Siegfried Bosch

Algebra

From the Viewpoint of Galois Theory

Birkhäuser

Siegfried Bosch
Mathematisches Institut
Westfälische Wilhelms-Universität
Münster, Germany

ISSN 1019-6242 ISSN 2296-4894 (electronic)
Birkhäuser Advanced Texts Basler Lehrbücher
ISBN 978-3-319-95176-8 ISBN 978-3-319-95177-5 (eBook)
https://doi.org/10.1007/978-3-319-95177-5

Library of Congress Control Number: 2018950547

Mathematics Subject Classification (2010): 12-01, 13-01, 14-01

Translation from the German language edition: *Algebra* by Siegfried Bosch, Copyright © Springer-Verlag GmbH Deutschland, 2013. All Rights Reserved. ISBN 978-3-642-39566-6

This book is published under the imprint Birkhäuser, www.birkhauser-science.com by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The material presented here can be divided into two parts. The first, sometimes referred to as abstract algebra, is concerned with the general theory of algebraic objects such as groups, rings, and fields, hence, with topics that are also basic for a number of other domains in mathematics. The second centers around Galois theory and its applications. Historically, this theory originated from the problem of studying algebraic equations, a problem that, after various unsuccessful attempts to determine solution formulas in higher degrees, found its complete clarification through the brilliant ideas of E. Galois. To convert Galois's approach into a comprehensible theory, in other words, to set up Galois theory, has taken quite a period of time. The reason is that simultaneously several new concepts of algebra were emerging and had to be developed as natural prerequisites. In fact, the study of algebraic equations has served as a motivating terrain for a large part of abstract algebra, and according to this, algebraic equations will be visible as a guiding thread throughout the book.

To underline this point, I have included at the beginning a historical introduction to the problem of solving algebraic equations. Later, every chapter begins with some introductory remarks on "Background and Overview," where I give motivation for the material that follows and where I discuss some of its highlights on an informal level. In contrast to this, the remaining "regular" sections (some of them optional, indicated by a star) go step by step, elaborating the corresponding subject in full mathematical strength. I have tried to proceed in a way as simple and as clear as possible, basing arguments always on "true reasons," in other words, without resorting to simplifying ad hoc solutions. The text should therefore be useful for "any" course on the subject and even for self-study, certainly since it is essentially self-contained, up to a few prerequisites from linear algebra. Each section ends with a list of specially adapted exercises, some of them printed in *italics* to signify that there are solution proposals in the appendix.

On many occasions, I have given courses on the subject of this book, usually in units of two for consecutive semesters. In such courses I have addressed the "standard program" consisting of the unstarred sections. The latter yield a well-founded and direct access to the world of algebraic field extensions, with the fundamental theorem of Galois theory as a first milestone. Also let me point out that group theory has been split up into an elementary part in Chapter 1 and a more advanced part later in Chapter 5 that is needed for the applications

of Galois theory. Of course, if preferred, Chapter 5 can be covered immediately after Chapter 1. Finally, the optional starred sections complement the standard program or, in some cases, provide a first view on nearby areas that are more advanced. Such sections are particularly well suited for seminars.

The first versions of this book appeared in German as handouts for my students. They were later compiled into a book on algebra that appeared in 1993. I'm deeply indebted to my students and colleagues for their valuable comments and suggestions. All this found its way into later editions. The present English edition is a translation and critical revision of the eighth German edition of 2013. Here my thanks go to my colleague and friend Alan Huckleberry, with whom I discussed several issues of the English translation, as well as to Birkhäuser and its editorial team for the smooth editing and publishing procedure.

Münster, May 2018 Siegfried Bosch

Contents

Introduction: On the Problem of Solving Algebraic Equations 1

1 Elementary Group Theory 9
1.1 Groups 10
1.2 Cosets, Normal Subgroups, Factor Groups 15
1.3 Cyclic Groups 20

2 Rings and Polynomials 23
2.1 Polynomial Rings in One Variable 26
2.2 Ideals 32
2.3 Ring Homomorphisms, Factor Rings 35
2.4 Prime Factorization 41
2.5 Polynomial Rings in Several Variables 51
2.6 Zeros of Polynomials 57
2.7 A Theorem of Gauss 59
2.8 Criteria for Irreducibility 65
2.9 Theory of Elementary Divisors* 67

3 Algebraic Field Extensions 83
3.1 The Characteristic of a Field 85
3.2 Finite and Algebraic Field Extensions 87
3.3 Integral Ring Extensions* 94
3.4 Algebraic Closure 100
3.5 Splitting Fields 107
3.6 Separable Field Extensions 111
3.7 Purely Inseparable Field Extensions 119
3.8 Finite Fields 123
3.9 Beginnings of Algebraic Geometry* 126

4 Galois Theory 133
4.1 Galois Extensions 135
4.2 Profinite Galois Groups* 142
4.3 The Galois Group of an Equation 153
4.4 Symmetric Polynomials, Discriminant, Resultant* 162
4.5 Roots of Unity 176

4.6 Linear Independence of Characters . . . 186
4.7 Norm and Trace . . . 188
4.8 Cyclic Extensions . . . 194
4.9 Multiplicative Kummer Theory* . . . 200
4.10 General Kummer Theory and Witt Vectors* . . . 205
4.11 Galois Descent* . . . 224

5 More Group Theory . . . 231
5.1 Group Actions . . . 232
5.2 Sylow Groups . . . 237
5.3 Permutation Groups . . . 245
5.4 Solvable Groups . . . 249

6 Applications of Galois Theory . . . 255
6.1 Solvability of Algebraic Equations . . . 256
6.2 Algebraic Equations of Degree 3 and 4* . . . 264
6.3 Fundamental Theorem of Algebra . . . 272
6.4 Compass and Straightedge Construction . . . 275

7 Transcendental Field Extensions . . . 283
7.1 Transcendence Bases . . . 284
7.2 Tensor Products* . . . 290
7.3 Separable, Primary, and Regular Extensions* . . . 301
7.4 Differential Calculus* . . . 311

Appendix: Solutions to Exercises . . . 323

Literature . . . 355

Glossary of Notation . . . 357

Index . . . 361

Introduction

On the Problem of Solving Algebraic Equations

The word *algebra* is of Arabic origin (ninth century AD) and means doing calculations on equations, such as combining different terms of the equation, or changing terms by suitable manipulations on both sides of the equation. Here an equation is meant as a relation between known quantities, so-called coefficients, and unknown quantities or variables, whose possible value is to be determined by means of the equation. In algebra one is mostly interested in polynomial equations, for example of type

$$2x^3 + 3x^2 + 7x - 10 = 0,$$

where x stands for the unknown quantity. Such an equation will be referred to as an *algebraic* equation for x. Its *degree* is given by the exponent of the highest power of x that actually occurs in the equation. Algebraic equations of degree 1 are called *linear*. The study of these or, more generally, of systems of linear equations in finitely many variables, is a central problem in *linear algebra*.

On the other hand, *algebra* in the sense of the present book is about algebraic equations of higher degree in one variable. In today's language, this is the theory of field extensions together with all its abstract concepts, including those of group-theoretic nature that, in their combination, make possible a convenient and comprehensive treatment of algebraic equations. Indeed, even on an "elementary" level, modern algebra is much more influenced by abstract methods and concepts than one is used to from other areas, for example from analysis. The reason becomes apparent if we follow the problem of solving algebraic equations from a historical viewpoint, as we will briefly do in the following.

In the beginning, algebraic equations were used essentially in a practical manner, to solve certain numerical "exercises." For example, a renowned problem of ancient Greece (c. 600 BC – 200 AD) is the problem on the duplication of the cube. Given a cube of edge length 1, it asks to determine the edge length of a cube of double volume. In other words, the problem is to solve the algebraic equation $x^3 = 2$, which is of degree 3. Today the solution would be described by $x = \sqrt[3]{2}$. However, what is $\sqrt[3]{2}$ if only rational numbers are known? Since it was not possible to find a rational number whose third power is 2, one had to content oneself with approximate solutions and hence sufficiently good approximations of $\sqrt[3]{2}$. On the other hand, the duplication of the cube is a problem of

geometric nature. Hence, it suggests to try a geometric solution if other computational methods do not work. On many occasions, we find in ancient Greece constructions by compass and straightedge, for example of Euclid, that rely on intersection points of lines and circles with objects of the same type. But by applying such a technique, it is still not possible to construct $\sqrt[3]{2}$, as we know today; cf. Section 6.4. Since constructions by compass and straightedge could not always lead to the desired solution, one also finds constructions in terms of more complicated curves in ancient Greece.

Once it is accepted that for the solution of algebraic equations, say with rational coefficients, one needs the process of taking nth roots for variable n, besides the "rational" operations of addition, subtraction, multiplication, and division, we can pose the question whether a repeated application of these operations will be sufficient for calculating the solutions from the coefficients. This is the fundamental question on the *solvability of algebraic equations by radicals*. For example, algebraic equations of degree 1 and 2 are solvable by radicals:

$$\begin{aligned} x^1 + a = 0 \qquad &\Longleftrightarrow \qquad x = -a, \\ x^2 + ax^1 + b = 0 \qquad &\Longleftrightarrow \qquad x = -\frac{a}{2} \pm \sqrt{\frac{a^2}{4} - b}. \end{aligned}$$

The solvability of quadratic equations was basically already known to the Babylonians (from the end of the third millennium BC on), using elementary geometric methods, even if in specific examples that are conveyed, square roots were mainly taken from square numbers. From the ninth century AD on, after the Babylonian and the Greek periods had finished, the solution of quadratic equations was further refined by Arabian mathematicians. They also worked on the solvability of cubic equations and of equations of higher degree, however, without any noteworthy contribution to the subject.

The sensational discovery that cubic equations are solvable by radicals was achieved only around 1515 by the Italian S. del Ferro. He considered an equation of type $x^3 + ax = b$ for $a, b > 0$ and found as its solution

$$x = \sqrt[3]{\frac{b}{2} + \sqrt{\left(\frac{b}{2}\right)^2 + \left(\frac{a}{3}\right)^3}} + \sqrt[3]{\frac{b}{2} - \sqrt{\left(\frac{b}{2}\right)^2 + \left(\frac{a}{3}\right)^3}}.$$

Although he knew that before him, generations of mathematicians had worked on this problem without success, del Ferro kept his findings secret, without publishing them. However, we know about his work from the *Ars Magna*, some sort of schoolbook that was published by G. Cardano in 1545. Cardano had heard about del Ferro's solution formula in an indirect way and then was able to work it out by himself. Furthermore, he realized that as a rule, equations of degree 3 should have three solutions. It is remarkable in his work that Cardano was less hesitant than his contemporaries to use negative numbers. Also, there are some first signs by him of the use of complex numbers. Finally, his student L. Ferrari discovered after 1545 that algebraic equations of degree 4 are solvable by radicals; see Section 6.1 for the corresponding formulas.

During the next two centuries, there was only little progress on the solvability of algebraic equations. F. Viète discovered the connection between the coefficients of an equation and its solutions, which carries his name. From today's viewpoint this is a triviality if we use the decomposition of polynomials into linear factors. Furthermore, there was already a certain understanding about the multiplicities of solutions, including the idea that an algebraic equation of degree n should always have n solutions, counted with multiplicities, just as examples clearly show in ideal cases. However, it must be pointed out that the latter finding was only rather vague, since the nature of solutions, say real or complex, or even hypercomplex (neither real nor complex), was not made precise. At that time, there were also several attempts, unsuccessful though, for example by G. W. Leibniz, to solve algebraic equations of degree 5 and higher by radicals.

Finally, a certain consolidation of the situation was taking place by means of the *fundamental theorem of algebra*. The first ideas of its verification appeared in 1746 by J. d'Alembert, while further proofs of varying strength were carried out by L. Euler in 1749, by J. L. Lagrange in 1772, as well as later in 1799 by C. F. Gauss in his thesis. The theorem asserts that every nonconstant polynomial of degree n with complex coefficients admits precisely n complex zeros, counted with multiplicities, or in other words, that every such polynomial can be written as a product of linear factors. Even if the fundamental theorem of algebra did not directly contribute to the problem of solving algebraic equations by radicals, it nevertheless gave an answer to the question of where to look for solutions of such equations with rational, real, or complex coefficients. On this basis, further progress was achieved, particularly by Lagrange. In 1771 he subjected the solvability of algebraic equations of degree 3 and 4 to a complete revision and observed, among other things, that the cube roots in del Ferro's formula must be chosen in accordance with the side condition

$$\sqrt[3]{\frac{b}{2}+\sqrt{\left(\frac{b}{2}\right)^2+\left(\frac{a}{3}\right)^3}}\cdot\sqrt[3]{\frac{b}{2}-\sqrt{\left(\frac{b}{2}\right)^2+\left(\frac{a}{3}\right)^3}}=-\frac{a}{3}.$$

As a result, not nine possible values were obtained, but only the true solutions x_1, x_2, x_3 of the equation $x^3 + ax = b$ under consideration. However, more important was the detection that on choosing a nontrivial third root of unity ζ, i.e., a complex number $\zeta \neq 1$ satisfying $\zeta^3 = 1$, the expression

$$(x_1 + \zeta x_2 + \zeta^2 x_3)^3$$

takes only *two* different values on permuting the x_i and thus must satisfy a quadratic equation (with coefficients from the considered number domain, for example the rational numbers). In this way, the sums $x_{\pi(1)} + \zeta x_{\pi(2)} + \zeta^2 x_{\pi(3)}$, for any permutation π, can be determined by solving a quadratic equation and subsequently extracting a cube root. In particular, since x_1, x_2, x_3 can be calculated from these sums by means of rational operations, the solvability of the equation $x^3 + ax = b$ by radicals becomes clear. In a similar manner Lagrange characterized the solvability by radicals of algebraic equations of degree 4, in which

also in this case, permutations of the solutions play a major role. In this way, Lagrange introduced for the first time group-theoretic arguments into the discussion, an approach that eventually led Galois to a complete characterization of the solvability by radicals for algebraic equations of arbitrary degree.

Proceeding like Lagrange, Gauss studied, in 1796, the solutions of the equation $x^p - 1 = 0$ for prime numbers $p > 2$, relying on preparative work by A. T. Vandermonde. The corresponding permutations of the solutions give rise to groups again, "cyclic" ones in this case. Furthermore, the methods of Gauss led to new insight on the question of which regular polygons with a given number n of sides can be obtained in terms of compass and straightedge constructions. Around this time there were also studies by P. Ruffini, rendered more precise by N. H. Abel in 1820, showing that the "generic equation" of degree n is not solvable by radicals for $n \geq 5$.

After such a number of partial results that were obtained mainly through a systematic application of group arguments, the time seemed to be ripe for a full clarification of the problem on the solvability of algebraic equations. This culminating step was successfully accomplished by E. Galois, with his brilliant ideas in the years 1830–1832. To a much greater extent than Abel, it was Galois who had very precise ideas about enlarging number domains, for example the rational numbers, by adding solutions of algebraic equations; from today's point of view, it concerned a prestage of the notion of a field, as well as the technique of adjoining algebraic elements. Galois also introduced the notion of irreducibility for algebraic equations. Furthermore, he proved the primitive element theorem for the splitting field L of an algebraic equation $f(x) = 0$ with simple solutions, i.e., for the field generated by all solutions $x_1, \ldots, x_r$ of such an equation. The theorem asserts that there is an irreducible algebraic equation $g(y) = 0$ such that L contains all solutions $y_1, \ldots, y_s$ of this equation and, in addition, is obtained from the coefficient domain by adjoining any single one of the solutions y_j. Now it was Galois's idea to represent the x_i in an obvious way as functions of y_1, say $x_i = h_i(y_1)$, and then to replace y_1 by an arbitrary element y_j. As he showed, the elements $h_i(y_j)$, $i = 1, \ldots, r$, represent again all solutions of $f(x) = 0$. Furthermore, substituting y_1 by y_j gives rise to a permutation π_j of the x_i, and it follows that the π_j form a group, indeed, the "Galois group" of the equation $f(x) = 0$ as we say today.

Based on these facts, Galois was led to the fundamental observation that the subfields of the splitting field L correspond in a certain way to the subgroups of the corresponding Galois group G, a result that we nowadays formulate in a refined way as the "fundamental theorem of Galois theory." Finally, making use of this knowledge, Galois was able to show that the equation $f(x) = 0$ is solvable by radicals precisely when the group G admits a chain of subgroups $G = G_0 \supset \ldots \supset G_n = \{1\}$, where in each case, G_{i+1} is a normal subgroup of G_i such that the factor group G_i/G_{i+1} is cyclic. We could continue now discussing Galois theory in greater detail, but let us refer to the later Sections 4.1, 4.3, 4.8, and 6.1 instead.

In any case, as we have seen, the delicate problem of solving algebraic equations by radicals, which is quite easy to formulate, was fully clarified by Galois, due to his unconventional new ideas. In particular, one can now understand why, over many centuries, mathematicians were denied access to the problem. The solution does not consist of a comprehensible condition on the coefficients of the equation under consideration, say in terms of a formula. On the contrary, even to be formulated it requires a new language, more precisely, a new way of thinking in combination with new concepts, that could only be established in a long process of trial and error and of studying examples. Also we have to point out that the true benefit of Galois's investigations does not concern so much his contribution to the solvability by radicals of algebraic equations, but instead, consists in the fundamental correspondence between algebraic equations and their associated "Galois" groups. Indeed, the fundamental theorem of Galois theory provides a means to characterize the "nature" of solutions of arbitrary algebraic equations in terms of group-theoretic properties. In view of this fact, the task of solving specific algebraic equations by radicals has largely lost its original significance.

And how was Galois's contribution perceived by his contemporaries? To give an impression, we take a brief look at Galois's life; see also [11], Section 7. Evariste Galois was born in 1811 near Paris and died in 1832 at the age of only 20 years. Already during his schooldays he looked at papers of Lagrange and wrote a first small treatise on continued fractions. Twice he tried to join the renowned *Ecole Polytechnique* in Paris, but was not able to pass the entrance examination, so that finally, he had to settle for the *Ecole Normale*. Here he began his studies in 1829, at the age of 18. In the same year he submitted a first *Mémoire* to the *Académie des Sciences* concerning the solution of algebraic equations. However, the manuscript did not receive any attention and was even lost, as was a second one that he submitted a week later. After another *Mémoire* had suffered the same fate in 1830, Galois made a final attempt in early 1831, submitting his paper on the solvability of algebraic equations by radicals that today is judged to be his most prominent work. This time it was refereed, but declined for reasons of immaturity and incomprehensibility. Disappointed that he could not find any recognition in mathematics, Galois turned his attention to the political events of his epoch. Due to his new activities he was several times arrested and eventually condemned to imprisonment. Finally, in May 1832 he was provoked to fight a duel, where he met his death. However, to preserve his work for posterity just in case he did not survive, Galois wrote a letter to a friend during the night before the duel, in which he put together his pioneering discoveries in programmatic form. Although this program was published in 1832, the significance of Galois's studies was not immediately recognized. One may speculate about the reason, but two facts are certainly responsible for this. Firstly, Galois was an unknown young mathematician, besides that with a dubious history. On the other hand, the characterization of the solvability of algebraic equations made such an inapproachable impression that no one among Galois's contemporaries was prepared to accept this as a serious solution to

the problem. Also note that Lagrange, whose important preparative work was mentioned before, had died in 1813.

We do not want to describe in full detail how Galois's ideas eventually made their way to recognition and esteem. A major point is certainly the fact that J. Liouville, about 10 years after Galois's death, came across his work and was able to publish a part of it in 1846. In fact, during the second half of the nineteenth century a phase began in which, among other things, one was concerned with understanding and polishing Galois's ideas. Soon the problem of solvability of algebraic equations by radicals was reduced to its actual size. The problem was of extreme importance only because it had opened the door to an even more wide-ranging classification of irrational numbers, including transcendence aspects. Already in 1844 Liouville could establish the existence of transcendental numbers in a constructive way, a result that G. Cantor obtained more rigorously in 1874 using a countability argument. Similar studies of this type concern the transcendence proofs for e in 1873 by Ch. Hermite [8], as well as for π in 1882 by F. Lindemann [13]. Furthermore, some transcendence problems of general type were addressed in 1910 by E. Steinitz in his paper [15].

Through Galois's work, it became apparent that focusing on single algebraic equations was somehow cumbersome. It was better to be flexible and to consider, so to speak, several equations at the same time, possibly also for different coefficient domains. This new insight led to the study of so-called algebraic field extensions, replacing single equations as considered before. The first to really follow this plan in Galois theory was R. Dedekind in his lectures 1855–1858 in Göttingen. In particular, he interpreted Galois groups as automorphism groups of fields and not only as groups that permute the solutions of an algebraic equation. Another significant improvement of the theory is due to L. Kronecker, who published in 1887 the construction principle for algebraic field extensions that is named after him. In this way, it became possible to set up Galois theory without relying on the fundamental theorem of algebra, and thereby to free it from the physical presence of complex numbers, for example in order to adapt it to finite fields.

Taking into account all these developments, we are already quite close to the concepts that are still followed in the theory of field extensions today. Of course, there have been further completions, ameliorations, and simplifications that were essentially presented within the framework of books on the subject. Worth mentioning are—in historical order—the publications of H. Weber [17], B. L. van der Waerden [16], E. Artin [1], [2], as well as further pioneering books by N. Bourbaki [5] and S. Lang [12]. Even though the theory may nowadays seem to be "completed," appearing in "optimal" shape, I would like to encourage the reader to remember from time to time the arduous journey the problem of solving algebraic equations has made from its beginnings on. Only if one bears in mind the enormous difficulties that had to be overcome will one understand and appreciate the fascinating solutions that mathematicians have found in difficult struggles over the course of centuries.

However, we do not want to give the impression that the investigation of algebraic equations has come to an end today. On the contrary, it has found its natural continuation in the study of systems of algebraic equations in several unknown quantities, within the fields of algebraic geometry, see [3], and number theory. Also concerning this setting, we can mention a problem that is easy to formulate, but which has resisted the attacks of mathematicians for a rather long period of time. It was solved only in the recent past, in the years 1993/94 by A. Wiles with the help of R. Taylor. We are alluding to *Fermat's last theorem*, a conjecture stating that the equation $x^n + y^n = z^n$ does not admit a nontrivial integer solution for $n \geq 3$. It is said that Fermat, around 1637, had noted this conjecture in the margin of his copy of Diophantus's *Arithmetica* (c. 250 AD), adding that he had a truely marvelous demonstration for it, which, however, the margin was too narrow to contain.

1. Elementary Group Theory

Background and Overview

There are two important reasons for considering groups in this book. On the one hand, the notion of a group exhibits a fundamental mathematical structure that is found, for example, in rings, fields, vector spaces, and modules, in which one interprets the inherent addition as a law of composition. All groups of this type are commutative or, as we also will say, abelian, referring to the mathematician N. H. Abel. On the other hand, there are groups originating from another source, such as the so-called Galois groups related to the work of E. Galois. These groups will be of central interest for us, serving as a key tool for the investigation of algebraic equations. From a simplified point of view, Galois groups are permutation groups, i.e., groups whose elements describe bijective transformations (self-maps) on sets like $\{1, \ldots, n\}$.

The main feature of a group G is its law of composition that assigns to a pair of elements $g, h \in G$ a third element $g \circ h \in G$, called the product or, in the commutative case, the sum of g and h. Such laws of composition are always around when one is doing calculations with numbers. But for a long time there was no need to pay special attention to the properties of these laws, since the latter were judged to be "evident." Therefore, one can understand that up to the beginning of the seventeenth century, negative results from computations, for example from subtractions, were perceived as being "suspicious," due to the fact that negative numbers did not have any precise meaning yet. However, from the nineteenth century on, the notion of a group began to take shape, notably when laws of composition were applied to objects that could not be interpreted as numbers anymore. For example, permutation groups played an important role in the attempts to solve algebraic equations. Since the related groups consist of only finitely many elements, it was still possible to formulate the group axioms without explicitly mentioning "inverse elements," an approach that does not extend to the infinite case; cf. Exercise 3 of Section 1.1. An explicit postulation of "inverse elements," and thereby an axiomatic characterization of groups from today's point of view, emerged only at the end of the nineteenth century in the works of S. Lie and H. Weber. Prior to this, Lie, when studying his "transformation groups," had still tried to derive the existence of inverse elements from the remaining group axioms, however without success.

In the present chapter we will explain some of the basics on groups, material that most readers will certainly be familiar with already. In addition to the definition of a group, we concentrate on normal subgroups, associated factor groups, as well as on cyclic groups. Already at this stage one can notice the lasting impact that the problem of solving algebraic equations, and in particular, Galois theory, have exercised on the development of groups. For example, the notion of a normal subgroup is strongly related to the fundamental theorem of Galois theory 4.1/6.[1] Indeed, this theorem states, among other things, that an intermediate field E of a finite Galois extension L/K is normal over K in the sense of 3.5/5 if and only if the subgroup of the Galois group $\mathrm{Gal}(L/K)$ corresponding to E is normal. Also note that referring to Proposition 1.2/3 as the theorem of Lagrange is inspired by group-theoretic arguments that Lagrange introduced when working on the solution of algebraic equations.

More involved results on groups, and in particular, permutation groups, that are of special interest from the viewpoint of Galois theory, will be presented in Chapter 5. In addition, let us mention the fundamental theorem of finitely generated abelian groups 2.9/9, which provides a classification of such groups. Its proof will be carried out within the context of elementary divisors.

1.1 Groups

Let M be a set and $M \times M$ the Cartesian product with itself. An (*inner*) *law of composition* on M is a map $M \times M \longrightarrow M$. In many cases we will write the image of a given pair $(a, b) \in M \times M$ as a "product" $a \cdot b$ or ab. Thus, in terms of elements, the law of composition on M is characterized by the assignment $(a, b) \longmapsto a \cdot b$. The law is said to be

associative if $(ab)c = a(bc)$ for all $a, b, c \in M$, and

commutative if $ab = ba$ for all $a, b \in M$.

An element $e \in M$ is called a *unit element* or a *neutral element* with respect to the law of composition on M if $ea = a = ae$ holds for all $a \in M$. Such a unit element e is uniquely determined by its defining property; usually we will write 1 instead of e. A set M with a law of composition $\sigma\colon M \times M \longrightarrow M$ is called a *monoid* if σ is associative and M admits a unit element with respect to σ.

For a monoid M and elements $a_1, \ldots, a_n \in M$, the product

$$\prod_{i=1}^{n} a_i := a_1 \cdot \ldots \cdot a_n$$

is defined. Note that a special bracketing on the right-hand side is not necessary, since the law of composition is assumed to be associative (use an intelligent inductive argument to prove this). Empty products are not excluded: we set

$$\prod_{i=1}^{0} a_i := e = \text{ unit element.}$$

[1] As an example, note that "4.1/6" refers to the sixth numbered item in Section 4.1.

For $a \in M$ and an exponent $n \in \mathbb{N}$, the nth power a^n is defined in the usual way.[2] Note that $a^0 = e$, due to the convention on empty products. An element $b \in M$ is called an *inverse* of a given element $a \in M$ if $ab = e = ba$. Then b, if it exists, is uniquely determined by a. Indeed, if $ab' = e = b'a$ for some $b' \in M$, then

$$b = eb = b'ab = b'e = b'.$$

If an element $a \in M$ admits an inverse, it is denoted by a^{-1}.

Definition 1. *A* group *is a monoid G such that every element of G admits an inverse. More explicitly, this means we are given a set G with a law of composition $G \times G \longrightarrow G$, $(a, b) \longmapsto ab$, such that*:

(i) *The law is* associative, *i.e., we have $(ab)c = a(bc)$ for $a, b, c \in G$.*

(ii) *There exists a* unit element, *i.e., an element $e \in G$ such that $ea = a = ae$ for all $a \in G$.*

(iii) *Every element $a \in G$ admits an* inverse, *i.e., an element $b \in G$ such that $ab = e = ba$.*

The group is called commutative *or* abelian *if the law is commutative, i.e., if*

(iv) *$ab = ba$ for all $a, b \in G$.*

Remark 2. *In Definition* 1 *it is enough to require the following weaker conditions instead of* (ii) *and* (iii):

(ii′) *There is a* left neutral element *in G, i.e., an element $e \in G$ satisfying $ea = a$ for all $a \in G$.*

(iii′) *For each $a \in G$ there is a* left inverse *in G, i.e., an element $b \in G$ such that $ba = e$.*

For a verification of the fact that conditions (ii′) and (iii′) in conjunction with (i) are sufficient for defining a group, we refer to Exercise 1 below and to its solution given in the appendix.

In dealing with abelian groups, the law of composition will usually be noted additively, i.e., we write $a + b$ instead of $a \cdot b$ and $\sum a_i$ instead of $\prod a_i$, as well as $n \cdot a$ instead of an nth power a^n. Accordingly, $-a$ instead of a^{-1} will denote the inverse of an element a, and 0 (*zero element*) instead of e or 1 will be the neutral element. Here are some examples of monoids and groups:

(1) $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, equipped with the usual addition, are abelian groups.

(2) $\mathbb{Q}^*$, $\mathbb{R}^*$, $\mathbb{C}^*$, equipped with the usual multiplication, are abelian groups; the same is true for $\mathbb{Q}_{>0} = \{x \in \mathbb{Q}\,;\, x > 0\}$ and $\mathbb{R}_{>0} = \{x \in \mathbb{R}\,;\, x > 0\}$. More generally, we can look at matrix groups from linear algebra like SL_n and GL_n, taking coefficients in $\mathbb{Q}, \mathbb{R}$, or $\mathbb{C}$. For $n > 1$, the latter groups fail to be commutative.

[2] $\mathbb{N}$ is the set of natural numbers *including* 0.

(3) $\mathbb{N}$ equipped with the usual addition, $\mathbb{N}$, $\mathbb{Z}$ with the usual multiplication, are commutative monoids, but fail to be groups.

(4) For a set X, let $S(X)$ be the set of bijective maps $X \longrightarrow X$. Then the composition of maps makes $S(X)$ a group. This group is not abelian if X consists of at least three elements. In the special case $X = \{1, \ldots, n\}$, we put $\mathfrak{S}_n := S(X)$ and call it the *symmetric group of degree n* or the *group of permutations* of the integers $1, \ldots, n$. Quite often a permutation $\pi \in \mathfrak{S}_n$ is described explicitly in the form

$$\begin{pmatrix} 1 & \ldots & n \\ \pi(1) & \ldots & \pi(n) \end{pmatrix},$$

where $\pi(1), \ldots, \pi(n)$ are the images under π. By counting all ordered combinations of $1, \ldots, n$, we see that $\mathfrak{S}_n$ consists of precisely $n!$ elements.

(5) Let X be a set and G a group. We write $G^X := \mathrm{Map}(X, G)$ for the set of all maps $X \longrightarrow G$; it is canonically a group. Indeed, given $f, g \in G^X$, the product $f \cdot g$ is defined by $(f \cdot g)(x) := f(x) \cdot g(x)$ for x varying over X. Thus, $f \cdot g$ is obtained by multiplying values of f and g with respect to the law of composition on G. We call G^X the *group of G-valued functions on X*. In the same way, the group $G^{(X)}$ can be considered; it consists of all maps $f\colon X \longrightarrow G$ satisfying $f(x) = 1$ for almost all $x \in X$ (i.e., for all $x \in X$, up to finitely many exceptions). The groups G^X and $G^{(X)}$ are commutative if G is commutative. Furthermore, G^X coincides with $G^{(X)}$ if X consists of only finitely many elements.

(6) Let X be an index set and $(G_x)_{x \in X}$ a family of groups. The set-theoretic product $\prod_{x \in X} G_x$ becomes a group if we define the composition of two elements $(g_x)_{x \in X}$, $(h_x)_{x \in X} \in \prod_{x \in X} G_x$ componentwise via

$$(g_x)_{x \in X} \cdot (h_x)_{x \in X} := (g_x \cdot h_x)_{x \in X}.$$

The group $\prod_{x \in X} G_x$ is called the *direct product* of the groups G_x, $x \in X$. In the special case $X = \{1, \ldots, n\}$, the direct product is usually denoted by $G_1 \times \ldots \times G_n$. If all groups G_x are copies of one and the same group G, then we have $\prod_{x \in X} G_x = G^X$, using the notation of the preceding example. In addition, if X is finite, say $X = \{1, \ldots, n\}$, one writes G^n instead of G^X or $G^{(X)}$.

Definition 3. *Let G be a monoid. A subset $H \subset G$ is called a* submonoid *if H satisfies the following conditions*:

(i) $e \in H$,

(ii) $a, b \in H \Longrightarrow ab \in H$.

If G is a group, H is called a subgroup *of G if in addition,*

(iii) $a \in H \Longrightarrow a^{-1} \in H$.

In particular, a subgroup of a group G is a submonoid that is closed under the process of taking inverses.

In defining a subgroup $H \subset G$, condition (i) can be weakened by simply requiring $H \neq \emptyset$, since this implies $e \in H$ using (ii) and (iii). Of course, similar reasoning is not possible for monoids. Every group G admits $\{e\}$ and G as *trivial subgroups*. Given $m \in \mathbb{Z}$, the set $m\mathbb{Z}$ consisting of all integral multiples of m is a subgroup of the additive group $\mathbb{Z}$. We will see in 1.3/4 that all subgroups of $\mathbb{Z}$ are of this type. More generally, every element a of a group G gives rise to a so-called *cyclic subgroup* of G. It consists of all powers a^n, $n \in \mathbb{Z}$, where we put $a^n = (a^{-1})^{-n}$ for $n < 0$; see also Section 1.3.

Definition 4. *Let G, G' be monoids with corresponding unit elements e and e'. A* monoid homomorphism $\varphi\colon G \longrightarrow G'$ *is a map φ from G to G' such that*

(i) $\varphi(e) = e'$,

(ii) $\varphi(ab) = \varphi(a)\varphi(b)$ *for all $a, b \in G$.*

Furthermore, if G, G' are groups, φ is called a group homomorphism.

