E_f/k)| < \infty$ since $k \leqslant E_f$ is finite. In particular, using the Isomorphism Theorems for Groups, Lagrange's theorem, the fundamental theorem of Galois theory for finite extensions and the multiplicativity of the degrees, we get

$$[G : H] = [G/N : H/N] = \frac{|G/N|}{|H/N|} = \frac{[E_f : k]}{[E_f : L]} = [L : k].$$

Suppose $H \lhd G$. We will show that $k \leqslant L$ is a Galois extension. It is obviously separable by the transitivity of separable extensions. It remains to show that it is normal. Let $a \in L \setminus k$ and $m = m(a, k)$ be its minimal polynomial. $k \leqslant E$ is Galois; let $b \in E$ be another root of m. We will show that $b \in L$. From the Isomorphism Extension Theorem, there is some $\sigma \in \mathrm{Gal}(E/k)$ such that $\sigma(a) = b$. Since $H \lhd G$, we have $\sigma^{-1}\tau\sigma \in H$ for every $\tau \in H$. Therefore

$$\tau(b) = \tau\sigma(a) = \sigma\ \underbrace{\sigma^{-1}\tau\sigma}_{\in \mathrm{Gal}(E/L)}\ (\underbrace{a}_{\in L}) = \sigma(a) = b.$$

In other words, $b \in \mathrm{Fix}_E(H) = \mathrm{Fix}_E\left(\mathrm{Gal}(E/L)\right) = L$.

For the contrary, suppose that $k \leqslant L$ is Galois. Then

$$\vartheta : G = \mathrm{Gal}(E/k) \to \mathrm{Gal}(L/k) : \sigma \mapsto \sigma|_L$$

is a surjective group homomorphism with $\ker \vartheta_L = H \lhd G$ (using analogous arguments as before, when we defined ϑ_{E_f}).

By the 1st Isomorphism Theorem, there is an isomorphism

$$v : G/H = \mathrm{Gal}(E/k)/\mathrm{Gal}(E/L) \xrightarrow{\simeq} \mathrm{Gal}(L/k).$$

We will show that v is a homeomorphism.

The basic open sets of $\mathrm{Gal}(L/k)$ as a subspace of G are

$$\mathcal{B}' = \{\tau\,\mathrm{Gal}(L/K)\}$$

where $k \leqslant K \leqslant L$ with $k \leqslant K$ being a finite Galois extension and $\tau \in \mathrm{Gal}(L/k)$. For every such basic open set, $\mathrm{Gal}(E/K) \in \mathcal{N}$ since $E \in \mathcal{L}$ and therefore

$$\vartheta^{-1}\big(\tau\,\mathrm{Gal}(L/K)\big) = \sigma\,\mathrm{Gal}(E/K)$$

for some $\sigma \in \mathrm{Gal}(E/k)$ that extends τ. So ϑ is continuous.

Moreover, ϑ is closed. Indeed, G is compact and $\mathrm{Gal}(L/k)$ is Hausdorff (because G is), so any closed subset of the compact G is compact and is therefore mapped through the continuous ϑ to a compact subset of the Hausdorff space $\mathrm{Gal}(L/k)$. Recall from General Topology that *any compact subset of a Hausdorff space is compact* and the assertion follows.

Now it follows at once that the isomorphism

$$v : G/H = \mathrm{Gal}(E/k)/\mathrm{Gal}(E/L) \xrightarrow{\simeq} \mathrm{Gal}(L/k)$$

induced by ϑ is a homeomorphism by the definition of the quotient topology on G/H. ◇

[...] geometry is the art of reasoning well from badly drawn figures; however, these figures, if they are not to deceive us, must satisfy certain conditions; the proportions may be grossly altered, but the relative positions of the different parts must not be upset.

Jules Henri Poincaré (1854-1912).

2 THE CLASSIFICATION OF COVERING SPACES

Similar to that of Galois Theory, the next classification theorem we are pursuing focuses on the interplay between two fundamental notions of Algebraic Topology: *covering spaces* and *fundamental groups*. Covering spaces' first appearance was in the works of Georg Friedrich Bernhard Riemann (1826-1866), in the theory of analytic functions of a complex variable. Fundamental groups on the other hand, which were introduced by Poincaré in his monumental paper *"Analysis Citus"* [31] (the starting point of Algebraic Topology), were initially used as a tool to classify topological spaces.

Despite their different origins, the two concepts share a deep connection. One of the aspects of this connection is the main theme of this chapter: *for certain topological spaces, we can determine and classify up to isomorphism all of their possible covering spaces using subgroups of their fundamental group*. Sounds familiar? It should!

Our aim again is not a thorough investigation but a quick development of some basic notions that will help us reach the next classification theorem. We therefore assume the same familiarity with Abstract Algebra and General Topology as the previous chapter. Some elementary notions regarding parameterizations of curves are also assumed to be known.

2.1 THE FUNDAMENTAL GROUP

INTRODUCTION The theory of the fundamental group provides a way of reducing difficult topological problems into simpler, algebraic ones. This is done by associating each topological space with an appropriately chosen group, the *fundamental group* of the space. Then, certain problems concerning the space can be translated into

problems concerning the fundamental group which are in some cases more tractable.

Fundamental groups can also be used to study continuous functions. Every continuous function between two topological spaces induces a group homomorphism between their respective fundamental groups. Again, studying the induced homomorphism can tell us a lot about the function.

From this perspective, fundamental groups constitute powerful tools in the study of topological spaces. Unfortunately, we will not see them in action. This section contains only some basic definitions and calculations of fundamental groups in a few, very simple cases; just the absolute essentials for what follows.

Convention. For the rest of this chapter we denote the closed unit interval $[0,1] \subseteq \mathbb{R}$ as I. To avoid repetition, we will occasionally refer to *topological spaces* simply as *spaces* and we will interchange among the terms *function*, *map* and *mapping*. To save words, all subsets of $\mathbb{R}^n$ are assumed to have the usual topology induced by the Euclidean metric for all $n \in \mathbb{N}$.

Before continuing, we recall a result from Topology which we will soon have to invoke.

Lemma 2.1.1 (Glueing Lemma)**.** *Suppose that X and Y are topological spaces such that $X = A \cup B$ for some closed subsets $A, B \subseteq X$. If $f : A \to Y$ and $g : B \to Y$ are continuous functions such that $f(a) = g(a)$ for all $a \in A \cap B$, then the mapping*

$$h : X \to Y : \; h(x) = \begin{cases} f(x), & \text{if } x \in A \\ g(x), & \text{if } x \in B \end{cases}$$

is also continuous.

Proof. Let C be a *closed* subset of Y. Since

$$h^{-1}(C) = f^{-1}(C) \cup g^{-1}(C),$$

and f, g are continuous, $f^{-1}(C) \cup g^{-1}(C)$ is also closed in X and so is $h^{-1}(C)$. Therefore, h is continuous. ◇

HOMOTOPY CLASSES OF LOOPS Let X be an arbitrary topological space. A **path** in X is a *continuous* mapping $f : I \to X$. The point $f(0)$ is called the **initial** point of the path and $f(1)$ its **terminal**; together they are the **endpoints** of the path. If the endpoints of a path coincide, that is $f(0) = f(1) = x_0$ for some $x_0 \in X$, then f is called a **closed path** or a **loop** (**based**) **at** x_0.

If $f : I \to X$ is a path, its image $f(I)$ is a (**topological**) **curve** in X which we will denote by C_f. We must be careful not to confuse a path $f : I \to X$ with the curve C_f. A path is a *parameterization* of the curve C_f, while C_f is just a set of points in X.

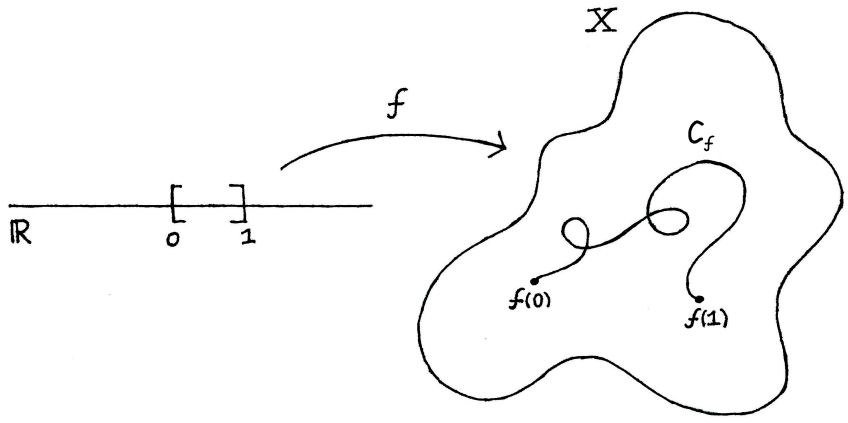


Figure 2.1: A path is a parameterization of the curve C_f

Example 2.1.2. Let X be a topological space and $x_0 \in X$. The **constant path** or **constant loop at** x_0 is defined to be

$$c_{x_0} : I \to X : c_{x_0}(s) = x_0 \ \forall s \in I.$$

Example 2.1.3. Let $f : I \to X$ be a path. We define its **inverse path**, denoted by f^{-1}, to be

$$f^{-1} : I \to X : f^{-1}(s) = f(1 - s).$$

Note that f and f^{-1} parameterize the *same* curve $C_f = C_{f^{-1}}$ in X yet they are *different* parameterizations of C_f hence *distinct* paths; f^{-1} traverses C_f in the opposite direction of f, from $f(1)$ to $f(0)$.

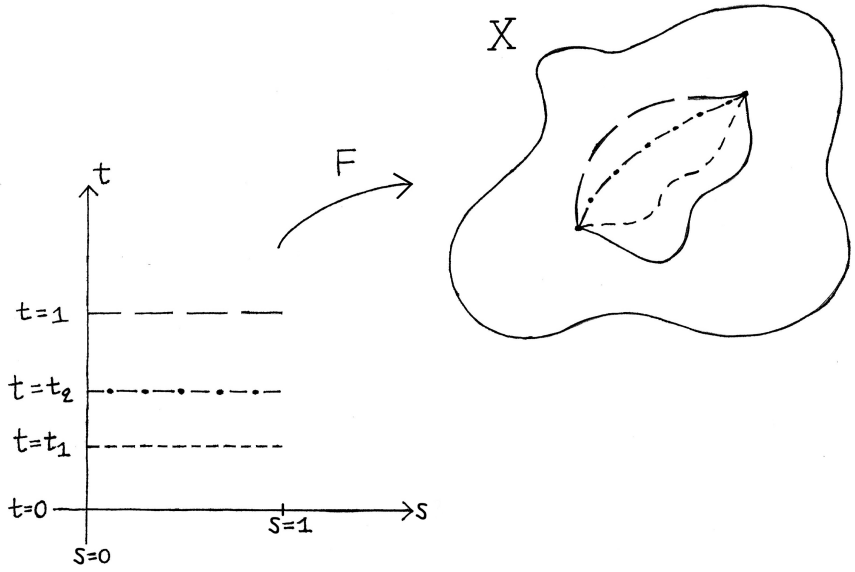


Figure 2.2: Homotopy as a deformation of paths

Definition 2.1.4. A **homotopy** of two paths $f_0, f_1 : I \to X$ whose *endpoints coincide* is a *continuous* map

$$F : I \times I \to X$$

such that

$$\begin{cases} F(s,0) = f_0(s) & \forall s \in I \\ F(s,1) = f_1(s) & \forall s \in I \\ F(0,t) = f_0(0) = f_1(0) & \forall t \in I \\ F(1,t) = f_0(1) = f_1(1) & \forall t \in I. \end{cases} \tag{2.1}$$

In particular, a homotopy between two *loops* $f_0, f_1 : I \to X$ based at $x_0 \in X$ is a continuous function $F : I \times I \to X$ such that

$$\begin{cases} F(s,0) = f_0(s) & \forall s \in I \\ F(s,1) = f_1(s) & \forall s \in I \\ F(0,t) = F(1,t) = x_0 & \forall t \in I. \end{cases} \tag{2.2}$$

If a homotopy exists, the paths are called **homotopic**. We use the notation $f_0 \sim f_1$ to indicate that f_0 and f_1 are homotopic, or $F : f_0 \sim f_1$ if we want to refer to the homotopy F as well.

We can think of the second coordinate t as representing time and the homotopy as a continuous (with respect both to s and t) deformation of path f_0 to path f_1. The first two conditions say that at $t = 0$ we begin with f_0 and, continuously deforming it, we obtain f_1 at $t = 1$; at any given time $t_0 \in I$, f_0 has been deformed to $f_{t_0} \equiv F(s, t_0) : I \to X$. The last two conditions ensure that, during the deformation, the endpoints remain fixed.

Example 2.1.5. Let X be $\mathbb{R}^n$ for some $n \in \mathbb{N}$. *Any* two paths $f_0, f_1 : I \to X$ whose endpoints match are homotopic. The required homotopy is the **linear homotopy**

$$L : I \times I \to X : \; L(s,t) = (1-t)f_0(s) + tf_1(s)$$

which moves each point $f_0(s) \in C_{f_0}$ continuously along the straight line that connects it with $f_1(s)$. It is immediate that the function L is continuous (using the algebra of continuous functions) and satisfies all conditions of (2.1).

In particular, given some point $x_0 \in X$, *all loops in X based at x_0 are homotopic*.

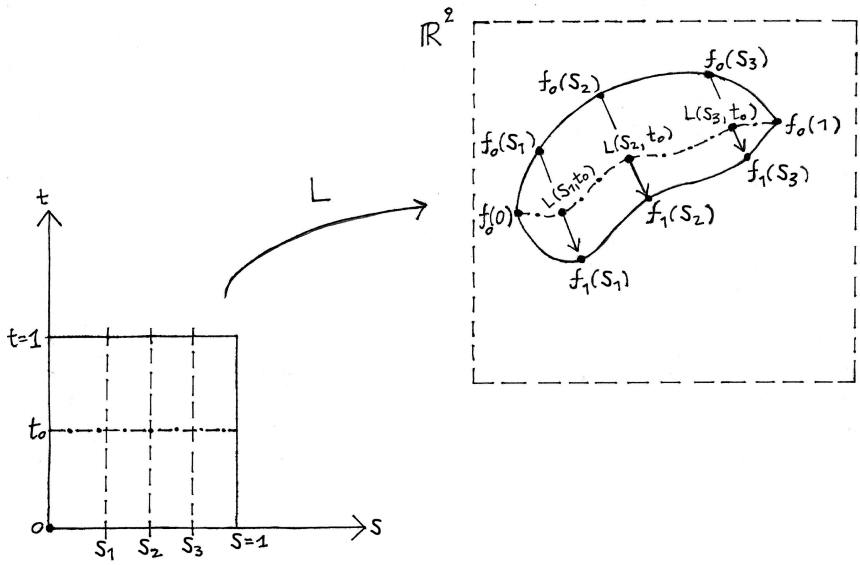


Figure 2.3: A linear homotopy of paths in $\mathbb{R}^2$

Counterexample 2.1.6. Suppose X is the usual *punctured plane* $\mathbb{R}^2 \setminus \{(0,0)\}$. In this case, there exist non homotopic loops because of the hole at $(0,0)$. Indeed, the loop

$$f : I \to X : s \mapsto \big(\cos(2\pi s), \sin(2\pi s)\big)$$

based at $(1,0)$ (for which $C_f = \mathrm{S}^1$) is *not* homotopic to the constant loop $c_{(1,0)}$. Intuitively, any *continuous* deformation from S^1 to $c_{(1,0)}$ requires the path f to pass through the origin, which is not in X.

Lemma 2.1.7. *Homotopy is an equivalence relation.*

Proof. The map $F(s,t) = f(s)$ is a homotopy from f to itself and if $F(s,t)$ is a homotopy from f to g then $G(s,t) = F(s,1-t)$ is a homotopy from g to f. Lastly, given homotopies $F : f \sim g$ and $G : g \sim h$, the mapping

$$H(s,t) = \begin{cases} F(s,2t), & t \in [0, \frac{1}{2}] \\ G(s,2t-1), & t \in [\frac{1}{2}, 1] \end{cases}$$

is a homotopy $H : f \sim h$. All these functions satisfy conditions (2.1) and are continuous; in particular, H is continuous by the glueing lemma 2.1.1. ◇

We denote the equivalence class of f under $\sim$ as $[f]$ and call it the **homotopy class** of f. So, by definition

$$f \sim g \iff [f] = [g].$$

When there is no danger of confusion however, we shall denote the homotopy class of f with the same symbol f, i.e. $[f] \equiv f$.

THE FUNDAMENTAL GROUP Given two paths in X, $f, g : I \to X$, we define their **product** $f \cdot g \equiv fg$ to be the path

$$fg : I \to X : \quad (fg)(s) = \begin{cases} f(2s), & s \in [0, \frac{1}{2}] \\ g(2s-1), & s \in [\frac{1}{2}, 1] \end{cases}$$

which is just the first path followed by the second, both traversed in double speed so that the domain of their product can be I. Let's consider the spacial case where both f and g are loops based at

the same basepoint. In that case the map fg is continuous (by the glueing lemma 2.1.1) with the property that $fg(0) = fg(1)$, hence a loop. We can extend this definition to the case of homotopy classes. We define the product of two *homotopy classes of loops* as

$$[f][g] = [fg]. \tag{2.3}$$

This is a well defined binary operation as the next lemma suggests.

Lemma 2.1.8. *Given four loops in X such that $[f_0] = [f_1]$ and $[g_0] = [g_1]$, we have $[f_0][g_0] = [f_1][g_1]$.*

Proof. Given homotopies $F : f_0 \sim f_1$ and $G : g_0 \sim g_1$, the mapping

$$H : I \times I \to X : \; H(s,t) = \begin{cases} F(2s,t), & t \in [0, \frac{1}{2}] \\ G(2s-1,t), & t \in [\frac{1}{2}, 1] \end{cases}$$

is continuous by the glueing lemma 2.1.1 and satisfies conditions (2.2); it is a homotopy $H : f_0 g_0 \sim f_1 g_1$. ◇

Let X be a topological space, $x_0 \in X$ and let $\pi_1(X, x_0)$ denote the set of all homotopy classes loops based at x_0, i.e.

$$\pi_1(X, x_0) = \{[f] \,|\, f : I \to X : f(0) = f(1) = x_0\}.$$

The previous lemma assures that the product defined in (2.3) is a well defined binary operation on $\pi_1(X, x_0)$.

Proposition 2.1.9. *The set $\pi_1(X, x_0)$ is a group under the product of loops as defined in (2.3).*

Proof. First, we must prove associativity:

$$\begin{aligned} ([f][g])[h] = [f]([g][h]) &\iff [fg][h] = [f][gh] \\ &\iff [(fg)h] = [f(gh)] \\ &\iff (fg)h \sim f(gh) \end{aligned}$$

for all $[f], [g], [h] \in \pi_1(X, x_0)$. The mapping

$$F(s,t) = \begin{cases} f(\frac{4s}{1+t}), & s \in [0, \frac{t+1}{4}] \\ g(4s-t-1), & s \in [\frac{t+1}{4}, \frac{t+2}{4}] \\ h(\frac{4s-t-2}{2-t}), & s \in [\frac{t+2}{4}, 1] \end{cases}$$

is a homotopy $F : (fg)h \sim f(gh)$. As for the identity element, it is easy to see that the constant loop on x_0 plays this role, i.e.

$$\begin{aligned}[c_{x_0}][f] = [f][c_{x_0}] = [f] &\iff [c_{x_0}f] = [fc_{x_0}] = [f] \\ &\iff c_{x_0}f \sim fc_{x_0} \sim f\end{aligned}$$

for all $[f] \in \pi_1(X, x_0)$. Indeed, the map

$$G : I \times I \to X : \; G(s,t) = \begin{cases} x_0, & s \in [0, \frac{t}{2}] \\ f(\frac{2s-t}{2-t}), & s \in [\frac{t}{2}, 1] \end{cases}$$

is a homotopy $G : c_{x_0}f \sim f$ (the other homotopy is similar). It remains to check for inverses. Recall that the inverse of a loop f is the loop $f^{-1}(s) = f(1-s)$. The reason we called the loop $f(1-s)$ the *inverse* of f is that

$$\begin{aligned}[f][f^{-1}] = [c_{x_0}] = [f^{-1}][f] &\iff [ff^{-1}] = [c_{x_0}] = [f^{-1}f] \\ &\iff f^{-1}f \sim c_{x_0} \sim ff^{-1}\end{aligned}$$

for all $[f] \in \pi_1(X, x_0)$. In other words, $[f]^{-1} = [f^{-1}]$. Indeed, the function

$$H : I \times I \to X : \; H(s,t) = \begin{cases} f\big(2s(1-t)\big), & s \in \big[0, \frac{1}{2}\big] \\ f\big((2-2s)(1-t)\big), & s \in \big[\frac{1}{2}, 1\big] \end{cases}$$

is a homotopy $H : ff^{-1} \sim c_{x_0}$ (the other homotopy is again similar). As with previous proofs, the verification of conditions (2.2) is straightforward and continuity is a simple application of the glueing lemma 2.1.1. ◇

Definition 2.1.10. The group $\pi_1(X, x_0)$ is called the **fundamental group** of X based at x_0.