Remark 5. *A map $\varphi\colon G \longrightarrow G'$ between groups is a group homomorphism if and only if $\varphi(ab) = \varphi(a)\varphi(b)$ for all $a, b \in G$.*

Proof. We conclude $\varphi(e) = e'$ from $\varphi(e) = \varphi(ee) = \varphi(e)\varphi(e)$. □

Remark 6. *Let $\varphi\colon G \longrightarrow G'$ be a group homomorphism. Then inverse elements satisfy $\varphi(a^{-1}) = (\varphi(a))^{-1}$ for all $a \in G$.*

Proof. $e' = \varphi(e) = \varphi(aa^{-1}) = \varphi(a)\varphi(a^{-1})$. □

A group homomorphism $\varphi\colon G \longrightarrow G'$ is called an *isomorphism* if φ admits an inverse, i.e., if there exists a group homomorphism $\psi\colon G' \longrightarrow G$ such that $\psi \circ \varphi = \mathrm{id}_G$ and $\varphi \circ \psi = \mathrm{id}_{G'}$, for id_G and $\mathrm{id}_{G'}$ the identity maps on G and G'. Note that a group homomorphism is an isomorphism if and only if it is bijective. Injective (resp. surjective) group homomorphisms $G \longrightarrow G'$ are called *monomorphisms* (resp. *epimorphisms*). An *endomorphism* of G is a homomorphism $G \longrightarrow G$, an *automorphism* of G is an isomorphism $G \longrightarrow G$.

Let $\varphi\colon G \longrightarrow G'$ and $\psi\colon G' \longrightarrow G''$ be group homomorphisms. Then the composition $\psi \circ \varphi\colon G \longrightarrow G''$ is a group homomorphism again. Moreover, given a group homomorphism $\varphi\colon G \longrightarrow G'$, we can consider the subgroups

$$\ker \varphi = \big\{g \in G\,;\, \varphi(g) = 1\big\} \subset G \qquad (\textit{kernel} \text{ of } \varphi)$$

as well as

$$\operatorname{im} \varphi = \varphi(G) \subset G' \qquad (\textit{image} \text{ of } \varphi).$$

Note that φ is injective if and only if $\ker \varphi = \{1\}$. We continue by listing some examples of homomorphisms.

(1) Let G be a monoid. Fixing an element $x \in G$, the map

$$\varphi\colon \mathbb{N} \longrightarrow G, \qquad n \longmapsto x^n,$$

defines a monoid homomorphism when $\mathbb{N}$ is considered a monoid with respect to the usual addition. If G is a group, we obtain in the same way a group homomorphism

$$\varphi\colon \mathbb{Z} \longrightarrow G, \qquad n \longmapsto x^n,$$

where we put $x^n := (x^{-1})^{-n}$ for $n < 0$. On the other hand, it is clear that each monoid homomorphism $\varphi\colon \mathbb{N} \longrightarrow G$, resp. each group homomorphism $\varphi\colon \mathbb{Z} \longrightarrow G$, must be of this type; just put $x = \varphi(1)$.

(2) Let G be a group and $S(G)$ the corresponding group of all bijective maps from G to itself. For $a \in G$, let $\tau_a \in S(G)$ be the *left translation* by a on G, i.e., the map

$$\tau_a\colon G \longrightarrow G, \qquad g \longmapsto ag.$$

Then

$$G \longrightarrow S(G), \qquad a \longmapsto \tau_a,$$

defines an injective group homomorphism, and we can identify G with its image in $S(G)$, thereby interpreting G as a subgroup of $S(G)$. In particular, every group consisting of n elements can be viewed as a subgroup of the symmetric group $\mathfrak{S}_n$, a result generally known as Cayley's theorem.

Similarly as before, we can define *right translations* on G. Also these can be used to construct an injective group homomorphism $G \longrightarrow S(G)$; see Exercise 4 below.

(3) Let G be an abelian group, and fix $n \in \mathbb{N}$. Then

$$G \longrightarrow G, \qquad g \longmapsto g^n,$$

is a group homomorphism.

(4) Let G be a group, and fix $a \in G$. Then

$$\varphi_a\colon G \longrightarrow G, \qquad g \longmapsto aga^{-1},$$

is a so-called *inner automorphism* of G. The set $\mathrm{Aut}(G)$ of all automorphisms of G is a group under the composition of automorphisms in terms of maps, and the map $G \longrightarrow \mathrm{Aut}(G)$, $a \longmapsto \varphi_a$, is a group homomorphism.

(5) The exponential function defines a group isomorphism

$$\mathbb{R} \xrightarrow{\sim} \mathbb{R}_{>0}, \qquad x \longmapsto \exp(x).$$

Of course, to verify this we must use the properties of the exponential function known from analysis, notably the functional equation

$$\exp(x + y) = \exp(x) \cdot \exp(y).$$

Exercises

1. *Give a proof of Remark* 2.

2. *The exponential function gives rise to an isomorphism between the additive group* $\mathbb{R}$ *and the multiplicative group* $\mathbb{R}_{>0}$. *Check whether there can exist a similar isomorphism between the additive group* $\mathbb{Q}$ *and the multiplicative group* $\mathbb{Q}_{>0}$.

3. For a monoid G, consider the following conditions:

   (i) G is a group.

   (ii) For $a, x, y \in G$, each of the equations $ax = ay$ and $xa = ya$ implies $x = y$.

   Then (i) $\Longrightarrow$ (ii). Show that the reverse implication holds for finite monoids G, but not for arbitrary monoids G.

4. Let G be a group. Analogously to the notation of left translations, introduce right translations on G and use them to construct an injective group homomorphism $G \longrightarrow S(G)$.

5. Let X be a set and consider a subset $Y \subset X$. Show that the group $S(Y)$ can be viewed canonically as a subgroup of $S(X)$.

6. Let G be a finite abelian group. Show that $\prod_{g \in G} g^2 = 1$.

7. Let G be a group such that $a^2 = 1$ for all $a \in G$. Show that G is abelian.

8. Consider a group G together with subgroups $H_1, H_2 \subset G$. Show that $H_1 \cup H_2$ is a subgroup of G if and only if $H_1 \subset H_2$ or $H_2 \subset H_1$ holds.

1.2 Cosets, Normal Subgroups, Factor Groups

Let G be a group and $H \subset G$ a subgroup. A *left coset* of H in G is a subset of G of type

$$aH := \{ah\,;\, h \in H\},$$

where $a \in G$.

Proposition 1. *Any two left cosets of* H *in* G *have the same cardinality*[3] *and are disjoint if they do not coincide. In particular,* G *is the disjoint union of all left cosets of* H *in* G.

Proof. For each $a \in G$, the left translation $H \longrightarrow aH$, $h \longmapsto ah$, is bijective. Therefore, all left cosets of H in G have the same cardinality. The second assertion is a consequence of the following lemma:

Lemma 2. *Let* aH *and* bH *be left cosets of* H *in* G. *Then the following conditions are equivalent*:

(i) $aH = bH$.
(ii) $aH \cap bH \neq \emptyset$.
(iii) $a \in bH$.
(iv) $b^{-1}a \in H$.

[3] Two sets X, Y are said to have the *same cardinality* if there exists a bijection $X \longrightarrow Y$.

Proof. The implication (i) $\Longrightarrow$ (ii) is trivial, since $H \neq \emptyset$. Next, assume (ii). There exists an element $c \in aH \cap bH$, say $c = ah_1 = bh_2$, where $h_1, h_2 \in H$. This means that $a = bh_2h_1^{-1} \in bH$, and we see that (iii) holds. Multiplication by b^{-1} and b shows that (iii) is equivalent to (iv). Finally, assuming (iv), we get $a \in bH$ and hence $aH \subset bH$. On the other hand, the inverse of $b^{-1}a \in H$ must be contained in H as well: thus $a^{-1}b \in H$. Similarly as before we conclude that $bH \subset aH$ and therefore $aH = bH$. □

All elements of a left coset aH are called *representatives* of this coset. In particular, a is a representative of aH, and we see from Lemma 2 that $a'H = aH$ for every representative $a' \in aH$. The set of left cosets of H in G is denoted by G/H. Analogously one defines the set $H\backslash G$ of all *right cosets* of H in G, i.e., of all subsets of type

$$Ha = \{ha\,;\, h \in H\},$$

where $a \in G$. It is easily checked that the bijection

$$G \longrightarrow G, \qquad g \longmapsto g^{-1},$$

maps a left coset aH bijectively onto the right coset Ha^{-1} and thereby defines a bijective map

$$G/H \longrightarrow H\backslash G, \qquad aH \longmapsto Ha^{-1}.$$

In particular, Proposition 1 and Lemma 2 (with the obvious modifications in Lemma 2) are valid for right cosets as well. The number of elements in G/H, resp. $H\backslash G$, is called the *index* $(G : H)$ of H in G. Writing $\operatorname{ord} G$ for the number of elements of a group G and calling it the *order* of G, we can conclude from Proposition 1 the following corollary:

Corollary 3 (Theorem of Lagrange). *Let G be a finite group and H a subgroup of G. Then*

$$\operatorname{ord} G = \operatorname{ord} H \cdot (G : H).$$

Definition 4. *A subgroup $H \subset G$ is called a* normal subgroup *of G if $aH = Ha$ for all $a \in G$, i.e., if for each element $a \in G$ the associated left and right cosets of H in G coincide. If such is the case, the coset $aH = Ha$ given by a is referred to as the* residue class *of a modulo H.*

The condition $aH = Ha$ can be rewritten as $aHa^{-1} = H$. Note that a subgroup $H \subset G$ is normal as soon as we have $aHa^{-1} \subset H$ for all $a \in G$ (alternatively: $H \subset aHa^{-1}$ for all $a \in G$). Indeed, $aHa^{-1} \subset H$ is equivalent to $aH \subset Ha$, and likewise, $a^{-1}Ha \subset H$ to $Ha \subset aH$. Moreover, observe that every subgroup of a commutative group is normal.

Remark 5. *The kernel of any group homomorphism $\varphi\colon G \longrightarrow G'$ is a normal subgroup in G.*

Proof. $\ker\varphi$ is a subgroup of G, and we get $a\cdot(\ker\varphi)\cdot a^{-1}\subset\ker\varphi$ for all $a\in G$ from 1.1/6. □

Now, starting with a normal subgroup $N\subset G$, we want to look at the reverse problem of constructing a group homomorphism $\varphi\colon G\longrightarrow G'$ whose kernel coincides with N. To do this we introduce a suitable group structure on the set of residue classes G/N and define φ as the projection $\pi\colon G\longrightarrow G/N$, assigning to an element $a\in G$ the corresponding residue class aN. As a technical tool we define the product of two subsets $X,Y\subset G$ by

$$X\cdot Y:=\{x\cdot y\in G\,;\,x\in X,y\in Y\}.$$

Then, using the fact that N is normal in G, we can write for $a,b\in G$,

$$(aN)\cdot(bN)=\{a\}\cdot(Nb)\cdot N=\{a\}\cdot(bN)\cdot N=\{ab\}\cdot(NN)=(ab)N.$$

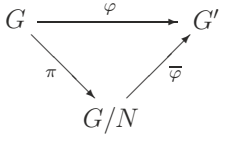
As we see, the product of the cosets aN and bN with representatives a and b is a coset again, namely the one $(ab)N$ with representative ab. Now, considering this product as a law of composition "$\cdot$" on G/N, we conclude immediately from the properties of G being a group that G/N is a group again; $N=1N$ is the unit element in G/N, and $a^{-1}N$ is the inverse of $aN\in G/N$. Furthermore, it is clear that the map

$$\pi\colon G\longrightarrow G/N,\qquad a\longmapsto aN,$$

the *canonical projection* from G to G/N, is a surjective group homomorphism satisfying $\ker\pi=N$. The group G/N is called the *factor group* or the *residue class group* of G modulo N.

For many applications it is important to know that the group homomorphism $\pi\colon G\longrightarrow G/N$ satisfies a so-called *universal property* that characterizes G/N up to canonical isomorphism:

Proposition 6 (Fundamental theorem on homomorphisms). *Let $\varphi\colon G\longrightarrow G'$ be a group homomorphism and $N\subset G$ a normal subgroup such that $N\subset\ker\varphi$. Then there exists a unique group homomorphism $\overline{\varphi}\colon G/N\longrightarrow G'$ satisfying $\varphi=\overline{\varphi}\circ\pi$, i.e., such that the diagram*



is commutative. Furthermore,

$$\operatorname{im}\overline{\varphi}=\operatorname{im}\varphi,\qquad\ker\overline{\varphi}=\pi(\ker\varphi),\qquad\ker\varphi=\pi^{-1}(\ker\overline{\varphi}),$$

and it follows that $\overline{\varphi}$ is injective if and only if $N=\ker\varphi$.

Proof. If $\overline{\varphi}$ exists, then

$$\overline{\varphi}(aN) = \overline{\varphi}\big(\pi(a)\big) = \varphi(a)$$

for $a \in G$ and we see that $\overline{\varphi}$ is unique. On the other hand, we can try to set $\overline{\varphi}(aN) = \varphi(a)$ when defining $\overline{\varphi}$. However, for this to work well, it is necessary to know that $\varphi(a)$ is independent of the choice of the representative $a \in aN$. To justify this, assume $aN = bN$ for two elements $a, b \in G$. Then we have $b^{-1}a \in N \subset \ker\varphi$ and thus $\varphi(b^{-1}a) = 1$, which yields $\varphi(a) = \varphi(b)$. That $\overline{\varphi}$ is a group homomorphism follows from the definition of the group law on G/N, or in other words, from the fact that π is an epimorphism. This settles the existence of $\overline{\varphi}$.

Finally, the equation $\ker\varphi = \pi^{-1}(\ker\overline{\varphi})$ follows from the fact that φ is the composition of $\overline{\varphi}$ and π. Moreover, we can conclude $\operatorname{im}\overline{\varphi} = \operatorname{im}\varphi$ as well as $\ker\overline{\varphi} = \pi(\ker\varphi)$ from the surjectivity of π. □

Corollary 7. *If $\varphi\colon G \longrightarrow G'$ is a surjective group homomorphism, then G' is canonically isomorphic to $G/\ker\varphi$.*

As an application of Proposition 6, we want to prove the so-called *isomorphism theorems* for groups.

Proposition 8 (First isomorphism theorem). *Let G be a group, $H \subset G$ a subgroup, and $N \subset G$ a normal subgroup of G. Then HN is a subgroup of G admitting N as a normal subgroup, and $H \cap N$ is a normal subgroup of H. The canonical homomorphism*

$$H/H \cap N \longrightarrow HN/N$$

is an isomorphism.

Proof. Using the fact that N is normal in G, one easily shows that HN is a subgroup of G. Furthermore, N is normal in HN, since it is normal in G. Now consider the composition of homomorphisms

$$H \hookrightarrow HN \xrightarrow{\pi} HN/N,$$

where π is the canonical projection. It is surjective and has $H \cap N$ as its kernel. Therefore, $H \cap N$ is a normal subgroup in H, and the induced homomorphism

$$H/H \cap N \longrightarrow HN/N$$

is an isomorphism, due to Proposition 6 or Corollary 7. □

Proposition 9 (Second isomorphism theorem). *Let G be a group and let N, H be normal subgroups of G satisfying $N \subset H \subset G$. Then N is normal in H as well, and one can view H/N as a normal subgroup of G/N. Furthermore, the canonical group homomorphism*

$$\big(G/N\big)\big/\big(H/N\big) \longrightarrow G/H$$

is an isomorphism.

Proof. To begin with, let us explain how to view H/N as a subgroup of G/N. Look at the group homomorphism

$$H \hookrightarrow G \xrightarrow{\pi} G/N,$$

where π is the canonical projection. Since this homomorphism admits N as kernel, it induces by Proposition 6 a monomorphism $H/N \hookrightarrow G/N$. Thus, we can identify H/N with its image in G/N.

Next observe that the kernel of the canonical projection $G \longrightarrow G/H$, which is H, contains N as a normal subgroup. Therefore, using Proposition 6, the projection $G \longrightarrow G/H$ induces an epimorphism $G/N \longrightarrow G/H$ whose kernel is normal in G/N and coincides with the image of H under the projection $G \longrightarrow G/N$; the latter image was identified with H/N before. Now, applying Proposition 6 or Corollary 7 again, we see that $G/N \longrightarrow G/H$ gives rise to an isomorphism

$$\big(G/N\big)\big/\big(H/N\big) \xrightarrow{\sim} G/H.$$

□

Exercises

1. *Let G be a group and H a subgroup of index 2. Show that H is normal in G. Is the same assertion true in the case that H is of index 3 in G?*

2. *Let G be a group and $N \subset G$ a normal subgroup. Give an alternative construction of the factor group G/N. Proceed as follows: Consider the* set *$X = G/N$ of all left cosets of N in G and show that there is a group homomorphism $\varphi\colon G \longrightarrow S(X)$ such that* $\ker\varphi = N$.

3. Let X be a set, $Y \subset X$ a subset, G a group, and G^X the group of G-valued functions on X. Let $N := \{f \in G^X\,;\, f(y) = 1 \text{ for all } y \in Y\}$. Show that N is a normal subgroup of G^X satisfying $G^X/N \simeq G^Y$.

4. Let $\varphi\colon G \longrightarrow G'$ be a group homomorphism. Show:
   (i) If $H \subset G$ is a subgroup, then $\varphi(H)$ is a subgroup in G'. The corresponding assertion for normal subgroups is valid only if φ is surjective.
   (ii) If $H' \subset G'$ is a subgroup (resp. normal subgroup) in G', the same is true for $\varphi^{-1}(H') \subset G$.

5. Let G be a finite group, and let $H_1, H_2 \subset G$ be subgroups satisfying $H_1 \subset H_2$. Show that $(G : H_1) = (G : H_2) \cdot (H_2 : H_1)$.

6. Let G be a group and $N \subset G$ a normal subgroup satisfying the following maximality condition: If $H \subsetneq G$ is a proper subgroup containing N, then it coincides with N. Show for all subgroups $H_1, H_2 \subset G$ satisfying $H_1 \neq \{1\} \neq H_2$ and $H_1 \cap N = H_2 \cap N = \{1\}$ that H_1 is isomorphic to H_2.

1.3 Cyclic Groups

For a group G and a subset $X \subset G$, define H as the intersection of all subgroups of G containing X. Then H is a subgroup of G, in fact the (unique) smallest subgroup of G containing X. We say that H is *generated* by X or, if H coincides with G, that G is *generated* by X. The subgroup $H \subset G$ can be described in more explicit terms. It consists of all elements

$$x_1^{\varepsilon_1} \cdot \ldots \cdot x_n^{\varepsilon_n},$$

where $x_1, \ldots, x_n \in X$ and $\varepsilon_1, \ldots, \varepsilon_n \in \{1, -1\}$, with n varying over $\mathbb{N}$. Clearly, the elements of this type form the smallest subgroup of G containing X and thus by definition constitute the subgroup $H \subset G$.

For the moment we are interested only in the case in which X consists of a single element $x \in G$. The subgroup generated by x in G is denoted by $\langle x \rangle$, and its description simplifies to the following:

Remark 1. *Let x be an element of a group G. Then the subgroup $\langle x \rangle \subset G$ generated by x in G consists of all powers x^n, $n \in \mathbb{Z}$. In other words, $\langle x \rangle$ coincides with the image of the group homomorphism*

$$\mathbb{Z} \longrightarrow G, \qquad n \longmapsto x^n,$$

where $\mathbb{Z}$ means the additive group of all integers. In particular, $\langle x \rangle$ is commutative.

Definition 2. *A group G is called* cyclic *if it is generated by a single element. This is equivalent to the fact that there exists a surjective group homomorphism $\mathbb{Z} \longrightarrow G$.*

Observe that for a commutative group G with additively written law of composition, the map $\mathbb{Z} \longrightarrow G$ from Remark 1 is given by $n \longmapsto n \cdot x$, where $n \cdot x$ is to be interpreted as the n-fold sum of x for $n \geq 0$ and as the $(-n)$-fold sum of $-x$ for $n < 0$. In particular, the additive group $\mathbb{Z}$ is generated by the element $1 \in \mathbb{Z}$ and therefore is cyclic. It is called the *free cyclic group*; its order is infinite. On the other hand, given $m \in \mathbb{Z}$, the subgroup $m\mathbb{Z}$ of all integral multiples of m is cyclic, since it is generated by $m = m \cdot 1$. The factor group $\mathbb{Z}/m\mathbb{Z}$ is cyclic as well, generated by the residue class $1 + m\mathbb{Z}$. If $m \neq 0$, say $m > 0$, then $\mathbb{Z}/m\mathbb{Z}$ is called the *cyclic group of order m*. Indeed, $\mathbb{Z}/m\mathbb{Z}$, where $m > 0$, consists of precisely m elements, namely the residue classes $0 + m\mathbb{Z}, \ldots, (m-1) + m\mathbb{Z}$. In the following we want to show that $\mathbb{Z}$ and the groups of type $\mathbb{Z}/m\mathbb{Z}$ are the only cyclic groups, up to isomorphism. Due to the fundamental theorem on homomorphisms (in the version of 1.2/7) we see that a group G is cyclic if and only if there exists an isomorphism $\mathbb{Z}/H \xrightarrow{\sim} G$, for H a (normal) subgroup of $\mathbb{Z}$. Therefore, in order to determine all cyclic groups it is enough to determine all subgroups of $\mathbb{Z}$.

Proposition 3. *Let G be a cyclic group. Then*:

$$G \simeq \begin{cases} \mathbb{Z}, & \text{if } \operatorname{ord} G = \infty, \\ \mathbb{Z}/m\mathbb{Z}, & \text{if } \operatorname{ord} G = m < \infty. \end{cases}$$

In particular, the groups $\mathbb{Z}$ and $\mathbb{Z}/m\mathbb{Z}$ for integers $m > 0$ are the only cyclic groups, up to isomorphism.

As we have seen before, to prove the proposition it is enough to establish the following lemma:

Lemma 4. *Let $H \subset \mathbb{Z}$ be a subgroup. Then there exists an integer $m \in \mathbb{Z}$ such that $H = m\mathbb{Z}$. In particular, every subgroup of $\mathbb{Z}$ is cyclic.*

Proof. We may assume $H \neq 0$, i.e., that H is different from the zero subgroup of $\mathbb{Z}$ given by the zero element. Then H must contain positive integers; let m be the smallest among these. We claim that $H = m\mathbb{Z}$, where clearly, $m\mathbb{Z} \subset H$. To show the reverse inclusion, let $a \in H$. Using Euclidean division of a by m, there are integers $q, r \in \mathbb{Z}$, $0 \leq r < m$, such that $a = qm + r$. Then $r = a - qm$ belongs to H. However, since all positive integers in H are greater than or equal to m, we must have $r = 0$. Thus, $a = qm \in m\mathbb{Z}$ and therefore $H \subset m\mathbb{Z}$. All in all, we get $H = m\mathbb{Z}$. □

Proposition 5. (i) *Every subgroup H of a cyclic group G is itself cyclic.*

(ii) *If $\varphi\colon G \longrightarrow G'$ is a group homomorphism, where G is cyclic, then* $\ker\varphi$ *and* $\operatorname{im}\varphi$ *are cyclic.*

Proof. It follows immediately from the definition of cyclic groups that the image of a cyclic group under a group homomorphism $\varphi\colon G \longrightarrow G'$ is cyclic. Since $\ker\varphi$ is a subgroup of G, it remains to verify assertion (i). Therefore, let G be cyclic and let $H \subset G$ be a subgroup. Furthermore, let $\pi\colon \mathbb{Z} \longrightarrow G$ be an epimorphism. Then $\pi^{-1}(H)$ is a subgroup of $\mathbb{Z}$ and therefore cyclic by Lemma 4. But then H is cyclic, since it is the image of $\pi^{-1}(H)$ with respect to π, and assertion (i) follows. □

Let G be a group. The *order* $\operatorname{ord} a$ of an element $a \in G$ is defined as the order of the cyclic group generated by a in G. As we know already, $\varphi\colon \mathbb{Z} \longrightarrow G$, $n \longmapsto a^n$, yields an epimorphism from $\mathbb{Z}$ onto the cyclic subgroup $H \subset G$ that is generated by a. If $\ker\varphi = m\mathbb{Z}$ and G is finite, then necessarily $m \neq 0$, say $m > 0$, and H is isomorphic to $\mathbb{Z}/m\mathbb{Z}$. Thus, m is the smallest positive integer satisfying $a^m = 1$, and we see that H consists of the (distinct) elements $1 = a^0, a^1, \ldots, a^{m-1}$. In particular, $\operatorname{ord} a = m$.

Proposition 6 (Fermat's little theorem). *Let G be a finite group, $a \in G$. Then* $\operatorname{ord} a$ *divides* $\operatorname{ord} G$ *and we have* $a^{\operatorname{ord} G} = 1$.

Proof. Apply the theorem of Lagrange 1.2/3 to the cyclic subgroup of G that is generated by a. □

Corollary 7. *Let G be a finite group such that $p := \operatorname{ord} G$ is prime. Then G is cyclic, $G \simeq \mathbb{Z}/p\mathbb{Z}$, and every element $a \in G$, $a \neq 1$, is of order p. In particular, every such element a generates G.*

Proof. For each element $a \in G$, $a \neq 1$, consider the cyclic subgroup $H \subset G$ generated by a. Then $\operatorname{ord} a = \operatorname{ord} H$ is different from 1 and, according to Proposition 6, a divisor of $p = \operatorname{ord} G$. Since p is prime, we get $\operatorname{ord} a = \operatorname{ord} H = p$. Therefore, $H = G$, i.e., G is generated by a and hence is cyclic. Furthermore, G is isomorphic to $\mathbb{Z}/p\mathbb{Z}$, due to Proposition 3. □

Exercises

1. *For $m \in \mathbb{N} - \{0\}$ consider the set $G_m := \{0, 1, \ldots, m-1\}$ and define a law of composition on it via*

$$a \circ b := \text{ the remainder of } a + b \text{ with respect to division by } m.$$

   Give a direct argument showing that "$\circ$" constitutes a group law on G_m and that the resulting group is isomorphic to $\mathbb{Z}/m\mathbb{Z}$.

2. *Determine all subgroups of $\mathbb{Z}/m\mathbb{Z}$ for $m \in \mathbb{N} - \{0\}$.*

3. Consider $\mathbb{Z}$ as an additive subgroup of $\mathbb{Q}$ and show:
   (i) Every element in $\mathbb{Q}/\mathbb{Z}$ is of finite order.
   (ii) The factor group $\mathbb{Q}/\mathbb{Z}$ admits for each $n \in \mathbb{N} - \{0\}$ precisely one subgroup of order n, and it is cyclic.

4. Let $m, n \in \mathbb{N} - \{0\}$. Show that the groups $\mathbb{Z}/mn\mathbb{Z}$ and $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ are isomorphic if and only if m and n are relatively prime. In particular, a product of two finite cyclic groups whose orders are relatively prime is itself cyclic.

5. Let $\varphi\colon \mathbb{Z}^n \longrightarrow \mathbb{Z}^n$ be an endomorphism of the n-fold product of the additive group $\mathbb{Z}$, where $n \in \mathbb{N}$. Show that φ is injective if and only if $\mathbb{Z}^n / \operatorname{im} \varphi$ is finite. *Hint*: Consider the homomorphism of $\mathbb{Q}$-vector spaces $\varphi_{\mathbb{Q}}\colon \mathbb{Q}^n \longrightarrow \mathbb{Q}^n$ attached to φ.

2. Rings and Polynomials

Background and Overview

A *ring* is an abelian additive group R that is equipped with an additional multiplication, just like the ring $\mathbb{Z}$ of integers. More specifically, it is required that R be a monoid with respect to the multiplication and that the multiplication be distributive over the addition. We will always assume that the multiplication of a ring is *commutative*, except for a few occasions in Section 2.1. If the nonzero elements of a ring form an (abelian) group under the multiplication, the ring is actually a *field*. In principle, the definition of a ring goes back to R. Dedekind. For Dedekind, rings were motivated by questions in number theory involving intergral elements in algebraic number fields, or in other words, by the study of algebraic equations with *integer* coefficients. However, we will deal with rings of integral algebraic numbers only occasionally. More important for us are fields serving as coefficient domains for algebraic equations, as well as polynomial rings over fields. These are of fundamental importance in studying algebraic equations, and in particular algebraic field extensions. In the following, let us have a first look at polynomials.

If we want to solve an algebraic equation

$$(*) \qquad x^n + a_1 x^{n-1} + \ldots + a_n = 0,$$

say with coefficients $a_1, \ldots, a_n$ in a field K, we can try to view the symbol x as a quantity that is "variable." More precisely, we consider the expression $f(x) = x^n + a_1 x^{n-1} + \ldots + a_n$ as a function assigning to an element x the value given by $f(x)$. Then, of course, we have to determine the zeros of the function $f(x)$. On the other hand, to be strict, we must fix the domain where x may vary, for example K itself or, if $K = \mathbb{Q}$, also the real or complex numbers. We say that $f(x)$ is a *polynomial function* in x or, by abuse of language, a *polynomial* in x.

However, finding out about a suitable domain of definition that is big enough to contain "all" zeros of f is a basic problem. From a historical point of view, the fundamental theorem of algebra is a good device to settle this point. It implies for every subfield $K \subset \mathbb{C}$ that all solutions of $(*)$ that can appear in extension fields over K may be viewed as complex numbers. Therefore it is appropriate to interpret $f(x)$ in this case as a polynomial function on $\mathbb{C}$. Problems of a different kind arise when one is considering algebraic equations with coefficients

from a finite field $\mathbb{F}$; cf. 2.3/6 or Section 3.8 for the definition of such fields. For example, if $\mathbb{F}$ consists of the elements $x_1, \ldots, x_q$, then

$$g(x) = \prod_{j=1}^{q}(x - x_j) = x^q + \ldots + (-1)^q x_1 \ldots x_q$$

is a polynomial function that vanishes on all of $\mathbb{F}$, although not all of its "coefficients" are zero. Thereby we see, depending on the domain of definition, that it is not always possible to reconstruct the coefficients of the equation $(*)$ if the attached polynomial function $f(x)$ is known as a map only.

To avoid such difficulties one refrains from the idea that a polynomial might be a *function* on a certain domain of definition. Instead one tries to implement the following two aspects. The first is that there should be a one-to-one correspondence between polynomials and their "coefficient" sequences. On the other hand, one likes to retain the possibility of relating polynomials to functions, in such a way that polynomials can be evaluated at elements of certain fields (or rings) extending the given domain of coefficients. To achieve this we define a polynomial with coefficients $a_0, \ldots, a_n$ as a formal sum $f = \sum_{j=0}^{n} a_j X^j$, where, in down-to-earth terms, this just means that f is identified with the sequence of its coefficients $a_0, \ldots, a_n$. If the coefficient domain is a field (or a ring), we can add and multiply polynomials in the usual way by applying the conventional rules formally. In this way, all polynomials with coefficients in a field K form a ring $K[X]$. Also note that we can evaluate such polynomials $f \in K[X]$ at elements x belonging to arbitrary extension fields (or rings) $K' \supset K$; just substitute the variable X by x and consider the resulting expression $f(x)$ as an element of K'. In particular, we can talk about zeros of f in K'. We will study this formalism more closely for polynomials in one variable in 2.1 and for polynomials in several variables in 2.5.

Now the problem of solving algebraic equations with coefficients in a field K can be phrased in a slightly more precise way as follows: determine the zeros in suitable extension fields K' containing K for monic polynomials with coefficients in K, i.e., for polynomials of type $f = X^n + a_1 X^{n-1} + \ldots + a_n \in K[X]$. There is one reduction step that should be applied if possible. If the polynomial f can be written as the product of two polynomials $g, h \in K[X]$, i.e., $f = gh$, then to specify the zeros of f it is enough to specify the zeros of g and h separately. The reason is that we have $f(x) = (gh)(x) = g(x)h(x)$ for $x \in K'$, as is verified without difficulty. Since the latter equation has to be read in a field, we see that f vanishes at x precisely when g or h vanishes at this point. Therefore, to simplify the problem, one should try to reduce the algebraic equation $f(x) = 0$ to equations of lower degree, by factoring f in $K[X]$ into a product of monic polynomials of lower degree. If that is impossible, then f as well as the algebraic equation $f(x) = 0$ are said to be *irreducible*.

In particular, the preceding considerations show that factorizations of polynomials should be studied. We will do this in 2.4. Starting out from the fact that there is a so-called Euclidean division in polynomial rings over fields, i.e., a division process with remainder, we will show that the theorem of unique prime

factorization is valid in $K[X]$, just as it is in the ring $\mathbb{Z}$ of integers. Hence, we conclude that every monic polynomial admits a unique factorization into irreducible monic polynomials. Further considerations in 2.7 and 2.8 will deal with criteria for irreducibility and thereby with the question of how to decide whether a given polynomial $f \in K[X]$ is irreducible.

There is another reason why prime factorizations in polynomial rings $K[X]$ are of special interest. To explain this in more detail, let us briefly touch upon the notion of *ideals*, a concept that belongs to the basics of ring theory; it will be dealt with in 2.2. An ideal $\mathfrak{a}$ of a ring R is an additive subgroup of R such that $ra \in \mathfrak{a}$ for all $r \in R$ and all $a \in \mathfrak{a}$. In many respects ideals behave like normal subgroups of groups. For example, we can construct the residue class ring $R/\mathfrak{a}$ of a ring R by an ideal $\mathfrak{a} \subset R$, prove the fundamental theorem on homomorphisms, and so on; cf. 2.3. Ideals appeared in mathematics at the end of the nineteenth century, alongside attempts to extend the theorem on unique prime factorization from the ring of integers $\mathbb{Z}$ to more general rings of algebraic integers. When it was realized that this was impossible in the general case, one started looking at factorizations into so-called *ideal numbers*. However, finally it was Dedekind who observed that instead of factorizations of single elements, one should rather concentrate on factorizations of certain subsets, which he called ideals, of the given ring. In this way, Dedekind proved in 1894 the theorem on the unique prime factorization for ideals in rings of algebraic integers. Today an integral domain, i.e., a nonzero ring without nontrivial zero divisors, is called a *Dedekind domain* if Dedekind's theorem holds for it.

For us it is important to know that the polynomial ring $K[X]$ over a field K is a *principal ideal domain*; this means that it is an integral domain and that every ideal $\mathfrak{a} \subset K[X]$ is of type (f), i.e., generated by a single element $f \in K[X]$. This result will be established in 2.4/3. Furthermore, we show that the theorem on unique prime factorization is valid in principal ideal domains. Investigations of this kind lead directly to the so-called *construction of Kronecker*, which will be discussed at length in 3.4/1. Given an irreducible algebraic equation $f(x) = 0$ with coefficients in a field K, the construction allows one to specify in a simple way an extension field K' that contains a solution. Indeed, set $K' = K[X]/(f)$, check that it is a field naturally extending K, and observe that the residue class $\overline{X}$ of $X \in K[X]$ solves the equation. Even if the construction does not provide any closer details on the structure of the field K', for example on the solvability by radicals, it nevertheless gives a valuable contribution to the question of the existence of solutions.

To illustrate the potential of principal ideal domains, we present at the end of the chapter in 2.9 the theory of elementary divisors, a topic that actually belongs to the domain of linear algebra. As a generalization of vector spaces over fields, we study "vector spaces" or, as one prefers to say, *modules* over rings and, in particular, over principal ideal domains.

2.1 Polynomial Rings in One Variable

Definition 1. *A* ring (admitting a unit element) *is a set R together with two (inner) laws of composition written as addition "+" and multiplication "·" such that the following conditions are satisfied*:

(i) *R is a commutative group with respect to addition.*

(ii) *R is a monoid with respect to multiplication, i.e., the multiplication is associative, and there exists a unit element in R.*

(iii) *The distributive laws hold, i.e.,*

$$(a+b)\cdot c = a\cdot c + b\cdot c, \qquad c\cdot(a+b) = c\cdot a + c\cdot b, \qquad \text{for } a,b,c \in R.$$

R is called commutative *if the multiplication is commutative.*[1]

On the right-hand sides of the distributive laws in (iii) we have refrained from introducing a special bracketing. Just as for computations with ordinary numbers, it is common that multiplication is granted a higher precedence than addition. For every ring, the zero element of the addition will be denoted by 0, the unit element of the multiplication by 1. Note that the case $1 = 0$ is not excluded; it characterizes the so-called *zero ring*, which consists of a single element 0. If no confusion is possible, the zero ring is denoted by 0 as well. For calculations in rings, one may use essentially the same rules as for calculations in terms of ordinary numbers, e.g.,

$$0\cdot a = 0 = a\cdot 0, \qquad (-a)\cdot b = -(ab) = a\cdot(-b), \qquad \text{for } a,b \in R.$$

However, note that from $ab = ac$, resp. $a\cdot(b-c) = 0$ (even for $a \neq 0$), we cannot necessarily conclude $b = c$. The latter equality can be obtained only when one is dealing with so-called integral domains (see below) or in the case that a admits an inverse element with respect to multiplication. Thus, caution is required when applying cancellation rules.

Let R be a ring. A subset $S \subset R$ is called a *subring* of R if S is a subgroup with respect to the addition on R and a submonoid with respect to the multiplication on R. In particular, using the laws of composition inherited from R, it is clear that S is a ring again. The pair $S \subset R$ is called a *ring extension*.

Given a ring R, we write

$$R^* = \{a \in R\,;\ \text{there exists } b \in R \text{ such that } ab = ba = 1\}$$

for the set of multiplicatively invertible elements; these are referred to as the *units* of R. It is easily checked that R^* is a group with respect to multiplication. R is called a *division ring* or a *skew field* if $R \neq 0$ and $R^* = R - \{0\}$, i.e., if $1 \neq 0$ and each nonzero element of R is a unit. In addition, if the multiplication

[1] Although we will refer to a few notions and examples of noncommutative rings in this section, we will generally assume that all rings are *commutative*, unless stated otherwise.

of R is commutative, R is called a *field*. An element a of a ring R is called a *zero divisor* if there exists an element $b \in R - \{0\}$ such that $ab = 0$ or $ba = 0$. Fields and skew fields do not admit any zero divisors, except for 0, the trivial zero divisor. Finally, a commutative ring R is called an *integral domain* if it is nonzero and does not admit nontrivial zero divisors. We give some examples of rings.

(1) $\mathbb{Z}$ is an integral domain whose group of units consists of the elements 1 and -1.

(2) $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ form fields, the Hamiltonian quaternions $\mathbb{H}$ a skew field. For completeness, let us recall the construction of $\mathbb{H}$. Start with a 4-dimensional $\mathbb{R}$-vector space V, say with a basis e, i, j, k. Set

$$\begin{gathered} e^2 = e, \quad ei = ie = i, \quad ej = je = j, \quad ek = ke = k, \\ i^2 = j^2 = k^2 = -e, \\ ij = -ji = k, \quad jk = -kj = i, \quad ki = -ik = j, \end{gathered}$$

and define the product of arbitrary elements in V by $\mathbb{R}$-linear extension. The resulting multiplication, together with the vector space addition, makes V a (noncommutative) ring $\mathbb{H}$, even a skew field with e as unit element. Identifying the field $\mathbb{R}$ of real numbers with $\mathbb{R}e$, we can view $\mathbb{R}$ as a subfield of $\mathbb{H}$, i.e., as a subring that is a field. In a similar way, we can interpret $\mathbb{C}$ as a subfield of $\mathbb{H}$.

(3) Let K be a field. Then $R = K^{n\times n}$, the set of all $(n \times n)$ matrices with coefficients in K defines a ring together with the ordinary addition and multiplication of matrices; its group of units is

$$R^* = \{A \in K^{n\times n}\,;\ \det A \neq 0\}.$$

Note that R is noncommutative for $n \geq 2$ and that in this case, R admits nontrivial zero divisors. More generally, we can state that the set of endomorphisms of a vector space V (or, alternatively, of an abelian group G) is a ring. Here the addition of endomorphisms is defined via the inherent addition on V resp. G, and the multiplication as composition of endomorphisms.

(4) Let X be a set and R a ring. Then R^X, the set of R-valued functions on X, becomes a ring if we set for $f, g \in R^X$,

$$\begin{aligned} f + g\colon X &\longrightarrow R, \qquad x \longmapsto f(x) + g(x), \\ f \cdot g\colon X &\longrightarrow R, \qquad x \longmapsto f(x) \cdot g(x). \end{aligned}$$

In particular, if $X = \{1, \ldots, n\} \subset \mathbb{N}$, we can view R^X as the n-fold Cartesian product $R^n = R \times \ldots \times R$, the ring structure on R^n being defined by

$$(*) \qquad \begin{aligned} (x_1, \ldots, x_n) + (y_1, \ldots, y_n) &= (x_1 + y_1, \ldots, x_n + y_n), \\ (x_1, \ldots, x_n) \cdot (y_1, \ldots, y_n) &= (x_1 \cdot y_1, \ldots, x_n \cdot y_n). \end{aligned}$$

The zero and the unit elements are given by $0 = (0, \ldots, 0)$ and $1 = (1, \ldots, 1)$. Furthermore, the equation $(1, 0, \ldots, 0) \cdot (0, 1, \ldots, 1) = 0$ shows for $n \geq 2$ that

R^n will generally admit nontrivial zero divisors, even if R itself is an integral domain. We call R^n the n-fold *ring-theoretic product* of R with itself. More generally, we can consider the ring-theoretic product

$$P = \prod_{x \in X} R_x$$

of a family of rings $(R_x)_{x \in X}$. Addition and multiplication on P are defined componentwise, just as in the formulas $(*)$. If the rings R_x are copies of one and the same ring R, then $\prod_{x \in X} R_x$ and R^X coincide naturally.

From now on we will restrict ourselves to commutative rings. Thus, unless stated otherwise, the term *ring* will always be used in the sense of a *commutative ring*. Starting out from such a ring R, we want to construct a ring extension $R[X]$, the so-called *polynomial ring* in a variable X and with coefficients in R. In terms of sets, we let $R[X] := R^{(\mathbb{N})}$, where as usual, $R^{(\mathbb{N})}$ stands for the set of all maps $f : \mathbb{N} \longrightarrow R$ satisfying $f(i) = 0$ for almost all $i \in \mathbb{N}$. Identifying a map $f : \mathbb{N} \longrightarrow R$ with its corresponding sequence $(f(i))_{i \in \mathbb{N}}$ of images in R, we can write

$$R^{(\mathbb{N})} = \big\{(a_i)_{i \in \mathbb{N}}\,;\; a_i \in R,\ a_i = 0 \text{ for almost all } i \in \mathbb{N}\big\}.$$

Now, in order to introduce a ring structure on $R^{(\mathbb{N})}$, define the addition componentwise, i.e., by

$$(a_i) + (b_i) := (a_i + b_i).$$

Note that in terms of maps $\mathbb{N} \longrightarrow R$, this corresponds to the usual addition as considered in example (4) above. Concerning the multiplication, we proceed differently and use a construct that is inspired by the multiplication of polynomial functions. In fact, we set

$$(a_i) \cdot (b_i) := (c_i),$$

where

$$c_i := \sum_{\mu + \nu = i} a_\mu b_\nu.$$

It can readily be checked that $R^{(\mathbb{N})}$ becomes a ring; the zero element is given by the sequence $(0, 0, 0, \ldots)$ and the unit element by the sequence $(1, 0, 0, \ldots)$. One writes $R[X]$ for the ring thus constructed, calling it the *ring of polynomials in one variable X over R*. The role of the "variable" X becomes more plausible if we write elements $(a_i) \in R[X]$ in the familiar polynomial form

$$\sum_{i \in \mathbb{N}} a_i X^i \qquad \text{or} \qquad \sum_{i=0}^{n} a_i X^i,$$

where n is large enough that $a_i = 0$ for $i > n$ and where X is given by the sequence $(0, 1, 0, 0, \ldots)$. In terms of polynomial sums, addition and multiplication in $R[X]$ are described by the familiar formulas

$$\sum_i a_i X^i + \sum_i b_i X^i = \sum_i (a_i + b_i) X^i,$$
$$\sum_i a_i X^i \cdot \sum_i b_i X^i = \sum_i \Big(\sum_{\mu+\nu=i} a_\mu \cdot b_\nu \Big) X^i.$$

Finally, to interpret R as a subring of $R[X]$, we view the elements of R as constant polynomials in $R[X]$, i.e., we use the map $R \hookrightarrow R[X]$, $a \longmapsto aX^0$, as an identification. This is permitted, since the injection respects ring structures on R and $R[X]$ and thus is a homomorphism of rings, as we will say.

To explain more closely the significance of the "variable" X, consider a ring extension $R \subset R'$ and a polynomial $f = \sum a_i X^i$ in $R[X]$. Then we can substitute the variable X by any element $x \in R'$ and thereby compute the value $f(x) = \sum a_i x^i$ of f at x. In particular, f gives rise to a well-defined map $R' \longrightarrow R'$, $x \longmapsto f(x)$, where

$$(f+g)(x) = f(x) + g(x), \qquad (f \cdot g)(x) = f(x) \cdot g(x)$$

for $f, g \in R[X]$. Notice that in order to establish the multiplicativity of the right-hand equation, we need the commutativity of the multiplication on R' or, what is enough, the permutability relation $ax = xa$ for $a \in R$, $x \in R'$. In particular, in doing computations in the polynomial ring $R[X]$, the "variable" X behaves like a universally variable quantity with the special property that equations in $R[X]$ are preserved when X is substituted by elements in R'.

Given a polynomial $f = \sum a_i X^i \in R[X]$, its ith coefficient a_i is called its *coefficient of degree* i. Moreover, the *degree* of f is defined by

$$\deg f := \max\{i \,;\, a_i \neq 0\},$$

where we assign the degree $-\infty$ to the zero polynomial 0. If $\deg f = n \geq 0$, then a_n is referred to as the highest or the leading coefficient of f. If it is 1, we say that f is *monic*. Each polynomial $f \in R[X] - \{0\}$ whose leading coefficient a_n is a unit can be transformed into a monic one by multiplication by the inverse a_n^{-1} of a_n.

Remark 2. *Consider the polynomial ring $R[X]$ in a variable X over a ring R and let $f, g \in R[X]$. Then*

$$\deg(f+g) \leq \max(\deg f, \deg g),$$
$$\deg(f \cdot g) \leq \deg f + \deg g,$$

and even $\deg(f \cdot g) = \deg f + \deg g$ *if R is an integral domain.*

Proof. The assertions are clear if f or g is zero. Therefore, assume $m = \deg f \geq 0$, as well as $n = \deg g \geq 0$, say $f = \sum a_i X^i$, $g = \sum b_i X^i$. Then we conclude that $a_i + b_i = 0$ for $i > \max(m, n)$ and hence $\deg(f+g) \leq \max(m, n)$. Similarly, we see that $\sum_{\mu+\nu=i} a_\mu b_\nu = 0$ for $i > m + n$ and therefore get $\deg(f \cdot g) \leq m + n$. Finally, if R is an integral domain, then $\deg f = m$ and $\deg g = n$ imply

that the coefficients a_m, b_n are nonzero and hence that $\sum_{\mu+\nu=m+n} a_\mu b_\nu = a_m b_n$, which is the coefficient of degree $m+n$ in $f \cdot g$, is nonzero. This shows that $\deg(f \cdot g) = m+n$. □

There are several properties of rings that a polynomial ring $R[X]$ inherits from its ring of coefficients R. As a simple example, we look at integral domains.

Remark 3. *Let R be an integral domain. Then the polynomial ring $R[X]$ is an integral domain as well. Furthermore, $(R[X])^* = R^*$.*

Proof. Use the relation $\deg(f \cdot g) = \deg f + \deg g$ from Remark 2. □

Finally, we want to establish *Euclidean division* for polynomial rings, a process that is particularly known from the ring of integers $\mathbb{Z}$. Euclidean division is used in 2.4 in order to show that polynomial rings over fields are unique prime factorization domains.

Proposition 4. *Let R be a ring and $g = \sum_{i=0}^{d} a_i X^i \in R[X]$ a polynomial whose leading coefficient a_d is a unit in R. Then, for each $f \in R[X]$, there exist unique polynomials $q, r \in R[X]$ such that*

$$f = qg + r, \qquad \deg r < d.$$

Proof. First observe that we have $\deg(qg) = \deg q + \deg g$ for arbitrary polynomials $q \in R[X]$, even if R is not an integral domain. Indeed, the leading coefficient a_d of g is a unit. So if q is of some degree $n \geq 0$ with leading coefficient c_n, then $c_n a_d \neq 0$. However, this is the leading coefficient of qg, and we conclude that $\deg(qg) = n + d$.

Next, to justify the uniqueness assertion, consider a polynomial $f \in R[X]$ admitting two decompositions of the desired type, say $f = qg + r = q'g + r'$. Then we get $0 = (q - q')g + (r - r')$ and, by the above argument,

$$\deg(q - q') + \deg g = \deg(r - r').$$

Since r and r' are of degree $< d$, the same is true for $r - r'$, and we see that $\deg(q-q')+\deg g < d$. However, this can be true only for $q = q'$, since $\deg g = d$. But then we must have $r = r'$ as well, and the uniqueness assertion is clear.

To derive the existence part of Euclidean division we proceed by induction on $n = \deg f$. If $\deg f < d$, set $q = 0$ and $r = f$. On the other hand, if we have $f = \sum_{i=0}^{n} c_i X^i$ as well as $c_n \neq 0$ and $n \geq d$, then

$$f_1 = f - c_n a_d^{-1} X^{n-d} g$$

is a polynomial of degree $< n$. By the induction hypothesis, this admits a decomposition $f_1 = q_1 g + r_1$ with polynomials $q_1, r_1 \in R[X]$, where $\deg r_1 < d$. Hence,

$$f = (q_1 + c_n a_d^{-1} X^{n-d})g + r_1$$

is a decomposition of f, as desired. □

The argument presented in the above proof can be used to explicitly carry out Euclidean division in polynomial rings $R[X]$, similarly to how this is done for the ring of integers $\mathbb{Z}$. To give an example, consider the polynomials

$$f = X^5 + 3X^4 + X^3 - 6X^2 - X + 1, \qquad g = X^3 + 2X^2 + X - 1$$

from $\mathbb{Z}[X]$:

$$\begin{array}{rrrrrl}
(X^5 & +3X^4 & +X^3 & -6X^2 & -X & +1) : (X^3 + 2X^2 + X - 1) = X^2 + X - 2 \\
X^5 & +2X^4 & +X^3 & -X^2 & & \\
\hline
 & X^4 & & -5X^2 & -X & \\
 & X^4 & +2X^3 & +X^2 & -X & \\
\hline
 & & -2X^3 & -6X^2 & & +1 \\
 & & -2X^3 & -4X^2 & -2X & +2 \\
\hline
 & & & -2X^2 & +2X & -1
\end{array}$$

In a first step we subtract $X^2 g$ from f, then in a second Xg from $f - X^2 g$, and in a third $-2g$ from $f - X^2 g - Xg$. We obtain $-2X^2 + 2X - 1$ as remainder, thus leading to the decomposition

$$f = (X^2 + X - 2)g + (-2X^2 + 2X - 1).$$

Finally, let us mention that the construction of the polynomial ring $R[X]$ can be generalized. For example, in 2.5 we will introduce polynomial rings in several variables. On the other hand, we can replace the set $R^{(\mathbb{N})}$ by $R^{\mathbb{N}}$, the set of *all* maps from $\mathbb{N}$ to R. Then, proceeding in the same way as in the case of polynomial rings, the ring $R[\![X]\!]$ of *formal power series* in one variable X over R is obtained. Its elements can be written as *infinite* series of type $\sum_{i=0}^{\infty} a_i X^i$.

Exercises

1. *Verify the relations* $0 \cdot a = 0$ *and* $(-a) \cdot b = -(a \cdot b)$ *for elements* a, b *of a ring* R.
2. *The polynomial ring* $R[X]$ *has been defined over commutative rings* R*. To what extent is it possible and makes sense to consider polynomial rings within the context of not necessarily commutative rings*?
3. Explicitly work out Euclidean division in $\mathbb{Z}[X]$, as specified in Proposition 4, for the following polynomials:
   (i) $f = 3X^5 + 2X^4 - X^3 + 3X^2 - 4X + 7, \quad g = X^2 - 2X + 1.$
   (ii) $f = X^5 + X^4 - 5X^3 + 2X^2 + 2X - 1, \quad g = X^2 - 1.$

4. Let K be a field and $g \in K[X]$ a polynomial in one variable of degree $d > 0$. Prove the existence of the so-called *g-adic expansion*: Given $f \in K[X]$, there are unique polynomials $a_0, a_1, \ldots \in K[X]$ of degree $< d$, where $a_i = 0$ for almost all i, such that $f = \sum_i a_i g^i$.
5. Let R be a ring containing a *nilpotent* element $a \neq 0$; nilpotent means that there is some $n \in \mathbb{N}$ such that $a^n = 0$. Show that the group of units R^* is a proper subgroup of the group of units $(R[X])^*$.
6. Determine the smallest subring of $\mathbb{R}$ containing $\mathbb{Q}$ and $\sqrt{2}$, and show that it is in fact a field.
7. Let R be a ring. Show that a formal power series $\sum a_i X^i \in R[\![X]\!]$ is a unit if and only if a_0 is a unit in R.
8. Prove that the Hamiltonian quaternions $\mathbb{H}$ from Example (2) form a skew field.

2.2 Ideals

Ideals of rings are basic in ring theory, just as normal subgroups are in group theory. However, an ideal is not necessarily a subring of its ambient ring, due to the fact that in nontrivial cases, it does not contain the unit element of multiplication.

Definition 1. *Let R be a ring. A subset $\mathfrak{a} \subset R$ is called an* ideal *in R if*:
(i) $\mathfrak{a}$ *is an additive subgroup of R.*
(ii) $r \in R$, $a \in \mathfrak{a} \Longrightarrow ra \in \mathfrak{a}$.

Every ring R contains the so-called *trivial* ideals, namely the *zero ideal* $\{0\}$, denoted by 0 as well, and the *unit ideal* R. These are the only ideals if R is a field. For given ideals $\mathfrak{a}, \mathfrak{b}$ of a ring R, the following ideals can be constructed:

$$\begin{aligned} \mathfrak{a} + \mathfrak{b} &:= \{a + b \,;\, a \in \mathfrak{a},\, b \in \mathfrak{b}\}, \\ \mathfrak{a} \cdot \mathfrak{b} &:= \Big\{ \sum_{i=1}^{<\infty} a_i b_i \,;\, a_i \in \mathfrak{a},\, b_i \in \mathfrak{b} \Big\}, \\ \mathfrak{a} \cap \mathfrak{b} &:= \{x \,;\, x \in \mathfrak{a} \text{ and } x \in \mathfrak{b}\}. \end{aligned}$$

Note that always $\mathfrak{a} \cdot \mathfrak{b} \subset \mathfrak{a} \cap \mathfrak{b}$. Furthermore, one can define the product of finitely many ideals similarly as before, as well as the sum and the intersection of arbitrarily many ideals. The sum $\sum \mathfrak{a}_i$ of a family of ideals $(\mathfrak{a}_i)_{i \in I}$ consists of all sums $\sum a_i$, where $a_i \in \mathfrak{a}_i$ and $a_i = 0$ for almost all $i \in I$. For $a \in R$, we call $Ra := \{ra \,;\, r \in R\}$ the *principal ideal generated by a*. More generally, for elements $a_1, \ldots, a_n \in R$, one can consider the ideal that is *generated by* $a_1, \ldots, a_n$, namely

$$(a_1, \ldots, a_n) := Ra_1 + \ldots + Ra_n = \{r_1 a_1 + \ldots + r_n a_n \,;\, r_1, \ldots, r_n \in R\}.$$

It is the smallest ideal in R containing $a_1, \ldots, a_n$, i.e., every ideal of R that contains the elements $a_1, \ldots, a_n$ will also contain the ideal $(a_1, \ldots, a_n)$. Analogously, we can consider the ideal generated in R by a family of elements $(a_i)_{i \in I}$ of R, namely the ideal $\sum_{i \in I} Ra_i$.

Definition 2. *Let $\mathfrak{a}$ be an ideal of a ring R. A family $(a_i)_{i \in I}$ of elements in $\mathfrak{a}$ is called a* system of generators *of $\mathfrak{a}$ if $\mathfrak{a} = \sum_{i \in I} Ra_i$, i.e., if $\mathfrak{a}$ coincides with the ideal generated by the family $(a_i)_{i \in I}$. We say that $\mathfrak{a}$ is* finitely generated *if $\mathfrak{a}$ admits a finite system of generators. Furthermore, $\mathfrak{a}$ is called a* principal ideal *if $\mathfrak{a}$ is generated by a single element, i.e., if there exists $a \in \mathfrak{a}$ such that $\mathfrak{a} = (a)$. If R is an integral domain and every ideal of R is principal, then R is called a* principal ideal domain.

The trivial ideals of a ring are always principal, since they are generated by the zero element 0, resp. the unit element 1. Moreover, the subgroups of type $m\mathbb{Z} \subset \mathbb{Z}$ constitute principal ideals in the ring of integers $\mathbb{Z}$. These are the only subgroups of $\mathbb{Z}$ by 1.3/4. Hence, there cannot exist any further ideals in $\mathbb{Z}$. Since $\mathbb{Z}$ is an integral domain, we have the following:

Proposition 3. *$\mathbb{Z}$ is a principal ideal domain.*

Generators of principal ideals are not unique, since they can always be altered by multiplication by a unit. In integral domains this is the only way to change generators of principal ideals:

Remark 4. *Let R be an integral domain. Then two principal ideals $\mathfrak{a} = (a)$ and $\mathfrak{b} = (b)$ of R coincide if and only if there is a unit $c \in R^*$ such that $b = ca$.*

Proof. If $\mathfrak{a} = \mathfrak{b}$, there is no loss of generality in assuming $\mathfrak{a} = \mathfrak{b} \neq 0$. Then $b \in \mathfrak{a}$ and there exists an element $c \in R$ such that $b = ca$. Likewise, we have $a \in \mathfrak{b}$ and therefore an element $c' \in R$ such that $a = c'b$. Then $b = ca = cc'b$, and thus

$$(1 - cc')b = 0.$$

Since R is an integral domain and b is nonzero due to $\mathfrak{b} \neq 0$, we get $cc' = 1$, and c is a unit. The reverse implication is trivial. □

Two elements a, b of a ring R are said to be *associated* to each other if there is a unit $c \in R^*$ such that $b = ca$. Therefore we can state that two elements of an integral domain generate the same principal ideal if and only if they are associated. Note that this assertion does not extend to more general rings; cf. Exercise 7 in Section 2.3.

Finally, let us consider the polynomial ring $\mathbb{Z}[X]$ as an example. The principal ideal generated by X is given by

$$(X) = \Big\{ \sum a_i X^i \in \mathbb{Z}[X] \,;\, a_0 = 0 \Big\},$$

and the one generated by 2 is

$$(2) = \Big\{ \sum a_i X^i \in \mathbb{Z}[X] \,;\, a_i \text{ is even for all } i \Big\}.$$

Since there does not exist any nonunit in $\mathbb{Z}[X]$ that admits 2 as well as X as multiples, we immediately see that

$$(2, X) = \Big\{ \sum a_i X^i \in \mathbb{Z}[X] \,;\, a_0 \text{ is even} \Big\}$$

is an ideal in $\mathbb{Z}[X]$ that cannot be principal. In particular, $\mathbb{Z}[X]$ fails to be a principal ideal domain, although it is an integral domain by 2.1/3.

Exercises

1. *Let $\mathfrak{a} = (a_1, \ldots, a_m)$ and $\mathfrak{b} = (b_1, \ldots, b_n)$ be ideals in a ring R. Specify systems of generators for the ideals $\mathfrak{a} + \mathfrak{b}$ and $\mathfrak{a} \cdot \mathfrak{b}$. Furthermore, discuss the case of the ideal $\mathfrak{a} \cap \mathfrak{b}$.*

2. *Examine in which cases the union of two ideals or, more generally, the union of a family of ideals of a ring R is itself an ideal.*

3. *Let K be a field. Consider $K^2 = K \times K$ as the ring-theoretic product of K with itself, but also as a K-vector space. Compare the notions of subrings, ideals, and subvector spaces in this example.*

4. Specify single generators of the following ideals in $\mathbb{Z}$:

$$(2) + (3), \qquad (4) + (6), \qquad (2) \cap (3), \qquad (4) \cap (6).$$

5. Let R be a ring, X a set, and $Y \subset X$ a subset. Examine which of the following subsets of R^X, the ring of all maps $X \longrightarrow R$, gives rise to a subring or an ideal:

$$\begin{aligned} M_1 &= \big\{ f \in R^X \,;\, f \text{ is constant on } Y \big\}, \\ M_2 &= \big\{ f \in R^X \,;\, f(Y) = 0 \big\}, \\ M_3 &= \big\{ f \in R^X \,;\, f(y) \neq 0 \text{ for all } y \in Y \big\}, \\ M_4 &= \big\{ f \in R^X \,;\, f(y) = 0 \text{ for almost all } y \in Y \big\}. \end{aligned}$$

   Which conditions on Y ensure that we get principal ideals in the ideal cases?

6. Let R be a ring. Show that

$$\big\{ a \in R \,;\, \text{there is an integer } n \in \mathbb{N} \text{ such that } a^n = 0 \big\}$$

   gives rise to an ideal in R (the so-called *radical* or *nilradical* of R).

7. Let K be a field. Determine all ideals of the ring of formal power series $K[\![X]\!]$. (Use Exercise 7 from Section 2.1.)

2.3 Ring Homomorphisms, Factor Rings

The notion of a group homomorphism extends naturally to the setting of rings.

Definition 1. *Let R and R' be rings. A map $\varphi\colon R \longrightarrow R'$ is called a* ring homomorphism *if*:

(i) $\varphi(a+b) = \varphi(a)+\varphi(b)$ *for all $a, b \in R$, i.e., φ is a group homomorphism with respect to addition.*

(ii) $\varphi(a \cdot b) = \varphi(a) \cdot \varphi(b)$ *for all $a, b \in R$ and $\varphi(1) = 1$, i.e., φ is a monoid homomorphism with respect to multiplication.*

It is verified without difficulty that the composition of two ring homomorphisms yields a ring homomorphism. A ring homomorphism $\varphi\colon R \longrightarrow R'$ is called an *isomorphism* if φ admits an inverse, i.e., if there is a ring homomorphism $\psi\colon R' \longrightarrow R$ satisfying $\psi \circ \varphi = \mathrm{id}_R$ and $\varphi \circ \psi = \mathrm{id}_{R'}$. This is equivalent to the fact that φ is bijective. Injective (resp. surjective) ring homomorphisms $R \longrightarrow R'$ are also referred to as *monomorphisms* (resp. *epimorphisms*). An *endomorphism* of R is a homomorphism $R \longrightarrow R$, and an *automorphism* of R is an isomorphism $R \longrightarrow R$.

Remark 2. *Let $\varphi\colon R \longrightarrow R'$ be a ring homomorphism. Then*:

(i) $\ker\varphi = \{a \in R\,;\, \varphi(a) = 0\}$ *is an ideal in R.*

(ii) $\operatorname{im}\varphi = \varphi(R)$ *is a subring of R'.*

(iii) *φ induces a group homomorphism $R^* \longrightarrow R'^*$ between the multiplicative groups of units in R and R'.*

The assertions can be verified straightaway. Note that the image of a ring homomorphism $\varphi\colon R \longrightarrow R'$ will generally fail to be an ideal in R'. If R and R' are fields, ring homomorphisms $R \longrightarrow R'$ will also be referred to as *field homomorphisms*.

Remark 3. *Let K be a field and R a ring, $R \neq 0$. Then every homomorphism $\varphi\colon K \longrightarrow R$ is injective. In particular, every homomorphism of fields is injective.*

Proof. We know that $\ker\varphi$ is an ideal in K. It is even a proper ideal, since $\varphi(1) = 1 \neq 0$. But then $\ker\varphi = 0$, since a field does not contain any proper ideals, except for the zero ideal. □

For an arbitrary ring R, there is precisely one ring homomorphism $\mathbb{Z} \longrightarrow R$, namely the map given by $n \longmapsto n \cdot 1$. Here $n \cdot 1$ for $n \geq 0$ is to be interpreted as the n-fold sum of the unit element $1 \in R$, and likewise for $n < 0$ as the $(-n)$-fold sum of -1. The inclusion map $R \hookrightarrow R'$ attached to a ring extension $R \subset R'$ is a (trivial) example of a ring homomorphism. Furthermore, given such an extension, every element $x \in R'$ gives rise to a so-called *substitution homomorphism*

$$R[X] \longrightarrow R', \qquad f = \sum a_i X^i \longmapsto f(x) = \sum a_i x^i,$$

which is a ring homomorphism. Indeed, the process of evaluating polynomials $f, g \in R[X]$ at elements $x \in R'$ was already dealt with in 2.1, including the compatibilities $(f+g)(x) = f(x) + g(x)$ and $(f \cdot g)(x) = f(x) \cdot g(x)$ that are required for a ring homomorphism.

In the following, let R be a ring and $\mathfrak{a}$ an ideal in R. We want to adapt the construction of the factor group G/N of a group G by a normal subgroup N to the setting of rings and construct a so-called *factor ring* or *residue class ring* $R/\mathfrak{a}$, together with a surjective ring homomorphism $\pi\colon R \longrightarrow R/\mathfrak{a}$ satisfying $\ker \pi = \mathfrak{a}$. To begin with we define $R/\mathfrak{a}$ as an abelian group, just by viewing $\mathfrak{a}$ as a (normal) subgroup of the additive group of R. Then $R/\mathfrak{a}$ consists of all residue classes of type $x + \mathfrak{a}$ for $x \in R$, where addition in $R/\mathfrak{a}$ is given by the formula

$$(x + \mathfrak{a}) + (y + \mathfrak{a}) = (x + y) + \mathfrak{a}.$$

That this law of composition is well defined and makes $R/\mathfrak{a}$ an abelian group was shown in 1.2. Proceeding similarly with the multiplication, we define for residue classes $x + \mathfrak{a}$, $y + \mathfrak{a}$ in $R/\mathfrak{a}$ their product by

$$(x + \mathfrak{a}) \cdot (y + \mathfrak{a}) := (x \cdot y) + \mathfrak{a}.$$

Of course, we have to check that the product is well defined or, more precisely, that the residue class $(x \cdot y) + \mathfrak{a}$ is independent of the choice of the representatives x, y of the residue classes $x + \mathfrak{a}$ and $y + \mathfrak{a}$. To do this, assume $x' + \mathfrak{a} = x + \mathfrak{a}$ and thus $x' = x + a$ for some $a \in \mathfrak{a}$, as well as $y' + \mathfrak{a} = y + \mathfrak{a}$ and $y' = y + b$ for some $b \in \mathfrak{a}$. Then $x'y' = xy + ay' + xb \in (xy) + \mathfrak{a}$, and we see that

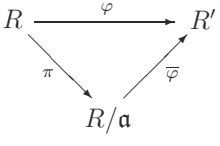
$$(xy) + \mathfrak{a} = (x'y') + \mathfrak{a}.$$

Therefore, the multiplication is well defined on $R/\mathfrak{a}$, and it is immediately clear that the ring properties of R carry over to $R/\mathfrak{a}$. Moreover, the canonical projection

$$\pi\colon R \longrightarrow R/\mathfrak{a}, \qquad x \longmapsto x + \mathfrak{a},$$

is a ring homomorphism satisfying $\ker \pi = \mathfrak{a}$, which, similarly as in 1.2/6, admits a universal property:

Proposition 4 (Fundamental theorem on homomorphisms). *Let $\varphi\colon R \longrightarrow R'$ be a ring homomorphism and $\mathfrak{a} \subset R$ an ideal satisfying $\mathfrak{a} \subset \ker \varphi$. Then there exists a unique ring homomorphism $\overline{\varphi}\colon R/\mathfrak{a} \longrightarrow R'$ such that the diagram*



is commutative. Furthermore,

$$\operatorname{im}\overline{\varphi} = \operatorname{im}\varphi, \qquad \ker\overline{\varphi} = \pi(\ker\varphi), \qquad \ker\varphi = \pi^{-1}(\ker\overline{\varphi}).$$

In particular, $\overline{\varphi}$ is injective if and only if $\mathfrak{a} = \ker\varphi$.