Example 2.1.11. Let X be $\mathbb{R}^n$ and $x_0 \in X$ arbitrary. By Ex. 2.1.5, any two loops based at x_0 are homotopic. In particular, every loop based at x_0 is homotopic to the constant loop c_{x_0} and therefore $\pi_1(X, x_0) = \{[c_{x_0}]\}$, the trivial group.

Definition 2.1.12. A topological space X is called **simply connected** if it is *path connected* and has trivial fundamental group.

Example 2.1.13. $\mathbb{R}^n$ is a simply connected space for all $n \in \mathbb{N}$. In particular, $\mathbb{R}$ is simply connected.

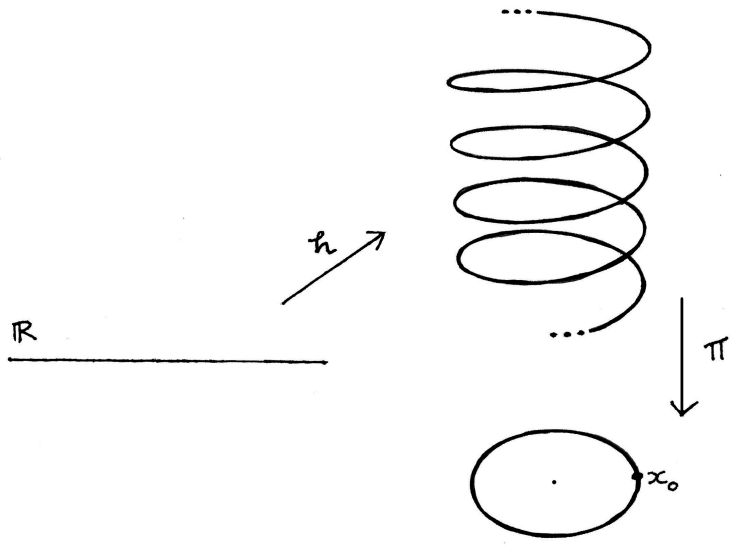


Figure 2.4: $\mathbb{R}$ and helix H over S^1

THE FUNDAMENTAL GROUP OF THE CIRCLE The next calculation will serve as a first (yet informal) encounter with covering spaces.

Example 2.1.14. We are going to calculate the fundamental group of the circle $S^1 \subseteq \mathbb{R}^2$ based at $x_0 = (1, 0)$ (we will later see that the choice of basepoint is actually irrelevant).

Consider the map

$$q : \mathbb{R} \to S^1 : \ x \mapsto \big(\cos(2\pi x), \sin(2\pi x)\big) \ \forall x \in \mathbb{R}.$$

To have a geometric image of this map (Fig. 2.4), we can factor it as $q = \pi \circ h$ where

$$h : \mathbb{R} \to H \subseteq \mathbb{R}^3 : t \mapsto \big(\cos(2\pi t), \sin(2\pi t), t\big)$$

is the usual parametrization of the helix H and

$$\pi : \mathbb{R}^3 \to S^1 \subseteq \mathbb{R}^2 : (x, y, z) \mapsto (x, y)$$

is the usual projection onto the first two coordinates.

This map has two crucial properties which will help us calculate $\pi_1(S^1, x_0)$ (and whose proof will be given in the next section in a much broader context).

1. (*Lifting of paths*) For every path $f : I \to \mathrm{S}^1$ and every $\widetilde{x}_0 \in \mathbb{R}$ with $q(\widetilde{x}_0) = x_0$, there is a *unique* path $\widetilde{f} : I \to \mathbb{R}$ such that $q \circ \widetilde{f} = f$ and $\widetilde{f}(0) = \widetilde{x}_0$. Such a path is called a **lift** of f.

Having established that liftings of paths exist and are unique, we can define the *degree* of a loop. Let f be a loop in S^1 based at x_0. Since

$$q^{-1}(x_0) = q^{-1}(1,0) = \mathbb{Z},$$

the lifting property guarantees that for every $k \in \mathbb{Z}$ there is a *unique lift* of f whose initial point is k. In particular, there is a unique lift of f that starts at $0 \in \mathbb{Z}$. We will denote this lift as $\widetilde{f}_o$. We define the **degree** of the loop f to be $\partial f = \widetilde{f}_o(1)$, i.e. the terminal point of $\widetilde{f}_o$. Note that

$$f(1) = x_0 \overset{q\widetilde{f}_o = f}{\Longrightarrow} q\widetilde{f}_o(1) = x_0 \implies \widetilde{f}_o(1) \in q^{-1}(x_0) = \mathbb{Z},$$

hence $\partial f \in \mathbb{Z}$. From the uniqueness of $\widetilde{f}_o$, the degree of a loop is well defined.

Our plan now is to extend this definition to homotopy classes of loops. We can define a degree function as above,

$$\partial : \pi_1(\mathrm{S}^1, x_0) \to \mathbb{Z} : [f] \mapsto \widetilde{f}_o(1),$$

but this time we cannot say for sure that this map is well defined. We must ensure that homotopic maps have the same degree. This where the second property comes into play.

2. (*Monodromy Theorem*) If $f, g : I \to \mathrm{S}^1$ are two *homotopic* paths ($f \sim g$) of S^1 and $\widetilde{f}, \widetilde{g}$ are two of their lifts in $\mathbb{R}$ with the same initial point $\widetilde{f}(0) = \widetilde{g}(0)$, then $\widetilde{f}(1) = \widetilde{g}(1)$.

It is now obvious that ∂ is a well defined map. In fact we have everything we need. This map is actually a group isomorphism!

∂ is 1-1. If $\partial f \neq \partial g \iff \widetilde{f}_o(1) \neq \widetilde{g}_o(1)$ then f cannot be homotopic to g. If it were, by the monodromy theorem we would have $\widetilde{f}_o(1) = \widetilde{g}_o(1)$.

∂ is onto. For every $k \in \mathbb{Z}$, the loop

$$f_k : I \to \mathbb{R} : t \mapsto \big(\cos(2k\pi t), \sin(2k\pi t)\big)$$

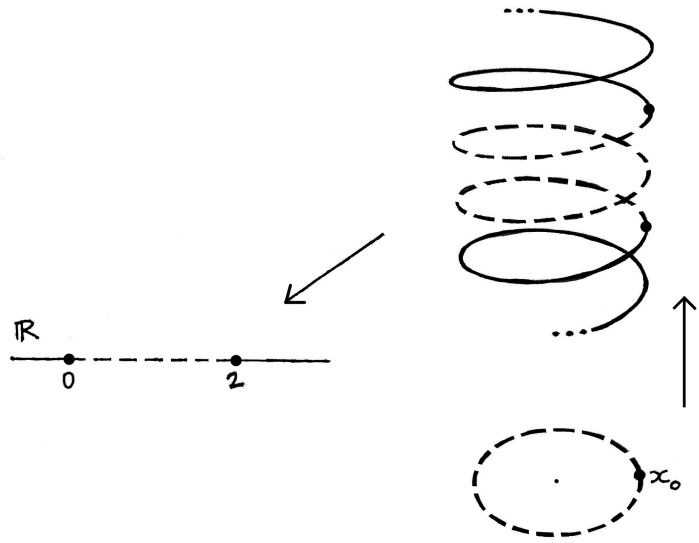


Figure 2.5: Lifting the loop that goes around the circle *twice*

has degree k. Indeed, it is immediate that $\widetilde{f_k} : I \to \mathbb{R} : t \mapsto kt$ is a lift of f_k whose initial point is 0 and so $\widetilde{f_k} = (\widetilde{f_k})_o$. Therefore

$$\partial f = (\widetilde{f_k})_o(1) = \widetilde{f_k}(1) = k.$$

∂ is a group homomorphism. We must show that

$$\partial([f][g]) = \partial[f] + \partial[g].$$

Indeed, it is easily seen that

$$\widetilde{fg} = \begin{cases} \widetilde{f}_o(2s), & s \in [0, \frac{1}{2}] \\ \widetilde{f}_o(1) + \widetilde{g}_o(2s-1), & s \in [\frac{1}{2}, 1] \end{cases}$$

is a lift of fg whose initial point is 0, so $\widetilde{fg} = (\widetilde{fg})_o$. Therefore

$$\partial([f][g]) = \partial[fg] = (\widetilde{fg})_o(1) = \widetilde{fg}(1) = \widetilde{f}_o(1) + \widetilde{g}_o(1) = \partial[f] + \partial[g]$$

and, as a result,

$$\pi(\mathrm{S}^1, x_0) \simeq \mathbb{Z}.$$

THE CHOICE OF BASEPOINT Before we continue we must address the elephant in the room.

Given a topological space X with $|X| \geqslant 2$ and two distinct elements $x \neq y \in X$, we can form the fundamental groups $\pi_1(X, x)$ and $\pi_1(X, y)$. Is there any relation between these groups? Does it matter which point do we choose? These are the questions we want to answer.

Recall from General Topology that a space X is called **path connected** if for every two points $x, y \in X$ there is *path* $\gamma : I \to X$ that *connects* them, i.e. $\gamma(0) = x$ and $\gamma(1) = y$. A space X is called **connected** if we *cannot* find two nonempty, disjoint, open subsets of X whose union is X. Connectedness is a weaker condition that path connectedness. *Every path connected space is connected* but the converse does *not* hold; the most famous counterexample perhaps is the **topologist's sine curve**

$$\mathrm{cl}_{\mathbb{R}^2}\{x \times \sin(1/x) : 0 < x \leqslant 1\}$$

which is connected but not path connected.

Proposition 2.1.15. *Let X be a path connected space, $x_0, y_0 \in X$ and $\gamma : I \to X$ a path such that $\gamma(0) = x_0$ and $\gamma(1) = y_0$. Then, the map*

$$\psi : \pi_1(X, x_0) \to \pi_1(X, y_0) : \ [f] \mapsto [\gamma]^{-1}[f][\gamma]$$

is a group isomorphism.

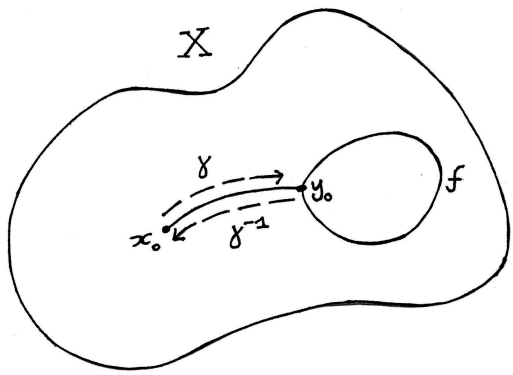


Figure 2.6: The basepoint is irrelevant in a *path connected* space

Proof. ψ is a group homomorphism since

$$\begin{aligned}\psi([f][g]) = \psi([fg]) = [\gamma]^{-1}[fg][\gamma] = [\gamma]^{-1}[f][g][\gamma] \\ = [\gamma]^{-1}[f][\gamma][\gamma]^{-1}[g][\gamma] = \psi([f])\psi([g]).\end{aligned}$$

It is straightforward that

$$\psi^{-1} : \pi_1(X, y_0) \to \pi_1(X, x_0) : \ [f] \mapsto [\gamma][f][\gamma]^{-1}$$

is also a homomorphism and a two sided inverse of ψ. Therefore ψ is an isomorphism. ◇

Counterexample 2.1.16. The hypothesis that X is path connected is essential. Consider for example the space $X = S^1 \sqcup Y \subseteq \mathbb{R}^2$ where $Y = \{(0,0)\}$. For any $x \in S^1$ we have $\pi_1(X, x) \simeq \mathbb{Z}$ while $\pi_1\big(X, (0,0)\big) \simeq \{1\}$.

For spaces that are not path connected, the fundamental group loses its power since, depending on the choice of the basepoint, there can be *more than one non isomorphic* fundamental groups. It is therefore natural in Algebraic Topology to restrict our attention to path connected spaces only.

Convention. For the rest of the chapter we will restrict our attention to path connected (hence connected) spaces only, whose fundamental group is unique up to isomorphism. *All* topological spaces are therefore assumed to be *path connected* unless explicitly stated otherwise.

THE INDUCED HOMOMORPHISM Let us see now how we can study continuous maps using fundamental groups.

Consider two topological spaces X, Y and a continuous mapping between them $h : X \to Y$. Suppose that $x \in X$ and $y \in Y$ are two points such that $h(x) = y$. We use the notation

$$h : (X, x) \to (Y, y)$$

to denote such a situation. If $f : I \to X$ is an arbitrary path in X then the composition

$$I \xrightarrow{f} X \xrightarrow{h} Y$$

is a path in Y since both h and f are continuous. In particular, if f is a loop based at x then $h \circ f$ is a loop based at $h(x) = y$. Therefore, the function h induces a map h_* between the fundamental groups $\pi_1(X, x)$ and $\pi_1(Y, y)$ defined as

$$h_* : \pi_1(X, x) \to \pi_1(Y, y) : \ [f] \mapsto [h \circ f]$$

Proposition 2.1.17. *The induced map h_* is a group homomorphism.*

Proof. The map h_* is well defined. Indeed, if $[f] = [g] \in \pi_1(X, x)$ and $F : f \sim g$ is a homotopy, then hF is a homotopy $hF : hf \sim hg$. It is continuous since h and F are, and satisfies conditions (2.2) because F does. Moreover,

$$\begin{aligned} h_*([f] \cdot [g]) = h_*([f \cdot g]) = [h \circ (f \cdot g)] = [h \circ f \cdot h \circ g] \\ = [h \circ f] \cdot [h \circ g] = h_*([f]) \cdot h_*([g]) \end{aligned}$$

so h_* is a group homomorphism. ◇

Two immediate yet crucial properties of the induced homomorphism are the following.

Lemma 2.1.18. *If X is a topological space and $x \in X$ then*

$$(id_X)_* = id_{\pi_1(X,x)}.$$

Lemma 2.1.19. *If $h_1 : (X, x) \to (Y, y)$ and $h_2 : (Y, y) \to (Z, z)$ are continuous functions between spaces then*

$$(h_1 \circ h_2)_* = h_{1*} \circ h_{2*}.$$

These two properties imply the next extremely important result.

Corollary 2.1.20. *The fundamental group is a topological invariant, i.e. homeomorphic spaces have isomorphic fundamental groups.*

Proof. If $h : (X, x) \to (Y, y)$ is a homeomorphism, then there is a homeomorphism $h^{-1} : (Y, y) \to (X, x)$ such that $hh^{-1} = \mathrm{id}_Y$ and $h^{-1}h = \mathrm{id}_X$. Using the above two properties on these relations we get

$$h_*^{-1}h_* = \mathrm{id}_{\pi_1(X,x)} \quad \text{and} \quad h_*h_*^{-1} = \mathrm{id}_{\pi_1(Y,y)}.$$

So h_* is a group homomorphism with inverse $(h_*)^{-1} = (h^{-1})_*$, hence an isomorphism. ◇

Counterexample 2.1.21. The converse does *not* hold. For example, both $\mathbb{R}$ and $\{r\}$ ($r \in \mathbb{R}$) have trivial fundamental groups, yet they are certainly *not* homeomorphic (just compare their cardinalities).

2.2 COVERING SPACES

COVERING SPACES We now define the second of the two central objects of our study, *covering spaces*. As with fundamental groups, we will restrict our attention only to the absolute essentials for what follows.

Definition 2.2.1. Let X be a topological space. A **covering space** (or simply a **covering**) of X is a pair $(\widetilde{X}, p)$ consisting of a topological space $\widetilde{X}$ along with a *continuous* and *surjective* function $p : \widetilde{X} \to X$ so that every $x \in X$ has an *open* neighborhood U_x that is **evenly covered**, that is, a neighborhood whose inverse image $p^{-1}(U_x)$ can be written as a *disjoint* union of (open) subsets $\{V_a\}_{a \in A}$ of $\widetilde{X}$ each homeomorphic to U_x through p, i.e. $p|_{V_a} : V_a \to U_x$ is a homeomorphism for every $a \in A$.

We call X the **base space** and p the **covering map**. The sets V_a are called the **sheets** of U_x. When there is no danger of confusion, we will denote a covering $(\widetilde{X}, p)$ simply as $\widetilde{X}$.

Convention. As with all topological spaces so far, covering spaces are also assumed to be path connected (hence connected). Thus we do not have to worry when considering their fundamental group.

Before looking at some examples of coverings, lets see an important property of covering maps.

Recall from General Topology that a **local homeomorphism** between two spaces $\phi : X \to Y$ is a map such that for every $x \in X$ there is an *open* neighborhood V_x of x so that $\phi(V_x)$ is *open* in Y and $\phi|_{V_x} : V_x \to \phi(V_x)$ is a *homeomorphism*. *Every local homeomorphism is a continuous and open map*.

Lemma 2.2.2. *Every covering map is a local homeomorphism.*

Proof. If $p : \widetilde{X} \to X$ is a covering and $\widetilde{x} \in \widetilde{X}$, then, by the definition of a covering map, $x = p(\widetilde{x}) \in X$ has an open neighborhood U_x

such that $p^{-1}(U_x) = \bigsqcup_{a\in A} V_a$, where each V_a is homeomorphic to U_x through p. Choose $a \in A$ such that $\widetilde{x} \in V_a$. Then V_a is an open neighborhood of $\widetilde{x}$, $p(V_a)$ is homeomorphic to U_x hence open in X, and $p|_{V_a}$ is a homeomorphism. ◇

Example 2.2.3. If X is a topological space, (X, id_X) is (trivially) a covering space of X.

Example 2.2.4. If $h : Y \to X$ is a *homeomorphism* between topological spaces, then (Y, h) is a covering of X.

Example 2.2.5. The set of real numbers $\mathbb{R}$ with the function

$$q : \mathbb{R} \to \mathrm{S}^1 : \ t \mapsto \big(\cos(2\pi t), \sin(2\pi t)\big) \ \forall t \in \mathbb{R}$$

is a covering of S^1.

Indeed, q is onto and continuous since cos and sin are. Furthermore, the sets

$$\begin{aligned} U_1 &= \{(x,y) \in \mathrm{S}^1 : x > 0\} \\ U_2 &= \{(x,y) \in \mathrm{S}^1 : x < 0\} \\ U_3 &= \{(x,y) \in \mathrm{S}^1 : y > 0\} \\ U_4 &= \{(x,y) \in \mathrm{S}^1 : y < 0\} \end{aligned}$$

constitute an *open cover* of S^1 such that

$$\begin{aligned} q^{-1}(U_1) &= \bigcup_{n=-\infty}^{\infty} \left(n - \frac{1}{4}, n + \frac{1}{4}\right) \\ q^{-1}(U_2) &= \bigcup_{n=-\infty}^{\infty} \left(n + \frac{1}{4}, n + \frac{3}{4}\right) \\ q^{-1}(U_3) &= \bigcup_{n=-\infty}^{\infty} \left(n, n + \frac{1}{2}\right) \\ q^{-1}(U_4) &= \bigcup_{n=-\infty}^{\infty} \left(n - \frac{1}{2}, n\right). \end{aligned}$$

In other words, for every $t \in \mathbb{R}$ there is an open neighborhood U_i (for some $i = 1, 2, 3, 4$) such that $t \in U_i$ and $q^{-1}(U_i)$ is a disjoint union of open subsets of $\mathbb{R}$. It is immediate that $q|_{U_i}$ is a homeomorphism for every $i = 1, 2, 3, 4$.