Corollary 5. *If $\varphi\colon R \longrightarrow R'$ is a surjective ring homomorphism, then R' is canonically isomorphic to $R/\ker\varphi$.*

To do the *proof of Proposition* 4 we apply 1.2/6 to the additive group of R. Then it remains only to check that the group homomorphism $\overline{\varphi}\colon R/\mathfrak{a} \longrightarrow R'$ obtained from 1.2/6 is in fact a ring homomorphism. However, since $\overline{\varphi}$ is characterized by the equation

$$\overline{\varphi}(x + \mathfrak{a}) = \varphi(x), \qquad x \in R,$$

this assertion is clear. □

Also note that the isomorphism theorems 1.2/8 and 1.2/9, which were obtained as corollaries to the fundamental theorem on homomorphisms 1.2/6, carry over without difficulty from the setting of groups to the present setting of rings, either in a direct way or as a consequence of Proposition 4; just replace normal subgroups by ideals.

The rings $\mathbb{Z}/m\mathbb{Z}$ viewed in 1.3 as abelian groups are natural examples of residue class rings. In particular, for $m > 0$ we see that $\mathbb{Z}/m\mathbb{Z}$ is a ring consisting of m elements.

Proposition 6. *The following conditions are equivalent for $m \in \mathbb{Z}$, $m > 0$:*

(i) *m is a prime number.*
(ii) *$\mathbb{Z}/m\mathbb{Z}$ is an integral domain.*
(iii) *$\mathbb{Z}/m\mathbb{Z}$ is a field.*

Proof. For any $x \in \mathbb{Z}$, let us denote by $\overline{x} \in \mathbb{Z}/m\mathbb{Z}$ the attached residue class modulo $m\mathbb{Z}$. To begin with, assume condition (i), i.e., that m is a prime number. Then $m > 1$ and $\mathbb{Z}/m\mathbb{Z}$ is nonzero. Now if $\overline{a} \cdot \overline{b} = 0$ for two integers $a, b \in \mathbb{Z}$, then $ab \in m\mathbb{Z}$, and we see, for example by looking at the prime factorizations of a, b, and ab, that m divides a or b. Thereby we get $a \in m\mathbb{Z}$ or $b \in m\mathbb{Z}$, i.e., $\overline{a} = 0$ or $\overline{b} = 0$, and $\mathbb{Z}/m\mathbb{Z}$ is an integral domain, as required in (ii).

Next we can conclude for every $\overline{a} \in \mathbb{Z}/m\mathbb{Z} - \{0\}$ from (ii) that the map

$$\mathbb{Z}/m\mathbb{Z} \longrightarrow \mathbb{Z}/m\mathbb{Z}, \qquad \overline{x} \longmapsto \overline{a} \cdot \overline{x},$$

is injective and, in fact, bijective, since $\mathbb{Z}/m\mathbb{Z}$ is finite. In particular, the unit element $\overline{1}$ of $\mathbb{Z}/m\mathbb{Z}$ belongs to the image of this map, and it follows that $\overline{a}$ admits a multiplicative inverse in $\mathbb{Z}/m\mathbb{Z}$. But then $\mathbb{Z}/m\mathbb{Z}$ is field, and we get (iii).

Finally, assume that $\mathbb{Z}/m\mathbb{Z}$ is a field as in (iii) or, more generally, an integral domain. This implies $\mathbb{Z}/m\mathbb{Z} \neq 0$ and therefore $m > 1$. To show that m is prime, consider a divisor $d \in \mathbb{N}$ of m and an equation $m = da$. Then $\overline{d} \cdot \overline{a} = 0$, and we get $\overline{d} = 0$ or $\overline{a} = 0$, since $\mathbb{Z}/m\mathbb{Z}$ is an integral domain. In the first case, m

divides d, and hence $d = m$. In the second, m divides a, so that $a = m$ and $d = 1$. Thus, m admits only itself and 1 as divisors and therefore is prime. □

In particular, we have seen that the ring $\mathbb{Z}/p\mathbb{Z}$, for a prime p, is a field consisting of p elements; it is denoted by $\mathbb{F}_p$. More generally, using elementary number theory, one can show for integers $m > 1$ that the group of units $(\mathbb{Z}/m\mathbb{Z})^*$ consists of all residue classes $\overline{a}$ such that $a \in \mathbb{Z}$ is relatively prime to m. Next we want to view the assertion of Proposition 6 in a more general context.

Definition 7. *Let R be a ring.*

(i) *A proper ideal $\mathfrak{p} \subsetneq R$ is called a* prime ideal *if $ab \in \mathfrak{p}$ for elements $a, b \in R$ implies $a \in \mathfrak{p}$ or $b \in \mathfrak{p}$.*

(ii) *A proper ideal $\mathfrak{m} \subsetneq R$ is called a* maximal ideal *if $\mathfrak{m} \subset \mathfrak{a} \subset R$ for an ideal $\mathfrak{a} \subset R$ implies $\mathfrak{a} = \mathfrak{m}$ or $\mathfrak{a} = R$.*

For example, the zero ideal of a ring R is prime if and only if R is an integral domain.

Proposition 8. *Let R be a ring.*

(i) *An ideal $\mathfrak{p} \subset R$ is prime if and only if $R/\mathfrak{p}$ is an integral domain.*

(ii) *An ideal $\mathfrak{m} \subset R$ is maximal if and only if $R/\mathfrak{m}$ is a field.*

In particular, every maximal ideal is prime.

Proof. First of all, note that $\mathfrak{p}$ is a proper ideal in R if and only if the residue class ring $R/\mathfrak{p}$ is nonzero, similarly for $\mathfrak{m}$. Now assertion (i) is easy to verify. Look at residue classes $\overline{a}, \overline{b} \in R/\mathfrak{p}$ of elements $a, b \in R$. Then

$$a \cdot b \in \mathfrak{p} \quad \Longrightarrow \quad a \in \mathfrak{p} \text{ or } b \in \mathfrak{p}$$

is clearly equivalent to

$$\overline{a} \cdot \overline{b} = 0 \quad \Longrightarrow \quad \overline{a} = 0 \text{ or } \overline{b} = 0.$$

Furthermore, assertion (ii) is a consequence of the following two lemmas:

Lemma 9. *An ideal $\mathfrak{m} \subset R$ is maximal if and only if the zero ideal $0 \subset R/\mathfrak{m}$ is maximal.*

Lemma 10. *The zero ideal $0 \subset R$ of a ring R is maximal if and only if R is a field.*

Proof of Lemma 9. Let $\pi\colon R \longrightarrow R/\mathfrak{m}$ be the canonical projection. It is easily checked that the mappings

$$\begin{aligned} R \supset \quad \mathfrak{a} \quad &\longmapsto \pi(\mathfrak{a}) \subset R/\mathfrak{m}, \\ R \supset \pi^{-1}(\mathfrak{b}) &\longmapsfrom \quad \mathfrak{b} \quad \subset R/\mathfrak{m}, \end{aligned}$$

define a bijection between all ideals $\mathfrak{a} \subset R$ such that $\mathfrak{m} \subset \mathfrak{a} \subset R$ and the ideals $\mathfrak{b} \subset R/\mathfrak{m}$. The stated equivalence is an immediate consequence of this fact.

Alternatively, the assertion can be verified in a more direct way. Indeed, recall that $\mathfrak{m}$ is a proper ideal in R if and only if the residue class ring $R/\mathfrak{m}$ is nonzero. Now, a proper ideal $\mathfrak{m} \subset R$ is maximal if and only if we have $\mathfrak{m} + Ra = R$ for each $a \in R - \mathfrak{m}$, i.e., if and only if for each such a there exist elements $r \in R$ and $m \in \mathfrak{m}$ such that $ra + m = 1$. Using the projection $\pi\colon R \longrightarrow R/\mathfrak{m}$, this condition is true if and only if each $\overline{a} \in R/\mathfrak{m} - \{0\}$ admits a multiplicative inverse $\overline{r} \in R/\mathfrak{m}$ satisfying $\overline{r} \cdot \overline{a} = 1$, i.e., if and only if, finally, the zero ideal is maximal in $R/\mathfrak{m}$. □

Proof of Lemma 10. Assume that the zero ideal $0 \subset R$ is maximal and consider an element $a \in R - \{0\}$. Then $aR = R$, and there exists an element $b \in R$ such that $ab = 1$. Thus $R^* = R - \{0\}$, and R is a field. Conversely, that the zero ideal of a field is maximal is immediately clear. □

Propositions 6 and 8 give a complete overview of prime and maximal ideals in $\mathbb{Z}$:

Corollary 11. *An ideal in $\mathbb{Z}$ is prime if and only if it is of type $p\mathbb{Z}$ for a prime number p or for $p = 0$. An ideal in $\mathbb{Z}$ is maximal if and only if it is a nonzero prime ideal.*

Just use the fact that $\mathbb{Z}$ is a principal ideal domain by 2.2/3 and that the zero ideal of an integral domain is prime. To end the present section we want to establish the so-called *Chinese remainder theorem.*

Proposition 12. *Let R be a ring and $\mathfrak{a}_1, \ldots, \mathfrak{a}_n \subset R$ pairwise* coprime *ideals, i.e., assume that $\mathfrak{a}_i + \mathfrak{a}_j = R$ for $i \neq j$. Then, writing $\pi_i\colon R \longrightarrow R/\mathfrak{a}_i$ for the canonical projections, the homomorphism*

$$\varphi\colon R \longrightarrow R/\mathfrak{a}_1 \times \ldots \times R/\mathfrak{a}_n, \qquad x \longmapsto \big(\pi_1(x), \ldots, \pi_n(x)\big),$$

is surjective and satisfies $\ker \varphi = \mathfrak{a}_1 \cap \ldots \cap \mathfrak{a}_n$. *In particular, it induces an isomorphism*

$$R/\bigcap_{i=1}^{n} \mathfrak{a}_i \xrightarrow{\sim} \prod_{i=1}^{n} R/\mathfrak{a}_i,$$

where $\prod_{i=1}^n R/\mathfrak{a}_i = R/\mathfrak{a}_1 \times \ldots \times R/\mathfrak{a}_n$ is the ring-theoretic product of the residue class rings $R/\mathfrak{a}_i$.

Proof. To begin with, let us show for $j = 1, \ldots, n$ that the ideals $\mathfrak{a}_j$ and $\bigcap_{i \neq j} \mathfrak{a}_i$ are coprime in the sense that their sum yields R. To do this, fix an index j. Since $\mathfrak{a}_j$ and $\mathfrak{a}_i$ are coprime by assumption for each $i \neq j$, there are elements $a_i \in \mathfrak{a}_j$ and $a_i' \in \mathfrak{a}_i$ such that $a_i + a_i' = 1$. Hence, we get

$$1 = \prod_{i \neq j} (a_i + a_i') \in \mathfrak{a}_j + \prod_{i \neq j} \mathfrak{a}_i \subset \mathfrak{a}_j + \bigcap_{i \neq j} \mathfrak{a}_i$$

and therefore $\mathfrak{a}_j + \bigcap_{i \neq j} \mathfrak{a}_i = R$, as claimed.

In particular, there exist equations $d_j + e_j = 1$ for $j = 1, \ldots, n$ and elements $d_j \in \mathfrak{a}_j$, $e_j \in \bigcap_{i \neq j} \mathfrak{a}_i$, and we see that

$$\pi_i(e_j) = \begin{cases} 1 & \text{for} \quad i = j, \\ 0 & \text{for} \quad i \neq j. \end{cases}$$

This shows that φ is surjective. Indeed, look at an element $y = (y_1, \ldots, y_n)$ in $R/\mathfrak{a}_1 \times \ldots \times R/\mathfrak{a}_n$ and choose a π_i-preimage $x_i \in R$ of y_i for each i. Then

$$\varphi\Big(\sum_{i=1}^{n} x_i e_i\Big) = y.$$

Finally, the assertion about the kernel of φ is trivial, and the stated isomorphism is readily derived from the fundamental theorem on homomorphisms. □

If $\mathfrak{a}$ is an ideal of a ring R, two elements $x, y \in R$ are said to be *congruent modulo* $\mathfrak{a}$, written $x \equiv y \mod \mathfrak{a}$, if x and y give rise to the same residue class in $R/\mathfrak{a}$, i.e., if $x - y \in \mathfrak{a}$. In the case that $\mathfrak{a}$ is a principal ideal Ra, one also writes "mod a" instead of "mod $\mathfrak{a}$." Using such terminology, the surjectivity of the map φ in Proposition 12 can be expressed as follows: given $x_1, \ldots, x_n \in R$, there exists an element $x \in R$ such that $x \equiv x_i \mod \mathfrak{a}_i$ for $i = 1, \ldots, n$. Consequently, for the ring of integers $\mathbb{Z}$, the Chinese remainder theorem takes the following shape:

Corollary 13. *Let $a_1, \ldots, a_n \in \mathbb{Z}$ be integers that are pairwise relatively prime. Then the system of simultaneous congruences $x \equiv x_i \mod a_i$, $i = 1, \ldots, n$, is solvable for arbitrary integers $x_1, \ldots, x_n \in \mathbb{Z}$, and the solution x is unique modulo $a_1 \cdot \ldots \cdot a_n$. Therefore, the set of all solutions forms a residue class of type $x + a_1 \cdot \ldots \cdot a_n \mathbb{Z}$.*

Of course, it has to be checked that relatively prime integers $a, a' \in \mathbb{Z}$ satisfy the equations

$$(a, a') = (1) \quad \text{and} \quad (a \cdot a') = (a) \cap (a');$$

for details see 2.4/13. Also note that the proof of the Chinese remainder theorem provides a constructive method for solving systems of simultaneous congruences. Indeed, one determines integers $d_j \in (a_j)$, $e_j \in (\prod_{i \neq j} a_i)$ for $j = 1, \ldots, n$ that satisfy $d_j + e_j = 1$, for example using *Euclid's algorithm*; see 2.4/15. Then $x = \sum_{i=1}^{n} x_i e_i$ is a solution of the system $x \equiv x_i \mod a_i$, $i = 1, \ldots, n$, and all other solutions are obtained from this special one by adding a multiple of $\prod_{i=1}^{n} a_i$.

Exercises

1. *Consider a ring homomorphism* $\varphi\colon R \longrightarrow R'$ *and look for valid assertions about the images of ideals* $\mathfrak{a} \subset R$*, as well as on the preimages of ideals* $\mathfrak{a}' \subset R'$*. Examine the same question also for prime and maximal ideals.*

2. *For an element* x *of a ring* R*, consider the substitution homomorphism*

$$\varphi_x\colon R[X] \longrightarrow R, \qquad \sum a_i X^i \longmapsto \sum a_i x^i,$$

and describe the kernel of φ_x*. In particular, discuss the cases in which it is a prime ideal, resp. a maximal ideal in* $R[X]$*.*

3. Generalize the isomorphism theorems 1.2/8 and 1.2/9 to the setting of rings, by considering rings instead of groups and ideals instead of normal subgroups.

4. Let $\varphi\colon R \longrightarrow R'$ be a ring homomorphism. Show for an element $x \in R'$ that there is precisely one ring homomorphism $\Phi\colon R[X] \longrightarrow R'$ satisfying $\Phi|_R = \varphi$ and $\Phi(X) = x$. In particular, the set of ring homomorphisms $\Phi\colon R[X] \longrightarrow R'$ such that $\Phi|_R = \varphi$ is in one-to-one correspondence with the set of elements of R'.

5. Let R be an integral domain and $\Phi\colon R[X] \longrightarrow R[X]$ a ring homomorphism satisfying $\Phi|_R = \mathrm{id}_R$. Show that Φ is an automorphism if and only if there are elements $a \in R^*$ and $b \in R$ such that $\Phi(X) = aX + b$.

6. Let $\mathfrak{p}$ be a prime ideal of a ring R. Show that $\mathfrak{p}R[X]$, the ideal generated by $\mathfrak{p}$ in $R[X]$, is a prime ideal.

7. Let K be a field and $K[X,Y] = K[X][Y]$ the polynomial ring in two variables X and Y over K. Consider the residue class ring $R = K[X,Y]/(XY^2)$ and denote by $\overline{X}$, resp. $\overline{Y}$, the corresponding residue classes of X, resp. Y. Show that the elements $\overline{X}$ and $\overline{X} + \overline{X} \cdot \overline{Y}$ generate the same principal ideal in R, although they are not associated. *Hint*: Look at the ideal consisting of all elements $\overline{f} \in R$ such that $\overline{f} \cdot \overline{X} = 0$, resp. at the ideal of all elements $f \in K[X,Y]$ such that $fX \in (XY^2)$.

8. Let R be a ring. Show that $\{\sum a_i X^i \in R[X]\ ;\ a_1 = 0\}$ is a subring of $R[X]$ and that it is isomorphic to $R[X][Y]/(X^2 - Y^3)$.

2.4 Prime Factorization

Basic properties of the ring of integers $\mathbb{Z}$, and the polynomial ring $K[X]$ over a field K, are related to the fact that these rings admit a division with remainder, which is called *Euclidean division*. We want to look at general integral domains admitting such a division process and show that they belong to the class of principal ideal domains. For principal ideal domains, in turn, we will prove the existence and uniqueness of prime factorizations.

Definition 1. *An integral domain* R *is called a* Euclidean domain *if it admits a map* $\delta\colon R - \{0\} \longrightarrow \mathbb{N}$ *making possible* Euclidean division *in* R *in the following sense*:

Given elements $f, g \in R$, $g \neq 0$, there are elements $q, r \in R$ such that

$$f = qg + r, \quad \textit{where} \ \ \delta(r) < \delta(g) \ \ \textit{or} \ \ r = 0.$$

The map δ is referred to as a Euclidean function *of the Euclidean domain R.*

Every field is a Euclidean domain for trivial reasons, but there are some more interesting examples.

(1) $\mathbb{Z}$ is a Euclidean domain under the usual division with remainder. The map $\delta\colon \mathbb{Z}-\{0\} \longrightarrow \mathbb{N}$, $a \longmapsto |a|$, serves as a Euclidean function.

(2) The polynomial ring $K[X]$ over a field K is a Euclidean domain under the usual polynomial division introduced in 2.1/4; the map $\delta\colon K[X]-\{0\} \longrightarrow \mathbb{N}$, $f \longmapsto \deg f$, serves as a Euclidean function.

(3) The *ring of Gaussian integers* $\mathbb{Z}[i] := \{x + iy\,;\, x, y \in \mathbb{Z}\} \subset \mathbb{C}$ is a Euclidean domain, with Euclidean function

$$\delta\colon \mathbb{Z}[i]-\{0\} \longrightarrow \mathbb{N}, \qquad x + iy \longmapsto x^2 + y^2 = |x + iy|^2.$$

In order to characterize the division with remainder in $\mathbb{Z}[i]$, observe that the distance between adjacent points in $\mathbb{Z}[i]$ is at most $\sqrt{2}$. Thus, given $f, g \in \mathbb{Z}[i]$, $g \neq 0$, there exist integers $x, y \in \mathbb{Z}$ such that $|fg^{-1} - (x + iy)| \leq \frac{1}{2} \cdot \sqrt{2} < 1$. Now setting $q := (x + iy)$ and $r := f - qg$, we get $|r| < |g|$ and therefore

$$f = qg + r, \quad \text{where} \quad \delta(r) < \delta(g) \text{ or } r = 0.$$

(4) Let $d \neq 0, 1$ be a square-free integer, which means that $d \in \mathbb{Z}$ does not admit a square of an integer > 1 as a divisor, and consider the following subring of $\mathbb{C}$:

$$R_d = \begin{cases} \mathbb{Z} + \sqrt{d} \cdot \mathbb{Z}, & \text{if } d \equiv 2, 3 \mod 4, \\ \mathbb{Z} + \frac{1}{2}\big(1 + \sqrt{d}\big) \cdot \mathbb{Z}, & \text{if } d \equiv 1 \mod 4. \end{cases}$$

For $d = -1$ we obtain the ring of Gaussian integers, as discussed before. The rings R_d are of special interest in number theory. One would like to know whether R_d, depending on d, is a unique factorization domain, i.e., whether the elements of R_d admit unique prime factorization. Since a Euclidean domain is a principal ideal domain and a principal ideal domain is a unique factorization domain, see Proposition 2 and Corollary 11 below, one starts by looking at values d such that R_d is Euclidean. As Euclidean function $\delta\colon R_d-\{0\} \longrightarrow \mathbb{N}$ one may try the so-called *norm* map given by $\delta(a + b\sqrt{d}) = |a^2 - b^2 d|$; see Section 4.7 for details on the norm. It can be shown that the norm map is a Euclidean function on R_d precisely for the following values of d:

$$d = -1, -2, -3, -7, -11,$$
$$d = 2, 3, 5, 6, 7, 11, 13, 17, 19, 21, 29, 33, 37, 41, 57, 73.$$

In particular, R_d is Euclidean and thus a unique factorization domain in these cases. Moreover, it is known for $d < 0$ that R_d is a unique factorization domain in precisely the following additional cases:

$$d = -19, -43, -67, -163.$$

On the other hand, there are several values $d > 0$ for which R_d is known to be a unique factorization domain, but not necessarily a Euclidean domain; see, for example, H. Hasse [7], §16.6.

Proposition 2. *Every Euclidean domain is a principal ideal domain.*

Proof. Proceeding as in 1.3/4, let $\mathfrak{a} \subset R$ be an ideal, where we may assume $\mathfrak{a} \neq 0$. Choose an element a of $\mathfrak{a} - \{0\}$ such that $\delta(a)$ is minimal with respect to the Euclidean function δ considered on R. We claim that $\mathfrak{a} = (a)$. Indeed, let $f \in \mathfrak{a}$ and decompose it in terms of Euclidean division, say $f = qa + r$, where $\delta(r) < \delta(a)$ or $r = 0$. Then we get $r = f - qa \in \mathfrak{a}$. However, due to the minimality of $\delta(a)$, we must have $r = 0$ and hence $f = qa \in (a)$. This shows that $\mathfrak{a} \subset (a)$. Since the reverse inclusion is trivial, we get $\mathfrak{a} = (a)$, and $\mathfrak{a}$ is principal. □

Corollary 3. *The rings $\mathbb{Z}$, $\mathbb{Z}[i]$, as well as the polynomial ring $K[X]$ over a field K, are Euclidean domains and hence principal ideal domains.*

Next we want to study prime factorizations in principal ideal domains. For elements x and y of an integral domain R we say that x *divides* y, writing $x \,|\, y$, if there is an element $c \in R$ such that $cx = y$, or equivalently, if $y \in (x)$. If such is not the case, i.e., if x does not divide y, we write $x \nmid y$.

Definition 4. *Let R be an integral domain and $p \in R$ a nonzero nonunit.*

(i) *p is called* irreducible *if a factorization of type $p = xy$ with $x, y \in R$ implies $x \in R^*$ or $y \in R^*$. Furthermore, p is called* reducible *if it is not irreducible.*

(ii) *p is called a* prime element *if from $p \,|\, xy$ with $x, y \in R$ we get $p \,|\, x$ or $p \,|\, y$, i.e., in other words, if the principal ideal (p) is prime.*

The irreducible elements of the ring of integers $\mathbb{Z}$ are given, up to sign, precisely by the usual prime numbers, while we can see for the polynomial ring $K[X]$ over a field K that particularly the linear polynomials of type $X - a$ for $a \in K$ are irreducible. Note that for $K = \mathbb{C}$ there are no further irreducible polynomials, up to associatedness, i.e., up to multiplication by nonzero constants from K; this will be a consequence of the fundamental theorem of algebra, to be proved in Section 6.3. However, over a general field K, there can exist irreducible polynomials of higher degree, such as the polynomial $X^2 + 1$ in $\mathbb{R}[X]$. Furthermore, we will see in Proposition 6 that the irreducible and prime

elements coincide in principal ideal domains, for example in the ring of integers $\mathbb{Z}$ and the polynomial ring $K[X]$ over a field K.

Remark 5. *Let R be an integral domain and $p \in R$ a nonzero nonunit.*
(i) *If (p) is a maximal ideal in R, then p is a prime element.*
(ii) *If p is a prime element, then p is irreducible.*

Proof. If (p) is a maximal ideal in R, then it is a prime ideal as well, see 2.3/8, and it follows that p is a prime element. This verifies assertion (i). To establish (ii), assume that p is a prime element. Then, if $p = xy$ for some $x, y \in R$, we get $p \,|\, x$ or $p \,|\, y$, since p is prime. Assuming $p \,|\, x$, there exists an element $c \in R$ such that $pc = x$ and hence $p = xy = pcy$. Since R is an integral domain, we must have $cy = 1$ and therefore $y \in R^*$. Thus, p is irreducible. □

For principal ideal domains, the assertions we have just proved can be sharpened substantially; see also 2.3/6.

Proposition 6. *Let R be a principal ideal domain and $p \in R$ a nonzero nonunit. The following conditions are equivalent*:
(i) *p is irreducible.*
(ii) *p is a prime element.*
(iii) *(p) is a maximal ideal in R.*

Proof. In view of Remark 5, it remains to show that (i) implies (iii). Therefore, assume that p is irreducible and let $\mathfrak{a}$ be an ideal in R satisfying $(p) \subset \mathfrak{a} \subset R$, say $\mathfrak{a} = (a)$, since R is a principal ideal domain. Then there is an element $c \in R$ such that $p = ac$, and p being irreducible implies $a \in R^*$ or $c \in R^*$. Hence, we get $\mathfrak{a} = R$ in the first case and $\mathfrak{a} = (p)$ in the second. This shows that (p) is a maximal ideal in R. □

Using the preceding result, it is quite easy to derive the existence of prime factorizations in principal ideal domains. In fact, it is enough to look at factorizations into irreducible elements.

Proposition 7. *Let R be a principal ideal domain. Then every nonzero nonunit $a \in R$ is a product of prime elements.*[2]

Proof. Fix an element $a \in R - (R^* \cup \{0\})$. If a is irreducible (and thereby prime), nothing has to be shown. Otherwise, decompose a into the product bc of two nonunits in R. If one of the factors b and c is not yet irreducible, we can further decompose b or c and so forth. To obtain the desired assertion of the proposition it remains only to show that the procedure stops after finitely many steps and thus yields a factorization of a into irreducible and thereby prime factors. Concerning the rings that are of special interest to us, such as $\mathbb{Z}$ and

[2] A product of elements of a ring is always meant to be a *finite* product.

$K[X]$ for a field K, this is immediately clear. Indeed, in $\mathbb{Z}$ we have $|b|, |c| < |a|$ for a factorization of a into nonunits b, c. Similarly, we get $\deg b, \deg c < \deg a$ in $K[X]$, as follows from 2.1/2. Hence, recursively decomposing a into a product of nonunits, the absolute value, resp. the degree of factors, strictly decreases every time we further decompose a factor into a product of nonunits. Therefore, it becomes clear that the procedure must stop after finitely many steps.

However, in order to establish Proposition 7 in its full generality, we add here a general argument, valid for principal ideal domains R, that implies the existence of factorizations of a into (finite) products of irreducible elements. The following auxiliary assertion is needed:

Lemma 8. *Every principal ideal domain R is* Noetherian, *i.e., every ascending chain of ideals $\mathfrak{a}_1 \subset \mathfrak{a}_2 \subset \ldots \subset R$ becomes stationary in the sense that there is some $n \in \mathbb{N}$ such that $\mathfrak{a}_i = \mathfrak{a}_n$ for all $i \geq n$.*

The assertion is easy to verify. Since the union of an ascending chain of ideals is itself an ideal, we can consider $\mathfrak{a} = \bigcup_{i \geq 1} \mathfrak{a}_i$ as an ideal in R; it is a principal ideal, say $\mathfrak{a} = (a)$. However, since $a \in \mathfrak{a}$, there is some $n \in \mathbb{N}$ such that $a \in \mathfrak{a}_n$, and we get $(a) \subset \mathfrak{a}_n \subset \mathfrak{a} = (a)$. As a result, the chain of ideals $\mathfrak{a}_1 \subset \mathfrak{a}_2 \subset \ldots$ becomes stationary at $\mathfrak{a}_n$.

Now we can prove Proposition 7 for general principal ideal domains R. Let S be the set of all principal ideals in R admitting a generator $a \in R - (R^* \cup \{0\})$ such that a does not allow a finite factorization into irreducible elements. We have to show that $S = \emptyset$. Assuming $S \neq \emptyset$, we conclude from Lemma 8 the existence of a maximal element in S, i.e., of an element $\mathfrak{a} \in S$ such that every strict inclusion $\mathfrak{a} \subsetneq \mathfrak{b}$ of ideals in R implies that $\mathfrak{b}$ cannot belong to S. Now let $\mathfrak{a} = (a)$ be such a maximal element of S. Then the generating element a must be reducible, say $a = a_1 a_2$ for nonunits $a_1, a_2 \in R$. As a consequence, we get strict inclusions

$$(a) \subsetneq (a_1), \qquad (a) \subsetneq (a_2),$$

and we see that (a_1) and (a_2) cannot belong to S. In particular, a_1 and a_2 admit factorizations into irreducible elements, and the same is true for the product $a = a_1 a_2$, in contradiction to $(a) \in S$. Therefore, $S = \emptyset$, and the assertion of Proposition 7 is clear. □

Next we want to show that prime factorizations as considered in Proposition 7 are essentially unique.

Lemma 9. *Let R be an integral domain. For an element $a \in R$, consider factorizations*

$$a = p_1 \ldots p_r = q_1 \ldots q_s$$

into prime elements p_i and irreducible elements q_j. Then $r = s$, and one can renumber the q_j in such a way that p_i is associated to q_i for $i = 1, \ldots, r$.

Proof. Since $p_1 \,|\, q_1 \dots q_s$ and p_1 is prime, there exists an index j such that $p_1 \,|\, q_j$. Renumbering the q_j, we may assume $j = 1$. Hence, using the fact that q_1 is irreducible, there is an equation $q_1 = \varepsilon_1 p_1$ for a unit ε_1, and we can conclude that

$$p_2 \dots p_r = \varepsilon_1 q_2 \dots q_s.$$

Continuing inductively, the desired assertion follows. □

Proposition and Definition 10. *For an integral domain R, the following conditions are equivalent*:

(i) *Every element $a \in R - (R^* \cup \{0\})$ can be uniquely written, up to associatedness and order, as a product of irreducible elements.*

(ii) *Every element $a \in R - (R^* \cup \{0\})$ is a product of prime elements.*

An integral domain R satisfying the equivalent conditions (i) *and* (ii) *is called a* factorial ring *or a* unique factorization domain. *Alternatively, we say that the elements of R admit* unique prime factorization.

An element of a unique factorization domain is irreducible if and only if it is prime.

Proof. First, assuming condition (i), we want to show that every irreducible element of R is prime. To do this, let $a \in R$ be irreducible and fix $x, y \in R$ such that $a \,|\, xy$. We have to show that $a \,|\, x$ or $a \,|\, y$, where we may assume that x and y are nonunits. Now let $x = x_1 \dots x_r$ and $y = y_1 \dots y_s$ be factorizations into irreducible elements as provided by (i). Then we get $a \,|\, (x_1 \dots x_r y_1 \dots y_s)$, and the uniqueness assertion of (i) implies that a, being irreducible, is associated to one of the elements x_i, y_j. In particular, we get $a \,|\, x$ or $a \,|\, y$. Hence, a is prime.

As a by-product, this argument settles the implication from (i) to (ii), while the reverse follows from Lemma 9. Indeed, due to Remark 5, any factorization into prime elements is a factorization into irreducible elements.

Finally, if R is a unique factorization domain, it satisfies condition (i), and thus all irreducible elements are prime. The converse of this follows from Remark 5 again. □

Now the assertion of Proposition 7 can be phrased in a new way:

Corollary 11. *Every principal ideal domain is a unique factorization domain.*

Fields are unique factorization domains for trivial reasons. But also the rings $\mathbb{Z}$, $\mathbb{Z}[i]$, as well as the polynomial ring $K[X]$ over a field K, are unique factorization domains, since they are Euclidean and thus principal ideal domains. We will show in 2.7/1 that the polynomial ring $R[X]$ over a unique factorization domain R is itself a unique factorization domain. For example, we thereby see that the polynomial ring $\mathbb{Z}[X]$ is a unique factorization domain, although it is not a principal ideal domain. The same holds for the polynomial ring $K[X, Y] := K[X][Y]$ in two variables X and Y over a field K.

For unique factorization domains R, it is quite common to write a product of associated prime elements as a power of one of them, of course adjusted by a unit. In this way, one considers prime factorizations of type

$$a = \varepsilon p_1^{\nu_1} \dots p_r^{\nu_r},$$

for a unit ε, certain exponents $\nu_1, \dots, \nu_r$, and pairwise nonassociated prime elements $p_1, \dots, p_r$. Then, formally speaking, every element $a \in R - \{0\}$ admits such a prime factorization (with exponents $\nu_i = 0$ if a is a unit). To further standardize prime factorizations, one can fix a system P of representatives of the prime elements in R, i.e., a subset $P \subset R$ that contains precisely one element from each class of mutually associated prime elements. In this way, prime factorizations in R take the form

$$a = \varepsilon \prod_{p \in P} p^{\nu_p(a)},$$

where now $\varepsilon \in R^*$ and the exponents $\nu_p(a) \in \mathbb{N}$ are unique. Of course, we have $\nu_p(a) = 0$ for almost all $p \in P$, so that the product is actually finite. For the ring of integers $\mathbb{Z}$ it is common to define P as the set of all (positive) prime numbers, while in the polynomial ring $K[X]$ over a field K one takes for P the set of all monic irreducible (or prime) polynomials, i.e., the set of all irreducible polynomials whose leading coefficient is 1.

In the following we want to discuss the notions of the greatest common divisor and the least common multiple, which are often used within the context of unique factorization domains. To do this, consider an integral domain R and fix elements $x_1, \dots, x_n \in R$. An element $d \in R$ is called a *greatest common divisor* of $x_1, \dots, x_n$ if:

(i) $d \,|\, x_i$ for $i = 1, \dots, n$, i.e., d is a common divisor of all the x_i.

(ii) If $a \in R$ is a common divisor of the x_i, i.e., if $a \,|\, x_i$ for $i = 1, \dots, n$, then $a \,|\, d$.

Such a greatest common divisor d is unique up to associatedness, if it exists, and we will write $d = \gcd(x_1, \dots, x_n)$. If $d = 1$, then $x_1, \dots, x_n$ are said to be *coprime*.