Example 2.2.6. An important class of coverings are obtained via group actions. We say that **a group** G **acts on a topological space** X if G acts on the set X and the action is compatible to the topology of X in the sense that for each $g \in G$, the permutation $\rho_g : X \to X$ of X induced by g is *continuous*. In that case, G acts on X via homeomorphisms, i.e., the permutation representation of the action is

$$\rho : G \to \mathrm{Homeo}(X) : g \mapsto \rho_g.$$

Indeed, for any $g \in G$, we know from the general theory of group actions that ρ_g is bijective and, by the definition of a group acting on a topological space, continuous. The map $\rho_{g^{-1}}$ is also bijective and continuous and

$$\rho_g \circ \rho_{g^{-1}} = \rho_{g^{-1}} \circ \rho_g = \mathrm{id}_X.$$

Therefore ρ_g is a homeomorphism of X.

We say that the action is **properly discontinuous** if every $x \in X$ has an open neighborhood V_x such that $g_1 V_x \cap g_2 V_x = \emptyset$ for every $g_1 \neq g_2 \in G$. In that case the stabilizer of every element is trivial and therefore action is faithful.

Since G acts on the set X, the orbits of X are a partition of X into disjoint sets. We denote by X/G **the set of orbits of** X **under** G. We give X/G the quotient topology induced by the projection

$$\pi : X \to X/G : x \mapsto \mathcal{O}(x)$$

where $\mathcal{O}(x) = \{gx : g \in G\}$ is the orbit of $x \in X$. In other words, the topology of X/G is

$$\tau_{X/G} = \{V \subseteq X/G : \pi^{-1}(V) \subseteq X \text{ open}\}.$$

If the action of G on X is properly discontinuous, then (X, π) is a covering of X/G. The map π is by definition surjective and continuous since X/G was given the quotient topology. It is also open. Indeed, for each open subset U of X, the set

$$\begin{aligned}
\pi^{-1}\big(\pi(U)\big) &= \{x \in X : \pi(x) \in \pi(U)\} \\
&= \{x \in X : \mathcal{O}(x) = \mathcal{O}(y) \text{ for some } y \in U\} \\
&= \{x \in X : x = gy \text{ for some } y \in U \text{ and } g \in G\} \\
&= \bigcup_{g \in G} gU = \bigcup_{g \in G} \rho_g(U)
\end{aligned}$$

is a union of open subsets of X (U is open and ρ_g a homeomorphism) hence open in X. But X/G has the quotient topology so $\pi(U)$ is open in X/G.

It remains to show that every element in X/G has an open neighborhood U whose inverse image $\pi^{-1}(U)$ is a disjoint union of open subsets of X each homeomorphic to U. Take an orbit in $\mathcal{O} \in X/G$ and some element $x \in X$ which belongs in it, i.e. $\mathcal{O} = \mathcal{O}(x)$. Take the neighborhood V_x of x described in the definition of a properly discontinuous action. Since π is open, $U = \pi(V_x)$ is open and it is the required open neighborhood of $\mathcal{O}(x)$ in X/G. Indeed,

$$\pi^{-1}\big(\pi(V_x)\big) = \bigsqcup_{g \in G} gV_x$$

which is a union of open subsets of X that is disjoint because the action is properly discontinuous. Furthermore, each gV_x is homeomorphic to $U = \pi(V_x)$ through the restriction

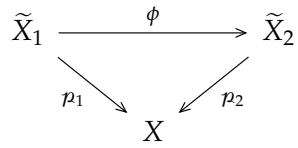
$$\pi|_{gV_x} : gV_x \to \pi(V_x).$$

Indeed, each $\pi|_{gV_x}$ is continuous and open as the restriction of the continuous function π in the open subset gV_x of X; it is clearly surjective and, since the action is properly discontinuous, injective as well. So $\pi|_{gV_x}$ is a homeomorphism.

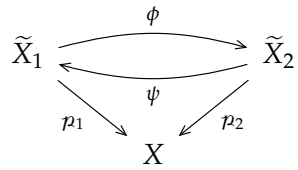
The above examples suggest that a space X can have more than one covering. So the natural question to ask is whether we find *all* possible coverings of a topological space, up to isomorphism.

COVERING SPACE ISOMORPHISMS But first things first. We must define what we mean when we say that two covering spaces are isomorphic.

Definition 2.2.7. Let $(\widetilde{X}_1, p_1)$ and $(\widetilde{X}_2, p_2)$ be two coverings of a space X. A **morphism** from $(\widetilde{X}_1, p_1)$ to $(\widetilde{X}_2, p_2)$ is a *continuous* map $\phi : \widetilde{X}_1 \to \widetilde{X}_2$ that makes the following diagram commute, i.e. $p_2 \circ \phi = p_1$.

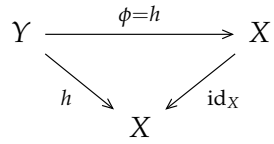


A morphism ϕ of covering spaces as above is called an **isomorphism** if there is a morphism $\psi : \widetilde{X}_2 \to \widetilde{X}_1$ such that $\psi\phi = \mathrm{id}_{\widetilde{X}_1}$ and $\phi\psi = \mathrm{id}_{\widetilde{X}_2}$.



We write $(\widetilde{X}_1, p_1) \cong (\widetilde{X}_2, p_2)$ or simply $\widetilde{X}_1 \cong \widetilde{X}_2$ to indicate that the two coverings are isomorphic. An isomorphism from a covering space to itself is called a **deck transformation** or, more charmingly, a **Deckbewegung** (in the German literature).

Example 2.2.8. Let X, Y be arbitrary but *homeomorphic* spaces and $h : Y \to X$ a homeomorphism. The coverings (Y, h) and (X, id_X) of X are isomorphic. Indeed, if we take $\phi = h$ in the definition then the diagram



obviously commutes.

The following results are immediate.

Lemma 2.2.9. *Covering space isomorphism is an equivalence relation.*

Lemma 2.2.10. *A morphism of covering spaces is an isomorphism if and only if it is a homeomorphism.*

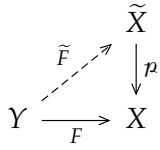
Lemma 2.2.11. *Suppose X is a topological space. The set of deck transformations $A(\widetilde{X}, p)$ of a covering space $(\widetilde{X}, p)$ of X form a group under composition.*

LIFTING PROPERTIES We will now see that the two crucial properties we used to calculate the fundamental group of the circle (Ex. 2.1.14) are consequences of the fact that $(\mathbb{R}, q)$ is a covering of S^1.

Definition 2.2.12. Let X, Y be two sets, $(\widetilde{X}, p)$ a covering of X and $F : Y \to X$ a *continuous* function. A **lift** (or **lifting**) of F in $\widetilde{X}$ is a map

$$\widetilde{F} : Y \to \widetilde{X}$$

that makes the following diagram commute.



For the rest of this section fix X an arbitrary topological space, $x_0 \in X$ and $(\widetilde{X}, p)$ an arbitrary covering of X.

We are primarily interest at two kinds of lifts at this point. Lifts of *paths* (where $Y = I$) and lifts of *homotopies* (where $Y = I \times I$). Before proving the main results, let us recall a lemma from General Topology that we are going to need.

Lemma 2.2.13 (Lebesgue's[1] Number Lemma)**.** *Suppose X is a compact metric space and $\mathcal{U}$ is an open cover of X. Then there is some $\delta > 0$ so that every subset of X with diameter less than δ is contained in an element of $\mathcal{U}$.*

Proposition 2.2.14 (Lifting of Paths)**.** *For every path $f : I \to X$ and every $\widetilde{x}_0 \in \widetilde{X}$ with $p(\widetilde{x}_0) = x_0$, there is a unique lift $\widetilde{f} : I \to \widetilde{X}$ of f such that $\widetilde{f}(0) = \widetilde{x}_0$.*

Proof. Let $f : I \to X$ be a path and $\widetilde{x}_0 \in \widetilde{X}$ an element such that $p(\widetilde{x}_0) = x_0$. We are going to construct $\widetilde{f}$ step by step. By definition of the covering space, for every $x \in X$ we can find an open subset U_x of X such that $x \in U_x$ and

$$p^{-1}(U_x) = \bigcup_{a \in A} V_a$$

[1] Named after Henri Léon Lebesgue (1875–1941)

where $\{V_a\}_{a\in A}$ is a disjoint family of open subsets of $\widetilde{X}$ so that $p|_{V_a} : V_a \to U_x$ is a homeomorphism for every $a \in A$. So $\{U_x : x \in X\}$ is an open cover of X and consequently, since f is continuous, $\{f^{-1}(U_x) : x \in X\}$ is an open cover of I. But I is compact. From Lebesgue's Lemma, we can find a partition

$$0 = s_0 < s_1 < s_2 < \ldots < s_{k-1} < s_k = 1$$

of $I = [0,1]$ such that

$$[s_i, s_{i+1}] \subseteq f^{-1}(U_x) \iff f([s_i, s_{i+1}]) \subseteq U_x \text{ for some } x \in X$$

for all $i = 0, 1, \ldots, k-1$. Now we can define $\widetilde{f}$ on each $[s_i, s_{i+1}]$. We define

$$\widetilde{f}(0) = \widetilde{x}_0$$

and if $\widetilde{f}(s_i) \in V_a$ for some $a \in A_x$ for some $x \in X$, then

$$\widetilde{f}(s) = (p|_{V_a})^{-1}\big(f(s)\big) \quad \forall s \in (s_i, s_{i+1}].$$

Since $p|_{V_a}$ is a homeomorphism, $\widetilde{f}$ is continuous on $[s_i, s_{i+1}]$ for every $i = 0, 1, \ldots, k-1$ hence, by the glueing lemma, continuous on I. The fact that $\widetilde{f}$ is a lift of f is immediate from the way we constructed it.

It remains to show that $\widetilde{f}$ is unique. Let $\widetilde{f}_1, \widetilde{f}_2$ be two lifts of f starting at $\widetilde{x}_0$. We will show that $\widetilde{f}_1(s) = \widetilde{f}_2(s)\ \forall s \in I$. We have

$$\widetilde{f}_1(0) = \widetilde{x}_0 = \widetilde{f}_2(0)$$

and if $\widetilde{f}_1(s) = \widetilde{f}_2(s)$ for every $0 \leqslant s \leqslant s_i$ then $\widetilde{f}_1(s) = \widetilde{f}_2(s)$ for every $s_i \leqslant s \leqslant s_{i+1}$. Indeed, if

$$\widetilde{f}_1(s) = (p|_{V_a})^{-1}\big(f(s)\big) \quad \forall s \in (s_i, s_{i+1}]$$

and

$$\widetilde{f}_2(s) = (p|_{V_b})^{-1}\big(f(s)\big) \quad \forall s \in (s_i, s_{i+1}]$$

then $a = b$ since both $\widetilde{f}_1([s_i, s_{i+1}])$ and $\widetilde{f}_2([s_i, s_{i+1}])$ are connected, the family $\{V_a\}_{a\in A_x}$ is disjoint and $\widetilde{f}_1(s_i) = \widetilde{f}_2(s_i)$. ◇

Proposition 2.2.15 (Lifting of Homotopies). *For every homotopy $F : I^2 \to X$ with $F(0,0) = x_0$ and every $\widetilde{x}_0 \in \widetilde{X}$ with $p(\widetilde{x}_0) = x_0$, there is a unique lift $\widetilde{F} : I^2 \to \widetilde{X}$ such that $\widetilde{F}(0,0) = \widetilde{x}_0$.*

Proof. The proof is similar to that of the previous proposition. We first define

$$\widetilde{F}(0,0) = \widetilde{x}_0.$$

By the previous proposition, we can extend $\widetilde{F}$ on $I \times \{0\}$ and $\{0\} \times I$. Now, as before, we can extend $\widetilde{F}$ on I^2 by defining it step by step in rectangles of the form $[t_j, t_{j+1}] \times [s_i, s_{i+1}]$. Uniqueness is also proven using similar arguments. ◇

An immediate corollary is the second crucial property we used in Ex. 2.1.14.

Corollary 2.2.16 (Monodromy Theorem)**.** *If $f, g : I \to X$ are two homotopic paths ($H : f \sim g$) of X and $\widetilde{f}, \widetilde{g}$ are two of their lifts in $\widetilde{X}$ such that $\widetilde{f}(0) = \widetilde{g}(0) = x_0$, then $\widetilde{f}(1) = \widetilde{g}(1)$.*

Proof. By the homotopy lifting property, there is a unique lift $\widetilde{H}$ of H such that $\widetilde{H}(0,0) = x_0$. It is immediate that both $\widetilde{f}$ and $\widetilde{H}(s,0)$ are lifts of f starting at x_0 so

$$\widetilde{f}(s) = \widetilde{H}(s,0) \ \forall s \in I$$

by the path lifting property. Similarly

$$\widetilde{g}(s) = \widetilde{H}(s,1) \ \forall s \in I.$$

Now $\widetilde{H}(1,t)$ is a lift of the constant path $c_{H(1,t)}$ starting at $\widetilde{H}(1,0)$. As a lift of a *constant* path, $\widetilde{H}(1,t)$ is itself constant, so

$$\widetilde{H}(1,t) = \widetilde{H}(1,0) \ \forall t \in I.$$

Therefore, $\widetilde{f}(1) = \widetilde{H}(1,0) = \widetilde{H}(1,1) = \widetilde{g}(1)$. ◇

2.3 GALOIS THEORY OF COVERINGS

We have all the tools we need to state our next classification theorem which concerns fundamental groups and covering spaces. As we mentioned, a topological space can have more than one covering and a natural question we can ask is whether there is a way to determine all possible coverings of a given space. It turns out that for topological spaces that satisfy some nice (not too restrictive) conditions, there is a way to find all possible coverings using subgroups of their fundamental groups.

THE CORRESPONDENCE We must first establish a correspondence between the coverings of a space and the subgroups of its fundamental group.

Let X be a topological space, $x_0 \in X$ and $(\widetilde{X}, p)$ a covering of X. For every element $\widetilde{x}_0 \in p^{-1}(x_0)$ we have a continuous function

$$p : (\widetilde{X}, \widetilde{x}_0) \to (X, x_0)$$

which induces a map

$$p_* : \pi_1(\widetilde{X}, \widetilde{x}_0) \to \pi_1(X, x_0).$$

Proposition 2.3.1. *The indunced map p_* is a group monomorphism.*

Proof. We have already seen that p_* is a group homomorphism. It remains to show that p_* is 1-1.

If $[\widetilde{f}], [\widetilde{g}] \in \pi_1(\widetilde{X}, \widetilde{x}_0)$ such that

$$p_*[\widetilde{f}] = p_*[\widetilde{g}] \iff \exists\, H : p\widetilde{f} \sim p\widetilde{g},$$

then, by the homotopy lifting property 2.2.15,

$$\exists\, \widetilde{H} : \widetilde{f} \sim \widetilde{g} \iff [\widetilde{f}] = [\widetilde{g}].$$

Therefore p_* is indeed 1-1. ◇

Thus, for *path connected* spaces, we obtain a correspondence

$$\begin{aligned} \{\text{(path connected) coverings of } X\} &\rightleftarrows \{\text{subgroups of } \pi_1(X, x_0)\} \\ \widetilde{X} &\mapsto \pi_1(\widetilde{X}, \widetilde{x}_0) \end{aligned}$$

where $\widetilde{x}_0 \in p^{-1}(x_0)$. This correspondence is *not* well defined though. For a choice $\widetilde{x}_0 \neq \widetilde{x}_1 \in p^{-1}(x_0)$, $p_*\big(\pi_1(\widetilde{X}, \widetilde{x}_0)\big) \neq p_*\big(\pi_1(\widetilde{X}, \widetilde{x}_1)\big)$ in general.

Recall from Group Theory that two subgroups H_0, H_1 of a group G are said to be **conjugate** if there is some element $g \in G$ such that $g^{-1}H_0g = H_1$. We will see that the subgroups $H_0 = p_*\big(\pi_1(\widetilde{X}, \widetilde{x}_0)\big)$ and $H_1 = p_*\big(\pi_1(\widetilde{X}, \widetilde{x}_1)\big)$ of $G = \pi_1(X, x_0)$ are conjugate.

$\widetilde{X}$ is path connected. If $\widetilde{\gamma} : I \to \widetilde{X}$ is a path from $\widetilde{x}_0$ to $\widetilde{x}_1$, i.e. $\widetilde{\gamma}(0) = \widetilde{x}_0$ and $\widetilde{\gamma}(1) = \widetilde{x}_1$, then $p \circ \widetilde{\gamma} : I \to X$ is path from

$p(\widetilde{\gamma}(0)) = p(\widetilde{x}_0) = x_0$ to $p(\widetilde{\gamma}(1)) = p(\widetilde{x}_1) = x_0$, that is, a loop in X based at x_0. Choose $g = [p \circ \widetilde{\gamma}] \in G = \pi_1(X, x_0)$.

It is immediate that $g^{-1}H_0g \subseteq H_1$. Indeed, for any element $p_*[\widetilde{f}] \in H_0 = p_*\big(\pi_1(\widetilde{X}, \widetilde{x}_0)\big)$,

$$\begin{aligned}
[p \circ \widetilde{\gamma}]^{-1} \cdot p_*[\widetilde{f}] \cdot [p \circ \widetilde{\gamma}] &= [(p \circ \widetilde{\gamma})^{-1}] \cdot [p \circ \widetilde{f}] \cdot [p \circ \widetilde{\gamma}] \\
&= [\big(p \circ (\widetilde{\gamma}^{-1})\big) \cdot \big(p \circ \widetilde{f}\big) \cdot \big(p \circ \widetilde{\gamma}\big)] \\
&= [p \circ (\widetilde{\gamma}^{-1} \cdot \widetilde{f} \cdot \widetilde{\gamma})] \\
&= p_*([\widetilde{\gamma}^{-1} \cdot \widetilde{f} \cdot \widetilde{\gamma}]) \in H_1 = p_*\big(\pi_1(\widetilde{X}, \widetilde{x}_1)\big)
\end{aligned}$$

since $\widetilde{\gamma}^{-1} \cdot \widetilde{f} \cdot \widetilde{\gamma}$ is a loop based at $\widetilde{x}_1$.

Similarly, $gH_1g^{-1} \subseteq H_0$ which implies that $H_1 \subseteq g^{-1}H_0g$. Therefore $g^{-1}H_0g = H_1$ and so the subgroups H_0 and H_1 of G are indeed conjugate.

Moreover, if $H = s^{-1}H_0s$ is a subgroup conjugate to H_0 in G, where $s \in \pi_1(X, x_0)$, then, from the path lifting property, there is a lift $\widetilde{s} : I \to \widetilde{X}$ of s with $\widetilde{s}(0) = \widetilde{x}_0$. It is immediate from the discussion above that

$$H = p_*\Big(\pi_1\big(\widetilde{X}, \widetilde{s}(1)\big)\Big).$$

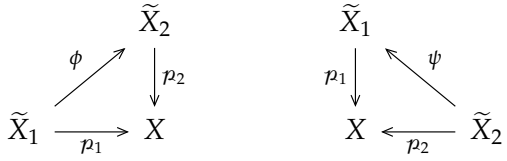
Thus, for *path connected spaces*, we have a *well defined* correspondence

$$\begin{aligned}
\{\text{coverings of } X\} &\rightleftarrows \{\text{conjugacy classes of subgroups of } \pi_1(X, x_0)\} \\
\widetilde{X} &\mapsto \pi_1(\widetilde{X}, \widetilde{x}_0)
\end{aligned}$$

We set out to understand for which (path connected) topological spaces the above correspondence is a bijection between the set of all coverings of the given space and the set of all subgroups of its fundamental group.

UNIQUENESS First, we examine whether the correspondence is injective.

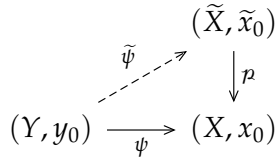
If we want to decide whether two coverings of an arbitrary space are isomorphic, we need to find continuous maps ϕ, ψ as described in Definition 2.2.7. Since both $\widetilde{X}_1$ and $\widetilde{X}_2$ are coverings, the maps ϕ and ψ can be considered as *liftings* of p_1 and p_2 respectively.



So far we have seen liftings of paths and homotopies but the maps p_1, p_2 are neither paths nor homotopies. We want to find some necessary and sufficient conditions for liftings of covering functions to exist. The first step is to examine conditions under which liftings of arbitrary maps exist.

Recall from General Topology that a space is called **locally path connected** if for every $y \in Y$ and every open neighborhood $\mathcal{N}_y$ of y, there is a *path connected open set* $\mathcal{U}_y$ such that $y \in \mathcal{U}_y \subseteq \mathcal{N}_y$.

Proposition 2.3.2. *Suppose X is a topological space, fix an element $x_0 \in X$ and let $p : (\widetilde{X}, \widetilde{x}_0) \to (X, x_0)$ be a covering of X. Moreover, let Y be a path connected and locally path connected topological space, $y_0 \in Y$ and $\psi : (Y, y_0) \to (X, x_0)$ a continuous map.*

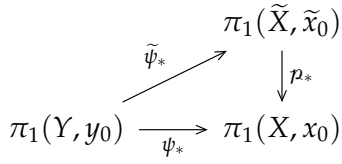


A lift $\widetilde{\psi}$ of ψ exists if and only if

$$\psi_*\big(\pi_1(Y, y_0)\big) \subseteq p_*\big(\pi_1(\widetilde{X}, \widetilde{x}_0)\big).$$

Furthermore, if such $\widetilde{\psi}$ exists, it is unique.

Proof. ($\Rightarrow$) The one direction is simple and we only need 2.3.1. If such lift exists, we obtain the following commutative diagram



which implies that

$$\psi_*\big(\pi_1(Y,y_0)\big) = p_*[\widetilde{\phi}_*\big(\pi_1(Y,y_0)\big)] \subseteq p_*\big(\pi_1(\widetilde{X},\widetilde{x}_0)\big).$$

($\Leftarrow$) For the contrary, suppose that $\psi_*\big(\pi_1(Y,y_0)\big) \subseteq p_*\big(\pi_1(\widetilde{X},\widetilde{x}_0)\big)$. Define

$$\widetilde{\psi} : (Y,y_0) \to (X,x_0)$$

as follows. For each $y \in Y$ take a path $f_y : I \to Y$ in Y that starts at $f_y(0) = y_0$ and ends at $f_y(1) = y$; such a path exists since Y is path connected. Then

$$I \xrightarrow{f_y} Y \xrightarrow{\psi} X$$

is a path in X that starts at $\psi f_y(0) = \psi(y_0) = x_0$. From the path lifting property, there is a unique lift

$$\widetilde{\psi f_y} : I \to \widetilde{X},$$

i.e. $p\widetilde{\psi f_y} = \psi f_y$, that starts at $\widetilde{\psi f_y}(0) = \widetilde{x}_0$. Define

$$\widetilde{\psi} : (Y,y_0) \to (X,x_0) : \widetilde{\psi}(y) = \widetilde{\psi f_y}(1).$$

First we must show that this map is well defined. Suppose that $\alpha,\beta : I \to Y$ are two *homotopic* paths, say $H : \alpha \sim \beta$ that start at $\alpha(0) = \beta(0) = y_0$ and end at $\alpha(1) = \beta(1) = y$. Then $\psi H : \psi\alpha \sim \psi\beta$ and by the Modoromy Theorem, $\widetilde{\psi\alpha}(1) = \widetilde{\psi\beta}(1)$. So choosing a path hotopic to f_y does not affect $\widetilde{\psi}$. Next let $\alpha,\beta : I \to Y$ be two (not necessarily homotopic) paths that start at $\alpha(0) = \beta(0) = y_0$ and end at $\alpha(1) = \beta(1) = y$. Then the product

$$\alpha\beta^{-1} : I \to Y : \alpha\beta^{-1}(s) = \begin{cases} \alpha(2s), & s \in [0,\frac{1}{2}] \\ \beta^{-1}(2s-1) = \beta(2-2s), & s \in [\frac{1}{2},1] \end{cases}$$

is a loop based at $\alpha\beta^{-1}(0) = \alpha\beta^{-1}(1) = y_0$. By the hypothesis,

$$\psi_*(\alpha\beta^{-1}) \in p_*\big(\pi_1(\widetilde{X},\widetilde{x}_0)\big).$$

In other words, there is an element of $\pi_1(\widetilde{X},\widetilde{x}_0)$ that is send to $\alpha\beta^{-1}$ through p_*, i.e. a class of loops in $\widetilde{X}$ based at $\widetilde{x}_0$ that are lifts of the homotopy class of $\alpha\beta^{-1}$. Since ψ_* is a homomorphism,

$$\psi_*(\alpha\beta^{-1}) = \psi_*(\alpha)\psi_*(\beta)^{-1}$$

and the product of the lifts of $\psi_*(\alpha)$ and $\psi_*(\beta)^{-1}$ is a lift of $\psi_*(\alpha\beta^{-1})$, i.e. a loop based at $\widetilde{x}_0$; therefore, the lifts of $\psi_*(\alpha)$ and $\psi_*(\beta)^{-1}$ must have the same endpoints and so must the lifts of $\psi_*(\alpha)$ and $\psi_*(\beta)$. In other words, the lifts $\widetilde{\psi\alpha}$ and $\widetilde{\psi\beta}$ of $\psi\alpha$ and $\psi\beta$ respectively, have the property

$$\widetilde{\psi\alpha}(0) = \widetilde{\psi\beta}(0) \quad \text{and} \quad \widetilde{\psi\alpha}(1) = \widetilde{\psi\beta}(1).$$

Therefore, $\widetilde{\psi}$ is well define.

Obviously $\widetilde{\psi}$ is a lift of ψ since

$$p\widetilde{\psi}(y) = p\widetilde{\psi f_y}(1) \overset{p\widetilde{\psi f_y}=\psi f_y}{=} \psi f_y(1) \overset{f_y(1)=y}{=} \psi(y).$$

It remains to show that $\widetilde{\psi}$ is continuous. Take some open set $U \subseteq \widetilde{X}$ and some $y \in \widetilde{\psi}^{-1}(U)$. In this case U is an open neighborhood of $\widetilde{\psi}(y)$. We will show that $\widetilde{\psi}^{-1}(U) \subseteq Y$ is open by showing that there is an open neighborhood of y that is contained in $\widetilde{\psi}^{-1}(U)$.

p is open so $p(U)$ is an open neighborhood of $p(\widetilde{\psi}(y)) = \psi(y)$. Consider the open neighborhood $U_{\psi(y)}$ of $\psi(y)$ in the definition of a covering space and take $U' = U_{\psi(y)} \cap p(U)$.

Note that since $U_{\psi(y)}$ is evenly covered, so is U' (if the sheets of $U_{\psi(y)}$ are $\{Z_a\}$ then the sheets of U' are $\{Z_a \cap U_{\psi(y)}\}$). Therefore, if $p^{-1}(U') = \bigcup_{a\in A} V_a$ then, since $\widetilde{\psi}(y) \in U'$, there is some $a \in A$ such that $\widetilde{\psi}(y) \in V_a$. Take $W = U \cap V_a$. W is by its definition an open neighborhood of $\widetilde{\psi}(y)$.

Using again the fact that p is open, $p(W) \subseteq X$ is open and since ψ is continuous, $\psi^{-1}(p(W))$ is open in Y. Obviously $y \in \psi^{-1}(p(W))$ since $\psi(y) = p(\widetilde{\psi}(y)) \in p(W)$. Note that $p(W) \subseteq p(U) \subseteq p(U')$ and $p(W)$ is therefore evenly covered since $p(U')$ is.

We now use the fact that Y is locally path connected. So there is some path connected open set $V \subseteq Y$ such that $y \in V \subseteq \psi^{-1}(p(W))$.

Finally, we claim that $\widetilde{\psi}(V) \subseteq U$. Clearly the element $\widetilde{\psi}(y)$ of $\widetilde{\psi}(V)$ is in U. Let $y' \in V$. Since V is path connected, there is a path $f : I \to V$ from y to y'. By definition,

$$\widetilde{\psi}(y') = \widetilde{\psi f_y}(1) = \widetilde{\psi f}(1)$$

where $\widetilde{\psi f}$ is the unique lift of the path ψf that starts at $\widetilde{\psi}(y)$. Now

$$f(I) \subseteq V \Rightarrow \psi f(I) \subseteq \psi(V) \subseteq p(W) \Rightarrow \widetilde{\psi f} \in p^{-1}(p(W)).$$

But as we mentioned, $p(W)$ is evenly covered, say $p(W) = \bigcup_{j\in J} W_j$, with the W_j's being disjoint and for some $k \in J$, $W = W_k$. Since $\widetilde{\psi f}(0) = \widetilde{\psi}(y) \in W$, this implies that $\widetilde{\psi f}(1) = \widetilde{\psi}(y') \in W \subseteq U$. Therefore $\widetilde{\psi}(V) \subseteq U$, hence $y \in V \subseteq \widetilde{\psi}^{-1}(U)$. So $\widetilde{\psi}$ is indeed continuous. ◇

Note that Y was also assumed to be *locally path connected*.

Convention. From now on, all coverings are assumed to be *locally path connected* as well.

Recall however that *if $f : X \to Y$ is a local homeomorphism then X is locally path connected if and only if $f(X)$ is*. Since covering maps are *surjective* local homeomorphisms, restricting our attention to locally path connected coverings means that we consider locally path connected base spaces as well.

Therefore we assume that *all* topological spaces are locally path connected.

Corollary 2.3.3. *Suppose X is a topological space and fix some $x_0 \in X$. An isomorphism ϕ between two coverings $p_1 : (\widetilde{X}_1, \widetilde{x}_1) \to (X, x_0)$ and $p_2 : (\widetilde{X}_2, \widetilde{x}_2) \to (X, x_0)$ such that $\phi(\widetilde{x}_1) = \widetilde{x}_2$ exists if and only if*

$$p_{1*}\big(\pi_1(\widetilde{X}_1, \widetilde{x}_1)\big) = p_{2*}\big(\pi_1(\widetilde{X}_2, \widetilde{x}_2)\big)$$

If such an isomorphism exists, it is unique.

Proof. Take $(\widetilde{X}, \widetilde{x}_0) = (\widetilde{X}_1, \widetilde{x}_1)$ and $(Y, y_0) = (\widetilde{X}_2, \widetilde{x}_2)$ in the previous proposition to find liftings ϕ and ψ of p_1 and p_2 respectively. These are the required isomorphisms. ◇

The above criterion, however useful, is only concerned with the existence of a *specific* isomorphism; the one which preserves the given basepoints. What about the existence of an isomorphism in general?

Corollary 2.3.4. *Let X be a topological space and $x_0 \in X$. Two coverings $(\widetilde{X}_1, p_1)$ and $(\widetilde{X}_2, p_2)$ of X are isomorphic if, and only if, for any two elements $\widetilde{x}_1 \in \widetilde{X}_1$ and $\widetilde{x}_2 \in \widetilde{X}_2$ such that $p_1(\widetilde{x}_1) = p_2(\widetilde{x}_2) = x_0$, the subgroups $p_{1*}\big(\pi_1(\widetilde{X}_1, \widetilde{x}_1)\big)$ and $p_*\big(\pi_2(\widetilde{X}_2, \widetilde{x}_2)\big)$ of $\pi_1(X, x_0)$ are conjugate.*

Proof. Immediate from the above Corollary and the fact that changing the basepoint of a covering space amounts to going to a conjugate subgroup of $\pi_1(X, x_0)$. ◇

EXISTENCE We now come to the existence part of the correspondence. We want to examine if the correspondence is surjective; that is, given a (path connected and locally path connected) space X, some element $x_0 \in X$ and a *conjugacy class of subgroups* of $\pi_1(X, x_0)$, is there a (path connected and locally path connected) covering $(\widetilde{X}, p)$ of X so that $p_*(\pi_1(\widetilde{X}, \widetilde{x}_0))$ belongs to the given conjugacy class for some $\widetilde{x}_0 \in p^{-1}(x_0)$?

Example 2.3.5. The conjugacy class to which the whole group belongs is

$$\{g\pi_1(X, x_0)g^{-1} : \ g \in \pi_1(X, x_0)\} = \{\pi_1(X, x_0)\}$$

This case is trivial; the required covering is (X, id_X) or, more generally, (Y, h) where $h : Y \to X$ is a homeomorphism. This covering is called the **trivial covering** of X. Note that from Corollary 2.3.4, the trivial covering is unique up to isomorphism.

Example 2.3.6. The other extreme case, the conjugacy class of the trivial subgroup,

$$\{g\{1\}g^{-1} : \ g \in \pi_1(X, x_0)\} = \{1\},$$

is not only non-trivial but, as we shall see, it is fundamental to our study. In this case we search for a covering of X whose fundamental group is trivial. Since we have limited ourselves to path connected spaces, we search for a *simply connected* covering of X.

Definition 2.3.7. A *simply connected* covering of a space X is called a **universal cover**.

Example 2.3.8. $(\mathbb{R}, q)$ is a universal cover of S^1.

An immediate consequence of Corollary 2.3.4 is that

Proposition 2.3.9. *If the universal cover of a space exists, it is unique up to covering space isomorphism.*

The reason it is called *universal* cover is justified by the following results.

Proposition 2.3.10. *If $(\widetilde{X}_1, p_1)$ and $(\widetilde{X}_2, p_2)$ are two locally path connected coverings of a space X and $\phi : \widetilde{X}_1 \to \widetilde{X}_2$ is a morphism of covering spaces, then $(\widetilde{X}_1, \phi)$ is a covering of $\widetilde{X}_2$.*

Corollary 2.3.11. *If (Y, q) is a universal cover of X and $(\widetilde{X}, p)$ is any other covering then there is a morphism of covering spaces $\phi : Y \to \widetilde{X}$ and (Y, ϕ) is a covering of $\widetilde{X}$.*

Since the above results are of no immediate interest to us, we refer the reader to the bibliography. The important thing to note is that the universal cover *covers every other covering*, hence its name.

We will see that if X has a universal cover, we can use it to construct coverings whose fundamental groups belong to any given conjugacy class of subgroups of $\pi_1(X, x_0)$. Thus the first step is to *determine which spaces have a universal cover*.

Definition 2.3.12. A topological space X is called **semilocally simply connected** if for every $x \in X$ we can find an open neighborhood U_x of x so that every loop in U_x is homotopic to a constant loop or, equivalently, if the inclusion $i : U_x \to X$ induces the trivial homomorphism $i_* : \pi_1(U_x, x) \to \pi_1(X, x)$.

Proposition 2.3.13. *Suppose X is a path connected, locally path connected space. Then X has a universal cover if and only if it is semilocally simply connected.*

Proof. ($\Rightarrow$) The one direction if fairly simple. If (Y, q) is a universal cover of X and $x \in X$, then the neighborhood U_x of x described in the definition of a covering space is the required neighborhood in the definition of a semilocally simply connected space.

($\Leftarrow$) For the contrary, fix some $x_0 \in X$. Define

$$Y = \{[f] : f \text{ is a path in } X \text{ starting at } x_0\},$$

that is, Y is the set of all homotopy classes of paths strarting at $x_0 \in X$, and

$$q : Y \to X : [f] \mapsto q([f]) = f(1),$$

i.e., q sends each homotopy class of paths starting at x_0 to the common terminal point. Observe that q is surjective since X is path connected.

Let us define a topology on Y. Since X is locally path connected and semilocally simply connected, we can find a basis

$$\mathcal{B} = \{U_i : i \in I\}$$

for the topology of X consisting of path connected sets such that $i_* : \pi_1(U_i, x) \to \pi_1(X, x)$ is trivial. For every $[f] \in Y$ and every $U \in \mathcal{B}$ with $f(1) \in U$, define

$$U_f = \{[g] \in Y : [g] = [f][\gamma] \text{ for some path } \gamma : I \to U : \gamma(0) = f(1)\}.$$

The elements of U_f are paths that first traverse f or some homotopic path of f, and then some path in U *which starts at* $f(1)$. The set

$$\mathcal{B}' = \{U_f : [f] \in Y,\ U \in \mathcal{B} : f(1) \in U\}$$

is a basis for a topology on Y. Indeed, for any two U_f, V_g and any $h \in U_f \cap V_g$, $h(1) \in U \cap V$. Find a basic open set $W \in \mathcal{B}$ with the property $h(1) \in W \subseteq U \cap V$. Then $W_h \subseteq U_f \cap V_g$.

Before continuing, lets establish two important properties of the map q. First of all, for every $[f] \in Y$ and every $U \in \mathcal{B}$ with $f(1) \in U$, the restriction

$$q|_{U_f} : U_f \to U$$

is bijective. If $[g], [h] \in U_f$ such that $q|_{U_f}([g]) = q|_{U_f}([h])$ then $[g] = [h]$. Indeed,

$$[g], [h] \in U_f \implies [g] = [f][f_1] \text{ and } [h] = [f][f_2]$$

for some $f_1, f_2 : I \to U$ with $f_1(0) = f_2(0) = f(1)$. So if

$$q|_{U_f}([g]) = q|_{U_f}([h]) \implies g(1) = h(1) \implies f_1(1) = f_2(1)$$

in which case we have two paths $f_1, f_2 : I \to U$ with the same ends. Given that $i_* : \pi_1(U_i, x) \to \pi_1(X, x)$ is trivial, $[f_1] = [f_2]$ so $[g] = [h]$ and $q|_{U_f}$ is injective. Moreover, every U is path connected, so for every $x \in U$, there is a path $f_3 : I \to U$ from x to $f(1) \in U$. Then the image of $[g] = [f][f_3] \in U_f$ through $q|_{U_f}$ is x and therefore $q|_{U_f}$ is surjective.

Secondly, it is clear that if $U \in \mathcal{B}$ and $x \in U$ then

$$q^{-1}(U) = \bigcup_{\lambda \in \Lambda} U_{f_\lambda} \tag{2.4}$$

where the union is taken over all path classes $[f_\lambda]$ in X from x_0 to x.

Given these results, we can show that q is a local homeomorphism. Take some $[f] \in Y$ and the open neighborhood U_f of $[f]$ for some $U \in \mathcal{B}$ with $f(1) \in U$. Now $q|_{U_f}$ is bijective, therefore $q(U_f) = q|_{U_f}(U_f) = U$ is open in X. It remains to show that $q|_{U_f}$ is a homeomorphism. We showed that is bijective. From (2.4) q is continuous, hence so is the restriction $q|_{U_f}$. Lastly, any open subset V of U_f is also open in Y since U_f is open in Y; so it can be written as a union of sets $V_g \in \mathcal{B}'$ with $V_g \subseteq U_f$. Therefore, using the fact that $q|_{U_f}$ is injective,

$$q|_{U_f}(V) = q|_{U_f}\left(\bigcup_g V_g\right) = \bigcup_g q|_{U_f}(V_g) = \bigcup_g q|_{V_g}(V_g) = \bigcup_g V$$

which is open in X. To summarize, $q|_{U_f}$ is bijective, continuous and open, hence a homeomorphism. The choice of $[f] \in Y$ was arbitrary, therefore q is a local homeomorphism.