An element $v \in R$ is called a *least common multiple* of $x_1, \dots, x_n$ if:

(i) $x_i \,|\, v$ for $i = 1, \dots, n$, i.e., v is a common multiple of all the x_i.

(ii) If $a \in R$ is a common multiple of the x_i, i.e., if $x_i \,|\, a$ for $i = 1, \dots, n$, then $v \,|\, a$.

Similarly as before, such a least common multiple v is unique up to associatedness, if it exists, and we write $v = \operatorname{lcm}(x_1, \dots, x_n)$. As usual, one proves the following result:

Proposition 12. *Let R be a unique factorization domain and let P be a system of representatives of all prime elements in R. If*

$$x_i = \varepsilon_i \prod_{p \in P} p^{\nu_p(x_i)}, \qquad i = 1, \dots, n,$$

are prime factorizations of elements $x_1, \dots, x_n \in R$, then $\gcd(x_1, \dots, x_n)$ and $\operatorname{lcm}(x_1, \dots, x_n)$ exist and are given by

$$\gcd(x_1, \dots, x_n) = \prod_{p \in P} p^{\min(\nu_p(x_1), \dots, \nu_p(x_n))},$$
$$\operatorname{lcm}(x_1, \dots, x_n) = \prod_{p \in P} p^{\max(\nu_p(x_1), \dots, \nu_p(x_n))},$$

up to associatedness.

In principal ideal domains the greatest common divisor and the least common multiple can be characterized in terms of ideals:

Proposition 13. *Let $x_1, \dots, x_n$ be elements of an integral domain R.*

(i) *If $(x_1, \dots, x_n)$, the ideal generated by the x_i in R, is principal, say generated by an element $d \in R$, then $d = \gcd(x_1, \dots, x_n)$.*

(ii) *If $(x_1) \cap \dots \cap (x_n)$ is a principal ideal, say generated by an element $v \in R$, then $v = \operatorname{lcm}(x_1, \dots, x_n)$.*

Proof. (i) Assume $(x_1, \dots, x_n) = (d)$. Then we get $x_i \in (d)$ and therefore $d \,|\, x_i$ for all i. Moreover, due to $d \in (x_1, \dots, x_n)$, there is an equation $d = \sum_{i=1}^n a_i x_i$ with suitable coefficients $a_i \in R$. In particular, every common divisor of the x_i is a divisor of d as well, which shows that $d = \gcd(x_1, \dots, x_n)$.

(ii) Assume $\bigcap_{i=1}^n (x_i) = (v)$. Then v belongs to each of the ideals (x_i) and hence is a common multiple of all the x_i. Now if a is another common multiple of the x_i, we have $a \in (x_i)$ for all i and hence $a \in \bigcap_{i=1}^n (x_i) = (v)$. This means that $v \,|\, a$, and we get $v = \operatorname{lcm}(x_1, \dots, x_n)$. □

The above characterization of the greatest common divisor and the least common multiple in terms of ideals can be used to derive the following special version of the Chinese remainder theorem 2.3/12:

Corollary 14. *Let R be a principal ideal domain and $a = \varepsilon p_1^{\nu_1} \dots p_r^{\nu_r}$ a prime factorization of some element $a \in R$, where ε is a unit and the prime elements p_i are pairwise nonassociated. Then the Chinese remainder theorem* 2.3/12 *provides a canonical isomorphism*

$$R/(a) \xrightarrow{\sim} R/(p_1^{\nu_1}) \times \dots \times R/(p_n^{\nu_n}).$$

Proof. Using Proposition 13 in conjunction with Proposition 12, the ideals $(p_1^{\nu_1}), \dots, (p_r^{\nu_r})$ are pairwise coprime in R, since $\gcd(p_i^{\nu_i}, p_j^{\nu_j}) = 1$ for $i \neq j$. Likewise, we have $(a) = \bigcap_{i=1}^r (p_i^{\nu_i})$, since $a = \operatorname{lcm}(p_1^{\nu_1}, \dots, p_r^{\nu_r})$. □

For Euclidean domains R, there exists a constructive process to determine the greatest common divisor of two given elements $x, y \in R$, the so-called *Euclidean algorithm*. Applying this algorithm iteratively and using relations of

type $\gcd(x,y,z) = \gcd(\gcd(x,y),z)$, the process can even be used to determine the greatest common divisor of any number of elements in R.

Proposition 15 (Euclidean algorithm). *Let R be a Euclidean domain. For two elements $x, y \in R-\{0\}$ consider the sequence $z_0, z_1, \ldots \in R$, which is inductively given by*

$$z_0 = x,$$
$$z_1 = y,$$
$$z_{i+1} = \begin{cases} \text{the remainder of } z_{i-1} \text{ with respect to division by } z_i \text{ if } z_i \neq 0, \\ 0 \text{ otherwise.} \end{cases}$$

Then $z_i = 0$ for almost all $i \in \mathbb{N}$. Furthermore, $z_n = \gcd(x,y)$, where $n \in \mathbb{N}$ is the smallest index satisfying $z_{n+1} = 0$.

Proof. Let $\delta\colon R-\{0\} \longrightarrow \mathbb{N}$ be the Euclidean function considered on R and fix an index $i > 0$ such that $z_i \neq 0$. According to the definition of the sequence $z_0, z_1, \ldots$, there is an equation of type

$$z_{i-1} = q_i z_i + z_{i+1},$$

where $\delta(z_{i+1}) < \delta(z_i)$ or $z_{i+1} = 0$. Therefore, the sequence of integers $\delta(z_i)$ is strictly decreasing for $i > 0$, at least as long as z_i is nonzero, and thus $\delta(z_i)$ is defined. In particular, z_i will be nonzero only for finitely many indices $i \in \mathbb{N}$, and there is a smallest index $n \in \mathbb{N}$ such that $z_{n+1} = 0$. Then $n > 0$, since $z_0 \neq 0 \neq z_1$. Now consider the equations

$$\begin{aligned}
&(\mathrm{E}_0) && z_0 = q_1 z_1 + z_2, \\
&\vdots && \vdots \\
&(\mathrm{E}_{n-2}) && z_{n-2} = q_{n-1} z_{n-1} + z_n, \\
&(\mathrm{E}_{n-1}) && z_{n-1} = q_n z_n.
\end{aligned}$$

We get $z_n \mid z_{n-1}$ from (E_{n-1}), and furthermore $z_n \mid z_{n-2}$ from (E_{n-2}), and so forth, until we end up with $z_n \mid z_1$ and $z_n \mid z_0$. In particular, z_n is a common divisor of x and y. If $a \in R$ is another common divisor of x and y, we get $a \mid z_2$ from (E_0), and furthermore $a \mid z_3$ from (E_1), and so forth, until we finitely obtain $a \mid z_n$. Hence, z_n is the greatest common divisor of x and y, as claimed. □

Not only does the Euclidean algorithm make it possible to determine the greatest common divisor d of two elements x, y in a Euclidean domain; beyond this, it yields a representation of this divisor as a linear combination $d = ax+by$. Indeed, in the above proof we get from (E_{n-2}) a representation of $d = z_n$ as a linear combination of z_{n-2}, z_{n-1}, and furthermore using (E_{n-3}), as a linear combination of z_{n-3}, z_{n-2}. Continuing like this, (E_0) finally leads to a representation

of d as a linear combination of $x = z_0$ and $y = z_1$. Let us add that such a representation is needed when one is explicitly solving simultaneous congruences; see 2.3/13 and the explanations following it. Of course, the mere existence of such solutions was already established in Proposition 13 within the context of general principal ideal domains.

Finally, let us refer to some applications of the results dealt with in the present section. We can once more conclude from 2.3/8 and Proposition 6 that the residue class ring $\mathbb{Z}/p\mathbb{Z}$ for an integer $p \in \mathbb{Z}$, $p > 0$, is a field if and only if p is a prime number. Likewise, for a field K and a polynomial $f \in K[X]$, the residue class ring $L = K[X]/(f)$ modulo the principal ideal generated by f is a field if and only if f is irreducible. Furthermore, it is easily seen (cf. the proof of 3.4/1) that the residue class of X in L becomes a zero of f. Just view K as a subfield of L via the canonical homomorphism $K \longrightarrow L$ (which is injective by 2.3/3) and, similarly, f as a polynomial with coefficients in L. Later, in 3.4/1, we will use this construction, which goes back to L. Kronecker, in order to construct, for a given polynomial $f \in K[X]-K$ that does not admit zeros in K, an extension field L such that f acquires a zero in L. To give some simple examples, consider the canonical isomorphism

$$\mathbb{R}[X]/(X^2+1) \simeq \mathbb{C},$$

obtained by applying the fundamental theorem on homomorphisms to the substitution homomorphism

$$\mathbb{R}[X] \longrightarrow \mathbb{C}, \qquad \sum a_n X^n \longmapsto \sum a_n i^n,$$

which maps X to the complex number i. In a similar way one shows that

$$\mathbb{R}[X]/(X-a) \simeq \mathbb{R}$$

for arbitrary $a \in \mathbb{R}$.

Exercises

1. *Determine all rings R such that the polynomial ring $R[X]$ is a principal ideal domain.*
2. *For principal ideal domains, we can conclude from Proposition* 13 *that the greatest common divisor as well as the least common multiple of two elements can be characterized in terms of ideal theory. Check whether the same is true for unique factorization domains.*
3. Prove that the subring $R = \mathbb{Z} + \sqrt{-5} \cdot \mathbb{Z} \subset \mathbb{C}$ is not a unique factorization domain. To do this, consider the factorizations $6 = 2 \cdot 3 = (1+\sqrt{-5}) \cdot (1-\sqrt{-5})$ and show that the elements 2, 3, $(1+\sqrt{-5})$, $(1-\sqrt{-5})$ are irreducible and pairwise nonassociated. Check whether the stated elements are prime.
4. Let K be a field and $R = K[X][Y]/(X^2 - Y^3)$ the integral domain from Exercise 8 in 2.3. Show that the residue classes $\overline{X}$ and $\overline{Y}$ of $X, Y \in K[X][Y]$ are irreducible, but not prime.

5. Let G be a cyclic group of finite order. Show for elements $a, b \in G$ that the subgroup generated by a and b in G is of order $\operatorname{lcm}(\operatorname{ord} a, \operatorname{ord} b)$.
6. Show that $2 = (1+i)(1-i)$ is the prime factorization of 2 in $\mathbb{Z}[i]$.
7. Use the Euclidean algorithm to determine the greatest common divisor of the following polynomials in $\mathbb{Q}[X]$:
$$f = X^3 + X^2 + X - 3, \qquad g = X^6 - X^5 + 6X^2 - 13X + 7.$$
8. Determine all irreducible polynomials of degree ≤ 3 of the polynomial ring $\mathbb{F}_2[X]$, where $\mathbb{F}_2$ is the field consisting of two elements.
9. For a prime number $p \in \mathbb{N}$, consider the following subset of the field $\mathbb{Q}$ of rational numbers:
$$\mathbb{Z}_p := \{0\} \cup \Big\{ \frac{x}{y} \in \mathbb{Q}\,;\, x, y \in \mathbb{Z} - \{0\} \text{ such that } \nu_p(x) - \nu_p(y) \geq 0 \Big\}.$$
Show that $\mathbb{Z}_p$ is a subring of $\mathbb{Q}$, a principal ideal domain, but not a field. Specify all units as well as all prime elements of $\mathbb{Z}_p$.
10. Show that a ring R is Noetherian (in the sense that every ascending chain of ideals $\mathfrak{a}_1 \subset \mathfrak{a}_2 \subset \ldots \subset R$ becomes stationary) if and only if every ideal in R admits a finite system of generators.

2.5 Polynomial Rings in Several Variables

In 2.1 we introduced the polynomial ring $R[X]$ in one variable X over a ring R. Iterating the construction, we could define the polynomial ring in n variables $X_1, \ldots, X_n$ over R:
$$R[X_1, \ldots, X_n] := \big(\ldots \big((R[X_1])[X_2] \big) \ldots \big)[X_n].$$
However, a more elegant way is to generalize the definition of 2.1 such that it applies to the case of several variables. In fact, we will define for a commutative monoid M the "polynomial ring" $R[M]$ in such a way that we can interpret M as the (multiplicative) monoid of all "monomials" in $R[M]$. In doing so, the polynomial ring $R[X]$ in one variable X is obtained by taking $M = \mathbb{N}$, the polynomial ring $R[X_1, \ldots, X_n]$ in n variables $X_1, \ldots, X_n$ by taking $M = \mathbb{N}^n$, and the polynomial ring $R[\mathfrak{X}]$ in a family of variables $\mathfrak{X} = (X_i)_{i \in I}$, indexed by an arbitrary index set I, by taking $M = \mathbb{N}^{(I)}$. In each case we take on $\mathbb{N}$, $\mathbb{N}^n$, and $\mathbb{N}^{(I)}$ the (componentwise) addition as the law of composition.

In the following let M be an arbitrary commutative monoid whose law of composition is written *additively*. Then we define $R[M]$ by
$$R[M] = R^{(M)} = \big\{ (a_\mu)_{\mu \in M}\,;\, a_\mu \in R,\ a_\mu = 0 \text{ for almost all } \mu \big\},$$
together with the laws of composition given by
$$(a_\mu)_{\mu \in M} + (b_\mu)_{\mu \in M} := (a_\mu + b_\mu)_{\mu \in M}, \qquad (a_\mu)_{\mu \in M} \cdot (b_\mu)_{\mu \in M} := (c_\mu)_{\mu \in M},$$

where

$$c_\mu = \sum_{\lambda+\nu=\mu} a_\lambda \cdot b_\nu.$$

It is verified without difficulty that $R[M]$ becomes a ring under these laws. In particular, for the monoid of natural numbers $M = \mathbb{N}$ we rediscover the polynomial ring $R[X]$ in one variable X, as defined in 2.1. However, also in the remaining cases we can use a polynomial notation for the elements of $R[M]$. Indeed, for $\mu \in M$ consider $X^\mu := (\delta_{\mu,\lambda})_{\lambda\in M}$ as an element of $R[M]$, where $\delta_{\mu,\lambda}$ is Kronecker's symbol, which is given by $\delta_{\mu,\lambda} = 1$ for $\mu = \lambda$ and $\delta_{\mu,\lambda} = 0$ for $\mu \neq \lambda$. We call X^μ the *monomial* in $R[M]$ that is attached to μ. Using this notation, the elements of $R[M]$ can be written as sums of type $\sum_{\mu\in M} a_\mu X^\mu$, where the coefficients $a_\mu \in R$ are unique and, of course, zero for almost all $\mu \in M$. Just as for polynomials in one variable X, addition and multiplication are expressed by the well-known formulas

$$\sum_{\mu\in M} a_\mu X^\mu + \sum_{\mu\in M} b_\mu X^\mu = \sum_{\mu\in M} (a_\mu + b_\mu) X^\mu,$$
$$\sum_{\mu\in M} a_\mu X^\mu \cdot \sum_{\mu\in M} b_\mu X^\mu = \sum_{\mu\in M} \Big(\sum_{\lambda+\nu=\mu} a_\lambda \cdot b_\nu \Big) X^\mu.$$

As usual, the zero polynomial $0 = \sum_{\mu\in M} 0 \cdot X^\mu$ serves as the zero element, and likewise, X^0 serves as the unit element of $R[M]$, where the exponent 0 indicates the neutral element of the monoid M. Also note that R is naturally a subring of $R[M]$. Just identify the elements $a \in R$ with their corresponding "constant" polynomials aX^0. The polynomial ring $R[M]$ admits the following universal property:

Proposition 1. *Let $\varphi\colon R \longrightarrow R'$ be a ring homomorphism and $\sigma\colon M \longrightarrow R'$ a monoid homomorphism, where we view R' as a monoid with respect to the ring multiplication. Then there exists a unique ring homomorphism $\Phi\colon R[M] \longrightarrow R'$ satisfying $\Phi|_R = \varphi$ and $\Phi(X^\mu) = \sigma(\mu)$ for all $\mu \in M$.*

Proof. To verify the uniqueness assertion, consider an element $\sum_{\mu\in M} a_\mu X^\mu$ in $R[M]$. If there exists a homomorphism Φ satisfying the stated conditions, we must have

$$\Phi\Big(\sum a_\mu X^\mu\Big) = \sum \Phi(a_\mu X^\mu) = \sum \Phi(a_\mu)\Phi(X^\mu) = \sum \varphi(a_\mu)\sigma(\mu).$$

Conversely, to establish the existence we can define Φ via the preceding equation. The properties of a ring homomorphism are easily checked; just use the facts that φ is a ring homomorphism and σ a monoid homomorphism. □

The property of polynomial rings proved in Proposition 1 is called a *universal property*, since $R[M]$ thereby appears, so to speak, as a master object from which all similar constructs combining coefficients in R and monomials in M

are derived via homomorphisms. In particular, the universal property uniquely characterizes $R[M]$ up to canonical isomorphism. In more detail, this means the following. Start out from a ring extension $R \subset S$ and a monoid homomorphism $\iota\colon M \longrightarrow S$, where S is viewed as a monoid under the ring multiplication, and assume that the mapping property stated in Proposition 1 is given, i.e., that for each ring homomorphism $\psi\colon R \longrightarrow R'$ and each monoid homomorphism $\tau\colon M \longrightarrow R'$ with R' as multiplicative monoid, there is a unique ring homomorphism $\Psi\colon S \longrightarrow R'$ such that $\Psi|_R = \psi$ and $\Psi \circ \iota = \tau$. Then the extensions $R \subset R[M]$ and $R \subset S$ are canonically isomorphic.

We want to briefly justify this, using the general argument that applies to any universal property. If we consider the homomorphisms $R \hookrightarrow S$ and $\iota\colon M \longrightarrow S$, the universal property of $R[M]$ yields a ring homomorphism $\Phi\colon R[M] \longrightarrow S$ that extends the identity on R and furthermore, satisfies $\Phi(X^\mu) = \iota(\mu)$ for all $\mu \in M$. On the other hand, by the universal property of S, the monoid homomorphism $M \longrightarrow R[M]$, $\mu \longmapsto X^\mu$, leads to a ring homomorphism $\Psi\colon S \longrightarrow R[M]$ extending the identity on R and satisfying $\Psi(\iota(\mu)) = X^\mu$ for all $\mu \in M$. Hence, $\Phi \circ \Psi$ and the identity map constitute two ring homomorphisms $S \longrightarrow S$ extending the identity on R and leaving $\iota(\mu)$ fixed for all $\mu \in M$. Now the uniqueness part of the universal mapping property on S yields $\Phi \circ \Psi = \mathrm{id}$ and likewise $\Psi \circ \Phi = \mathrm{id}$, using the same property on $R[M]$. It follows that Φ and Ψ are isomorphisms.

Now we want to consider the cases $M = \mathbb{N}^n$ and $M = \mathbb{N}^{(I)}$, thereby looking at polynomial rings in the stricter sense of the word. First let $M = \mathbb{N}^n$. We define the ith "variable" X_i, $1 \leq i \leq n$, by $X^{(0,\ldots,0,1,0,\ldots,0)}$, where the symbol 1 of the exponent is placed at position i. Then, for $\mu = (\mu_1, \ldots, \mu_n) \in \mathbb{N}^n$ we have $X^\mu = X_1^{\mu_1} \ldots X_n^{\mu_n}$, and the elements of $R[\mathbb{N}^n]$ can be written in a more explicit way as sums:

$$\sum_{(\mu_1,\ldots,\mu_n)\in\mathbb{N}^n} a_{\mu_1,\ldots,\mu_n} X_1^{\mu_1} \ldots X_n^{\mu_n},$$

where the coefficients $a_{\mu_1,\ldots,\mu_n} \in R$ are unique and, of course, zero for almost all indices $(\mu_1, \ldots, \mu_n)$. Instead of $R[\mathbb{N}^n]$ we use the notation $R[X_1, \ldots, X_n]$ or $R[X]$ and view $X = (X_1, \ldots, X_n)$ as a family of variables. Similarly, we proceed with monoids of type $M = \mathbb{N}^{(I)}$, where I is an arbitrary index set. Let ε_i for $i \in I$ specify the element of $\mathbb{N}^{(I)}$ whose components are all 0, except for 1 at position i. Then, setting $X_i = X^{\varepsilon_i}$, $i \in I$, we have $X^\mu = \prod_{i\in I} X_i^{\mu_i}$ for $\mu = (\mu_i)_{i\in I} \in \mathbb{N}^{(I)}$. Note that almost all factors of such a product are trivial, so that it is actually a finite product. In particular, the elements of $R[\mathbb{N}^{(I)}]$ can be written as sums of type

$$\sum_{\mu\in\mathbb{N}^{(I)}} a_\mu \prod_{i\in I} X_i^{\mu_i}$$

with coefficients $a_\mu \in R$ that are unique. Instead of $R[\mathbb{N}^{(I)}]$ we also use the notation $R[X_i\,;\, i \in I]$ or $R[\mathfrak{X}]$ for $\mathfrak{X} = (X_i)_{i\in I}$. The elements of $R[\mathfrak{X}]$ are polynomials in *finitely* many variables $X_{i_1}, \ldots, X_{i_n}$, and we can view $R[\mathfrak{X}]$ as the union of all subrings of type $R[X_{i_1}, \ldots, X_{i_n}]$, where the set $\{i_1, \ldots, i_n\}$

varies over all finite subsets of I. In particular, computations involving only finitely many elements of $R[\mathfrak{X}]$ can always be carried out in a polynomial ring in finitely many variables.

For simplicity, we will restrict ourselves in the following to the case of polynomial rings of type $R[X_1, \ldots, X_n]$, although the results we prove below remain valid mutatis mutandis also for arbitrary sets of variables. Let us add that basically, polynomial rings in infinitely many variables will be used only for the construction of algebraically closed fields in 3.4. Furthermore, observe, either by direct computation or by applying Proposition 1 (see also Exercise 3), that there are canonical isomorphisms of type

$$R[X_1, \ldots, X_n] \simeq \big(R[X_1, \ldots, X_{n-1}]\big)[X_n]$$

for $n > 0$, using the convention $R[X_1, \ldots, X_{n-1}] = R$ for $n = 1$. These isomorphisms allow one in many cases to inductively reduce problems on polynomials in several variables to the case of one variable.

Proposition 2. *If R is an integral domain, then for finitely many variables $X_1, \ldots, X_n$, the polynomial ring $R[X_1, \ldots, X_n]$ is also an integral domain.*

Proof. We have already seen in 2.1/3 that the proposition is true in the case of one variable. But then, using the isomorphism

$$R[X_1, \ldots, X_n] \simeq \big(R[X_1, \ldots, X_{n-1}]\big)[X_n],$$

the general case follows by induction.

Alternatively, one can use a direct argument to show that the product of two nonzero polynomials

$$f = \sum a_\mu X^\mu, \quad g = \sum b_\nu X^\nu \quad \in R[X_1, \ldots, X_n]$$

is nonzero if R is an integral domain. Indeed, introduce the lexicographic order on $\mathbb{N}^n$, i.e., we write $\mu < \mu'$ for indices

$$\mu = (\mu_1, \ldots, \mu_n), \quad \mu' = (\mu'_1, \ldots, \mu'_n) \quad \in \mathbb{N}^n,$$

if for some i, $1 \leq i \leq n$, we have

$$\mu_1 = \mu'_1, \quad \ldots, \quad \mu_{i-1} = \mu'_{i-1}, \quad \mu_i < \mu'_i.$$

Now choose $\overline{\mu} \in \mathbb{N}$ maximal (with respect to the lexicographic order) among all μ such that $a_\mu \neq 0$, as well as $\overline{\nu}$ maximal such that $b_\nu \neq 0$. Then the coefficient of the monomial $X^{\overline{\mu}+\overline{\nu}}$ in fg equals $a_{\overline{\mu}} b_{\overline{\nu}}$. In particular, if R is an integral domain, we have $a_{\overline{\mu}} b_{\overline{\nu}} \neq 0$ and hence $fg \neq 0$. □

Given an index $\mu = (\mu_1, \ldots, \mu_n) \in \mathbb{N}^n$, we write $|\mu| := \mu_1 + \ldots + \mu_n$ and call this number the *degree* of μ. Furthermore, for a polynomial $f = \sum a_\mu X^\mu$

in $R[X_1, \ldots, X_n]$, we call $f_i := \sum_{|\mu|=i} a_\mu X^\mu$ for $i \in \mathbb{N}$ the *homogeneous part of f of degree i*. In particular, f may be interpreted as the sum of its homogeneous parts, i.e., $f = \sum_{i=0}^{\infty} f_i$. We call f *homogeneous* if f equals one of its homogeneous parts or, more precisely, *homogeneous of degree i* if $f = f_i$. A homogeneous polynomial $f \neq 0$ is always homogeneous of a unique degree $i \geq 0$, whereas the zero polynomial is homogeneous of *every* degree $i \geq 0$. Furthermore,

$$\deg f = \max\{i \in \mathbb{N}\,;\, f_i \neq 0\} = \max\{|\mu|\,;\, a_\mu \neq 0\}$$

is called the *total degree* of f, with the convention that $\deg f := -\infty$ for $f = 0$. Note that the total degree of a polynomial in a single variable coincides with the degree as defined in 2.1. Moreover, there is the following analogue of 2.1/2:

Proposition 3. *Let $f, g \in R[X_1, \ldots, X_n]$ be polynomials with coefficients in a ring R. Then*

$$\deg(f+g) \leq \max(\deg f, \deg g),$$
$$\deg(f \cdot g) \leq \deg f + \deg g,$$

and $\deg(f \cdot g) = \deg f + \deg g$ if R is an integral domain.

Proof. The estimate for $\deg(f+g)$ becomes clear if we decompose polynomials in $R[X_1, \ldots, X_n]$ into the sums of their homogeneous parts. Furthermore, if $\deg f = r$ and $\deg g = s$, and if $f = \sum_{i=0}^{r} f_i$, $g = \sum_{i=0}^{s} g_i$ are the decompositions into homogeneous parts, we obtain, assuming $r, s \geq 0$,

$$f \cdot g = f_r \cdot g_s + (\text{homogeneous terms of degree } < r+s),$$

where $f_r \cdot g_s$ equals the homogeneous part of degree $r+s$ in $f \cdot g$. This shows that $\deg(f \cdot g) \leq \deg f + \deg g$. If R is an integral domain, then $f_r, g_s \neq 0$ implies $f_r g_s \neq 0$, due to Proposition 2, so that the degree of fg is $r+s$. □

Corollary 4. *If R is an integral domain, then*

$$\big(R[X_1, \ldots, X_n]\big)^* = R^*.$$

Next, we want to adapt the universal property of polynomial rings, which uniquely characterizes these rings up to canonical isomorphism, especially to polynomial rings of type $R[X_1, \ldots, X_n]$. Since a monoid homomorphism $\sigma\colon \mathbb{N}^n \longrightarrow R'$ is uniquely determined by the images of the canonical "generators" of $\mathbb{N}^n$, namely of the elements of type $(0, \ldots, 0, 1, 0, \ldots, 0)$, we can derive the following version of Proposition 1:

Proposition 5. *Let $\varphi\colon R \longrightarrow R'$ be a ring homomorphism and consider finitely many elements $x_1, \ldots, x_n \in R'$. Then there exists a unique ring homomorphism $\Phi\colon R[X_1, \ldots, X_n] \longrightarrow R'$ satisfying $\Phi|_R = \varphi$ and $\Phi(X_i) = x_i$ for $i = 1, \ldots, n$.*

Writing $x = (x_1, \dots, x_n)$ and $x^\mu = x_1^{\mu_1} \dots x_n^{\mu_n}$ for $\mu \in \mathbb{N}^n$ in the situation of the preceding proposition, we can describe the homomorphism Φ by

$$\Phi\colon R[X_1, \dots, X_n] \longrightarrow R', \qquad \sum a_\mu X^\mu \longmapsto \sum \varphi(a_\mu) x^\mu,$$

similarly as in the case of a single variable. We call Φ a *substitution homomorphism*, since the tuple x is substituted for X. In particular, if R is a subring of R' and $\varphi\colon R \hookrightarrow R'$ the canonical inclusion, then the image under Φ of a polynomial $f = \sum a_\mu X^\mu \in R[X_1, \dots, X_n]$ will generally be denoted by $f(x) = \sum a_\mu x^\mu$. If $f(x) = 0$, we call x a *zero* of f. Moreover, we use the notation

$$R[x] := \Phi\big(R[X_1, \dots, X_n]\big) = \Big\{ \sum a_\mu x^\mu \,;\, a_\mu \in R,\, a_\mu = 0 \text{ for almost all } \mu \Big\}$$

for the image of $R[X_1, \dots, X_n]$ with respect to Φ. Then $R[x]$, or in more explicit terms $R[x_1, \dots, x_n]$, is the smallest subring of R' that contains R and all components $x_1, \dots, x_n$ of x. Suggestively, we call $R[x]$ the ring of polynomials in x (or better, of all polynomial expressions in x) with coefficients in R.

Substitution homomorphisms will play an important role later on. As a typical example, let us mention the notion of *transcendence*.

Definition 6. *Let $R \subset R'$ be a ring extension and $x = (x_1, \dots, x_n)$ a system of elements in R'. Then x is called* algebraically independent *or* transcendental *over R if for a system of variables $X = (X_1, \dots, X_n)$ the ring homomorphism $R[X] \longrightarrow R'$, $f \longmapsto f(x)$, is injective and thus induces an isomorphism $R[X] \xrightarrow{\sim} R[x]$. Otherwise, x is called* algebraically dependent.

In particular, any system $x = (x_1, \dots, x_n)$ that is transcendental over R, admits the same properties, as does a system of variables. For example, we have already pointed out in the introduction that each of the numbers e and $\pi \in \mathbb{R}$, well-known from analysis, is transcendental over $\mathbb{Q}$; proofs for this fact go back to Ch. Hermite [8] and F. Lindemann [13].

Finally, let us refer to the *reduction of coefficients* of polynomials, a process that, formally speaking, belongs to the subject of substitution homomorphisms as well. If $\mathfrak{a} \subset R$ is an ideal and $\varphi\colon R \longrightarrow R/\mathfrak{a}$ the canonical homomorphism, we can apply Proposition 5 and consider the homomorphism $\Phi\colon R[X] \longrightarrow (R/\mathfrak{a})[X]$ that extends φ and maps X to X. We say that Φ reduces the coefficients of polynomials in $R[X]$ modulo the ideal $\mathfrak{a}$. For example, the homomorphism $\mathbb{Z}[X] \longrightarrow \mathbb{Z}/(p)[X]$ for a prime number p transforms polynomials with integer coefficients to polynomials with coefficients in the finite field $\mathbb{F}_p = \mathbb{Z}/(p)$.

Exercises

1. *The polynomial ring $R[M]$ with coefficients in a ring R has been defined for commutative monoids M. If we want to extend the definition to not necessarily commutative monoids, which new phenomena have to be paid attention to?*

2. *Examine how far the results of the present section on polynomial rings in finitely many variables $R[X_1, \dots, X_n]$ can be generalized to polynomial rings in arbitrary sets of variables $R[\mathfrak{X}]$.*

3. *For two monoids M, M', consider their Cartesian product $M \times M'$ as a monoid with componentwise law of composition. Show that there is a canonical ring isomorphism $R[M][M'] \xrightarrow{\sim} R[M \times M']$.*

4. Let R be a ring. Consider $\mathbb{Z}$, as well as $\mathbb{Z}/m\mathbb{Z}$ for $m > 0$, as monoids with respect to addition and show that

$$R[\mathbb{Z}] \simeq R[X, Y]/(1 - XY), \qquad R[\mathbb{Z}/m\mathbb{Z}] \simeq R[X]/(X^m - 1).$$

5. Let K be a field and $f \in K[X_1, \dots, X_n]$ a homogeneous polynomial of total degree $d > 0$. Show that for every prime factorization $f = p_1 \dots p_r$, the factors p_i are homogeneous as well.

6. Consider the polynomial ring $R[X_1, \dots, X_n]$ in n variables over a ring $R \neq 0$ and show that the number of monomials of total degree $d \in \mathbb{N}$ in $R[X_1, \dots, X_n]$ is

$$\binom{n + d - 1}{n - 1}.$$

7. Let K be a field and $\varphi\colon K[X_1, \dots, X_m] \longrightarrow K[X_1, \dots, X_n]$ a ring isomorphism such that $\varphi|_K = \mathrm{id}_K$. Show that $m = n$.

2.6 Zeros of Polynomials

Let K be a field and $f \in K[X]$ a nonzero polynomial in a variable X. Then, if $\alpha \in K$ is a zero of f, the linear polynomial $X - \alpha$ divides f. Indeed, Euclid's division of f by $X - \alpha$ leads to an equation

$$f = q \cdot (X - \alpha) + r,$$

where $\deg r < 1$ and hence $r \in K$. Furthermore, substituting X by α yields $r = 0$. We say that α is a *zero of multiplicity* r if $X - \alpha$ appears in the prime factorization of f with a power precisely r. Therefore, looking at degrees, we can assert the following:

Proposition 1. *Let K be a field and $f \in K[X]$ a polynomial of degree $n \geq 0$. Then, counting multiplicities, f admits at most n zeros in K. The number of these zeros is precisely n if and only if all factors of the prime factorization of f in $K[X]$ are linear.*

In particular, we thereby see that a polynomial of degree $\leq n$ for some $n \in \mathbb{N}$ equals the zero polynomial as soon as it has more than n zeros. Therefore, if K is an infinite field, the equality $f = 0$ (zero polynomial) is equivalent to the

fact that $f(\alpha) = 0$ for all $\alpha \in K$ (resp. for all α from a given infinite subset of K). On the other hand, for a finite field $\mathbb{F}$, the polynomial

$$f = \prod_{a \in \mathbb{F}} (X - a) \in \mathbb{F}[X]$$

is nonzero and satisfies $f(\alpha) = 0$ for all $\alpha \in \mathbb{F}$.

There is a simple criterion for the existence of multiple zeros. To formulate it consider the map

$$D\colon K[X] \longrightarrow K[X], \qquad \sum_{i=0}^{n} c_i X^i \longmapsto \sum_{i=1}^{n} i c_i X^{i-1},$$

which is defined just like the usual differentiation of real polynomials (interpret ic_i as the i-fold sum of c_i with itself). Note that D is not a ring homomorphism; it is a so-called *derivation*, i.e., D satisfies the following rules for $a, b \in K$, $f, g \in K[X]$:

$$D(af + bg) = aD(f) + bD(g), \qquad D(fg) = fD(g) + gD(f).$$

We will mostly write f' instead of Df, calling this the *first derivative of* f; see also 7.4.

Proposition 2. *Let $f \in K[X]$, $f \neq 0$, be a polynomial with coefficients in a field K. A zero $\alpha \in K$ of f is a multiple zero (i.e., a zero of multiplicity ≥ 2) if and only if $(f')(\alpha) = 0$.*

Proof. If α is a zero of f of multiplicity $r \geq 1$, then there is a factorization of type $f = (X - \alpha)^r g$ for some $g \in K[X]$ satisfying $g(\alpha) \neq 0$. Since

$$f' = (X - \alpha)^r g' + r(X - \alpha)^{r-1} g,$$

we see that $(f')(\alpha) = 0$ is equivalent to $r \geq 2$. □

Corollary 3. *An element $\alpha \in K$ is a multiple zero of a nonzero polynomial $f \in K[X]$ if and only if α is a zero of* $\gcd(f, f')$.

For example, if p is a prime number, the polynomial $f = X^p - X \in \mathbb{F}_p[X]$ does not admit multiple zeros. Indeed, we have $f' = -1$, since the p-fold sum $p \cdot 1$ of the unit element $1 \in \mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ is zero.

Exercises

1. *Let K be a field consisting of infinitely many elements and $f \in K[X_1, \ldots, X_n]$ a polynomial vanishing at all points of K^n. Show that $f = 0$, i.e., f is the zero polynomial.*

2. Let K be a field. Show for $n \in \mathbb{N}$, $n > 1$, that the multiplicative group K^* contains at most $n-1$ elements of order n.
3. Let K be a field. Show that the polynomial ring $K[X]$ contains infinitely many monic prime polynomials. Furthermore, if each nonconstant polynomial from $K[X]$ admits at least one zero in K, show that K consists of infinitely many elements.
4. Let K be a field and let $f = X^3 + aX + b \in K[X]$ be a polynomial admitting a factorization into linear factors in $K[X]$. Show that the zeros of f are distinct if and only if the "discriminant" $\Delta = -4a^3 - 27b^2$ is nonzero.

2.7 A Theorem of Gauss

The purpose of the present section is to prove the following basic result on unique factorization domains:

Proposition 1 (Gauss). *Let R be a unique factorization domain. Then the polynomial ring in one variable $R[X]$ is also a unique factorization domain.*

There are some immediate consequences:

Corollary 2. *If R is a unique factorization domain, then the same is true for the polynomial ring $R[X_1, \ldots, X_n]$.*

Corollary 3. *For any field K, the polynomial ring $K[X_1, \ldots, X_n]$ is a unique factorization domain.*

In particular, there exist unique factorization domains that are not principal ideal domains; just look at the polynomial ring $K[X, Y]$ in two variables X, Y over a field K, or at the polynomial ring in one variable $\mathbb{Z}[X]$ over the ring of integers $\mathbb{Z}$.

For the proof of Proposition 1 we need some preparations. To begin with, we construct the *field of fractions* $Q(R)$ of an integral domain R, taking the construction of rational numbers in terms of fractions of integers as a guide. Therefore, consider the set of pairs

$$M = \big\{(a, b)\,;\, a \in R,\ b \in R - \{0\}\big\}$$

and define an equivalence relation "$\sim$" on it by setting

$$(a, b) \sim (a', b') \quad \Longleftrightarrow \quad ab' = a'b.$$

The conditions of an equivalence relation are easily checked, namely

reflexivity: $(a, b) \sim (a, b)$ for all $(a, b) \in M$,
symmetry: $(a, b) \sim (a', b') \implies (a', b') \sim (a, b)$,
transitivity: $(a, b) \sim (a', b'),\ (a', b') \sim (a'', b'') \implies (a, b) \sim (a'', b'')$.

For example, to justify the transitivity we argue as follows:

$$\begin{aligned}(a,b)\sim(a',b') \quad &\Longrightarrow \quad ab'=a'b \quad \Longrightarrow \quad ab'b''=a'bb'',\\ (a',b')\sim(a'',b'') \quad &\Longrightarrow \quad a'b''=a''b' \quad \Longrightarrow \quad a'bb''=a''bb',\end{aligned}$$

so that

$$(a,b)\sim(a',b'),\ (a',b')\sim(a'',b'') \quad \Longrightarrow \quad ab'b''=a''bb'.$$

However, the last equation yields $ab''=a''b$, and thus $(a,b)\sim(a'',b'')$, since R is an integral domain.

Now observe that the equivalence relation "$\sim$" defines a partition of M into equivalence classes; let

$$Q(R)=M/\sim$$

be the corresponding set of classes. The equivalence class attached to a pair $(a,b)\in M$ is denoted by $\frac{a}{b}\in Q(R)$, using the notion of fractions. Observe that due to the definition of "$\sim$", we have

$$\frac{a}{b}=\frac{a'}{b'} \quad \Longleftrightarrow \quad ab'=a'b.$$

It is easily checked that $Q(R)$ is a field under the addition and multiplication of fractions

$$\frac{a}{b}+\frac{a'}{b'}=\frac{ab'+a'b}{bb'}, \qquad \frac{a}{b}\cdot\frac{a'}{b'}=\frac{aa'}{bb'},$$

where one shows as usual that the laws "$+$" and "$\cdot$" are well defined. We call $Q(R)$ the *field of fractions* of R. Furthermore,

$$R\longrightarrow Q(R), \qquad a\longmapsto\frac{a}{1},$$

is an injective ring homomorphism, which allows us to view R as a subring of $Q(R)$. For example, taking $R=\mathbb{Z}$, the associated field of fractions is $Q(\mathbb{Z})=\mathbb{Q}$, the field of rational numbers. If K is a field and X a variable, the field of fractions $Q(K[X])$ is called the *field of rational functions* in the variable X with coefficients in K, and is denoted by $K(X)=Q(K[X])$. Analogously, one considers rational function fields $K(X_1,\ldots,X_n)=Q(K[X_1,\ldots,X_n])$ in finitely many variables $X_1,\ldots,X_n$ and, more generally, function fields $K(\mathfrak{X})=Q(K[\mathfrak{X}])$ in a system of variables $\mathfrak{X}=(X_i)_{i\in I}$.

The above construction of fields of fractions attached to an integral domain extends to a more general setting. Fixing a ring R (that may have nontrivial zero divisors), consider a multiplicative system $S\subset R$, i.e., a multiplicative submonoid of R. Then, similarly as before, we can define the *ring of fractions* (in general it will not be a field)

$$S^{-1}R=\Big\{\frac{a}{s}\,;\ a\in R,\ s\in S\Big\},$$

where, due to possible nontrivial zero divisors in R, one works relative to the following equivalence relation:

$$\frac{a}{s} = \frac{a'}{s'} \quad \Longleftrightarrow \quad \text{there exists some } s'' \in S \text{ such that } as's'' = a'ss''.$$

We use R_S as a shorthand notation for $S^{-1}R$ and call this ring the *localization* of R by S. However, observe that the canonical map $R \longrightarrow S^{-1}R$ might not be injective, since it might have a nontrivial kernel. The kernel consists of all elements $a \in R$ such that there exists an element $s \in S$ satisfying $as = 0$. If R is an integral domain (this is the main case to be considered in the following), we have $Q(R) = S^{-1}R$, where $S := R - \{0\}$.

Remark 4. *Let R be a unique factorization domain, and P a system of representatives of the prime elements in R. Then every fraction $\frac{a}{b} \in Q(R)^*$ admits a unique factorization*

$$\frac{a}{b} = \varepsilon \prod_{p \in P} p^{\nu_p},$$

where $\varepsilon \in R^*$ and $\nu_p \in \mathbb{Z}$ with $\nu_p = 0$ for almost all p. In particular, $\frac{a}{b}$ belongs to R if and only if $\nu_p \geq 0$ for all $p \in P$.

Proof. Using the prime factorizations of a and b, the existence of the stated factorization follows. Furthermore, the uniqueness is a consequence of the uniqueness of prime factorizations in R, at least if one considers factorizations of $\frac{a}{b}$ satisfying $\nu_p \geq 0$ for all p. However, we can reduce to this case by multiplying $\frac{a}{b}$ by suitable nonzero elements of R. □

If we have $x = \frac{a}{b}$ in the situation of Remark 4, we write more explicitly $\nu_p(x)$ instead of ν_p, adding $\nu_p(0) := \infty$ as a convention. Then the uniqueness assertion of Remark 4 implies

$$\nu_p(xy) = \nu_p(x) + \nu_p(y)$$

for $x, y \in Q(R)$. Moreover, considering polynomials $f = \sum a_i X^i \in Q(R)[X]$ in one variable X, we set

$$\nu_p(f) := \min_i \nu_p(a_i),$$

where $f = 0$ is equivalent to $\nu_p(f) = \infty$ (for some, and hence all, $p \in P$). Also note that f belongs to $R[X]$ as soon as $\nu_p(f) \geq 0$ for all $p \in P$.

In order to show that the polynomial ring over a unique factorization domain is of the same type, we need a key fact about the function $\nu_p(\cdot)$:

Lemma 5 (Gauss). *Let R be a unique factorization domain and $p \in R$ a prime element. Then $\nu_p(\cdot)$ satisfies the following relation for elements $f, g \in Q(R)[X]$:*

$$\nu_p(fg) = \nu_p(f) + \nu_p(g).$$

Proof. As mentioned above, the stated relation holds for constant polynomials, i.e., for $f, g \in Q(R)$ and hence also for $f \in Q(R)$ and arbitrary polynomials $g \in Q(R)[X]$.

To deal with the general case we may assume $f, g \neq 0$. Furthermore, due to the preceding consideration, we are allowed to multiply f and g by constants from $Q(R)^*$. In particular, representing the coefficients of f as fractions of elements in R, we can multiply f by the least common multiple of all denominators. Proceeding in the same way with g, we are reduced to the case that f and g are polynomials with coefficients in R. Moreover, we can divide f by the greatest common divisor of all its coefficients, likewise for g, and thereby assume

$$f, g \in R[X], \qquad \nu_p(f) = 0 = \nu_p(g).$$

Then it remains to show that $\nu_p(fg) = 0$. To do this, consider the homomorphism

$$\Phi\colon R[X] \longrightarrow \big(R/pR\big)[X]$$

reducing coefficients. The kernel $\ker \Phi$ consists of all polynomials in $R[X]$ whose coefficients are divisible by p, i.e.,

$$\ker \Phi = \big\{ f \in R[X] \,;\, \nu_p(f) > 0 \big\}.$$

Since $\nu_p(f) = 0 = \nu_p(g)$, we get $\Phi(f), \Phi(g) \neq 0$. Now R/pR is an integral domain, and the same is true for $(R/pR)[X]$, by 2.1/3. Therefore, we conclude that

$$\Phi(fg) = \Phi(f) \cdot \Phi(g) \neq 0,$$

and in particular, $\nu_p(fg) = 0$. □

Corollary 6. *Let R be a unique factorization domain and $h \in R[X]$ a monic polynomial. Assume that there is a factorization $h = f \cdot g$ into monic polynomials $f, g \in Q(R)[X]$. Then necessarily $f, g \in R[X]$.*

Proof. We have $\nu_p(h) = 0$, as well as $\nu_p(f), \nu_p(g) \leq 0$ for every prime element $p \in R$, due to the fact that h, f, and g are monic. Furthermore, Gauss's lemma yields

$$\nu_p(f) + \nu_p(g) = \nu_p(h) = 0$$

and hence $\nu_p(f) = \nu_p(g) = 0$ for all p. However, this means that $f, g \in R[X]$, as claimed. □

A polynomial $f \in R[X]$ with coefficients in a unique factorization domain R is called *primitive* if the greatest common divisor of all its coefficients is 1 or, equivalently, if $\nu_p(f) = 0$ for all prime elements $p \in R$. For example, monic polynomials in $R[X]$ are primitive. Moreover, proceeding similarly as in the proof of Corollary 6, we can conclude for a polynomial $h \in R[X]$ and a factorization $h = f \cdot g$ with $f \in R[X]$ primitive and $g \in Q(R)[X]$ that g is already contained in $R[X]$.

In the following we will frequently use the fact that every nonzero polynomial $f \in Q(R)[X]$ can be written as $f = a\tilde{f}$ with a constant $a \in Q(R)^*$ and a primitive polynomial $\tilde{f} \in R[X]$. Just set

$$a = \prod_{p \in P} p^{\nu_p(f)}, \qquad \tilde{f} = a^{-1} f,$$

where P is a system of representatives of the prime elements in R.

Now we are able to prove the result of Gauss announced at the beginning of the present section. As a by-product, we will get a characterization of the prime elements in $R[X]$.

Proposition 7 (Gauss). *Let R be a unique factorization domain. Then the polynomial ring $R[X]$ is a unique factorization domain as well. A polynomial $q \in R[X]$ is prime if and only if*:

(i) *q is prime in R, or*

(ii) *q is primitive in $R[X]$ and prime in $Q(R)[X]$.*

In particular, a primitive polynomial $q \in R[X]$ is prime in $R[X]$ if and only if it is prime in $Q(R)[X]$.

Proof. Let q be a prime element in R. Then R/qR is an integral domain and the same is true for $R[X]/qR[X] \simeq (R/qR)[X]$. From this we conclude that q is prime also in $R[X]$.

Next, consider a primitive polynomial $q \in R[X]$ such that q is prime in $Q(R)[X]$. In order to show that q is prime even in $R[X]$, consider polynomials $f, g \in R[X]$ such that $q \,|\, fg$ in $R[X]$. Then we have $q \,|\, fg$ in $Q(R)[X]$ as well. Since q is prime in $Q(R)[X]$, it divides one of the two factors, say $q \mid f$, and there exists $h \in Q(R)[X]$ such that $f = qh$. Now apply Gauss's lemma to the latter equation. Since q is primitive, we get for every prime element $p \in R$ that

$$0 \leq \nu_p(f) = \nu_p(q) + \nu_p(h) = \nu_p(h),$$

and hence $h \in R[X]$. In particular, $q \,|\, f$ in $R[X]$, and it follows that q is prime in $R[X]$.

It remains to show that $R[X]$ is a unique factorization domain and that every prime element in $R[X]$ is of type (i) or (ii). To achieve this it is clearly enough to show that every nonzero nonunit $f \in R[X]$ admits a factorization into prime elements of type (i) and (ii). Let us establish the latter fact. Write $f = a\tilde{f}$, where $a \in R$ is the greatest common divisor of all coefficients of f and hence $\tilde{f}$ is primitive. Since a is a product of prime elements in R, it is enough to show that the primitive polynomial $\tilde{f}$ is a product of primitive polynomials in $R[X]$ that are prime in $Q(R)[X]$. Let $\tilde{f} = c\tilde{f}_1 \dots \tilde{f}_r$ be a factorization into prime elements from $Q(R)[X]$, for a constant $c \in Q(R)^*$. Choosing c suitably, we may assume that all $\tilde{f}_i$ are primitive in $R[X]$. Then Gauss's lemma implies for every prime element $p \in R$ that

$$\nu_p(\tilde{f}) = \nu_p(c) + \nu_p(\tilde{f}_1) + \dots + \nu_p(\tilde{f}_r),$$

and since

$$\nu_p(\tilde{f}) = \nu_p(\tilde{f}_1) = \dots = \nu_p(\tilde{f}_r) = 0,$$

that $\nu_p(c) = 0$; this means that c is a unit in R. Now replacing $\tilde{f}_1$ by $c\tilde{f}_1$, we see that $\tilde{f}$ is a product of prime elements of the desired type. □

Exercises

1. *Let R be a unique factorization domain and $\Phi\colon R[X] \longrightarrow R[X]$ a ring automorphism that restricts to an automorphism $\varphi\colon R \longrightarrow R$. Compare $\nu_p(f)$ and $\nu_{\varphi(p)}(\Phi(f))$ for polynomials $f \in R[X]$ and prime elements $p \in R$, and check whether $\Phi(f)$ is primitive when f is primitive. Show for $a \in R$ that a polynomial f is primitive if and only if $f(X + a)$ is primitive.*

2. *Consider a unique factorization domain R with field of fractions K and with a system P of representatives of its prime elements. For $f \in K[X]-\{0\}$ denote by $a_f := \prod_{p\in P} p^{\nu_p(f)}$ the "content" of f. Formulate the assertion of Gauss's lemma (Lemma 5) in an equivalent way using the notion of content.*

3. Consider the rational function field $K(X)$ in one variable X over a field K, as well as the polynomial ring $K(X)[Y]$ for a variable Y. Let $f(Y), g(Y) \in K[Y]$ be coprime with $\deg f(Y) \cdot g(Y) \geq 1$. Show that $f(Y) - g(Y)X$ is irreducible in $K(X)[Y]$.

4. Let R be a unique factorization domain. Show:

   (i) For a multiplicative system $S \subset R$, the ring of fractions $S^{-1}R$ is a unique factorization domain again. How are the prime elements of R related to those of $S^{-1}R$?

   (ii) For prime elements $p \in R$ set $R_p := S_p^{-1}R$, where $S_p = R\text{–}(p)$. A polynomial $f \in R[X]$ is primitive if and only if the induced polynomial $f_p \in R_p[X]$ is primitive for every prime element $p \in R$.

5. *Universal property of rings of fractions*: Let R be a ring and $S \subset R$ a multiplicative system. Show for every ring homomorphism $\varphi\colon R \longrightarrow R'$ satisfying $\varphi(S) \subset R'^*$ that there exists a unique ring homomorphism $\overline{\varphi}\colon S^{-1}R \longrightarrow R'$ such that $\varphi = \overline{\varphi} \circ \tau$; here $\tau\colon R \longrightarrow S^{-1}R$ denotes the canonical homomorphism given by $a \longmapsto \frac{a}{1}$.

6. *Partial fraction decomposition*: Let $f, g \in K[X]$ be polynomials with coefficients in a field K, where g is monic with prime factorization $g = g_1^{\nu_1} \dots g_n^{\nu_n}$ and pairwise nonassociated prime polynomials $g_1, \dots, g_n$. Show that in the field of fractions $K(X) = Q(K[X])$ there is a unique decomposition

$$\frac{f}{g} = f_0 + \sum_{i=1}^{n} \sum_{j=1}^{\nu_i} \frac{c_{ij}}{g_i^j}$$

   with polynomials $f_0, c_{ij} \in K[X]$, where $\deg c_{ij} < \deg g_i$. In particular, if the prime factors g_i are linear, the c_{ij} are of degree 0 and thus are constant. *Hint*: Prove the existence of a decomposition $fg^{-1} = f_0 + \sum_{i=1}^{n} f_i g_i^{-\nu_i}$ such that $g_i \nmid f_i$ and $\deg f_i < \deg g_i^{\nu_i}$. Then apply the g_i-adic expansion to the f_i; cf. Exercise 4 from 2.1.

2.8 Criteria for Irreducibility

Let R be a unique factorization domain and $K = Q(R)$ its field of fractions. In the following we want to discuss some techniques for checking whether a given polynomial $f \in K[X] - \{0\}$ is irreducible or prime, which are the same in unique factorization domains, due to 2.4/10. Depending on f, there is always a constant $c \in K^*$ such that $\tilde{f} = cf$ is a primitive polynomial in $R[X]$. Furthermore, we can conclude from the result of Gauss 2.7/7 that f and $\tilde{f}$ are irreducible in $K[X]$ if and only if $\tilde{f}$ is irreducible in $R[X]$. In this way, the irreducibility of polynomials in $K[X]$ can be reduced to the irreducibility of primitive polynomials in $R[X]$.

Proposition 1 (Eisenstein's criterion). *Let R be a unique factorization domain and $f = a_n X^n + \ldots + a_0 \in R[X]$ a primitive polynomial of degree > 0. Assume there is a prime element $p \in R$ such that*

$$p \nmid a_n, \qquad p \mid a_i \text{ for } i < n, \qquad p^2 \nmid a_0.$$

Then f is irreducible in $R[X]$ and hence, by 2.7/7, also in $Q(R)[X]$.

Proof. Suppose f is reducible in $R[X]$. Then there is a factorization

$$f = gh, \quad \text{say} \quad g = \sum_{i=0}^{r} b_i X^i,\ h = \sum_{i=0}^{s} c_i X^i,$$

where $r + s = n$, and $r > 0$, $s > 0$. Furthermore, from our assumption on f, we conclude that

$$\begin{aligned} a_n &= b_r c_s \neq 0, & p &\nmid b_r, & p &\nmid c_s, \\ a_0 &= b_0 c_0, & p &\mid b_0 c_0, & p^2 &\nmid b_0 c_0, \end{aligned}$$

and we may assume $p \mid b_0$, $p \nmid c_0$. Now let $t < r$ be maximal such that $p \mid b_\tau$ for $0 \leq \tau \leq t$. Setting $b_i = 0$ for $i > r$ and $c_i = 0$ for $i > s$, we get

$$a_{t+1} = b_0 c_{t+1} + \ldots + b_{t+1} c_0,$$

where a_{t+1} is not divisible by p. Indeed, $b_0 c_{t+1}, \ldots, b_t c_1$ are divisible by p, due to the definition of t, and $b_{t+1} c_0$ is not. But then we must have $t + 1 = n$, due to our assumption on f, and therefore $r = n$, $s = 0$, which is in contradiction to $s > 0$. □

Next we want to discuss the *reduction test* for irreducibility.

Proposition 2. *Let R be a unique factorization domain, $p \in R$ a prime element, and $f \in R[X]$ a polynomial of degree > 0 whose leading coefficient is not divisible by p. Furthermore, let $\Phi\colon R[X] \longrightarrow R/(p)[X]$ be the canonical homomorphism reducing coefficients* mod p. *Then*:

If $\Phi(f)$ is irreducible in $R/(p)[X]$, then the same is true for f in $Q(R)[X]$. If in addition, f is primitive, then it is irreducible in $R[X]$ as well.

Proof. First, assume that $f \in R[X]$ is primitive. Then if f is reducible, there is a factorization $f = gh$ in $R[X]$, where $\deg g > 0$ and $\deg h > 0$. Furthermore, p cannot divide the leading coefficient of g or h, since p does not divide the leading coefficient of f. Now we have

$$\Phi(f) = \Phi(g)\Phi(h),$$

where $\Phi(g)$ and $\Phi(h)$ are nonconstant polynomials in $R/(p)[X]$, and it follows that $\Phi(f)$ is reducible. Thus, by contraposition, $\Phi(f)$ irreducible implies that f is irreducible in $R[X]$.

To deal with the general case, write $f = c \cdot \tilde{f}$ for a constant $c \in R$ and a primitive polynomial $\tilde{f} \in R[X]$; note that p cannot divide c or the leading coefficient of $\tilde{f}$. Then if $\Phi(f)$ is irreducible, the same is true for $\Phi(\tilde{f})$, and it follows, as we just have seen, that $\tilde{f}$ is irreducible in $R[X]$. Applying the result of Gauss 2.7/7, we conclude that $\tilde{f}$ and hence also f are irreducible in $Q(R)[X]$. □

Let us add that alternatively, Eisenstein's criterion can be obtained as a consequence of the above reduction test for irreducibility. Indeed, let us place ourselves in the situation of Proposition 1. If there is a factorization $f = gh$ for polynomials $g, h \in R[X]$ of degree $< n$, we can apply the reduction homomorphism $\Phi\colon R[X] \longrightarrow R/(p)[X]$, thereby obtaining an equation $\overline{a}_n X^n = \Phi(f) = \Phi(g)\Phi(h)$. We claim that $\Phi(g)$ and $\Phi(h)$ are nontrivial powers of X, up to constant factors from $R/(p)$. To justify the claim, interpret the preceding factorization in the polynomial ring $k[X]$ over the field of fractions k of $R/(p)$, which is a unique factorization domain. In particular, the constant parts of g and h will be divisible by p, and it follows that the constant part of f is divisible by p^2, which is in contradiction to the assumption on f.

We want to add some examples showing how to apply the above irreducibility criteria:

(1) Let k be a field and $K := k(t)$ the field of rational functions in a variable t over k. Then the polynomial $X^n - t \in K[X]$ is irreducible for $n \geq 1$. Indeed, $R := k[t]$ is a unique factorization domain with field of fractions K. Furthermore, $t \in R$ is a prime element, and $X^n - t$ is a primitive polynomial in $R[X]$ such that Eisenstein's criterion can be applied for $p := t$.

(2) Let $p \in \mathbb{N}$ be a prime number. We claim that $f(X) = X^{p-1} + \ldots + 1$ is irreducible in $\mathbb{Q}[X]$. To justify this we apply Eisenstein's criterion to the polynomial $f(X+1)$, using the fact that $f(X+1)$ is irreducible if and only if the same is true for $f(X)$. Note that

$$f(X) = \frac{X^p - 1}{X - 1},$$
$$f(X+1) = \frac{(X+1)^p - 1}{X} = X^{p-1} + \binom{p}{1} X^{p-2} + \ldots + \binom{p}{p-1}.$$

The conditions of Eisenstein's criterion are satisfied for $f(X+1)$, since we have $\binom{p}{p-1} = p$ and $p \mid \binom{p}{\nu}$ for $\nu = 1, \ldots, p-1$; just observe that

$$\binom{p}{\nu} = \frac{p(p-1)\ldots(p-\nu+1)}{1\ldots\nu}$$

admits for $\nu = 1, \ldots, p-1$ a prime factor p in the numerator, but not in the denominator and hence is divisible by p.

(3) $f = X^3 + 3X^2 - 4X - 1$ is irreducible in $\mathbb{Q}[X]$. Indeed, view f as a primitive polynomial in $\mathbb{Z}[X]$ and reduce coefficients mod 3. It remains to show that the polynomial

$$X^3 - X - 1 \in \mathbb{F}_3[X]$$

is irreducible. This is easily checked, since the polynomial does not admit zeros in $\mathbb{F}_3$. More generally, one can show (cf. Exercise 2 below) that the polynomial $X^p - X - 1$ is irreducible in $\mathbb{F}_p[X]$ for every prime number p.

Exercises

1. Show that the following polynomials are irreducible:
   (i) $X^4 + 3X^3 + X^2 - 2X + 1 \in \mathbb{Q}[X]$.
   (ii) $2X^4 + 200X^3 + 2000X^2 + 20000X + 20 \in \mathbb{Q}[X]$.
   (iii) $X^2Y + XY^2 - X - Y + 1 \in \mathbb{Q}[X, Y]$.
2. Let $p \in \mathbb{N}$ be a prime number. Show that the polynomial $g = X^p - X - 1$ is irreducible in $\mathbb{F}_p[X]$. *Hint*: Note that g is invariant under the automorphism $\tau\colon \mathbb{F}_p[X] \longrightarrow \mathbb{F}_p[X]$, $f(X) \longmapsto f(X+1)$, and study the action of τ on the prime factorization of g.

2.9 Theory of Elementary Divisors*

In the present section we want to generalize the concept of vector spaces over fields to modules over rings, our main objective being modules over principal ideal domains. For example, any abelian group can be viewed as a $\mathbb{Z}$-module, i.e., as a module over the ring $\mathbb{Z}$. In any case, the study of abelian groups, in particular the classification of finitely generated abelian groups, is a good motivation for developing the theory of elementary divisors, up to its central result, the structure theorem for finitely generated modules over principal ideal domains. This theorem contains the classification of finitely generated abelian groups as a special case and, beyond this, admits other interesting applications,

such as the existence of canonical forms for endomorphisms of finite-dimensional vector spaces; cf. Exercise 3. In the following we will prove the so-called elementary divisor theorem, which clarifies the structure of submodules of finite rank in free modules over principal ideal domains. As a corollary, we derive the just mentioned structure theorem for finitely generated modules over principal ideal domains.

Let A be a ring, general for the moment, but later assumed to be a principal ideal domain. An *A-module* is an abelian group M, together with an exterior multiplication

$$A \times M \longrightarrow M, \qquad (a, x) \longmapsto a \cdot x,$$

that satisfies the usual "vector space axioms"

$$\begin{aligned} a \cdot (x + y) &= a \cdot x + a \cdot y, \\ (a + b) \cdot x &= a \cdot x + b \cdot x, \\ a \cdot (b \cdot x) &= (ab) \cdot x, \\ 1 \cdot x &= x, \end{aligned}$$

for elements $a, b \in A$, $x, y \in M$. *Homomorphisms* between A-modules, also referred to as *A-homomorphisms*, are defined just as in the context of vector spaces, likewise *submodules* of an A-module M, as well as the *residue class module* M/N of an A-module M by a submodule N. Furthermore, the fundamental theorem on homomorphisms 1.2/6 remains valid in the module context. If we consider A as a module over itself, the ideals of A coincide with the submodules of A. Moreover, for any ideal $\mathfrak{a} \subset A$ we can view the residue class ring $A/\mathfrak{a}$ as an A-module.

As already mentioned, every abelian group G can naturally be viewed as a $\mathbb{Z}$-module. Just define the product map $\mathbb{Z} \times G \longrightarrow G$, $(a, x) \longmapsto ax$, by $ax = \sum_{i=1}^{a} x$ for $a \geq 0$ and $ax = -(-a)x$ for $a < 0$. On the other hand, every $\mathbb{Z}$-module M gives rise to an abelian group G by forgetting about the $\mathbb{Z}$-multiplication on M. It is easily verified that in this way, abelian groups and $\mathbb{Z}$-modules correspond bijectively to each other and that the correspondence extends to homomorphisms, subgroups, and submodules, as well as to residue class groups and residue class modules. To give another example, consider a vector space V over a field K, together with a K-endomorphism $\varphi\colon V \longrightarrow V$. Then V becomes a module over the polynomial ring in one variable $K[X]$ if we define the multiplication by

$$K[X] \times V \longrightarrow V, \qquad \Big(\sum a_i X^i, v\Big) \longmapsto \sum a_i \varphi^i(v).$$

On the other hand, for every $K[X]$-module V we can consider its underlying K-vector space together with the K-endomorphism $\varphi\colon V \longrightarrow V$ that is given by multiplication by X. Also in this case, pairs of type (V, φ), consisting of a K-vector space V and a K-endomorphism $\varphi\colon V \longrightarrow V$, correspond bijectively to $K[X]$-modules.

For a module M and a family of submodules $M_i \subset M$, $i \in I$, their *sum* is defined as usual by

$$M' = \sum_{i\in I} M_i = \Big\{ \sum_{i\in I} x_i \,;\, x_i \in M_i,\, x_i = 0 \text{ for almost all } i \in I \Big\}.$$

If every $x \in M'$ admits a representation $x = \sum_{i\in I} x_i$ with elements $x_i \in M_i$ that are unique, we call M' the *direct sum* of the M_i, writing $M' = \bigoplus_{i\in I} M_i$ in this case. For example, a sum $M_1 + M_2$ of two submodules of M is direct if and only if $M_1 \cap M_2 = 0$. Furthermore, given a family of A-modules $(M_i)_{i\in I}$, we can naturally construct an A-module M that is the direct sum of the M_i. Indeed, let

$$M = \Big\{ (x_i)_{i\in I} \in \prod_{i\in I} M_i \,;\, x_i = 0 \text{ for almost all } i \Big\}$$

and identify M_i in each case with the submodule of M consisting of all families $(x_{i'})_{i'\in I}$, where $x_{i'} = 0$ for $i' \neq i$.

A family $(x_i)_{i\in I}$ of elements of an A-module M is called a *system of generators* of M if we have $M = \sum_{i\in I} Ax_i$. If M admits a finite system of generators, we say that M is *finitely generated*, or simply that M is a *finite* A-module.[3] Furthermore, the family $(x_i)_{i\in I}$ is called *free* or *linearly independent* if from an equation $\sum_{i\in I} a_i x_i = 0$ with coefficients $a_i \in A$ we can conclude that $a_i = 0$ for all $i \in I$. A free system of generators will also be referred to as a *basis*; in this case every element $x \in M$ admits a representation $x = \sum_{i\in I} a_i x_i$ with coefficients $a_i \in A$ that are unique, and we say that M is a *free* A-module. For example, A^n for $n \in \mathbb{N}$ is a free A-module, just as $A^{(I)}$ is for an arbitrary index set I.

If we consider a field K instead of a general ring A as coefficient domain, the theory of A-modules specializes to the theory of K-vector spaces. Furthermore, let us point out that computations in a module M over a ring A follow to a large extent the rules we are used to in vector spaces over fields. However, there is one major exception that has to be observed: from an equation $ax = 0$ for elements $a \in A$, $x \in M$ we cannot necessarily conclude that a or x vanishes, since even for $a \neq 0$ there might not exist an inverse a^{-1} in A. As a consequence, A-modules, even finitely generated ones, do not necessarily admit a basis. For example, if $\mathfrak{a} \subset A$ is a nontrivial ideal, then the residue class ring $A/\mathfrak{a}$ is an example of such an A-module that is not free.

From now on let A be an *integral domain*. An element x of an A-module M is called a *torsion element* if there exists an element $a \in A - \{0\}$ such that $ax = 0$. Due to the fact that A is an integral domain, the torsion elements constitute a submodule $T \subset M$, the so-called *torsion submodule* of M. If $T = 0$, we call M *torsion-free*; and if $T = M$, a *torsion module*. For example, every free module is torsion-free, and every finite abelian group, viewed as a $\mathbb{Z}$-module, is a torsion module. Further, we define the *rank* of an A-module M, denoted by $\operatorname{rank} M$,

[3] Observe the usage of language: in contrast to the notions of finite group, finite ring, and finite field, we do *not* require that a finite A-module consist of only finitely many elements.

as the supremum of all numbers n such that there exists a linearly independent system of elements $x_1, \ldots, x_n$ in M. In this way, the rank of a module is defined similarly to the dimension of a vector space. Note that M is a torsion module if and only if its rank is zero.