We now show that Y is connected. In particular we will show that there is a path from $[c_{x_0}] \in Y$ to any $[f] \in Y$. Let $[f] \in Y$. For every $s \in I$ define

$$f_s : I \to Y : f_s(t) = f(st).$$

It is immediate that $f_0 = c_{x_0}$ and $f_1 = f$. The required path is then

$$F : I \to Y : s \mapsto [f_s]$$

which starts at $F(0) = [f_0] = [c_{x_0}]$ and ends at $F(1) = [f_1] = [f]$. We only need to show that F is continuous, that is, for every $s \in I$ and every open basic neighborhood U_{f_s} of f_s there is an open interval J_1 open in I such that $s \in J_1 \subseteq I$ and $F(J_1) \subseteq U_{f_s}$. Take any $s \in I$ and any open basic neighborhood U_{f_s} of f_s. In particular U is a basic open neighborhood of $f_s(1)$. Since f_s is a path, hence continuous, and U is an open neighborhood of $f_s(1)$, there is an interval J_2 open in I such that $1 \in J_2 \subseteq I$ and $f_s(J_2) \subseteq U$. This interval is also the required interval J_1.

Finally, it remains to show that Y is simply connected, that is $\pi_1(Y, [c_{x_0}])$ is the trivial subgroup. Through the induced monomorphism

$$q_* : \pi_1(Y, [c_{x_0}]) \to \pi_1(X, x_0),$$

$\pi_1(Y, \widetilde{x}_0)$ is isomorphic to $q_*(\pi_1(Y, [c_{x_0}]))$ which is the stabilizer of $[c_{x_0}]$ for the action of $\pi_1(X, x_0)$ on $q^{-1}(x_0)$. Let's compute this stabilizer. Suppose $[f] \in \pi_1(X, x_0)$ such that

$$[f][c_{x_0}] = \widetilde{F}(1) = [c_{x_0}]$$

where $\widetilde{F}$ is a path in Y starting at $[c_{x_0}]$ and ending at $[c_{x_0}]$. So $\widetilde{F}$ is the constant path which is the lift of the constant path c_{x_0} in X. $\diamond$

The above proposition suggests one last restriction we must impose if we want to have a bijective correspondence; that is, we must require spaces to be semilocally simply connected.

Proposition 2.3.14. *Let X be a topological space which is path connected, locally path connected and semilocally simply connected and $x_0 \in X$. Then for any conjugacy class of subgroups of $\pi_1(X, x_0)$ there exists a covering of X whose fundamental group belongs to the given class.*

Proof. Suppose (Y, q) is the universal cover of X; since the universal cover is unique up to isomorphism, we can take Y to be the space constructed in the previous proposition. Take any conjugacy class of $\pi_1(X, x_0)$ and a subgroup $H \leqslant \pi_1(X, x_0)$ that belongs to the given class.

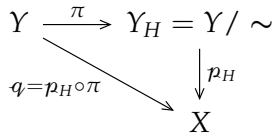
Define a relation $\sim \subseteq Y \times Y$ as

$$[f] \sim [g] \iff f(1) = g(1) \text{ and } [f][g]^{-1} \in H.$$

The fact that H is a subgroup implies that $\sim$ is an equivalence relation. Indeed, $[f] \sim [f]$ since $f(1) = f(1)$ and $[f][f]^{-1} = [ff^{-1}] = [c_{x_0}] \in H$ so $\sim$ is reflexive. Additionally, if $[f] \sim [g]$ then $f(1) = g(1)$ and $[f][g]^{-1} \in H$. Since H is a subgroup, $[g][f]^{-1} = ([f][g]^{-1})^{-1} \in H$ as well so $[g] \sim [f]$. Finally, given that $[f] \sim [g]$ and $[g] \sim [h]$ we have $f(1) = g(1) = h(1)$ and $[f][h]^{-1} = [f][g]^{-1}[g][h]^{-1} \in H$ because both $[f][g]^{-1}$ and $[g][h]^{-1}$ belong in H and H is a subgroup of $\pi_1(X, x_0)$.

Take the quotient space $Y_H = Y/\sim$ endowed with the quoetient topology induced by the projection $\pi : Y \to Y/\sim$, whose elements we shall write as $[[f]]$, and the map p_H induced by q, i.e.

$$p_H : Y_H \to X : [[f]] \mapsto f(1).$$



We will show that (Y_H, p_H) is a covering of X and that

$$p_{H*}\big(\pi_1(Y_H, y_0)\big) = H$$

for some $y_0 \in Y$.

First of all, the fact that (Y_H, p_H) is a covering of X is immediate since (Y, q) is a covering of X and $q = p_H \pi$. Since q is a continuous surjective map and $q = p_H \pi$, p_H is also continuous and surjective. Following the proof of the previous proposition, we can show that every $x \in X$ has an evenly covered neighborhood U_x through p_H. The only thing to note is that $[[f]] = [[g]]$ if and only if $[[f\gamma]] = [[g\gamma]]$ for some path $\gamma : I \to U_x$ starting at $f(1) = g(1)$. This immediate from the fact that $[f][g]^{-1} = [f][\gamma][\gamma]^{-1}[g]^{-1}$.

Lastly, $p_{H*}\big(\pi_1(Y_H, y_0)\big) = H$. Indeed, take $y_0 = [[c_{x_0}]] \in Y_H$ and obsverve that $[f] \in p_{H*}\big(\pi_1(Y_H, y_0)\big)$ if and only if $[f] = [p_H \circ \widetilde{f}]$ for some $[\widetilde{f}] \in \pi_1(Y_H, y_0)$. By definition, $\widetilde{f}$ is a lift of f in Y_H that starts at $y_0 = [[c_{x_0}]]$ and ends at $[[f]]$ since $[p_H \circ \widetilde{f}] = [f]$. So the image of $\widetilde{f}$ is f, a loop in X, if and only if $[[f]] = [[c_{x_0}]]$ if and only if, by definition, $[f][c_{x_0}]^{-1} = [f] \in H$. ◇

We have arrived at the next classification theorem.

Theorem 2.3.15 (Classification of Covering Spaces; part 1)**.** *If X is path connected, locally path connected and semilocally simply connected topological space and $x_0 \in X$, then there is a bijective correspondence*

$$\left\{\begin{array}{c}\textit{path connected,}\\ \textit{locally path connected}\\ \textit{covering spaces of } X\end{array}\right\} \rightleftarrows \left\{\begin{array}{c}\textit{conjugacy classes of}\\ \textit{subgroups of } \pi_1(X, x_0)\end{array}\right\}$$
$$\widetilde{X} \;\mapsto\; \pi_1(\widetilde{X}, \widetilde{x}_0)$$

between path connected and locally path connected coverings of X and conjugacy classes of subgroups of $\pi_1(X, x_0)$.

Proof. Immediate from Corollary 2.3.4 and Proposition 2.3.14. ◇

Observe the similarity of this theorem to Theorem 1.5.9. We can also prove an analogue of Theorem 1.5.10.

Definition 2.3.16. A covering space $(\widetilde{X}, p)$ of X is called a **Galois cover** if for every $x_0 \in X$ and every $\widetilde{x}_1, \widetilde{x}_2 \in p^{-1}(x_0)$ there is some $\sigma \in A(\widetilde{X}, p)$ such that $\sigma(\widetilde{x}_1) = \widetilde{x}_2$.

The name "Galois cover" comes from the similarity of the notions discussed to the notions of Galois Theory.

Recall from Group Theory that given a group G and a subgroup $H \leqslant G$, the **normalizer** of H in G is

$$N(H) = \{g \in G : g^{-1}Hg = H\}.$$

Theorem 2.3.17 (Classification of Covering Spaces; part 2)**.** *Suppose X is a path connected, locally path connected and semilocally simply connected space and $p : (\widetilde{X}, \widetilde{x}_0) \to (X, x_0)$ a path connected, locally path connected covering of X. Set $H = p_*(\pi_1(\widetilde{X}, \widetilde{x}_0))$. Then*

i) The covering is Galois if and only if H is normal in $G = \pi_1(X, x_0)$.

ii) $A(\widetilde{X}, p)$ is isomorphic to $N(H)/H$, where $N(H)$ is the normalizer of H in G.

iii) If $(\widetilde{X}, p)$ is Galois then $A(\widetilde{X}, p)$ is isomorphic to G/H. In particular, $A(Y, q)$ is isomorphic to G; here (Y, q) is the universal cover of X.

Proof. i) By the definition of the correspondence, the conjugacy class of H is exactly $\{p_*(\pi_1(\widetilde{X}, \widetilde{x})) : \widetilde{x} \in p^{-1}(x_0)\}$. Therefore $H \lhd G$ if and only if the above set is the singleton $\{H\}$ if and only if $p_*(\pi_1(\widetilde{X}, \widetilde{x}_1)) = p_*(\pi_1(\widetilde{X}, \widetilde{x}_2))$ for every $\widetilde{x}_1, \widetilde{x}_2 \in p^{-1}(x_0)$ if and only if, by Proposition 2.3.2, there is some $\sigma \in A(\widetilde{X}, p)$ such that $\sigma(\widetilde{x}_1) = \widetilde{x}_2$ for every $\widetilde{x}_1, \widetilde{x}_2 \in p^{-1}(x_0)$ which is the definition of the covering space being Galois.

ii) Define

$$\vartheta : N(H) \to A(\widetilde{X}, p)$$

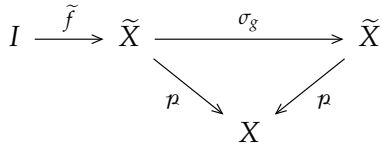
as follows. For each $[g] \in N(H)$, take a lift $\widetilde{g}$ of g starting at $\widetilde{x}_0$ and set $\widetilde{x}_1 = \widetilde{g}(1)$. Since $\widetilde{g}$ is a lift of g, $\widetilde{x}_0, \widetilde{x}_1 \in p^{-1}(x_0)$ and therefore

$$p_*(\pi_1(\widetilde{X}, \widetilde{x}_1)) = [g]^{-1}H[g] \overset{[g]\in N(H)}{=} H.$$

From Proposition 2.3.2, there is some $\sigma_g \in A(\widetilde{X}, p)$ with $\sigma_g(\widetilde{x}_0) = \widetilde{x}_1$. Set $\vartheta([g]) = \sigma_g$. Obviously ϑ is well defined since the choice of $\widetilde{x}_1$ depends only on the homotopy class $[g]$.

ϑ is a group homomorphism. Suppose $[g], [f] \in N(H)$ and $\widetilde{g}, \widetilde{f}$ are the lifts of g and f respectively starting at $\widetilde{x}_0$. Take their images $\sigma_g, \sigma_f \in A(\widetilde{X}, p)$. Since $\widetilde{g}(1) = \widetilde{x}_1 = \sigma_g(\widetilde{x}_0) = \sigma_g\widetilde{f}(0)$, we can define the path $\widetilde{g} \cdot (\sigma_g \circ \widetilde{f})$ which is a lift of $g \cdot f$. Indeed,

$$p(\widetilde{g} \cdot (\sigma_g \circ \widetilde{f})) = p\widetilde{f} \cdot p(\sigma_g \circ \widetilde{f}) \overset{p\circ\sigma_g = p}{=} g \cdot p\widetilde{f} = g \cdot f.$$



The endpoint of the lift $\widetilde{g} \cdot (\sigma_g \circ \widetilde{f})$ of gf is

$$\widetilde{g} \cdot (\sigma_g \circ \widetilde{f})(1) = \sigma_g \circ \widetilde{f}(1) = \sigma_g\sigma_f(\widetilde{x}_0) = \sigma_{gf}(\widetilde{x}_0).$$

Therefore $\sigma_{gf} = \sigma_g\sigma_f$ and ϑ is actually a homomorphism.

ϑ is onto. Let $\sigma \in A(\widetilde{X}, p)$ with $\sigma(\widetilde{x}_0) = \widetilde{x}_1$. Take a path $\widetilde{g} : I \to \widetilde{X}$ from $\widetilde{x}_0$ to $\widetilde{x}_1$. The composite $p \circ \widetilde{g}$ is a loop in X based at x_0 since $p\widetilde{g}(0) = p(\widetilde{x}_0) = x_0 = p(\widetilde{x}_1) = p\widetilde{g}(1)$. We will show that $[g] \in N(H)$, in which case $\vartheta([g]) = \sigma$ according to the definition of ϑ. Since $\sigma(\widetilde{x}_0) = \widetilde{x}_1$ and $\widetilde{g}$ is a lift of $g = p \circ \widetilde{g}$ from $\widetilde{x}_0$ to $\widetilde{x}_1$, we have

$$p_*\big(\pi_1(\widetilde{X}, \widetilde{x}_1)\big) = H \quad \text{and} \quad p_*\big(\pi_1(\widetilde{X}, \widetilde{x}_1)\big) = [g]^{-1}H[g].$$

Therefore $[g]^{-1}H[g] = H$ and $[g] \in N(H)$ as desired.

Lastly, we will compute the kernel of ϑ. We have

$$\begin{aligned}[g] \in \ker\vartheta &\Leftrightarrow \sigma_g = \mathrm{id}_{\widetilde{X}} \Leftrightarrow \widetilde{g}(0) = \widetilde{g}(1) \Leftrightarrow \widetilde{x}_0 = \widetilde{x}_1 \\ &\Leftrightarrow \text{the lift } \widetilde{g} \text{ of } g \text{ is a loop based at } \widetilde{x}_0 \\ &\Leftrightarrow [g] \in \operatorname{Im} p_* = H.\end{aligned}$$

Therefore $\ker\vartheta = H$ and the assertion follows from the First Isomorphism Theorem.

iii) Immediate from the above and the fact that the universal cover corresponds to the trivial subgroup.

◇

Part II

THE UNIFYING CONTEXT: GROTHENDIECK'S GALOIS THEORY

The similarity of the classification theorems of Galois Theory and Algebraic Topology indicate some sort of connection between them. Indeed, there is a deeper bond between them and it was Alexander Grothendieck (1928–2014) who understood and formulated it in his *Séminaire de Géométrie Algébrique* (*SGA1*, [12]). Grothendieck, inspired by the classification of covering spaces, established a similar classification theorem for *schemes* from which he deduced the classification theorems we have seen so far as special cases! His theory, known as *Galois Theory for Schemes* or *Grothendieck's Galois Theory*, provides a context within Algebraic Geometry in which Galois Theory and Algebraic Topology can be studied.

The introduction of the cipher 0 *or the group concept was general nonsense too, and mathematics was more or less stagnating for thousands of years because nobody was around to take such childish steps.*

Alexander Grothendieck.

3 GROTHENDIECK'S GALOIS THEORY

Having seen that the classification theorems of Galois Theory and Algebraic Topology are remarkably similar, our next step is to formulate the context in which the can be unified, that is, Grothendieck's Galois Theory. We are particulary interested in how classical Galois Theory can be studied through Grothendieck's approach.

We begin by first establishing the relevant notions from Category Theory. We then proceed to look at Galois Theory is seen from Grothendieck's perspective. Lastly, following a short discussion on schemes and other relevent notions from Algebraic Geometry, we state the main theorem of Grothendieck's Galois Theory in its full generality and see how it is related to Galois Theory as well. Unfortunately our study ends there. The interested reader who wishes to see a proof of this deep theorem and its relation to Algebraic Topology can consult the bibliography.

This chapter's exposition is of course faster and more advanced than the previous. Although we will state the relevant definitions and results from Category Theory and Algebraic Geometry we are going to need, we assume the reader has a good understanding of both principles.

3.1 CATEGORY THEORY ESSENTIALS

CATEGORIES Category Theory was introduced by Samuel Eilenberg (1913–1998) and Saunders MacLane (1909–2005) in their paper "General theory of natural equivalences" [11] for their foundational work on Algebraic Topology and Homological Algebra. Since then, Category Theory has evolved far beyond the context in which it was initially intended to be used. Our study of Category Theory is of

course much more limited. In this section we introduce only some of the fundamental notions of Category Theory we are going to need and see how they relate to our work so far.

Definition 3.1.1. A **category** $\mathcal{C}$ consists of

- A collection $\mathrm{obj}(\mathcal{C})$ whose elements are called **objects**. We shall use the letters $A, B, C, \ldots$ to denote objects. We write $A \in \mathcal{C}$ instead of $A \in \mathrm{obj}(\mathcal{C})$ to indicate that A is an object in $\mathcal{C}$.
- A collection $\mathcal{M}_{\mathcal{C}}$ which is a union of the form

$$\mathcal{M}_{\mathcal{C}} = \bigcup_{A,B \in \mathcal{C}} \mathrm{Mor}_{\mathcal{C}}(A, B)$$

  where for every $A, B \in \mathcal{C}$, $\mathrm{Mor}_{\mathcal{C}}(A, B)$ is a collection whose elements are called **morphisms** from A to B. We shall use the letters $f, g, h, \ldots$ to denote morphisms. We write $f : A \to B$ or $A \xrightarrow{f} B$ to indicate that $f \in \mathrm{Mor}_{\mathcal{C}}(A, B)$. The object A is called the **domain** of f and the object B the **codomain** of f.
- For every $A, B, C \in \mathcal{C}$, a **law of composition**

$$\begin{aligned} \circ : \mathrm{Mor}_{\mathcal{C}}(A, B) \times \mathrm{Mor}_{\mathcal{C}}(B, C) &\to \mathrm{Mor}_{\mathcal{C}}(A, C) \\ (f, g) &\mapsto g \circ f \equiv gf \end{aligned}$$

So that

i) The compositions are associative; that is, for every set of objects and every set of arrows between them

$$A \xrightarrow{f} B \xrightarrow{g} C \xrightarrow{h} D,$$

we have that $h(gf) = (hg)f$.

ii) For every object $A \in \mathcal{C}$, there is a morphism $1_A \in \mathrm{Mor}_{\mathcal{C}}(A, A)$, called the **identity** morphism on A, such that

$$1_A f = f \quad \text{and} \quad g 1_A = g$$

for every morphism f whose *codomain* is A and every morphism g whose *domain* is A.

Definition 3.1.2. Let $\mathcal{C}, \mathcal{D}$ be two categories. We say that $\mathcal{D}$ is a **subcategory** of $\mathcal{C}$ if

i) The objects of $\mathcal{D}$ are objects of $\mathcal{C}$, i.e. $\text{obj}(\mathcal{D}) \subseteq \text{obj}(\mathcal{C})$.

ii) For every two objects $A, B \in \mathcal{D}$, $\text{Mor}_{\mathcal{D}}(A, B) \subseteq \text{Mor}_{\mathcal{C}}(A, B)$ and the identity mapping of A in $\mathcal{D}$ is the identity mapping of A in $\mathcal{C}$.

iii) Composition in $\mathcal{D}$ is the composition in $\mathcal{C}$.

This abstract definition becomes clear with the first examples of categories which reveal that categories come up in every area of Mathematics.

Example 3.1.3. The most famous example is the *category* **Set** *of sets* whose objects are sets and morphisms are the function between sets.

Familiar examples from Algebra include the *category* **Grp** *of groups* whose elements are groups and morphisms are group homomorphisms, the *category* **Ab** *of abelian groups* whose objects are abelian groups and morphisms are group homomorphisms, the *category* **Rng** *of rings* whose objects are rings and morphisms are ring homomorphisms, the *category* $\mathbf{Vec}_k$ *of vector spaces over a field* k whose objects are vector spaces over k and morphisms are k-linear maps between them and the *category* ${}_R\mathbf{Mod}$ (resp. $\mathbf{Mod}_R$) *of left* (resp. right) *modules over a ring* R whose objects are left (resp. right) R-modules and morphisms are R-linear maps between them.

Familiar examples fom Analysis include the *category* **Top** *of topological spaces* whose objects are topological spaces and morphisms are the continuous maps, the *category of measurable spaces* whose objects are measurable spaces and morphisms are measurable functions between them and the *category of metric spaces* whose objects are metric spaces and morphisms are distant preserving functions.

An example from Geometry is *the category of smooth manifolds* whose objects are smooth manifolds and morphisms are C^∞ maps.

Example 3.1.4. The *category of fields* whose objects are fields and non-zero morphisms are the field extensions.

Example 3.1.5. Suppose k is a field. A **k-algebra** A is a ring that is also a k-vector space and the ring structure and vector space structure are compatible in the sense that

$$r(ab) = (ra)b = a(rb) \quad \forall r \in k, \forall a, b \in A.$$

A map $f : A \to B$ between two k-algebras is called an **k-algebra homomorphism** if it is an k-linear ring homomorphism. The **dimension of a k-algebra** is just its dimension as a k-vector space.

For example, every field extension of k can be viewed as a k-algebra; the compatibility condition is just a combination of commutativity and associativity of multiplication.

We can then form the *category of k-algebras* whose objects are k-algebras and morphisms are k-algebra homomorphisms.

Of particular interest will be the subcategory $\mathbf{Sep}_k$ of separable field extensions of k (which can also be viewed as k-algebras).

Example 3.1.6. The *category* $\mathbf{Top}_*$ *of pointed topological spaces* whose objects are pairs of the form (X, x) where X is a topological space and $x \in X$ and morphisms are the continuous maps that preserve the basepoint, i.e. $f : (X, x) \to (Y, y)$.

Example 3.1.7. Let G be a group. Recall that a **G-set** is a set X equipped with a (left) G-action and a function $f : X \to Y$ between two G-sets is called a **G-map** if $f(gx) = gf(x)$ for all $g \in G$ and $x \in X$. We can then form the *category of G-sets* whose objects are G-sets and morphisms are G-maps.

Example 3.1.8. Let G be a *topological group* and X a G-set that is also a topological space. We say that the action of G on X is **continuous** if the map $G \times X \to X$ is continuous. We can form the category **G-set** of sets equipped with a continuous G-action. The morphisms in this category are the continuous G-maps.

For our study, of particular interest is the subcategory $\mathbf{G\text{-}set}_f$ of *finite* sets equipped with a continuous G-action and continuous G-maps among them as well as the subcategory $\mathbf{G\text{-}set}_f^t$ of *finite* sets equipped with a continuous and *transitive* G-action and continuous G-maps among them.

The above categories are more or less familiar and we can surely think many more examples of categories whose objects are mathematical structures and morphisms are structure preserving maps. But the concept of a category is much more general.

Example 3.1.9. If $(X, \leqslant)$ is a partially ordered set, then X can be considered as a category. The objects of this category are the elements $x \in X$ and for every $x, y \in X$ the morphisms from x to y are

$$\mathrm{Mor}(x,y) = \begin{cases} \{(x,y)\}, & \text{if } x \leqslant y \\ \emptyset, & \text{otherwise.} \end{cases}$$

Example 3.1.10. Suppose $k \leqslant E$ is a Galois extension. The lattice of all subextensions of $k \leqslant E$ is a partially ordered set with

$$L_1 \leqslant L_2 \iff L_1 \subseteq L_2 \quad \forall\, k \leqslant L_1, L_2 \leqslant E$$

hence a category. When $L_1 \leqslant L_2$, the morphism $L_1 \overset{(L_1,L_2)}{\to} L_2$ can be considered to be the inclusion map. We shall denote this category by $\mathscr{L}$.

Example 3.1.11. Suppose again that $k \leqslant E$ is a Galois extension. Similarly to the previous example, the lattice of all *closed*[1] subgroups of $\mathrm{Gal}(E/F)$ can be viewed as a category. We shall denote this category by $\mathscr{H}$.

Example 3.1.12. Let (X, τ_X) be a topological space. The lattice of *open subsets* of X partially ordered by inclusion is a category.

Definition 3.1.13. Two objects A, B in a category $\mathcal{C}$ are called **isomorphic** if there are morphisms $f : A \to B$ and $g : B \to A$ such that $gf = 1_A$ and $fg = 1_B$.

FUNCTORS The philoshophy of Category Theory is, roughly speaking, that maps are more important than structures. It is therefore natural to consider *maps between categories*.

Definition 3.1.14. Suppose $\mathcal{C}, \mathcal{D}$ are two categories. A (**covariant**) **functor** $F : \mathcal{C} \to \mathcal{D}$ consists of

[1] Of course we could have considered all subgroups of $\mathrm{Gal}(E/k)$ but for our purposes, we restrict ourselves to closed subgroups.

- A map $F : \text{obj}(\mathcal{C}) \to \text{obj}(\mathcal{D})$ that sends objects A of $\mathcal{D}$ to objects $F(A)$ of $\mathcal{D}$.
- A map $F : \mathcal{M}_{\mathcal{C}} \to \mathcal{M}_{\mathcal{D}}$ that sends for every $A, B \in \mathcal{C}$, every morphism $f : A \to B$ to a morphism $F(f) : F(A) \to F(B)$.

So that

i) For every composable pair f, g of morphisms in $\mathcal{C}$,

$$F(g \circ f) = F(g) \circ F(f).$$

ii) For every object $A \in \mathcal{C}$,

$$F(1_A) = 1_{F(A)}.$$

The above definition says that (covariant) functors preserve the structure of a category, i.e. they preserve objects, the way the objects are related (morphisms), the composition and identity morphisms. But there is another kind of map between categories that also preserves these. Only this time, the relation among the objects of $\mathcal{C}$ (the morphisms) are preserved in a "different" way.

Definition 3.1.15. Suppose $\mathcal{C}, \mathcal{D}$ are two categories. A **contravariant functor** $F : \mathcal{C} \to \mathcal{D}$ consists of

- A map $F : \text{obj}(\mathcal{C}) \to \text{obj}(\mathcal{D})$ that sends objects A of $\mathcal{D}$ to objects $F(A)$ of $\mathcal{D}$.
- A map $F : \mathcal{M}_{\mathcal{C}} \to \mathcal{M}_{\mathcal{D}}$ that sends for every $A, B \in \mathcal{C}$, every morphism $f : A \to B$ to a morphism $F(f) : F(B) \to F(A)$.

So that

i) For every composable pair f, g of morphisms in $\mathcal{C}$,

$$F(g \circ f) = F(f) \circ F(g).$$

ii) For every object $A \in \mathcal{C}$,

$$F(1_A) = 1_{F(A)}.$$

Let us see some examples of functors we encountered so far.

Example 3.1.16. Let $F \leqslant E$ be a Galois extension and $\mathrm{Gal}(E/F)$ its Galois group. The map

$$\mathrm{Gal}(E/\bullet) : \mathscr{L} \to \mathscr{H}$$

is *contravariant* functor; it reverses the inclusions which are the morphisms in each category. Similarly,

$$\mathrm{Fix}_E(\bullet) : \mathscr{H} \to \mathscr{L}$$

is also a *contravariant* functor.

Example 3.1.17. Consider the category $\mathbf{Top}_{\mathrm{path},*}$ of *path connected*, pointed topological spaces and the category **Grp** of groups. The map

$$\pi_1 : \mathbf{Top}_{\mathrm{path},*} \to \mathbf{Grp}$$

that sends each path connected, pointed topological space (X, x_0) to its fundamental group $\pi_1(X, x_0)$ and each continuous function

$$f : (X, x_0) \to (Y, y_0)$$

to the induced group homomorphism

$$\pi_1(f) = f_* : \pi_1(X, x_0) \to \pi_1(Y, y_0)$$

is a covariant functor.

EQUIVALENT CATEGORIES The most crucial notion we want to define is when two categories are considered the same from the categorical point of view.

Definition 3.1.18. A functor $F : \mathcal{C} \to \mathcal{D}$ is **essentially surjective** if for every object B of $\mathcal{D}$ there is an object A of $\mathcal{C}$ such that $F(A) \simeq B$.

The above definition is self evident. An essentially surjective funtor is... essentially surjective! Since category theory does not distinguish among isomorphic objects, essential surjecivity is exactly the notion we need; it implies that the functor maps the classes of isomorphic objects in $\mathcal{C}$ onto the classes of isomorphic objects in $\mathcal{D}$.

Definition 3.1.19. A (covariant) functor $F : \mathcal{C} \to \mathcal{D}$ is called **fully faithful** if for every $A_1, A_2 \in \mathcal{C}$, $F : \mathrm{Mor}(A, B) \to \mathrm{Mor}(F(A), F(B))$ is bijective. We can similarly define the notion of a fully faithful contravariant functor.

Definition 3.1.20. Two categories $\mathcal{C}, \mathcal{D}$ are called **equivalent** (resp. **antiequivalent**) if there is a fully faithful, essentially surjective covariant (resp. contravariant) functor $F : \mathcal{C} \to \mathcal{D}$.

Example 3.1.21. The category of subextensions of a Galois extension $k \leqslant E$ is antiequivalent to the category of subgroups of $\mathrm{Gal}(E/k)$, when both categories are viewed as posets.

Example 3.1.22. The category of path connected, locally path connected coverings of a path connected, locally path connected ans semilocally simply connected topological space X is antiequivalent to the category of subgroups of its fundamental group $\pi_1(X, x_0)$.

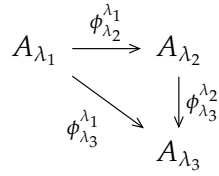
DIRECT LIMITS Another important notion from Category Theory which we will need when studying schemes is that of a *direct limit*.

Remark 3.1.23. Those with background in Category Theory will notice that what we called an inverse limit is what in Category Theory is usually called a *limit* and the direct limits we will now discuss are usually called *colimits*. The discussion that follows about the existence and uniqueness of direct limits ensures the existence and uniqueness of inverse limits by invoking the *Duality Principle*.

Definition 3.1.24. Let $\mathcal{C}$ be a category and $(\Lambda, \leqslant)$ an ordered set. An **inductive system** in $\mathcal{C}$ (**indexed by** Λ) is a covariant functor $\mathbb{A} : \Lambda \to \mathcal{C}$. Equivalently, an inductive system is a family $(A_\lambda)_{\lambda \in \Lambda}$ of objects in $\mathcal{C}$ together with a family $\phi_{\lambda_2}^{\lambda_1} : A_{\lambda_1} \to A_{\lambda_2}$ $\forall\, \lambda_1 \leqslant \lambda_2$, of morphisms that satisfy the following conditions.

i) $\phi_\lambda^\lambda = \mathrm{id}_{A_\lambda}$ $\forall\, \lambda \in \Lambda$, and

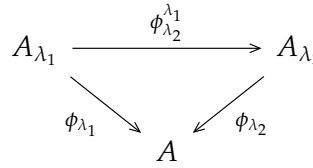
ii) $\phi_{\lambda_3}^{\lambda_2} \circ \phi_{\lambda_2}^{\lambda_1} = \phi_{\lambda_3}^{\lambda_1}$ $\forall\, \lambda_1 \leqslant \lambda_2 \leqslant \lambda_3$.



We shall write $(A_\lambda, \phi_{\lambda_2}^{\lambda_1})$ for an inductive system as above, omitting the indexes when there is no danger of confusion.

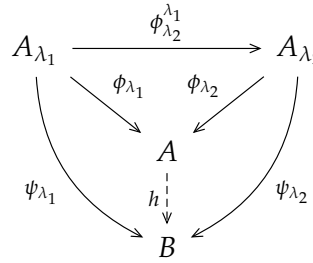
Definition 3.1.25. Suppose $(A_\lambda, \phi_{\lambda_2}^{\lambda_1})$ is an inductive system in $\mathcal{C}$. The **direct limit** of the system is a pair $(A, (\phi_\lambda)_{\lambda\in\Lambda})$ where $A \in \mathcal{C}$ and $\phi_\lambda \in \mathrm{Mor}_{\mathcal{C}}(A_\lambda, A)$, $\forall\, \lambda \in \Lambda$, if the following conditions hold.

i) $\phi_{\lambda_2} \circ \phi_{\lambda_2}^{\lambda_1} = \phi_{\lambda_1},\ \forall\, \lambda_1 \leqslant \lambda_2$



ii) For every pair (B, ψ_λ), where $B \in \mathcal{C}$ and $\psi_\lambda \in \mathrm{Mor}_{\mathcal{C}}(A_\lambda, B)$, $\forall\, \lambda \in \Lambda$, such that $\psi_{\lambda_2} \circ \phi_{\lambda_2}^{\lambda_1} = \psi_{\lambda_1},\ \forall\, \lambda_1 \leqslant \lambda_2$, there is *unique* $h \in \mathrm{Mor}_{\mathcal{C}}(A, B)$:

$$h \circ \phi_\lambda = \psi_\lambda,\ \forall\, \lambda \in \Lambda$$



We shall use the shorter notation (A, ϕ_λ) for a direct limit.

Proposition 3.1.26. *If the inductive system $(A_\lambda, \phi_{\lambda_2}^{\lambda_1})$ in $\mathcal{C}$ has direct limit (A, ϕ_λ), then this is uniquely determined up to isomorphism.*

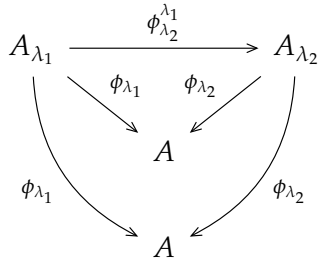
Proof. Suppose $(A, (\phi_\lambda)_{\lambda\in\Lambda})$ and $(B, (\psi_\lambda)_{\lambda\in\Lambda})$ are two direct limits. Because (A, ϕ_λ) is a direct limit and ψ_λ make the big triangle of the previous diagram commute, $\exists\,! h : A \to B$ such that

$$h \circ \phi_\lambda = \psi_\lambda,\ \forall\, \lambda \in \Lambda.$$

Similarly, (B, ψ_λ) is a direct limit, so $\exists\,! h' : B \to A$ such that

$$h' \circ \psi_\lambda = \phi_\lambda,\ \forall\, \lambda \in \Lambda.$$

Hence $(h' \circ h) \circ \phi_\lambda = \phi_\lambda,\ \forall\, \lambda \in \Lambda$.



In the above diagram, applying the definition of the direct limit for A, we get that $\exists\,! \xi : A \to A$ with $\xi \circ \phi_\lambda = \phi_\lambda,\ \forall\, \lambda \in \Lambda$. Since $1_A \circ \phi_\lambda = (h' \circ h) \circ \phi_\lambda = \phi_\lambda,\ \forall\, \lambda \in \Lambda$, we have that $h' \circ h = 1_A$. We can similarly prove that $h \circ h' = 1_B$. Therefore A and B are isomorphic. ◇

Proposition 3.1.27. *In the category* **Set** *of sets, every inductive system indexed by a directed set* $(\Lambda, \leqslant)$ *has direct limit.*

Proof. Let $(\Lambda, \leqslant)$ be a directed set and $(A_\lambda, \phi_{\lambda_2}^{\lambda_1})$ an inductive system of sets. Take the *disjoint* union

$$\bigsqcup_{\lambda\in\Lambda} A_\lambda$$

and define relation $\sim$ as follows. For every $a_1 \in A_{\lambda_1} \subseteq \sqcup A_\lambda$ and $a_2 \in A_{\lambda_2} \subseteq \sqcup A_\lambda$

$$a_1 \sim a_2 \iff \exists\, \lambda \geqslant \lambda_1, \lambda_2 : \phi_\lambda^{\lambda_1}(a_1) = \phi_\lambda^{\lambda_2}(a_2).$$

The relation $\sim$ is obviously reflexive and symmetric. It is transitive as well. Indeed, if

$$A_{\lambda_1} \ni a_1 \sim a_2 \in A_{\lambda_2} \ \text{ and } \ A_{\lambda_2} \ni a_2 \sim a_3 \in A_{\lambda_3}$$

then there are $\lambda \geqslant \lambda_1, \lambda_2$ and $\mu \geqslant \lambda_2, \lambda_3$ such that

$$\phi_\lambda^{\lambda_1}(a_1) = \phi_\lambda^{\lambda_2}(a_2) \text{ and } \phi_\lambda^{\lambda_2}(a_2) = \phi_\lambda^{\lambda_3}(a_3).$$

Since $(\Lambda, \leqslant)$ is directed, there is $k \geqslant \lambda, \mu$ (hance $k \geqslant \lambda_1, \lambda_3$) and we compute

$$\begin{aligned}\phi_k^{\lambda_1}(a_1) &= \phi_k^{\lambda} \circ \phi_\lambda^{\lambda_1}(a_1) = \phi_k^{\lambda} \circ \phi_\lambda^{\lambda_2}(a_2) = \phi_k^{\lambda_2}(a_2) = \phi_k^{\mu} \circ \phi_\mu^{\lambda_2}(a_2) \\ &= \phi_k^{\mu} \circ \phi_\mu^{\lambda_3}(a_3) = \phi_k^{\lambda_3}(a_3)\end{aligned}$$

that is, $a_1 \sim a_3$. Therefore $\sim$ is an equivalence relation.

Set

$$A = \bigsqcup_{\lambda \in \Lambda} A_\lambda / \sim$$

and if

$$q : \bigsqcup_{\lambda \in \Lambda} A \to A : a \mapsto [a]$$

is the canonical projection, then define

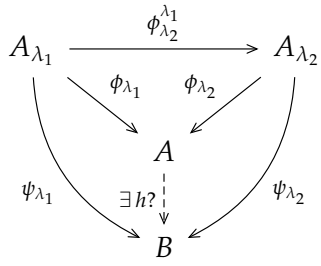
$$\phi_\lambda = q|_{A_\lambda} : A_\lambda \to A : a \mapsto [a] \ \forall \lambda \in \Lambda.$$

The pair (A, ϕ_λ) is the required direct limit. Indeed,

i) For every $\lambda_1 \leqslant \lambda_2$ and every $a_1 \in A_{\lambda_1}$ we have $a_1 \sim \phi_{\lambda_2}^{\lambda_1}(a_1)$ (take $\lambda = \lambda_2 \geqslant \lambda_1, \lambda_2$) so

$$\phi_{\lambda_2} \circ \phi_{\lambda_2}^{\lambda_1}(a_1) = [\phi_{\lambda_2}^{\lambda_1}(a_1)] = [a_1] = \phi_{\lambda_1}(a) \iff \phi_{\lambda_2} \circ \phi_{\lambda_2}^{\lambda_1} = \phi_{\lambda_1}.$$

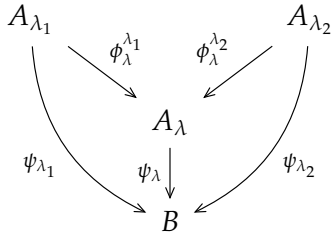
ii) If $(B, \psi_\lambda : A_\lambda \to B)$ is a pair where $B \in \mathcal{C}$ and $\psi_{\lambda_2} \circ \phi_{\lambda_2}^{\lambda_1} = \psi_{\lambda_1}$ for every $\lambda_1 \leqslant \lambda_2$ then we have the following commutative diagram.



Every $[a] \in A$ is the image of some $a \in A_\lambda$ through q. Define

$$h : A \to B : h([a]) = \psi_\lambda(a).$$

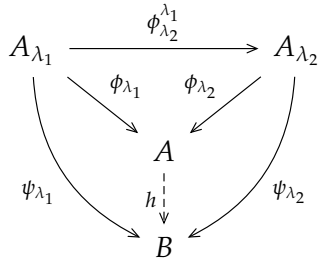
The map h is well defined. Indeed, if $[a_1] = [a_2]$ where $a_1 \in A_{\lambda_1}$, $a_2 \in A_{\lambda_2}$ then there is some $\lambda \geqslant \lambda_1, \lambda_2$ with $\phi_\lambda^{\lambda_1}(a_1) = \phi_\lambda^{\lambda_2}(a_2)$.



Therefore

$$\begin{aligned} h([a_1]) &= \psi_{\lambda_1}(a_1) = \psi_\lambda \circ \phi_\lambda^{\lambda_1}(a_1) = \psi_\lambda \circ \phi_\lambda^{\lambda_2}(a_2) \\ &= \psi_{\lambda_2}(a_2) = h([a_2]). \end{aligned}$$

From its definition, h makes the diagram



commute and is unique with this property.