Now let $S = A - \{0\}$ and consider the field of fractions $K = S^{-1}A$ of the integral domain A. For any given A-module M, we can construct the associated K-vector space $S^{-1}M$ by proceeding as in the case of rings of fractions in Section 2.7. Indeed, let $S^{-1}M$ be the set of all fractions of type $\frac{x}{s}$ for $x \in M$ and $s \in S$, where $\frac{x}{s}$ is identified with $\frac{x'}{s'}$ if there exists an element $s'' \in S$ such that $s''(s'x - sx') = 0$. Then $S^{-1}M$ becomes a K-vector space under the usual rules of fractional arithmetic, and it is verified without difficulty that the rank of M coincides with the dimension of $S^{-1}M$ as a K-vector space. Furthermore, the kernel of the canonical map $M \longrightarrow S^{-1}M$, $x \longmapsto \frac{x}{1}$, equals the torsion submodule $T \subset M$.

From now on we will always assume that A is a *principal ideal domain*. For technical reasons we need the notion of *length* of an A-module M, in particular in the case that M is a torsion module, which is defined as the supremum $l_A(M)$ of all numbers ℓ such that there is a chain of submodules

$$0 \subsetneq M_1 \subsetneq M_2 \subsetneq \ldots \subsetneq M_\ell = M$$

of length ℓ. For example, the zero module is of length 0, and the free $\mathbb{Z}$-module $\mathbb{Z}$ is of length ∞. If V is a vector space over a field K, then the length $l_K(V)$ coincides with the vector space dimension $\dim_K V$.

Lemma 1. (i) *Let A be a principal ideal domain and $a \in A$ an element with prime factorization $a = p_1 \ldots p_r$. Then $l_A(A/aA) = r$.*

(ii) *Let M be an A-module that is the direct sum of two submodules M' and M''. Then $l_A(M) = l_A(M') + l_A(M'')$.*

Proof. We start with assertion (ii). If there are chains of submodules

$$0 \subsetneq M'_1 \subsetneq M'_2 \subsetneq \ldots \subsetneq M'_r = M',$$
$$0 \subsetneq M''_1 \subsetneq M''_2 \subsetneq \ldots \subsetneq M''_s = M'',$$

then

$$0 \subsetneq M'_1 \oplus 0 \subsetneq M'_2 \oplus 0 \subsetneq \ldots \subsetneq M'_r \oplus 0 \subsetneq M'_r \oplus M''_1 \subsetneq M'_r \oplus M''_2 \subsetneq \ldots \subsetneq M'_r \oplus M''_s = M$$

is a chain of length $r+s$ in M. Consequently, we have $l_A(M) \geq l_A(M') + l_A(M'')$. To verify the opposite estimate, consider a chain of submodules

$$0 = M_0 \subsetneq M_1 \subsetneq M_2 \subsetneq \ldots \subsetneq M_\ell = M$$

and let $\pi'' \colon M' \oplus M'' \longrightarrow M''$ be the projection onto the second summand, so that $\ker \pi'' = M'$. Then we get $M_\lambda \cap M' \subsetneq M_{\lambda+1} \cap M'$ or $\pi''(M_\lambda) \subsetneq \pi''(M_{\lambda+1})$ for

$0 \leq \lambda < \ell$, and we can conclude that $\ell \leq l_A(M') + l_A(M'')$. Hence, assertion (ii) is clear.

Now assertion (i) is easy to justify. Renumbering the p_i, we can look at a prime factorization of type $a = \varepsilon p_1^{\nu_1} \dots p_s^{\nu_s}$ for a unit ε and pairwise nonassociated prime elements $p_1, \dots, p_s$, where $r = \nu_1 + \dots + \nu_s$. Then, due to the Chinese remainder theorem in the version of 2.4/14, we conclude that A/aA, as a ring, is isomorphic to the ring-theoretic product $\prod_{i=1}^{s} A/p_i^{\nu_i}A$. Moreover, thinking in terms of A-modules, the decomposition is interpreted from the additive point of view as the direct sum

$$A/aA \simeq A/p_1^{\nu_1}A \oplus \dots \oplus A/p_s^{\nu_s}A.$$

Therefore, using assertion (ii), which has already been proved, it is enough to consider the case $s = 1$, i.e., the case $a = p^\nu$ for a single prime element $p \in A$. The submodules of $A/p^\nu A$ correspond bijectively to the ideals $\mathfrak{a} \subset A$ satisfying $p^\nu \in \mathfrak{a}$, and since A is a principal ideal domain, bijectively to the divisors $p^0, p^1, \dots, p^\nu$ of p^ν. Since $p^{i+1}A$ is strictly contained in p^iA for all i, we get $l_A(A/p^\nu) = \nu$, which had to be shown. □

Next we turn to the proof of the *elementary divisor theorem*, which, as mentioned already, is a key result for the study of finitely generated modules over principal ideal domains and of finitely generated abelian groups.

Theorem 2. *Consider a finite free module F over a principal ideal domain A and a submodule $M \subset F$ of rank n. Then there exist elements $x_1, \dots, x_n \in F$ that are part of a basis of F, as well as coefficients $\alpha_1, \dots, \alpha_n \in A - \{0\}$ such that*:

(i) *$\alpha_1 x_1, \dots, \alpha_n x_n$ form a basis of M.*
(ii) *$\alpha_i \,|\, \alpha_{i+1}$ for $1 \leq i < n$.*

The elements $\alpha_1, \dots, \alpha_n$ are uniquely determined by M, up to associatedness, and are independent of the choice of the elements $x_1, \dots, x_n$. They are called the elementary divisors *of $M \subset F$.*

Remark 3. *In the situation of Theorem* 2, *the submodule $\bigoplus_{i=1}^{n} Ax_i \subset F$ is uniquely determined by M as the* saturation M_{sat} *of M in F, which consists of all elements $y \in F$ such that there exists an element $a \neq 0$ in A satisfying $ay \in M$. Furthermore, we have*

$$M_{\mathrm{sat}}/M \simeq \bigoplus_{i=1}^{n} A/\alpha_i A.$$

First, let us deduce Remark 3 from the existence assertion of Theorem 2. Clearly, we have $\alpha_n \cdot (\bigoplus_{i=1}^{n} Ax_i) \subset M$ and therefore $\bigoplus_{i=1}^{n} Ax_i \subset M_{\mathrm{sat}}$. Conversely, consider an element $y \in M_{\mathrm{sat}}$, where $ay \in M$ for some $a \in A - \{0\}$. Due

to the assertion of Theorem 2, we can enlarge the system $x_1, \ldots, x_n$ to a basis of F by adding elements $x_{n+1}, \ldots, x_r$. Now represent y as a linear combination of this basis, say $y = \sum_{j=1}^{r} a_j x_j$. Since $ay \in M$, we conclude that $aa_j = 0$, and in particular $a_j = 0$ for $j = n+1, \ldots, r$. Therefore, $y \in \bigoplus_{i=1}^{n} Ax_i$, and we get $M_{\mathrm{sat}} \subset \bigoplus_{i=1}^{n} Ax_i$, and thus in fact $\bigoplus_{i=1}^{n} Ax_i = M_{\mathrm{sat}}$. To justify the second assertion of Remark 3, consider the A-isomorphisms $A \xrightarrow{\sim} Ax_i$, $a \longmapsto ax_i$, for indices $i = 1, \ldots, n$ and observe that the ideal $\alpha_i A \subset A$ is mapped bijectively onto the submodule $A\alpha_i x_i \subset Ax_i$. Hence, $Ax_i/A\alpha_i x_i$ is isomorphic to $A/\alpha_i A$, and a direct sum analogue of this argument yields an isomorphism between $(\bigoplus_{i=1}^{n} Ax_i)/M$ and $\bigoplus_{i=1}^{n} A/\alpha_i A$. □

For the proof of Theorem 2 we need the notion of *content* for elements $x \in F$, denoted by $\mathrm{cont}(x)$. To define it, consider a basis $y_1, \ldots, y_r$ of F and represent x as a linear combination of the y_j with coefficients in A, say $x = \sum_{j=1}^{r} c_j y_j$. Then we set $\mathrm{cont}(x) = \gcd(c_1, \ldots, c_r)$. In this way, $\mathrm{cont}(x)$ does not specify a particular element of A, but rather a class of associated elements. Note that $\mathrm{cont}(0) = 0$, even if $F = 0$. To show that $\mathrm{cont}(x)$ is independent of the choice of the basis $y_1, \ldots, y_r$ of F, consider the A-module F^* of all A-homomorphisms $F \longrightarrow A$, i.e., of all linear functionals on F. It is easy to see that the elements of type $\varphi(x)$ for $\varphi \in F^*$ constitute an ideal in A, in fact a principal ideal (c), and we claim that $c = \mathrm{cont}(x)$. To justify this, choose an equation $\mathrm{cont}(x) = \sum_{j=1}^{r} a_j c_j$ with coefficients $a_j \in A$; cf. 2.4/13. Then, if $\varphi_1, \ldots, \varphi_r$ is the dual basis associated to $y_1, \ldots, y_r$, characterized by $\varphi_i(y_j) = 0$ for $i \neq j$ and $\varphi_i(y_i) = 1$, we get $\varphi(x) = \mathrm{cont}(x)$ for $\varphi = \sum_{j=1}^{r} a_j \varphi_j$. However, since $\mathrm{cont}(x) = \gcd(c_1, \ldots, c_r)$ is always a divisor of $\psi(x)$ for $\psi \in F^*$, we must have $c = \mathrm{cont}(x)$.

Let us list some properties of the notion of content that are used in the sequel.

Lemma 4. *In the situation of Theorem* 2 *the following assertions hold*:

(i) *Given* $x \in F$ *there exists* $\varphi \in F^*$ *such that* $\varphi(x) = \mathrm{cont}(x)$.
(ii) *For* $x \in F$ *and* $\psi \in F^*$ *we have* $\mathrm{cont}(x) \,|\, \psi(x)$.
(iii) *There exists an element* $x \in M$ *such that* $\mathrm{cont}(x) \,|\, \mathrm{cont}(y)$ *for all* $y \in M$.

Proof. Due to the considerations above, only assertion (iii) needs to be justified. To achieve this, look at the set of ideals of type $\mathrm{cont}(y) \cdot A$, where y varies over M. There is a maximal element among all these ideals, i.e., one that is not strictly contained in any of the ideals $\mathrm{cont}(y) \cdot A$, $y \in M$. Indeed, otherwise we could construct an infinite sequence of elements y_i in M such that

$$\mathrm{cont}(y_1) \cdot A \subsetneq \mathrm{cont}(y_2) \cdot A \subsetneq \ldots$$

is a strictly ascending chain of ideals, contradicting the fact that A is Noetherian; cf. 2.4/8. Therefore, we can find an element $x \in M$ such that $\mathrm{cont}(x) \cdot A$ is maximal in the sense just discussed. Furthermore, apply (i) and let $\varphi \in F^*$ satisfy $\varphi(x) = \mathrm{cont}(x)$. We want to show that

$$(*) \qquad \varphi(x) \,|\, \varphi(y) \text{ for all } y \in M.$$

To achieve this, consider an element $y \in M$ and let $d = \gcd(\varphi(x), \varphi(y))$. There are elements $a, b \in A$ such that $a\varphi(x) + b\varphi(y) = d$, and hence $\varphi(ax + by) = d$. Furthermore, we get $\operatorname{cont}(ax + by) \,|\, d$ from (ii), and even $\operatorname{cont}(ax + by) \,|\, \operatorname{cont}(x)$, since $d \,|\, \varphi(x)$. However, this implies $\operatorname{cont}(ax + by) = \operatorname{cont}(x)$, due to the maximality property of x. In particular, $\operatorname{cont}(x)$ is a divisor of d, and since $d \,|\, \varphi(y)$, even a divisor of $\varphi(y)$. This verifies $(*)$.

To prove $\operatorname{cont}(x) \,|\, \operatorname{cont}(y)$, it is enough by (i) to prove $\varphi(x) \,|\, \psi(y)$ for all $\psi \in F^*$. Since $\varphi(x) \,|\, \psi(x)$ by (ii), as well as $\varphi(x) \,|\, \varphi(y)$ by $(*)$, we may replace y by $y - \frac{\varphi(y)}{\varphi(x)}x$ and thereby assume $\varphi(y) = 0$. Furthermore, using these divisibility relations again, we can replace ψ by $\psi - \frac{\psi(x)}{\varphi(x)}\varphi$ and assume $\psi(x) = 0$. Now let $d = \gcd(\varphi(x), \psi(y))$, say $d = a\varphi(x) + b\psi(y)$ for $a, b \in A$. Then, since $\varphi(y) = 0$ and $\psi(x) = 0$, we get

$$(\varphi + \psi)(ax + by) = a\varphi(x) + b\psi(y) = d,$$

and thus $\operatorname{cont}(ax + by) \,|\, d$ by (ii). However, d divides $\varphi(x)$ by its definition. Therefore, we have $\operatorname{cont}(ax + by) \,|\, \varphi(x)$ and even $\operatorname{cont}(ax + by) = \varphi(x)$ by the maximality property of x. But then $\varphi(x) \,|\, d$, and we conclude that $\varphi(x) \,|\, \psi(y)$ as desired, since $d \,|\, \psi(y)$. □

Now we can do the *proof of Theorem* 2. In a first step we show that every submodule $M \subset F$ is free. This fact will then be used in a second step to derive the existence part of the theorem. In both cases we use induction on $n = \operatorname{rank} M$. So let us start by showing that M is free. If $n = 0$, we have $M = 0$, since M is torsion-free, and the assertion is clear. Now, assuming $n > 0$, choose an element $x \in M$ according to Lemma 4 (iii) that satisfies $\operatorname{cont}(x) \,|\, \operatorname{cont}(y)$ for all $y \in M$. Then there is a linear functional $\varphi \in F^*$ such that $\varphi(x) = \operatorname{cont}(x)$, cf. Lemma 4 (i), as well as a (unique) element $x_1 \in F$ such that $x = \varphi(x)x_1$. Setting $F' = \ker \varphi$ and $M' = M \cap F'$, we claim that

$$(*) \qquad F = Ax_1 \oplus F', \qquad M = Ax \oplus M'.$$

To justify the decomposition for M, consider an element $y \in M$ and write

$$y = \frac{\varphi(y)}{\varphi(x)}x + \Big(y - \frac{\varphi(y)}{\varphi(x)}x\Big).$$

Then the left-hand summand belongs to Ax. Indeed, we have $\varphi(x) \,|\, \varphi(y)$, since $\operatorname{cont}(x) \,|\, \operatorname{cont}(y)$ by the choice of x, and since $\operatorname{cont}(y) \,|\, \varphi(y)$ by Lemma 4 (ii). Moreover, the right-hand summand is contained in M', since it belongs to M as well as to $\ker \varphi$. In particular, the above decomposition of y shows that $M = Ax + M'$. Next, observe that we have $\varphi(x) \neq 0$, since $M \neq 0$ and thus that $Ax \cap M' = 0$. It follows that M is the direct sum of its submodules Ax and M'. In the same way one proves that F is the direct sum of the submodules Ax_1 and F'; just replace x by x_1 in the preceding argument and use $\varphi(x_1) = 1$.

From the decomposition $M = Ax \oplus M'$ we conclude that $\operatorname{rank} M' < n$, since $x \neq 0$. Then M' is free by the induction hypothesis, necessarily of rank $n-1$, and we see that M is free as well. This settles our first induction argument.

For the second induction we proceed in the same way, until we get to the decompositions $(*)$. From the first induction we know that F' is a free submodule of F. Thus, by the induction hypothesis, the existence part of Theorem 2 is available for the submodule $M' \subset F'$. In particular, there exist elements $x_2, \ldots, x_n \in F'$ that are part of a basis of F', as well as elements $\alpha_2, \ldots, \alpha_n \in A - \{0\}$ satisfying $\alpha_i \,|\, \alpha_{i+1}$ for $2 \leq i < n$ and with the property that $\alpha_2 x_2, \ldots, \alpha_n x_n$ form a basis of M'. It follows that $x_1, \ldots, x_n$ are part of a basis of $F = Ax_1 \oplus F'$, and that $\alpha_1 x_1, \ldots, \alpha_n x_n$ for $\alpha_1 := \varphi(x)$ form a basis of $M = Ax \oplus M'$. Thus, to derive the existence part of Theorem 2 it remains only to show that $\alpha_1 \,|\, \alpha_2$. To justify the latter divisibility consider a linear functional $\varphi_2 \in F^*$ satisfying $\varphi_2(x_2) = 1$. Then we get $\varphi(x) \,|\, \varphi_2(\alpha_2 x_2)$ and hence $\alpha_1 \,|\, \alpha_2$, since $\operatorname{cont}(x) \,|\, \operatorname{cont}(\alpha_2 x_2)$ by the choice of x and since $\operatorname{cont}(\alpha_2 x_2) \,|\, \varphi_2(\alpha_2 x_2)$ by Lemma 4 (ii). Thereby the existence part of Theorem 2 is clear.

It remains to prove the uniqueness of the α_i. In view of further applications, we do this in a slightly more general setting.

Lemma 5. *Let A be a principal ideal domain and let $Q \simeq \bigoplus_{i=1}^{n} A/\alpha_i A$ be an A-module, where $\alpha_1, \ldots, \alpha_n \in A - \{0\}$ are nonunits such that $\alpha_i \,|\, \alpha_{i+1}$ for $1 \leq i < n$. Then the elements $\alpha_1, \ldots, \alpha_n$ are uniquely determined by Q, up to associatedness.*

Proof. For technical reasons we invert the numbering of the elements α_i and consider two decompositions

$$Q \simeq \bigoplus_{i=1}^{n} A/\alpha_i A \simeq \bigoplus_{j=1}^{m} A/\beta_j A$$

such that $\alpha_{i+1} \,|\, \alpha_i$ for $1 \leq i < n$, as well as $\beta_{j+1} \,|\, \beta_j$ for $1 \leq j < m$. If there exists an index $k \leq \min\{m, n\}$ satisfying $\alpha_k A \neq \beta_k A$, we choose k minimal with this property. Since $\alpha_i A = \beta_i A$ for $1 \leq i < k$ and since all elements $\alpha_{k+1}, \ldots, \alpha_n$ are divisors of α_k, we can decompose $\alpha_k Q$ as follows:

$$\alpha_k Q \simeq \bigoplus_{i=1}^{k-1} \alpha_k \cdot \big(A/\alpha_i A\big) \simeq \bigoplus_{i=1}^{k-1} \alpha_k \cdot \big(A/\alpha_i A\big) \oplus \alpha_k \cdot \big(A/\beta_k A\big) \oplus \ldots \ .$$

Now use Lemma 1. Comparing both decompositions and using the fact that $l_A(Q) < \infty$, we see that $l_A(\alpha_k \cdot (A/\beta_k A)) = 0$. This means that $\alpha_k \cdot (A/\beta_k A) = 0$ and hence $\alpha_k A \subset \beta_k A$. Likewise, we get $\beta_k A \subset \alpha_k A$ and thus $\alpha_k A = \beta_k A$, contradicting our assumption on k. Therefore, we must have $\alpha_i A = \beta_i A$ for all indices i satisfying $1 \leq i \leq \min\{m, n\}$. Furthermore, if $m \leq n$, we conclude from Lemma 1 again that $\bigoplus_{i=m+1}^{n} A/\alpha_i A$ is of length 0 and hence vanishes. Consequently, $m = n$ and α_i is associated to β_i for $i = 1, \ldots, n$. $\square$

It remains to explain how to derive the uniqueness assertion of Theorem 2 from Lemma 5. To do this, assume in the situation of the theorem that we have elementary divisors $\alpha_1, \ldots, \alpha_n$ satisfying $\alpha_i \,|\, \alpha_{i+1}$, as well as $\beta_1, \ldots, \beta_n$ satisfying $\beta_i \,|\, \beta_{i+1}$ for $1 \leq i < n$. Then, according to Remark 3 (whose proof was based on the existence assertion of Theorem 2 and did not require uniqueness), we get an isomorphism

$$\bigoplus_{i=1}^{n} A/\alpha_i A \simeq \bigoplus_{i=1}^{n} A/\beta_i A.$$

Since A/aA vanishes for units $a \in A$, we can conclude from Lemma 5 that the nonunits among $\alpha_1, \alpha_2, \ldots$ coincide with the nonunits among $\beta_1, \beta_2, \ldots$, up to associatedness. Since the remaining α_i and β_i are units, we get $\alpha_i A = \beta_i A$ for $1 \leq i \leq n$, thereby ending the proof of Theorem 2. □

Next we want to give a more constructive characterization of elementary divisors, which will be of special interest for explicit computations.

Proposition 6. *Let A be a principal ideal domain, F a finite free A-module with basis $x_1, \ldots, x_r$, as well as $M \subset F$ a submodule of rank n with corresponding elementary divisors $\alpha_1, \ldots, \alpha_n$. Furthermore, let $z_1, \ldots, z_m \in M$ be a (not necessarily free) system of generators of M. For $j = 1, \ldots, m$ assume $z_j = \sum_{i=1}^{r} a_{ij} x_i$ for coefficients $a_{ij} \in A$, and let μ_t for $t = 1, \ldots, n$ be the greatest common divisor of all t-minors of the coefficient matrix $D = (a_{ij})$.*[4] *Then $\mu_t = \alpha_1 \ldots \alpha_t$, and in particular, $\alpha_1 = \mu_1$ as well as $\alpha_t \mu_{t-1} = \mu_t$ for $t = 2, \ldots, n$.*

In the present situation, the elements $\alpha_1, \ldots, \alpha_n$ are referred to as the elementary divisors of the matrix D.

Proof. To start with, let us verify the assertion for $t = 1$. Note that $(\alpha_1) \subset A$ is the ideal generated by all elements of type $\varphi(z)$ for $z \in M$ and $\varphi \in F^*$; this can be read from the assertion of Theorem 2 or from its proof. In particular, evaluate the linear functionals of the dual basis attached to $x_1, \ldots, x_r$ at the elements z_j. Thereby it is seen that the ideal (α_1) can also be generated by the coefficients a_{ij}. However, this means that α_1 is the greatest common divisor of all 1-minors of D.

To prove the assertion for arbitrary t it is convenient to use the *t-fold exterior power* $\bigwedge^t F$ of F. For our purposes it is enough to fix the basis $x_1, \ldots, x_r$ of F and to define $\bigwedge^t F$ as the free A-module with basis given by the symbols $x_{i_1} \wedge \ldots \wedge x_{i_t}$, where $1 \leq i_1 < \ldots < i_t \leq r$. Then, for a permutation $\pi \in \mathfrak{S}_t$, i.e., a bijective self-map of $\{1, \ldots, t\}$, we let

$$x_{i_{\pi(1)}} \wedge \ldots \wedge x_{i_{\pi(t)}} = (\operatorname{sgn} \pi) \cdot x_{i_1} \wedge \ldots \wedge x_{i_t},$$

where $\operatorname{sgn} \pi$ is the sign of the permutation π; cf. 5.3. To extend the notion of the t-fold "exterior product" $x_{i_1} \wedge \ldots \wedge x_{i_t}$ to arbitrary indices $i_1, \ldots, i_t \in \{1, \ldots, r\}$,

[4] The t-minors of D are the determinants of the $(t \times t)$ submatrices of D. Since D, viewed as an $(r \times m)$ matrix with coefficients in the field of fractions $Q(A)$, is of rank n, we have $n \leq \min(r, m)$.

we set $x_{i_1} \wedge \ldots \wedge x_{i_t} = 0$ for indices i_j that are not distinct. Then, finally, we can define the exterior product $z_1 \wedge \ldots \wedge z_t$ of arbitrary elements $z_1, \ldots, z_t \in F$ by A-multilinear extension. Due to its construction, this product is multilinear and alternating in its factors. For example, for elements of type $z_j = \sum_{i=1}^r a_{ij} x_i$ we get

$$\begin{aligned} z_1 \wedge \ldots \wedge z_t &= \Big(\sum_{i=1}^r a_{i1} x_i\Big) \wedge \ldots \wedge \Big(\sum_{i=1}^r a_{it} x_i\Big) \\ &= \sum_{i_1,\ldots,i_t=1}^r a_{i_1 1} \ldots a_{i_t t}\; x_{i_1} \wedge \ldots \wedge x_{i_t} \\ &= \sum_{1 \le i_1 < \ldots < i_t \le r} \Big(\sum_{\pi \in \mathfrak{S}_t} (\operatorname{sgn} \pi) \cdot a_{i_{\pi(1)} 1} \ldots a_{i_{\pi(t)} t}\Big)\; x_{i_1} \wedge \ldots \wedge x_{i_t}, \end{aligned}$$

where the coefficients $\sum_{\pi \in \mathfrak{S}_t} (\operatorname{sgn} \pi) \cdot a_{i_{\pi(1)} 1} \ldots a_{i_{\pi(t)} t}$ equal the t-minors of the coefficient matrix when $z_1, \ldots, z_t$ are represented as linear combinations of the basis $x_1, \ldots, x_r$. It should be noted that this computation can also be used to show that the above definition of $\bigwedge^t F$, together with the t-fold exterior product of elements in F, is naturally independent of the choice of the basis $x_1, \ldots, x_r$.

Now we turn back to the system of generators $z_1, \ldots, z_m$ of M that is given in the setting of the proposition, assuming for the moment that the z_i form a basis of M or, more specifically, that we have $z_i = \alpha_i x_i$ for $i = 1, \ldots, m$ and elements $\alpha_i \in A - \{0\}$, where $\alpha_i \mid \alpha_{i+1}$. Note that due to the elementary divisor theorem, such a setting can always be obtained for $m = n$ if we choose $x_1, \ldots, x_r$ as well as $z_1, \ldots, z_m$ in an appropriate way. Then we see that the t-fold exterior power $\bigwedge^t M$ is naturally a submodule of $\bigwedge^t F$. Indeed, the elements $x_{i_1} \wedge \ldots \wedge x_{i_t}$ for $1 \le i_1 < \ldots < i_t \le r$ are a basis of $\bigwedge^t F$, while the elements $\alpha_{i_1} \ldots \alpha_{i_t} x_{i_1} \wedge \ldots \wedge x_{i_t}$ for $1 \le i_1 < \ldots < i_t \le m$ are a basis of $\bigwedge^t M$. Then we discover the product $\alpha_1 \ldots \alpha_t$ as the first elementary divisor of the problem $\bigwedge^t M \subset \bigwedge^t F$, in accordance with the initial discussion for $t = 1$.

In the general setting of the proposition, the elements $z_1, \ldots, z_m$ constitute a system of generators of M that is not necessarily free. Nevertheless, it follows that the t-fold exterior products of type $z_{i_1} \wedge \ldots \wedge z_{i_t}$ for $1 \le i_1 < \ldots < i_t \le m$ generate the A-module $\bigwedge^t M$; use a computation as given above. Due to the consideration for $t = 1$ at the beginning of the proof, the first elementary divisor of the problem $\bigwedge^t M \subset \bigwedge^t F$ coincides with the greatest common divisor of all coefficients from A that are needed to represent the elements $z_{i_1} \wedge \ldots \wedge z_{i_t}$ as linear combinations of the basis $x_{i_1} \wedge \ldots \wedge x_{i_t}$, $1 \le i_1 < \ldots < i_t \le r$. However, as seen before, these coefficients are given by the t-minors of the matrix D, and it follows that the first elementary divisor attached to $\bigwedge^t M \subset \bigwedge^t F$ is given by μ_t. On the other hand, this elementary divisor has already been recognized as $\alpha_1 \ldots \alpha_t$, so that $\mu_t = \alpha_1 \ldots \alpha_t$. □

In the following we want to describe a constructive method that allows one to determine elementary divisors in the setting of Proposition 6 for the matrix $D = (a_{ij})$ or the submodule $M \subset F$. However, as a basic prerequisite, we have

to assume that A is a *Euclidean domain*. We consider A^m a free A-module with canonical basis $e_1, \ldots, e_m$ and look at the A-homomorphism

$$A^m \longrightarrow F, \qquad e_j \longmapsto z_j,$$

given by the matrix D relative to the bases $e_1, \ldots, e_m$ of A^m and $x_1, \ldots, x_r$ of F. Below we will show that we can transform D using elementary row and column transformations—i.e., by interchanging rows (resp. columns), as well as addition of a scalar multiple of a row (resp. column) to a further row (resp. column)—into a matrix of type

$$\begin{pmatrix} \alpha_1 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ 0 & \alpha_2 & \ldots & 0 & 0 & \ldots & 0 \\ . & . & \ldots & . & . & \ldots & . \\ 0 & 0 & \ldots & \alpha_n & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ . & . & \ldots & . & . & \ldots & . \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \end{pmatrix},$$

where $\alpha_i \mid \alpha_{i+1}$ for $1 \leq i < n$. These transformations can be interpreted as multiplication of D from the left or the right by invertible matrices $S \in A^{(r\times r)}$, resp. $T \in A^{(m\times m)}$. Then the resulting matrix SDT still describes the map f, of course relative to new bases $e'_1, \ldots, e'_m$ of A^m and $x'_1, \ldots, x'_r$ of F. In any case, M is generated by $\alpha_1 x'_1, \ldots, \alpha_n x'_n$ and the coefficients $\alpha_1, \ldots, \alpha_n$ are recognized as the elementary divisors of D, resp. $M \subset F$.

To find out about the necessary row and column transformations to be applied to the matrix D, we use the Euclidean function $\delta\colon A - \{0\} \longrightarrow \mathbb{N}$ of the Euclidean ring A. For $D = 0$ nothing has to be shown. Therefore, assume $D \neq 0$. It is our strategy to apply a series of elementary row and column transformations to D such that the minimum

$$d = \min\big\{\delta(a)\,;\, a \text{ is a coefficient } \neq 0 \text{ of } D\big\}$$

decreases step by step. Since δ takes values in $\mathbb{N}$, it is clear that the process must end after finitely many steps, and hence that d becomes minimal then. If $a \neq 0$ is a coefficient of the corresponding transformed matrix satisfying $\delta(a) = d$, we show with the help of Euclidean division that a divides all other coefficients of the matrix and hence that a is the first elementary divisor of D.

In more detail, we proceed as follows. By interchanging rows, resp. columns, of D we may assume $d = \delta(a_{11})$, i.e., that $\delta(a_{11})$ is minimal among all $\delta(a_{ij})$ such that $a_{ij} \neq 0$. If one of the elements of the first column, say a_{i1}, is not divisible by a_{11}, we apply Euclidean division by a_{11} to a_{i1}, say $a_{i1} = qa_{11} + b$, where $\delta(b) < \delta(a_{11})$ with $b \neq 0$, and subtract the q-fold first row from the ith row. As a result, the element b occurs at position $(i, 1)$, and we see that the minimum d of all δ-values of nonzero coefficients of D has decreased. To further reduce d we start the transformation process anew, until all elements of the first column are divisible by the coefficient at position $(1, 1)$.

In the same way we proceed with the elements of the first row. Since d assumes values in $\mathbb{N}$ and thus can decrease only finitely many times, we may assume after finitely many steps that every element of the first column as well as of the first row is a multiple of a_{11}. Adding suitable multiples of the first row to the remaining rows, we may assume $a_{i1} = 0$ for $i > 1$. In the same way we proceed with the first row, thereby achieving $a_{i1} = a_{1j} = 0$ for $i, j > 1$. In addition, we may assume that the minimum d coincides with $\delta(a_{11})$; otherwise we start the transformation process again from the beginning on.

If there should exist indices $i, j > 1$ such that $a_{11} \nmid a_{ij}$, we apply Euclidean division by a_{11} to a_{ij}, say $a_{ij} = qa_{11} + b$, where $b \neq 0$ and $\delta(b) < \delta(a_{11})$. Then we add the first row to the ith row and thereafter subtract the q-fold first column from the jth column. In this way, alongside other changes, a_{ij} is replaced by b, where now $\delta(b) < \delta(a_{11})$, and hence the minimum d of all δ-values of coefficients has decreased again. To continue, we start the transformation process anew from the beginning, until after finitely many steps we arrive at a matrix (a_{ij}), where now $a_{i1} = a_{1j} = 0$ for $i, j > 1$, as well as $a_{11} \mid a_{ij}$ for all $i, j > 1$. Then we consider the submatrix $(a_{ij})_{i,j>1}$ of $D = (a_{ij})$. Unless it is already zero, we can transform it in the same way that we did with D. Using an inductive argument, we arrive after finitely many steps at a matrix of the desired type in which the elementary divisors appear on the main diagonal and all other coefficients are zero.

Next we want to derive from the elementary divisor theorem the so-called *fundamental theorem of finitely generated modules over principal ideal domains*, whose assertion we split up into Corollaries 7 and 8. Note that in the following, A is a *principal ideal domain* again.

Corollary 7. *Let M be a finitely generated A-module and $T \subset M$ its corresponding torsion submodule. Then T is finitely generated, and there is a free submodule $F \subset M$ such that $M = T \oplus F$, where* $\operatorname{rank} M = \operatorname{rank} F$. *In particular, M is free if it does not admit torsion.*

Corollary 8. *Let M be a finitely generated torsion module over A, and let $P \subset A$ be a system of representatives of all prime elements of A. For $p \in P$ denote by*

$$M_p = \{x \in M \,;\, p^n x = 0 \text{ for some } n \in \mathbb{N}\}$$

the so-called submodule of p-torsion *in M. Then M decomposes into the direct sum*

$$M = \bigoplus_{p \in P} M_p,$$

where M_p is trivial for almost all $p \in P$. Furthermore, there exist natural numbers $1 \leq \nu(p,1) \leq \ldots \leq \nu(p, r_p)$ for each $p \in P$ such that

$$M_p \simeq \bigoplus_{j_p=1}^{r_p} A/p^{\nu(p,j_p)}A,$$

where $r_p = 0$ for almost all p. The numbers $r_p, \nu(p, j_p)$ are uniquely determined by the isomorphism

$$M \simeq \bigoplus_{p \in P} \bigoplus_{j_p=1}^{r_p} A/p^{\nu(p,j_p)}A.$$

Combining both results, we see that every finitely generated A-module M is isomorphic to a direct sum of type

$$A^d \oplus \bigoplus_{p \in P} \bigoplus_{j_p=1}^{r_p} A/p^{\nu(p,j_p)}A,$$

where the numbers d, r_p, and $\nu(p, j_p)$ as above are uniquely determined by M. This is the actual assertion of the fundamental theorem of finitely generated modules over principal ideal domains. Before we discuss its proof, let us formulate this theorem especially for finitely generated $\mathbb{Z}$-modules, where it is known as the *fundamental theorem of finitely generated abelian groups.*

Corollary 9. *Let G be a finitely generated abelian group, and let P be the set of prime numbers. Then G admits a decomposition into subgroups*

$$G = F \oplus \bigoplus_{p \in P} \bigoplus_{j_p=1}^{r_p} G_{p,j_p},$$

where F is finitely generated and free, say $F \simeq \mathbb{Z}^d$, where G_{p,j_p} is cyclic of p-power order, say $G_{p,j_p} \simeq \mathbb{Z}/p^{\nu(p,j_p)}\mathbb{Z}$ for $1 \leq \nu(p,1) \leq \ldots \leq \nu(p,r_p)$, and where r_p is zero for almost all $p \in P$. The numbers $d, r_p, \nu(p, j_p)$ are uniquely determined by G, and the same is true for the subgroups $G_p = \bigoplus_{j_p=1}^{r_p} G_{p,j_p}$.