◇

Remark 3.1.28. Direct limits in other familiar algebraic categories also exist. For example, direct limits of *groups* and *rings* exist. In the category **Rng** of rings for instance, if $[a_{\lambda_1}], [a_{\lambda_2}] \in A$ such that $a_{\lambda_1} \in A_{\lambda_1}$ and $a_{\lambda_2} \in A_{\lambda_2}$, we find some $\lambda_3 \geqslant \lambda_1, \lambda_2$ and define

$$\begin{aligned} [a_{\lambda_1}] + [a_{\lambda_2}] &= [\phi_{\lambda_3}^{\lambda_1}(a_{\lambda_1}) + \phi_{\lambda_3}^{\lambda_2}(a_{\lambda_2})] \\ [a_{\lambda_1}] \cdot [a_{\lambda_2}] &= [\phi_{\lambda_3}^{\lambda_1}(a_{\lambda_1}) \cdot \phi_{\lambda_3}^{\lambda_2}(a_{\lambda_2})]. \end{aligned}$$

It is easy to check that the above binary operations make A into a ring. The case for groups is similar.

3.2 GROTHENDIECK'S FORMULATION OF GALOIS THEORY

Grothendieck's Galois Theory for Schemes is a generalization of classical Galois Theory. In this section we will see how classical Galois Theory is seen from Grothendieck's point of view.

Classical Galois Theory classifies subextensions of a Galois extension $k \leqslant E$. The first generalization of Grothendieck's approach is to classify all separable extensions of k by considering the extension $k \leqslant k_{sep}$.

First we need a lemma.

Lemma 3.2.1. *Suppose G is a topological group and X is a finite G-set equipped with the discrete topology. Then the action is continuous if and only if* $\mathrm{stab}_G(x)$ *is an open subgroup of G for all $x \in X$.*

Proof. ($\Rightarrow$) Suppose that the action $G \times X \xrightarrow{\delta} X$ is continuous. For every $x \in X$ take the composition

$$G \xrightarrow{i_x} G \times X \xrightarrow{\delta} X$$

where $i_x : G \to G \times X : g \mapsto (g, x)$ is the canonical injection. i_x is obviously continuous and by hypothesis so is δ. Therefore their composite $\delta \circ i_x$ is continuous as well. For every $x \in X$,

$$\begin{aligned}\mathrm{stab}_G(x) &= \{g \in G : gx = x\} = \{g \in G : \delta(g, x) = x\} \\ &= \{g \in G : (\delta \circ i_x)(g) = x\} = (\delta \circ i_x)^{-1}(\{x\})\end{aligned}$$

which is open since $\{x\}$ is an open subset of the discrete space X.

($\Leftarrow$) For the contrary, suppose that the stabilizers of all elements are open subgroups of G. Let $x_0 \in X$. Now $\{x_0\}$ is an open subset of the discrete space X and its preimage through the action $\delta^{-1}(\{x_0\})$ is

$$\delta^{-1}(\{x_0\}) = \{(g, x) \in G \times X : gx = x_0\} = \bigsqcup_{x \in X} U_x$$

where

$$U_x = \{(g, x) \in G \times \{x\} : gx = x_0\} \subseteq G \times \{x\}.$$

Observe that $U_{x_0} = \mathrm{stab}_G(x_0) \times \{x_0\}$. This is a union of open subsets of $G \times X$. Indeed, each U_x is either empty or, if there is some $(h_x, x) \in U_x$ that is some $h_x \in G$ such that $h_x x = x_0$ then we can show that U_x and $\mathrm{stab}_G(x_0)$ are homeomorphic.

For this, define

$$\vartheta : \mathrm{stab}_G(x_0) \to U_x : g \mapsto (gh_x, x).$$

Observe that

$$\begin{aligned} g_1 = g_2 \in \mathrm{stab}_G(x_0) &\Leftrightarrow g_1 h_x = g_2 h_x \Leftrightarrow (g_1 h_x, x) = (g_2 h_x, x) \\ &\Leftrightarrow \vartheta(g_1) = \vartheta(g_2) \end{aligned}$$

therefore ϑ is well defined and injective.

ϑ is surjective as well. Let $(h', x) \in U_x$ so that $h'x = x_0$. Then $h' h_x^{-1} \in \mathrm{stab}_G(x_0)$. Indeed, $h_x x = x_0$ implies that $h_x^{-1} x_0 = x$. Therefore $h' h_x^{-1} x_0 = h' x = x_0$. It is now immediate that $\vartheta(h' h_x^{-1}) = (h', x)$.

The open subsets of $U_x \subseteq G \times \{x\}$ are of the form $U \times \{x\}$ where U is an open subset of G. Suppose $U \times \{x\}$ is an open subset of U_x. Then

$$\begin{aligned} \vartheta^{-1}(U \times \{x\}) &= \{g \in \mathrm{stab}_G(x_0) : \vartheta(g) = (gh_x, x) \in U \times \{x\}\} \\ &= \{g \in \mathrm{stab}_G(x_0) : gh \in U\} \\ &= \{g \in \mathrm{stab}_G(x_0) : g \in Uh^{-1}\} = \mathrm{stab}_G(x_0) \cap Uh^{-1}. \end{aligned}$$

But it is immediate that the all (left and right) cosets of U are homeomorphic to U through the map

$$U \to gU : x \mapsto gx$$

using the definition of a topological group. Therefore $\vartheta^{-1}(U \times \{x\})$ is open in $\mathrm{stab}_G(x_0)$ and ϑ is continuous.

Lastly, it is easy to see that ϑ is also open, hence a homeomorphism. Indeed, if $V \subseteq \mathrm{stab}_G(x_0)$ is open, then $\vartheta(V) = Vh_x \times \{x\}$ which is open in U_x by the above discussion and the definition of the product topology.

To sum up, we have shown that $\delta^{-1}(\{x_0\})$ is a union of open subsets of $G \times X$, hence open. Since the inverse images preserve unions, δ reverses open sets to open sets. Thus the action δ is continuous as required. ◇

Now onto the classification theorem.

Theorem 3.2.2 (Fundamental Theorem of Galois Theory in terms of separable algebras)**.** *Let k be a field and $G = \mathrm{Gal}(k_{sep}/k)$ its absolute Galois group. The category $\mathbf{Sep}_k$ of finite separable k-algebras is antiequivalent to the category $\mathbf{G\text{-}set}_f^t$ of finite sets equipped with a continuous and transitive G-action. Galois extensions of k give rise to G-sets that are isomorphic to finite quotients of G.*

Proof. We are going to break the proof of the above theorem into smaller parts. First we define a functor

$$F = \mathrm{Hom}_k(\square, k_{sep}) : \mathbf{Sep}_k \to \mathbf{G\text{-}set}_f^t$$

as follows. F sends a finite separable k-algebra L to $\mathrm{Hom}_k(L, k_{sep})$, the set of all k-algebra homomorphisms from L to k_{sep} and each morphism of separable k-algebras $\phi : L_1 \to L_2$ to

$$\begin{aligned} F(\phi) : \mathrm{Hom}_k(L_2, k_{sep}) &\to \mathrm{Hom}_k(L_1, k_{sep}) : \\ (L_2 \xrightarrow{f} k_{sep}) &\mapsto (L_1 \xrightarrow{\phi} L_2 \xrightarrow{f} k_{sep}). \end{aligned}$$

We will show that this is an essentially surjective, fully faithfull, contravariant functor.

Obviously $\mathrm{Hom}_k(L, k_{sep})$ is a finite set; if, using the primitive element theorem, we write $L = k(a)$ for some separable element $a \in L$, then the cardinality of $\mathrm{Hom}_k(L, k_{sep})$ equals $\partial m(a, k)$.

The absolute Galois group $G = \mathrm{Gal}(k_{sep}/k)$ acts on $\mathrm{Hom}_k(L, k_{sep})$ by compositions. That is,

$$\mathrm{Gal}(k_{sep}/k) \times \mathrm{Hom}_k(L, k_{sep}) \to \mathrm{Hom}_k(L, k_{sep}) : (\sigma, f) \mapsto \sigma \circ f. \quad (3.1)$$

$$L \xrightarrow{f} k_{sep} \xrightarrow{\sigma} k_{sep}$$

The fact that the above map defines an action is immediate. Indeed, if $L \xrightarrow{f} k_{sep} \xrightarrow{\tau} k_{sep} \xrightarrow{\sigma} k_{sep}$ then $(\sigma \circ \tau) \circ f = \sigma \circ (\tau \circ f)$ since the

composition of functions is associative, and if $L \xrightarrow{f} k_{sep} \xrightarrow{\text{id}} k_{sep}$ then $\text{id} \circ f = f$. We must show that the action is continuous. For this we need theprevious lemma.

For every $f \in \text{Hom}_k(L, k_{sep})$ we compute

$$\begin{aligned}\text{stab}_G(f) &= \{\sigma \in \text{Gal}(k_{sep}/k) : \sigma f = f\} \\ &= \{\sigma \in \text{Gal}(k_{sep}/k) : \sigma f(L) = f(L)\} \\ &= \{\sigma \in \text{Gal}(k_{sep}/k) : \sigma \text{ fixes } f(L)\} \\ &= \text{Gal}(k_{sep}/f(L)).\end{aligned}$$

Now $k \leqslant L$ is a finite separable extension and $f \in \text{Hom}_k(L, k_{sep})$. Therefore $k \leqslant f(L)$ is also a finite separable extension. From the Fundamental Theorem of Infinite Galois Theory,

$$[f(L) : k] < \infty \iff \text{Gal}(k_{sep}/f(L)) = \text{stab}_G(f) \text{ is open.}$$

Using the previous Lemma on the finite set $\text{Hom}_k(L, k_{sep})$ and the topological group G, the action is continuous.

The next step is to show that the action described in (3.1) is transitive.

Suppose $f, g \in \text{Hom}_k(L, k_{sep})$. Since L is a separable extension of k, $L = k(a)$ for some separable element $a \in L$ from the Primitive Element Theorem. By the Extension Theorem, there is some $\sigma \in G$ such that $\sigma(f(a)) = g(a)$ since both $f(a)$ an $g(a)$ must be roots of $m(a, k)$. Since f, g are determined by their images $f(a)$ and $g(a)$ respectively, that means that we can find some $\sigma \in G$ such that $\sigma f = g$, i.e. the action is transitive.

It is now immediate that F is a contravarinat functor.

Recall from Group Theory that if X is a transitive G-set then X and $G/\text{ stab}_G(x)$ are isomorphic as G-sets for any $x \in X$. Therefore $\text{Hom}_k(L, k_{sep})$ is isomorphic to the left coset space of some open subgroup of G. If $k \leqslant l$ is Galois, this coset space is a quotient of G by a normal subgroup from Infinite Galois Theory.

We will show that the functor F is essentially surjective.

Suppose S is a finite set equipped with a continuous and transitive G-action and $s \in S$. Since S is finite and the action is continuous, $\text{stab}_G(x)$ is open in G. Let

$$L = \text{Fix}_{k_{sep}}(\text{stab}_G(s)) \iff \text{stab}_G(s) = \text{Gal}(k_{sep}/L).$$

From Infinite Galois Theory, since $\mathrm{stab}_G(x)$ is open, $k \leqslant L$ is a finite extension. Obviously it is also separable. We will show that the G-sets S and $\mathrm{Hom}_k(L, k_{sep})$ are isomorphic.

Define

$$\vartheta : \mathrm{Hom}_K(L, k_{sep}) \to S$$

as follows. Since $\mathrm{Hom}_K(L, k_{sep})$ is transitive, ϑ is uniquelly determined by its value on some random $f \in \mathrm{Hom}_K(L, k_{sep})$. Take f to be the canonical inclusion $i : L \to k_{sep}$, and define

$$\vartheta(i) = s \implies \vartheta(\sigma i) = \sigma s \quad \forall \sigma \in G.$$

Observe that

$$\begin{aligned}\mathrm{stab}_G(i) &= \{\sigma \in G : \sigma i = i\} = \{\sigma \in G : \sigma(L) = L\} \\ &= \mathrm{Gal}(k_{sep}/L) = \mathrm{stab}_G(s).\end{aligned}$$

Therefore

$$\begin{aligned}\sigma i = \tau i &\Leftrightarrow \tau^{-1}\sigma i = i \Leftrightarrow \tau^{-1}\sigma \in \mathrm{stab}_G(i) \\ &\Leftrightarrow \tau^{-1}\sigma \in \mathrm{stab}_G(s) \Leftrightarrow \tau^{-1}\sigma s = s \Leftrightarrow \sigma s = \tau s.\end{aligned}$$

So ϑ is well defined.

Since both S and $\mathrm{Hom}_k(L, k_{sep})$ are transitive G-sets, the above observation yields an isomorphism of G-sets

$$\mathrm{Hom}_k(L, k_{sep}) \simeq G/\,\mathrm{stab}_G(i) = G/\,\mathrm{stab}_G(s) \simeq S.$$

Therefore the functor F is indeed essentially surjective.

Lastly, the functor F is fully faithful.

Suppose L, M are two finite separable extensions of k. We must show that the map

$$\begin{aligned}F : \mathrm{Mor}_{\mathbf{Sep}_k}(L, M) &\to \mathrm{Mor}_{\mathbf{G\text{-}set}_f^t}\left(\mathrm{Hom}_k(M, k_{sep}), \mathrm{Hom}_k(L, k_{sep})\right) \\ (L \xrightarrow{\phi} M) &\mapsto F(\phi)\end{aligned}$$

is bijective. Recall that

$$\begin{aligned}F(\phi) : \mathrm{Hom}_k(M, k_{sep}) &\to \mathrm{Hom}_k(L, k_{sep}) : F(\phi)(f) = f \circ \phi \\ (M \xrightarrow{f} k_{sep}) &\mapsto (L \xrightarrow{\phi} M \xrightarrow{f} k_{sep}).\end{aligned}$$

F is obviously injective; if $\phi \neq \psi \in \mathrm{Mor}_{\mathbf{Sep}_k}(L, M)$, that is $\phi(x) \neq \psi(x)$ for some $x \in L$, then $f \circ \phi \neq f \circ \psi$ for any $f : M \to k_{sep}$ such that $f(\phi(x)) \neq f(\psi(x))$, i.e. $F(\phi) \neq F(\psi)$. Such f always exists. It remains to show that F is surjective. Take some $\gamma \in \mathrm{Mor}_{\mathbf{G\text{-}set}^t_f}\big(\mathrm{Hom}_k(M, k_{sep}), \mathrm{Hom}_k(L, k_{sep})\big)$. Since both $\mathrm{Hom}_k(M, k_{sep})$ and $\mathrm{Hom}_k(M, k_{sep})$ are transitive G-sets, γ is determined by the image $\gamma(f)$ of some (any) $f \in \mathrm{Hom}_k(M, k_{sep})$. Moreover,

$$\begin{aligned} \sigma \in \mathrm{stab}_G(f) &\implies \sigma f = f \implies \gamma(\sigma f) = \gamma(f) \\ &\implies \sigma\gamma(f) = \gamma(f) \implies \sigma \in \mathrm{stab}_G(\gamma(f)) \\ &\implies \mathrm{stab}_G(f) \subseteq \mathrm{stab}_G(\gamma(f)). \end{aligned}$$

Taking the fixed fields of these subgroups we have

$$\mathrm{Fix}_{k_{sep}}(\mathrm{stab}_G(\gamma(f)) \subseteq \mathrm{Fix}_{k_{sep}}(\mathrm{stab}_G(f)).$$

As we have already seen,

$$\mathrm{stab}_G(f) = \quad \text{and} \quad \mathrm{stab}_G(\gamma(f)) = \mathrm{Gal}(k_{sep}/\gamma(f)(L)).$$

Therefore we have

$$\mathrm{Fix}_{k_{sep}}\big(\mathrm{Gal}(k_{sep}/f(M))\big) = \gamma(f)(L) \subseteq f(M) = \mathrm{Fix}_{k_{sep}}\big(\mathrm{Gal}(k_{sep}/\gamma f(L))\big)$$

Take $f^{-1} : f(M) \to M$ to be the inverse of f. Then $f^{-1} \circ \gamma(f) \in \mathrm{Mor}_k(L, M)$,

$$L \overset{\gamma(f)}{\to} \gamma(f)(L) \subseteq f(M) \subseteq k_{sep} \overset{f^{-1}}{\to} M,$$

and obviously $F(f^{-1}\gamma(f)) = \gamma$ as G-maps since they agree on f.

The final assertion follows from Infinite Galois Theory. ◇

Grothendieck's formulation however does not stop there. It generalizes the context and classifies *étale k-algebras* by considering sets on which G acts continuously but not neccesarily transitively.

Definition 3.2.3. A k-algebra L is called **étale** if it is isomorphic to a *finite* direct product of separable extensions of k.

Example 3.2.4. Every separable extension of k is an étale k-algebra. In particular, every Galois extension of k is étale.

In view of the above definition, we can form the category $\mathbf{Et}_k$ whose objects are étale k-algebras and morphisms are k-algebra homomorphisms. We can also form the category $\mathbf{FEt}_k$ of finite étale k-algebras and k-algebra homomorphisms. Now Grothendieck's formulation Galois Theory is the following.

Theorem 3.2.5 (Grothendieck's formulation of Galois Theory)**.** *Let k be a field and $G = \mathrm{Gal}(k_{sep}/k)$ its absolute Galois group. The category $\mathbf{FEt}_k$ of finite étale k-algebras is antiequivalent to the category $\mathbf{G\text{-}set}_f$ of finite sets equipped with a continuous G-action.*

Proof. The most part of the proof has already been done in the previous theorem. We again define a functor

$$F = \mathrm{Hom}_k(\square, k_{sep}) : \mathbf{FEt}_k \to \mathbf{G\text{-}set}_f$$

as before. Namely, F sends each étale k-algebra $A = \prod_{i=1}^n L_i$ to the set $\mathrm{Hom}_k(A, k_{sep})$ and each morphism of étale k-algebras $\phi : A \to B$ to $F(\phi)$ defined as before.

The first thing to note is that

$$\mathrm{Hom}_k(A, k_{sep}) = \bigsqcup_{i=1}^{n} \mathrm{Hom}_k(L_i, k_{sep}) \tag{3.2}$$

since each $f \in \mathrm{Hom}_k(A, k_{sep})$ induces an injection of exactly one of the L_i's into k_{sep}. Indeed, any injection of a product of more than one separable extensions of k into k_{sep}, e.g. $L_i \times L_j$, will produce zero divisors, e.g. $(1, 0)$ and $(0, 1)$, which is absurd.

Since each $\mathrm{Hom}_k(L_i, k_{sep})$ is finite, so is their finite dfisjoint union $\mathrm{Hom}_k(A, k_{sep})$. The absolute Galois group G acts on $\mathrm{Hom}_k(A, k_{sep})$ by compositions as before and this action is continuous by the same arguments as before.

Thus F is again a contravariant functor.

The second thing to note is that the action is not transitive. In fact, $\mathrm{Hom}_k(L_i, k_{sep})$ are the G-orbits (the transitive G-subsets) of $\mathrm{Hom}_k(A, k_{sep})$.

The fact that F is essentially surjective stems from the decomposition (3.2) of $\mathrm{Hom}_k(A, k_{sep})$ into a disjoint union. Any $f \in \mathrm{Hom}_k(A, k_{sep})$ belongs to some $\mathrm{Hom}_k(L_i, k_{sep})$ and we use the arguments of the previous theorem.

To prove that F is fully faithful, we must note one last thing. If $A = \prod_{i=1}^{n} L_i$ and $B = \prod_{j=1}^{m} K_j$ are two étale k-algebras, the set $\mathrm{Mor}_{\mathbf{FEt}_k}(A, B)$ can be decomposed into a union of sets of the form $\mathrm{Mor}_{\mathbf{Sep}_k}(L_i, K_j)$. We again apply the same arguments as before. $\diamond$

Remark 3.2.6. If instead of $k \leqslant k_{sep}$ we consider an arbitrary extension $k \leqslant E$, then we have that the category of finite k-algebras that are products of finite separable subextensions of $k \leqslant E$ is antiequivalent to the category of finite sets equipped with a continuous $\mathrm{Gal}(E/k)$-action. In particular, separable extensions correspond to transitive $\mathrm{Gal}(E/k)$-sets and Galois extensions to finite quotients of $\mathrm{Gal}(E/k)$.

3.3 GALOIS THEORY FOR SCHEMES

Grothendieck developed schemes as a generalization of *varieties* which, in classical Algebraic Geometry, are defined as sets of solutions of systems of polynomial equations. Indeed, every variety is a special kind of scheme. Our aim in this section to define schemes and other related notions which we need in order to formulate the last classification theorem.

To keep this dissertation short, we need to assume the reader is already familiar with the language of Algebraic Geometry. The first two parts of [39] should provide a good background for what is to follow.

THE SPECTRUM OF A RING

Definition 3.3.1. We define the **spectrum** of a ring R, denoted by $\mathrm{Spec}(R)$, to be the set of all prime ideals of R, i.e.

$$\mathrm{Spec}(R) = \{\mathfrak{p} : \mathfrak{p} \text{ prime ideal of } R\}.$$

Now we make $\mathrm{Spec}(R)$ into a topological space.

Proposition 3.3.2. *Suppose R is a ring and let $\mathrm{Spec}(R)$ be its spectrum. Define*

$$V(\mathfrak{a}) = \{\mathfrak{p} \in \mathrm{Spec}(R) : \mathfrak{p} \supseteq \mathfrak{a}\}.$$

Then

i) $V(0) = \mathrm{Spec}(R)$ *and* $V(R) = \emptyset$.

ii) For every collection $\{\mathfrak{a}_i\}_{i\in I}$ *of ideals of* R,

$$\bigcap_{i\in I} V(\mathfrak{a}_i) = V\big(\bigcup_{i\in I} \mathfrak{a}_i\big).$$

iii) For every two ideals $\mathfrak{a}, \mathfrak{b}$ *of* R,

$$V(\mathfrak{a}) \cup V(\mathfrak{b}) = V(\mathfrak{a} \cap \mathfrak{b}).$$

Proof. The first assertion is immediate; every prime ideal contains the zero ideal whilst there are no prime ideals containing R.

For the second assertion observe that

$$\mathfrak{p} \in \bigcap_{i\in I} V(\mathfrak{a}_i) \iff \mathfrak{p} \in V(\mathfrak{a}_i) \ \ \forall\, i \in I \iff \mathfrak{p} \supseteq \mathfrak{a}_i \ \ \forall\, i \in I$$
$$\iff \mathfrak{p} \supseteq \bigcup_{i\in I} \mathfrak{a}_i \iff \mathfrak{p} \in V\big(\bigcup_{i\in I} \mathfrak{a}_i\big).$$

For the last assertion, if $\mathfrak{p} \in V(\mathfrak{a}) \cup V(\mathfrak{b})$ then $\mathfrak{p}$ contains either one of them and therefore contains their intersection $\mathfrak{a} \cap \mathfrak{b}$; hence $\mathfrak{p} \in V(\mathfrak{a} \cap \mathfrak{b})$. On the other hand, if $\mathfrak{p} \in V(\mathfrak{a} \cap \mathfrak{b})$ but $\mathfrak{a} \nsubseteq \mathfrak{p}$ then there is some $a \in \mathfrak{a} \setminus \mathfrak{p}$. Choose $b \in \mathfrak{b}$. Since $\mathfrak{a}, \mathfrak{b}$ are ideals, $ab \in \mathfrak{a} \cap \mathfrak{b} \subseteq \mathfrak{p}$ and $\mathfrak{p}$ is prime. Therefore, $b \in \mathfrak{p}$ so $\mathfrak{b} \subseteq \mathfrak{p}$ and $\mathfrak{p} \in V(\mathfrak{b}) \subseteq V(\mathfrak{a}) \cup V(\mathfrak{b})$. ◇

Corollary 3.3.3. *With the hypotheses of the previous proposition,* $\mathrm{Spec}(R)$ *is a topological space whose closed sets are of the form* $V(\mathfrak{a})$ *with* $\mathfrak{a} \lhd R$.

Definition 3.3.4. The topology described above is called the **Zariski topology**.[2]

The open subsets $V(\mathfrak{a})^c$ of the Zariski topology are denoted by $D(\mathfrak{a})$. For every $f \in R$, we define

$$V(f) = \{\mathfrak{p} \in \mathrm{Spec}(R) : f \in \mathfrak{p}\} \subseteq \mathrm{Spec}(R).$$

Since $V(f) = V(\langle f \rangle)$, $V(f)$ are closed subsets of $\mathrm{Spec}(R)$. Therefore their complements

$$D(f) = V(f)^c = \{\mathfrak{p} \in \mathrm{Spec}(R) : f \notin \mathfrak{p}\} \subseteq \mathrm{Spec}(R)$$

[2] Named after Oscar Zariski (1899-1986) who was the first to introduce it.

are open subsets in $\mathrm{Spec}(R)$; these are called **distinguished open sets**. They are called distinguished because they form a basis for the Zariski topology. Indeed, for every ideal $\mathfrak{a}$ of R,

$$V(\mathfrak{a}) = \bigcap_{f \in \mathfrak{a}} V(f) \Rightarrow D(\mathfrak{a}) = \bigcup_{f \in \mathfrak{a}} D(f).$$

PRESHEAVES AND SHEAVES The idea of Algebraic Geometry is to study geometric objects using appropriately chosen algebraic ones. This is where Sheaf Theory comes into play and provides a way to assign algebraic objects such as groups and rings to topological spaces (which are geometric objects).

Definition 3.3.5. Let (X, τ_X) be a topological space. A **presheaf of sets** over X is a pair of families

$$(\{P(U)\}_{U \in \tau_X}, \{r_V^U : P(U) \to P(V)\}_{V \subseteq U \in \tau_X})$$

where for each $U \in \tau_X$, $P(U)$ is a *set* and for each $V \subseteq U \in \tau_X$ $r_V^U : P(U) \to P(V)$ is a function called **restriction map** such that

i) $r_U^U = \mathrm{id}_U$ for all $U \in \tau_X$ and

ii) $r_W^U = r_W^V \circ r_V^U$ for all $W \subseteq V \subseteq U \in \tau_X$.

We will write $(P(U), r_V^U)$ for a presheaf as above, to keep the notation clean and short.

Example 3.3.6. The prototypical example we have to bear in mind is the **presheaf of smooth functions**. Let M be a smooth manifold and τ_M be its topology. For each $U \in \tau_M$ let

$$\mathcal{C}^\infty(U) = \{f : U \to \mathbb{R} : f \text{ smooth}\},$$

and for every $V \subseteq U \in \tau_X$,

$$r_V^U : \mathcal{C}(U) \to \mathcal{C}(V) : f \mapsto f|_V$$

be the usual restriction map. Then $(\mathcal{C}^\infty(U), r_V^U)$ is a presheaf of sets.

Example 3.3.7. Let (X, τ_X) be a topological space. For every $U \in \tau_X$ let

$$\mathcal{C}(U) = \{f : U \to \mathbb{R} : f \text{ continuous}\},$$

and for every $V \subseteq U \in \tau_X$,

$$r_V^U : \mathcal{C}(U) \to \mathcal{C}(V) : f \mapsto f|_V$$

be the usual restriction map. Then $(\mathcal{C}(U), r_V^U)$ is a presheaf of sets, called the **presheaf of continuous functions**.

Example 3.3.8. Let (X, τ_X) be a topological space. For every $U \in \tau_X$ let

$$B(U) = \{g : U \to \mathbb{R} : g \text{ bounded}\},$$

and for every $V \subseteq U \in \tau_X$,

$$r_V^U : B(U) \to B(V) : g \mapsto g|_V$$

be the usual restriction map. Then $(B(U), r_V^U)$ is a presheaf of sets, called the **presheaf of bounded functions**.

Remark 3.3.9. Equivalently, a presheaf of sets can be defined as a *contravariant* functor

$$\mathcal{F} : (\tau_X, \leqslant) \to \mathbf{Set}$$

whose domain is the poset $(\tau_X, \leqslant)$ with the partial order given by $U \leqslant V \iff V \subseteq U$.

It is easy to see that every presheaf of sets induces such a functor $\mathcal{F}$ that sends each $U \in \tau_X$ to $P(U)$ and each morphism (U, V) to r_V^U; the axioms of a presheaf ensure that $\mathcal{F}$ is a functor.

And conversely, any *contravariant* functor $\mathcal{F} : (\tau_X, \leqslant) \to \mathbf{Set}$ induces a presheaf $\big(\mathcal{F}(U), \mathcal{F}(V, U)\big)$; the axioms of a functor ensure that this is indeed a presheaf of sets.

So when we speak of a presheaf $\mathcal{F}$ of a space, we refer to the functor $\mathcal{F}$.

As with the above example, the sets $P(U)$ of a presheaf are usually sets of functions defined on the open subset U and r_V^U are then the usual restriction maps.

It is a common occur for the sets $P(U)$ to have extra algebraic structure (such as rings or algebras). The relevant definition is given below.

Definition 3.3.10. Let (X, τ_X) be a topological space. A **presheaf of rings** is a contravariant functor $\mathcal{F} : (\tau_X, \leqslant) \to \mathbf{Rng}$.

Definition 3.3.11. If (X, τ_X) is a topological space, $x \in X$, and $\mathcal{F} \equiv (P(U), r_V^U)$ is a presheaf (either of sets or of rings) over X, then the **stalk of $\mathcal{F}$ at** x is the direct limit of the inductive system $(P(U), r_V^U)_{V \subseteq U \in \mathcal{N}_x^o}$ indexed by the **open neighborhoods $\mathcal{N}_x^o$ of** x.

Note that the open neighborhoods of x, ordered by inclusion, is a directed set, so the limit exists and is unique.

The presheaf of smooth functions has additional structure than plays an important role in the theory of smooth manifolds. In particular, *if two smooth functions defined on an open set U agree on every element of an open cover of U, then they are equal*. Another important property is the *glueing lemma for smooth maps*; *if U in an open set in X and we have smooth functions defined on the elements of an open cover so that they agree when their domains overlap, then we can glue them together and obtain a smooth function defined on U*.

We are thus led to the following definition.

Definition 3.3.12. Let (X, τ_X) be a topological space and $(P(U), r_V^U)$ a presheaf of sets over X. The presheaf $(P(U), r_V^U)$ is called a **sheaf** if for every $U \in \tau_X$ and every open cover $\{U_i\}_{i \in I}$ of U the following conditions hold.

i) If $f, g \in P(U)$ such that $f|_{U_i} = g|_{U_i}$ for every $i \in I$ then $f = g$.

ii) If $f_i \in P(U_i)$ for every $i \in I$ such that $r^{U_i}_{U_i \cap U_j}(f_i) = r^{U_j}_{U_i \cap U_j}(f_j)$ for every $i, j \in I$ with $U_i \cap U_j \neq \emptyset$ then there is some $f \in P(U)$ such that $r^U_{U_i}(f) = f_i$ for every $i \in I$.

Example 3.3.13. Thepresheaf of smooth functions is of course a sheaf of sets. The presheaf of continuous functions is also a sheaf of sets.

Counterexample 3.3.14. The presheaf of bounded functions is *not* a sheaf. Condition (ii) is not satisfied.

Definition 3.3.15. Similarly to the previous definition, we can define a **sheaf of rings**. We shall denote a sheaf of rings of a topological space X as $\mathcal{O}_X$.

Definition 3.3.16. A **ringed space** is a pair $(X, \mathcal{O}_X)$ where X is a topological space and $\mathcal{O}_X$ is a sheaf of rings over X. A **locally ringed space** is a ringed space whose stalks $\mathcal{O}_{X,x}$ are *local rings*.[3]

Definition 3.3.17. A **morphism of ringed spaces**

$$f : (X, \mathcal{O}_X) \to (Y, \mathcal{O}_Y)$$

is a *continuous* map $f : X \to Y$ such that the induced map

$$f_* : \mathcal{O}_Y(U) \to \mathcal{O}_X\big(f^{-1}(U)\big)$$

is a *ring homomorphism* for every $U \in \tau_Y$. A **morphism of locally ringed spaces** is a morphism of ringed spaces that also respects the maximal ideals, i.e. f_* sends maximal ideals to maximal ideals.

Example 3.3.18. We now come to the most important example. As we saw, the spectrum $\mathrm{Spec}(R)$ of a ring R is a topological space endowed with the Zariski topology. We will now construct an *sheaf of rings* $(P(U), r_V^U)$ over $\mathrm{Spec}(R)$. For every distinguished open set $D(f)$ of $\mathrm{Spec}(R)$ define $P(D(f))$ to be the localization of R at the multiplicative set of all functions that do not vanish outside of $V(f)$.

This is a *sheaf of rings* called the **structure sheaf**, it is denoted by $\mathcal{O}_{\mathrm{Spec}\, R}$ and $(\mathrm{Spec}(R), \mathcal{O}_{\mathrm{Spec}\, R})$ is a *locally ringed space*. For a detailed proof that this is actually a sheaf of rings, see [39], Theorem 4.1.2.

Definition 3.3.19. Let R be a ring. The ringed space consisting of the *spectrum* $\mathrm{Spec}(R)$ of R with the Zariski topology and the *structure sheaf*, $(\mathrm{Spec}(R), \mathcal{O}_{\mathrm{Spec}\, R})$, is called an **affine scheme**. A **morphism of affine schemes** is just a morphism of locally ringed spaces.

Definition 3.3.20. Suppose R is a ring. A **scheme** is a locally ringed space $(X, \mathcal{O}_X)$ such that every $x \in X$ has an open neighborhood U so that $(U, \mathcal{O}_X|_U)$ is an affine scheme.

A morphism of schemes is just a morphism of locally rigned spaces.

In view of the above definition, we have the *category* **Sch** *of schemes* whose objects are affine schemes and morphisms are the morphisms of affine schemes.

[3] Recall from Ring Theory that a (commutative) **local ring** is a ring that has a unique maximal ideal.

Definition 3.3.21. A morphism $p : \widetilde{X} \to X$ of schemes is called **finite étale** if there is a open cover of open affine subsets $\{\mathrm{Spec}(A_i)\}_{i\in I}$ of X such that $p^{-1}\big(\mathrm{Spec}(A_i)\big)$ is of the form $\mathrm{Spec}(B_i)$ for some étale A_i-algebra. If p is surjective, then $(\widetilde{X}, p)$ is a **finite étale cover**

We can think of finite étale morphisms as generalization of local homeomorphisms and finite étale coverings as generalization of covering spaces. Thus, we can form the category of finite étale coverings over a scheme X, denoted by $\mathbf{FEt}_X$.

The main theorem concerning schemes and étale coverings is the following. The interested reader can consult the bibliography for a proof.

Theorem 3.3.22 (Galois Theorem for Schemes)**.** *Let X be a connected scheme, i.e. a scheme that is connected as a topological space. Then there is a profinite group G uniquely determined up to isomorphism such that the category $\mathbf{FEt}_X$ of finite étale coverings of X is equivalent to the category $\mathbf{G\text{-}set}_f$ of finite sets on which G acts continuously.*

To retrieve Grothendieck's formulation of Galois Theory we must take $X = \mathrm{Spec}(k)$ and the profinite group is the absolute Galois group of k. The the finite étale coverings are exactly the étale k-algebras.

BIBLIOGRAPHY

[1] E. Artin. *Galois Theory*. Edited and supplemented with a selection on applications by Arthur N. Milgram. Mineola, New York: Dover Publications Inc., 1998.

[2] S. Axler. *Linear Algebra Done Right*. 3rd ed. Undergraduate Texts in Mathematics. Springer, 2015.

[3] J. R. Batista. *Field Extensions and Galois Theory*. 1st ed. Vol. 22. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1984.

[4] F. Borceux and G. Janelidze. *Galois Theories*. 1st ed. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2001.

[5] K. Brown. *The Primitive Element Theorem*. Cornell University, October 2010.

[6] K. Conrad. *Infinite Galois Theory*. 2020. A draft for a mini-course given at Connecticut Summer School in Number Theory.

[7] F. H. Croom. *Basic Concepts of Algebraic Topology*. 1st ed. Undergraduate Texts in Mathematics. Springer-Verlag New York, 1978.

[8] J. Dieudonné. *A History of Algebraic and Differential Topology, 1900 - 1960*. 1st ed. Modern Birkhäuser Classics. Birkhäuser Basel, 2009.

[9] R. Douady and A. Douady. *Algebra and Galois Theories*. 1st ed. Translated by U. Ray. Springer Cham, 2020.

[10] D. S. Dummit and R. M. Foote. *Abstract Algebra*. 3rd ed. John Wiley and Sons Inc., 2004.

[11] S. Eilenberg and S. MacLane. "General theory of natural equivalences". In: *Transactions of the Americal Mathematical Society* 58.2 (Sept. 1945), pp. 231–294.

[12] A. Grothendieck and M. Raynaud. *Revêtements Étales et Groupe Fondamental. Séminaire de Géométrie Algébrique du Bois Marie 1960/61 (SGA 1)*. 1st ed. Lecture Notes in Mathematics. Springer Berlin, Heidelberg, 1971. arXiv:math/0206203v2 [math.AG].

[13] A. Hatcher. *Algebraic Topology (2015 ed.)*.

[14] J. M. Howie. *Fields and Galois Theory*. 1st ed. Springer Undergraduate Mathematics Series. Springer-Verlag London, 2006.

[15] F. Jarvis. *Algebraic Number Theory*. 1st ed. Springer Undergraduate Mathematics Series. Springer International Publishing, 2014.

[16] G. Karpilovsky. *Topics in Field Theory*. Vol. 155. Mathematics Studies. North Holland, 1989.

[17] B. M. Kiernan. "The Development of Galois Theory from Lagrange to Artin". In: *Archive for History of Exact Sciences* 8.1/2 (Dec. 1971), pp. 40–154.

[18] C. Koppensteiner. *Notes for Math 532 - Algebraic Geometry I*.

[19] C. Kosniowski. *A First Course in Algebraic Topology*. Cambridge University Press, 1980.

[20] W. Krull. "Galoissche Theorie der unendlichen algebraischen Erweiterungen". In: *Mathematische Annalen* 100 (1928), pp. 687–698.

[21] E. Kunz. *Introduction to Commutative Algebra and Algebraic Geometry*. 1st ed. Modern Birkhäuser Classics. Birkhäuser New York.

[22] S. Lang. *Undergraduate Algebra*. 3rd ed. Vol. 211. Undergraduate Texts in Mathematics. Springer New York, 2005.

[23] T. Leinster. *Basic Category Theory*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2014. arXiv:1612.09375v1 [math.CT].

[24] H. W. Lenstra. *Galois Theory for Schemes*.

[25] W. S. Massey. *A Basic Course in Algebraic Topology*. 1st ed. Vol. 127. Graduate Texts in Mathematics. Springer-Verlag New York, 1991.

[26] J. S. Milne. *Fields and Galois Theory*.

[27] P. Morandi. *Field and Galois Theory*. 1st ed. Vol. 167. Graduate Texts in Mathematics. Springer-Verlag New York, 1996.

[28] Y. Moschovakis. *Notes on Set Theory*. 1st ed. Undergraduate Texts in Mathematics. Springer-Verlag New York, 1994.

[29] J. R. Munkres. *Topology*. 2nd ed. Prentice Hall, 2000.

[30] Ivan Niven. "A simple proof that π is irrational". In: *Bulletin of the American Mathematical Society* 53.6 (1947), pp. 509–509.

[31] H. Poincaré. "Analysis situs". In: *Journal de l' École Polytechnique* 1.1 (1895), pp. 1–123.

[32] H. Poincaré. *Papers on Topology. Analysis Situs and Its Five Supplements*. Translated by J. Stillwell.

[33] E. Riehl. *Category Theory in Context*. Dover Publications, 2016.

[34] S. Roman. *An Introduction to the Language of Category Theory*. 1st ed. Compact Textbooks in Mathematics. Birkhäuser, 2017.

[35] I. Stewart. *Galois Theory*. 4th ed. CRC Press, 2015.

[36] T. Szamuely. *Galois Groups and Fundamental Groups*. 1st ed. Vol. 117. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2009.

[37] T. Szamuely. *Galois Theory after Galois*. An informal survey of some modern aspects of Galois Theory.

[38] T. Szamuely. *Galois Theory: Past and Present*. 2010-11. Slides from a colloquium lecture on the occasion of the 200th anniversary of Galois' birth.

[39] R. Vakil. *The Rising Sea. Foundations of Algebraic Geometry*.