If G is a finitely generated torsion group, in the sense of a finitely generated torsion $\mathbb{Z}$-module, then G does not admit a free part and therefore consists of only finitely many elements, as we can conclude from Corollary 9. On the other hand, it is clear that every finite abelian group is a torsion group.

Let us turn now to the *proof of Corollary* 7. If $z_1, \ldots, z_r$ generate M as an A-module, we can define an A-homomorphism $f\colon A^r \longrightarrow M$ by mapping the canonical basis of A^r to $z_1, \ldots, z_r$. Then f is surjective, and we get an isomorphism $M \simeq A^r/\ker f$ from the fundamental theorem on homomorphisms. Next we apply the elementary divisor theorem to the submodule $\ker f \subset A^r$. Hence, there exist elements $x_1, \ldots, x_r$ forming a basis of A^r, as well as elements $\alpha_1, \ldots, \alpha_n \in A$, $n = \operatorname{rank}(\ker f)$, such that $\alpha_1 x_1, \ldots, \alpha_n x_n$ form a basis of $\ker f$. From this we get an isomorphism

$$M \simeq A^{r-n} \oplus \bigoplus_{i=1}^{n} A/\alpha_i A,$$

where $\bigoplus_{i=1}^{n} A/\alpha_i A$ corresponds to the torsion submodule $T \subset M$, and where A^{r-n} corresponds to a free submodule $F \subset M$, thus implying $M = T \oplus F$. Furthermore, $T \simeq \bigoplus_{i=1}^{n} A/\alpha_i A$ is finitely generated, thereby settling the proof of Corollary 7. □

To approach the *proof of Corollary* 8, assume that M is a torsion module. Then using the setting of the proof of Corollary 7, we see that M is isomorphic to the direct sum $\bigoplus_{i=1}^{n} A/\alpha_i A$. Now decompose the α_i into prime factors, say $\alpha_i = \varepsilon_i \prod_{p \in P} p^{\nu(p,i)}$ for units ε_i and exponents $\nu(p,i)$ that are trivial for almost all p. Applying the Chinese remainder theorem 2.4/14, we see that

$$A/\alpha_i A \simeq \bigoplus_{p \in P} A/p^{\nu(p,i)} A,$$

and hence that

$$M \simeq \bigoplus_{p \in P} \bigoplus_{i=1}^{n} A/p^{\nu(p,i)} A.$$

Clearly, the term $\bigoplus_{i=1}^{n} A/p^{\nu(p,i)} A$ of the latter decomposition corresponds to the submodule $M_p \subset M$ of p-torsion, which is unique; observe that the residue class of p in residue class rings of type $A/p'^r A$ for $p' \in P - \{p\}$ is always a unit. In particular, the above decomposition leads to a decomposition $M = \bigoplus_{p \in P} M_p$. Now, looking at the decomposition

$$M_p \simeq \bigoplus_{i=1}^{n} A/p^{\nu(p,i)} A,$$

we abandon all terms $A/p^{\nu(p,i)} A$ in which $\nu(p,i) = 0$, since these are trivial. Furthermore, renumbering the exponents $\nu(p,i)$ for fixed p in ascending order, say

$$M_p \simeq \bigoplus_{j_p=1}^{r_p} A/p^{\nu(p,j_p)} A,$$

where $1 \leq \nu(p,1) \leq \ldots \leq \nu(p,r_p)$, the existence assertion of Corollary 8 is clear, whereas the uniqueness is a consequence of Lemma 5. □

The methods and results treated in the present section rely in a fundamental way on the ideal-theoretic characterization 2.4/13 of the greatest common divisor, thus on a characterization that is valid in principal ideal domains, but not necessarily in unique factorization domains; cf. Section 2.4, Exercise 2. This is why it is not possible to extend the theory of elementary divisors to finitely generated modules over general unique factorization domains.

Exercises

Let A always be a principal ideal domain.

1. *Consider a decomposition $M = T \oplus F$ of a finitely generated A-module M into a torsion submodule T and a free submodule F. Discuss the uniqueness of such a decomposition. Study the same problem for a decomposition of type $M = M' \oplus M''$, where $M' \simeq A/p^r A$ and $M'' \simeq A/p^s A$ for a prime element $p \in A$.*

2. *A torsion-free A-module is free if it is finitely generated. Can we extend this result to arbitrary torsion-free A-modules?*

3. *Derive the theory of canonical forms for endomorphisms of finite-dimensional vector spaces from Corollary* 8.

4. Determine the elementary divisors of the following matrix:

$$\begin{pmatrix} 2 & 6 & 8 \\ 3 & 1 & 2 \\ 9 & 5 & 4 \end{pmatrix} \in \mathbb{Z}^{(3\times 3)}$$

5. Let $a_{11}, \ldots, a_{1n} \in A$ be elements such that $\gcd(a_{11}, \ldots, a_{1n}) = 1$. Show that there are elements $a_{ij} \in A$, $i = 2, \ldots, n$, $j = 1, \ldots, n$, such that the matrix $(a_{ij})_{i,j=1,\ldots,n}$ is invertible in $A^{(n\times n)}$.

6. Consider an A-homomorphism $f\colon L \longrightarrow M$ between finitely generated free A-modules. Show:
   (i) There exists a free submodule $F \subset L$ such that $L = \ker f \oplus F$.
   (ii) There exist bases $x_1, \ldots, x_m$ of L and $y_1, \ldots, y_n$ of M, as well as elements $\alpha_1, \ldots, \alpha_r \in A - \{0\}$, $r \leq \min\{m, n\}$, such that $f(x_i) = \alpha_i y_i$ for $i = 1, \ldots, r$ and $f(x_i) = 0$ for $i > r$. In addition, we can obtain the divisibility relations $\alpha_i \,|\, \alpha_{i+1}$ for $1 \leq i < r$.

7. Give a simple argument for extending the assertion of Theorem 2 to finite-rank submodules M of free A-modules F that are not necessarily of finite rank themselves.

3. Algebraic Field Extensions

Background and Overview

First let us indicate how algebraic equations are related to algebraic field extensions. We start with the simple case of an algebraic equation with rational coefficients, say $f(x) = 0$, where $f \in \mathbb{Q}[X]$ is a monic polynomial of degree ≥ 1. The problem of where to look for solutions of such an equation and how to use them in computations will be postponed for the moment, since we will assume that the fundamental theorem of algebra is already known. Thus, we will use the fact that f admits a zero α in $\mathbb{C}$, where now $f(\alpha) = 0$ has to be interpreted as an equation valid in $\mathbb{C}$. However, to better describe the "nature" of the zero α, one tries to construct a domain of numbers, as small as possible, where the equation $f(\alpha) = 0$ still makes sense. For example, such a domain is given by the smallest subring of $\mathbb{C}$ containing $\mathbb{Q}$ and α, hence by

$$\mathbb{Q}[\alpha] = \big\{ g(\alpha)\,;\, g \in \mathbb{Q}[X] \big\}.$$

Using the epimorphism $\varphi\colon \mathbb{Q}[X] \longrightarrow \mathbb{Q}[\alpha]$, $g \longmapsto g(\alpha)$, it is easily seen that $\mathbb{Q}[\alpha]$ is even a *field*. Indeed, $\mathbb{Q}[X]$ is a principal ideal domain, and hence $\ker\varphi$ is a principal ideal, say $\ker\varphi = (q)$. Of course, $f \in \ker\varphi$ shows that q is nonzero and therefore can be assumed to be monic in $\mathbb{Q}[X]$. Now the fundamental theorem on homomorphisms 2.3/5 implies that φ induces an isomorphism $\mathbb{Q}[X]/(q) \xrightarrow{\sim} \mathbb{Q}[\alpha]$, and it follows from 2.3/8 that q is a prime polynomial, the so-called *minimal polynomial* of α. If f is irreducible, we can even conclude that $q = f$ by a divisibility argument. In any case, the ideal (q) is maximal in $\mathbb{Q}[X]$ by 2.4/6, so that $\mathbb{Q}[\alpha] \simeq \mathbb{Q}[X]/(q)$ is indeed a field. We say that $\mathbb{Q}[\alpha]$ arises from $\mathbb{Q}$ by *adjoining* the zero α. In the same way, one can adjoin further zeros of f (or of other polynomials with coefficients in $\mathbb{Q}[\alpha]$) to $\mathbb{Q}[\alpha]$.

We can draw some important conclusions from these observations. First of all, we see that $\mathbb{Q}[\alpha]$, as a $\mathbb{Q}$-vector space, is of finite dimension and hence that $\mathbb{Q} \subset \mathbb{Q}[\alpha]$ is a *finite* field extension; cf. 3.2/6. Using a simple dimension argument from linear algebra, this implies that *every* element of $\mathbb{Q}[\alpha]$ may be viewed as the solution of an algebraic equation with coefficients in $\mathbb{Q}$, hence that $\mathbb{Q} \subset \mathbb{Q}[\alpha]$ is an *algebraic* field extension, as we will say; cf. 3.2/7. In particular, when looking at the extension $\mathbb{Q} \subset \mathbb{Q}[\alpha]$, we are dealing with a whole class of related algebraic equations simultaneously.

In the following we assume that the polynomial $f \in \mathbb{Q}[X]$ is *irreducible*; let $\alpha_1, \ldots, \alpha_n \in \mathbb{C}$ be its zeros. Then we obtain for every $i = 1, \ldots, n$ an isomorphism $\mathbb{Q}[\alpha_i] \simeq \mathbb{Q}[X]/(f)$, as constructed before, with α_i corresponding to the residue class of X modulo (f). In particular, given two indices i, j, there is an isomorphism $\sigma_{ij} \colon \mathbb{Q}[\alpha_i] \xrightarrow{\sim} \mathbb{Q}[\alpha_j]$ such that $\sigma_{ij}(\alpha_i) = \alpha_j$, and we see that all zeros of f are of equal priority. The isomorphisms σ_{ij} open up a first view on the Galois theory of the equation $f(x) = 0$. In the special case that the subfield $L = \mathbb{Q}[\alpha_i] \subset \mathbb{C}$ is independent of i, the σ_{ij} constitute (not necessarily distinct) automorphisms of L, these being the members of the Galois group of the equation $f(x) = 0$. In general, one considers instead of $\mathbb{Q}[\alpha_i]$ the so-called *splitting field* $L = \mathbb{Q}[\alpha_1, \ldots, \alpha_n]$ of f, which is constructed from $\mathbb{Q}$ by adjoining all zeros of f. One can show, using the primitive element theorem 3.6/12, that there is an irreducible polynomial $g \in \mathbb{Q}[X]$ with zeros $\beta_1, \ldots, \beta_r \in \mathbb{C}$ such that $L = \mathbb{Q}[\beta_j]$ for $j = 1, \ldots, r$. Then we are in the situation of the special case, as just considered, and we can define the Galois group of the equation $f(x) = 0$ as the corresponding group of the equation $g(x) = 0$.

Up to now we have restricted ourselves to field extensions of $\mathbb{Q}$. But how can we proceed when we want to replace $\mathbb{Q}$ by an arbitrary field K? In principle, no changes are necessary, as we will see in the present chapter. The only auxiliary tool we might need is a certain replacement of the fundamental theorem of algebra. To begin with, we study in 3.2 the relationship between finite and algebraic field extensions, without departing from special algebraic equations that are to be solved; a generalization of these considerations to the ring-theoretic level is presented in 3.3. Then, in 3.4, we approach the problem of constructing for an irreducible algebraic equation $f(x) = 0$, where $f \in K[X]$, an extension field L of K such that f admits a zero α in L. If L is such a field, we can consider the field $K[\alpha]$ as before; it is isomorphic to $K[X]/(f)$, since f is irreducible. On the other hand, we can define L simply by $K[X]/(f)$, observing that the residue class of X is a zero of f; this is *Kronecker's construction*; cf. 3.4/1. The construction allows us to successively adjoin zeros of polynomials to K. For example, if we have adjoined a zero α_1 of f to K, then f admits a factorization of type $f = (X - \alpha_1) f_1$ in $K[\alpha_1][X]$. In a next step, we can adjoin a zero α_2 of f_1 to $K[\alpha_1]$, and so on. In this way, we obtain after finitely many steps a splitting field L of f, i.e., an extension field of K, over which f factorizes completely into linear factors, and which is minimal in the sense that it is obtained by adjoining all zeros of f to K.

Although in principle, Kronecker's construction is sufficient for the study of algebraic equations, it is nevertheless desirable for several reasons to obtain a "true" substitute for the fundamental theorem of algebra. Therefore, we construct in 3.4 an *algebraic closure* $\overline{K}$ of K. In fact, we use a method of E. Artin that allows us to adjoin to K *all zeros* of the polynomials in $K[X]$ in one step. The field $\overline{K}$ is algebraic over K and has the property that every nonconstant polynomial in $\overline{K}[X]$ factorizes completely into linear factors. This construction makes it possible to talk about "the" zeros of f. Then, for example, the construction of splitting fields of a family of polynomials, as considered in 3.5,

no longer poses any problems. We thereby arrive at the notion of *normal field extensions*, a prestage of Galois extensions.

Finally, let us mention the phenomenon of *inseparability*, which occurs when one is dealing with fields K of characteristic > 0 instead of extension fields of $\mathbb{Q}$. The characteristic of K is defined as the smallest integer $p > 0$ such that $p \cdot 1 = 0$, where we put $p = 0$ if such a number does not exist; cf. 3.1. A polynomial $f \in K[X]$ is called *separable* if it admits only simple zeros (in an algebraic closure of K), and *purely inseparable* if it admits precisely one zero, the latter being of multiplicity $\deg f$. Irreducible polynomials over fields of characteristic 0 are always separable, while the same is not true over fields of characteristic > 0. Based on the corresponding notion for polynomials, we study separable algebraic field extensions in 3.6 and, as their counterpart, purely inseparable extensions in 3.7. Of special interest are the results 3.7/4 and 3.7/5, asserting that it is possible to split up algebraic field extensions into a separable and a purely inseparable part. As an example, we study in 3.8 special fields of characteristic > 0, namely finite fields.

The chapter closes in 3.9 with a first look at the beginnings of algebraic geometry, i.e., the theory of algebraic equations in several variables.

3.1 The Characteristic of a Field

Given a ring K, there exists a unique ring homomorphism

$$\varphi\colon \mathbb{Z} \longrightarrow K,$$

namely, the one characterized by $n \longmapsto n \cdot 1$. By the fundamental theorem on homomorphisms for rings 2.3/4, it gives rise to a monomorphism $\mathbb{Z}/\ker\varphi \hookrightarrow K$, where $\ker\varphi$ is a principal ideal in $\mathbb{Z}$ by 2.4/3. If K is an integral domain, for example a field, then $\mathbb{Z}/\ker\varphi$ is an integral domain as well, and it follows that $\ker\varphi$ is a prime ideal. Hence, $\ker\varphi$ is either the zero ideal or an ideal generated by a prime number p; cf. 2.3/11. Accordingly, 0 or p is called the *characteristic* of the integral domain or the field K.

Definition 1. *Let K be a field (or, more generally, an integral domain) and let $\varphi\colon \mathbb{Z} \longrightarrow K$ be the canonical ring homomorphism discussed before. Furthermore, let $p \in \mathbb{N}$ be a generating element of the principal ideal* $\ker\varphi$. *Then p is called the* characteristic *of K, and we write* $p = \operatorname{char} K$.

The fields $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ are all of characteristic 0, whereas for a prime number p the field $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ consisting of p elements is of characteristic p. A subring T of a field K is called a *subfield* if T is a field itself. Of course, we have $\operatorname{char} K = \operatorname{char} T$ in this case. Since the intersection of subfields of a field K is a subfield again, it follows that K contains a unique smallest subfield P, namely the one given by the intersection of all subfields in K. The field P is referred to as the *prime (sub)field* in K.

Proposition 2. *Let K be a field and $P \subset K$ its prime subfield. Then*:

(i) $\operatorname{char} K = p > 0 \Longleftrightarrow P \simeq \mathbb{F}_p$ *for p prime.*

(ii) $\operatorname{char} K = 0 \Longleftrightarrow P \simeq \mathbb{Q}$.

In particular, up to isomorphism, the only prime subfields that can occur, are $\mathbb{F}_p$ for a prime number p, as well as $\mathbb{Q}$.

Proof. We have $\operatorname{char} \mathbb{F}_p = p$, as well as $\operatorname{char} \mathbb{Q} = 0$. Since $\operatorname{char} P = \operatorname{char} K$, we get $\operatorname{char} K = p$ from $P \simeq \mathbb{F}_p$, and $\operatorname{char} K = 0$ from $P \simeq \mathbb{Q}$. This justifies in each of the cases (i) and (ii) the implication "$\Longleftarrow$".

To verify the reverse implications consider the canonical ring homomorphism $\varphi\colon \mathbb{Z} \longrightarrow K$; it factors through the prime subfield $P \subset K$, so that $\operatorname{im}\varphi \subset P$. If $\operatorname{char} K$ is a prime number p, we have $\ker\varphi = (p)$, and the image $\operatorname{im}\varphi \simeq \mathbb{Z}/(p)$ is a field by 2.3/6 or 2.4/6. Since P is the smallest subfield of K, we get $\operatorname{im}\varphi = P$, and hence $P \simeq \mathbb{F}_p$. Otherwise, if $\operatorname{char} K = 0$, we see that $\operatorname{im}\varphi$ is isomorphic to $\mathbb{Z}$. In particular, the field of fractions $Q(\operatorname{im}\varphi)$ is a subfield of P that is isomorphic to $\mathbb{Q}$, and it follows that $P = Q(\operatorname{im}\varphi) \simeq \mathbb{Q}$. □

Working in characteristic $p > 0$, we want to point out that the binomial expansion for p-powers of a sum of two elements takes a particularly simple form.

Remark 3. *Let p be a prime number and R an integral domain of characteristic p (or, more generally, a ring satisfying $p \cdot 1 = 0$). Then, for elements $a, b \in R$ and $r \in \mathbb{N}$, the binomial formula takes the following form*:

$$(a+b)^{p^r} = a^{p^r} + b^{p^r}, \qquad (a-b)^{p^r} = a^{p^r} - b^{p^r}.$$

Proof. Using induction on r, the assertion is easily reduced to the case $r = 1$. Now recall from example (2) at the end of Section 2.8 that the following divisibility relations hold:

$$p \,|\, \tbinom{p}{\nu} \quad \text{for} \quad \nu = 1, \ldots, p-1.$$

In particular, the specified binomial coefficients will vanish in R. Hence, the first of the asserted formulas follows for $r = 1$. If we use for p even, i.e., for $p = 2$, that $1 = -1$ holds in R, then also the second formula is clear. □

If K is a field of characteristic $p > 0$, then Remark 3 shows that the map

$$\sigma\colon K \longrightarrow K, \qquad a \longmapsto a^p,$$

respects the addition on K. Therefore, it defines a homomorphism of fields, the *Frobenius homomorphism* of K.

Exercises

1. *Can there exist homomorphisms between fields of different characteristic? Consider the same problem for integral domains.*

2. Does there exist a field consisting of six elements? Does there exist an integral domain consisting of six elements?

3. For a finite field K with multiplicative group K^*, consider $H = \{a^2\,;\, a \in K^*\}$ as a subgroup of K^* and show that

$$H = \begin{cases} K^* \text{ if } \operatorname{char} K = 2, \\ \text{a subgroup in } K^* \text{ of index 2 if} \operatorname{char} K > 2. \end{cases}$$

4. Let K be a field of characteristic > 0. Show that the Frobenius homomorphism $\sigma\colon K \longrightarrow K$ is an automorphism if K is finite. Check whether this assertion extends to the case that K is not necessarily finite.

5. Explicitly specify the Frobenius homomorphism on $\mathbb{F}_p$.

3.2 Finite and Algebraic Field Extensions

A pair of fields $K \subset L$, where K is a subfield of L, is called a *field extension*. More specifically, we will say that L is a field extending K or an extension field of K, or simply that L is a "field extension" of K. Given such a field extension, we can restrict the multiplication on L to a multiplication $K \times L \longrightarrow L$ and thereby view L as a K-vector space. For field extensions $K \subset L$ we will often use the notation L/K, at least if no confusion with factor group or factor ring constructions is possible. Dealing with field extensions L/K, we will also consider *intermediate fields*, i.e., subfields E such that $K \subset E \subset L$.

Definition 1. *Let $K \subset L$ be a field extension. Then the vector space dimension $[L : K] := \dim_K L$ is called the* degree *of L over K. The field extension is called* finite *or* infinite, *depending on whether $[L : K]$ is finite or infinite.*

Note that $L = K$ is obviously equivalent to $[L : K] = 1$.

Proposition 2 (Multiplicativity formula). *Let $K \subset L \subset M$ be field extensions. Then*

$$[M : K] = [M : L] \cdot [L : K].$$

Proof. If one of the degrees is infinite, the equation has to be interpreted in the obvious way. However, most interesting is the case in which both $[M : L]$ and $[L : K]$ are finite. In the latter case choose vector space bases $x_1, \ldots, x_m$ of L over K, as well as $y_1, \ldots, y_n$ of M over L. To verify the degree formula $[M : K] = [M : L] \cdot [L : K] = mn$ it is enough to check that the elements $x_i y_j$, $i = 1, \ldots m$, $j = 1, \ldots, n$, form a basis of M over K. In a first step we show that the linear independence of the x_i over K in conjunction with the linear independence of the y_j over L implies the linear independence of the $x_i y_j$ over K. To do this, consider coefficients $c_{ij} \in K$ satisfying $\sum_{ij} c_{ij} x_i y_j = 0$. Then we

can write the left-hand sum as a linear combination of the y_j with coefficients in L, namely

$$\sum_{j=1}^{n}\Big(\sum_{i=1}^{m} c_{ij}x_i\Big)y_j = 0.$$

Since the elements y_j are linearly independent over L, we get $\sum_i c_{ij}x_i = 0$ for all j. In the same way we conclude that $c_{ij} = 0$ for all i and j, since the x_i are linearly independent over K. Hence, it follows that the x_iy_j are linearly independent over K.

Just as easily we can see that the x_iy_j form a system of generators for M over K. Indeed, every element $z \in M$ admits a representation $z = \sum_{j=1}^{n} c_jy_j$ with coefficients $c_j \in L$, since the y_j form a system of generators for M over L. Further, for each j there is a representation $c_j = \sum_{i=1}^{m} c_{ij}x_i$ with coefficients $c_{ij} \in K$, since the x_i form a system of generators for L over K. Thereby we get

$$z = \sum_{j=1}^{n}\sum_{i=1}^{m} c_{ij}x_iy_j$$

and see that the x_iy_j form a system of generators, thus, altogether, a basis of M over K.

It remains to look at the case that at least one of the extensions M/L and L/K is not finite. In the first step of the proof we have shown for elements $x_1, \ldots, x_m \in L$ that are linearly independent over K, and for elements $y_1, \ldots, y_n \in M$ that are linearly independent over L, that the products x_iy_j are linearly independent over K. In other words, $[L : K] \geq m$ and $[M : L] \geq n$ imply $[M : K] \geq mn$. Therefore, $[M : K]$ is infinite as soon as one of the degrees $[M : L]$ and $[L : K]$ is infinite. □

Corollary 3. *If $K \subset L \subset M$ are field extensions such that $p = [M : K]$ is prime, then $L = K$ or $L = M$.*

Examples of finite field extensions of degree 2 are given by the extensions $\mathbb{R} \subset \mathbb{C}$ and $\mathbb{Q} \subset \mathbb{Q}[\sqrt{2}]$, where we view $\mathbb{Q}[\sqrt{2}]$ as a subring of $\mathbb{R}$. On the other hand, the extensions $\mathbb{Q} \subset \mathbb{R}$ as well as $K \subset K(X) = Q(K[X])$ for an arbitrary field K and a variable X are infinite.

Definition 4. *Given a field extension $K \subset L$, an element $\alpha \in L$ is called* algebraic *over K if it satisfies an algebraic equation over K, i.e., an equation*

$$\alpha^n + c_1\alpha^{n-1} + \ldots + c_n = 0$$

with coefficients $c_1, \ldots, c_n \in K$. This means that the kernel of the substitution homomorphism

$$\varphi\colon K[X] \longrightarrow L, \qquad g \longmapsto g(\alpha),$$

is nontrivial. Otherwise, α is called transcendental *over K.*

Furthermore, the extension field L is called algebraic *over K if every element $\alpha \in L$ is algebraic over K.*

For example, the nth root $\sqrt[n]{q} \in \mathbb{R}$ for some $q \in \mathbb{Q}$, $q \geq 0$, and $n \in \mathbb{N}\text{–}\{0\}$, is algebraic over $\mathbb{Q}$, since it is a zero of the polynomial $X^n - q \in \mathbb{Q}[X]$. Similarly, the complex number $e^{2\pi i/n}$ is an "nth root of unity" and therefore algebraic over $\mathbb{Q}$. However, in general it is not easy to decide whether a complex number z is algebraic over $\mathbb{Q}$, notably when z is constructed by means of methods from analysis; for example, see the transcendence problem for the numbers e and π that was mentioned in the introduction.

Remark 5. *Let $K \subset L$ be a field extension and $\alpha \in L$ an element that is algebraic over K. Then there exists a unique monic polynomial $f \in K[X]$ of smallest degree such that $f(\alpha) = 0$. The kernel of the substitution homomorphism*

$$\varphi\colon K[X] \longrightarrow L, \qquad g \longmapsto g(\alpha),$$

satisfies $\ker \varphi = (f)$*, so that in particular, f is prime and therefore irreducible. The polynomial f is called the* minimal polynomial *of α over K.*

Proof. Recall that $K[X]$ is a principal ideal domain by 2.4/3. Therefore, the ideal $\ker \varphi$ is generated by a polynomial $f \in K[X]$, where $f \neq 0$, since α is algebraic over K. Furthermore, such a generator is unique up to a multiplicative constant from K^*. Hence, if we require f to be monic, it becomes unique, and we see that f is the unique monic polynomial of smallest degree in $K[X]$ such that $f(\alpha) = 0$. Now observe that $\operatorname{im} \varphi$ is a subring of L and hence an integral domain. Furthermore, $\operatorname{im} \varphi$ is isomorphic to $K[X]/(f)$ by the fundamental theorem on homomorphisms 2.3/5. Therefore it follows that f is prime, and in particular, irreducible; cf. 2.3/8 and 2.4/6. □

Proposition 6. *Let $K \subset L$ be a field extension and $\alpha \in L$ algebraic over K with minimal polynomial $f \in K[X]$. Writing $K[\alpha]$ for the subring of L that is generated by α and K, i.e., for the image of the homomorphism $\varphi\colon K[X] \longrightarrow L$, $g \longmapsto g(\alpha)$, it follows that φ induces an isomorphism $K[X]/(f) \xrightarrow{\sim} K[\alpha]$.*

In particular, $K[\alpha]$ is a field, and in fact, a finite field extension of K of degree $[K[\alpha] : K] = \deg f$.

Proof. We have $K[\alpha] = \operatorname{im} \varphi \simeq K[X]/(f)$, due to the fundamental theorem on homomorphisms. Since $\ker \varphi = (f)$ is a nonzero prime ideal in $K[X]$, we conclude from 2.4/6 that (f) is a maximal ideal in $K[X]$. Therefore, $K[X]/(f)$ and hence $K[\alpha]$ are fields.

It remains to show that

$$\dim_K K[X]/(f) = \deg f.$$

Assume that $f = X^n + c_1 X^{n-1} + \ldots + c_n$ and hence $\deg f = n$. Furthermore, observe that Euclidean division by f is unique in $K[X]$ in the sense that for

every $g \in K[X]$ there are unique polynomials $q, r \in K[X]$ such that

$$g = qf + r, \qquad \deg r < n;$$

cf. 2.1/4. Thus, writing $\overline{X} \in K[X]/(f)$ for the residue class of $X \in K[X]$, it follows that every element of $K[X]/(f)$, viewed as a vector space over K, admits a unique representation as a linear combination with coefficients in K of the elements $\overline{X}^0, \ldots, \overline{X}^{n-1}$. Then $\overline{X}^0, \ldots, \overline{X}^{n-1}$ form a K-basis of $K[X]/(f)$, and using the isomorphism $K[\alpha] \simeq K[X]/(f)$, we can see that $\alpha^0, \ldots, \alpha^{n-1}$ form a K-basis of $K[\alpha]$. In particular, we get $\dim_K K[X]/(f) = \dim_K K[\alpha] = n$. □

Let us consider a simple example. Choosing a prime number p and an integer $n \in \mathbb{N} - \{0\}$, the nth root $\sqrt[n]{p} \in \mathbb{R}$ is algebraic over $\mathbb{Q}$, so that $\mathbb{Q}[\sqrt[n]{p}]$ is a finite field extension of $\mathbb{Q}$. The polynomial $f = X^n - p \in \mathbb{Q}[X]$ is irreducible by Eisenstein's criterion 2.8/1 and admits $\sqrt[n]{p}$ as a zero. Since it is monic, it must coincide with the minimal polynomial of $\sqrt[n]{p}$. Consequently, we get

$$\Big[\mathbb{Q}[\sqrt[n]{p}] : \mathbb{Q}\Big] = \deg f = n.$$

As a by-product we conclude that the extension $\mathbb{R}/\mathbb{Q}$ cannot be finite. Indeed, there are intermediate fields $\mathbb{Q}[\sqrt[n]{p}]$ of arbitrarily large degree.

Proposition 7. *Every finite field extension $K \subset L$ is algebraic.*

Proof. Let $[L : K] = n$ and consider an element $\alpha \in L$. Then the powers $\alpha^0, \ldots, \alpha^n$ give rise to a system of length $n+1$ and hence to a system in L that is linearly dependent over K. It follows that there exists a nontrivial equation

$$c_0\alpha^0 + \ldots + c_n\alpha^n = 0$$

with coefficients $c_i \in K$. Now let n' be maximal among all indices $i \in \{0, \ldots, n\}$ such that $c_i \neq 0$. Multiplying the equation by a suitable nonzero constant in K, we may assume that $c_{n'} = 1$ and thereby get an algebraic equation for α over K. □

At the end of the present section we will give an example of an algebraic field extension that is not finite. Thereby we see that the converse of Proposition 7 does not hold.

If $K \subset L$ is a field extension and $\mathfrak{A} = (\alpha_i)_{i \in I}$ a system of elements in L (or a subset of L), we can consider the subfield $K(\mathfrak{A}) \subset L$ that is generated by $\mathfrak{A}$ over K. It is the smallest subfield of L containing K as well as all elements α_i, i.e., $K(\mathfrak{A})$ is the intersection of all subfields of L that contain K and the α_i. Given a field extension $K \subset L$, there is always a system $\mathfrak{A}$ of elements in L such that $L = K(\mathfrak{A})$; for example, take for $\mathfrak{A}$ the system of all elements in L. We want to explicitly describe the subfield $K(\alpha_1, \ldots, \alpha_n) \subset L$ that is generated over K by finitely many elements $\alpha_1, \ldots, \alpha_n \in L$. Of course, it will contain the

ring $K[\alpha_1, \ldots, \alpha_n]$ of all polynomial expressions $f(\alpha_1, \ldots, \alpha_n)$ for polynomials $f \in K[X_1, \ldots, X_n]$ and then also its field of fractions, so that

$$K(\alpha_1, \ldots, \alpha_n) = Q\big(K[\alpha_1, \ldots, \alpha_n]\big).$$

Thereby we see that $K(\alpha_1, \ldots, \alpha_n)$ consists of all quotients of type

$$\frac{f(\alpha_1, \ldots, \alpha_n)}{g(\alpha_1, \ldots, \alpha_n)},$$

where $f, g \in K[X_1, \ldots, X_n]$, $g(\alpha_1, \ldots, \alpha_n) \neq 0$. For an arbitrary system $\mathfrak{A} = (\alpha_i)_{i \in I}$ of elements in L, the field $K(\mathfrak{A})$ can be described in a similar way, working with polynomials in $K[\mathfrak{X}]$ for a system of variables $\mathfrak{X} = (X_i)_{i \in I}$. Alternatively, we can interpret $K(\mathfrak{A})$ as the union of all subfields of L of type $K(\alpha_{i_1}, \ldots, \alpha_{i_s})$, where $i_1, \ldots, i_s \in I$.

Definition 8. *A field extension $K \subset L$ is called* simple *if there exists an element $\alpha \in L$ such that $L = K(\alpha)$. The degree $[K(\alpha) : K]$ is referred to as the* degree *of α over K.*

A field extension $K \subset L$ is called finitely generated *if there exist finitely many elements $\alpha_1, \ldots, \alpha_n \in L$ such that $L = K(\alpha_1, \ldots, \alpha_n)$.*

Proposition 9. *Let $L = K(\alpha_1, \ldots, \alpha_n)$ be a finitely generated field extension of K. Assume that $\alpha_1, \ldots, \alpha_n$ are algebraic over K. Then:*

(i) $L = K(\alpha_1, \ldots, \alpha_n) = K[\alpha_1, \ldots, \alpha_n]$.

(ii) *L is a finite, and in particular algebraic, field extension of K.*

Proof. We conclude by induction on n. The case $n = 1$ was already dealt with in Proposition 6. Therefore, let $n > 1$. We may assume by the induction hypothesis that $K[\alpha_1, \ldots, \alpha_{n-1}]$ is a finite field extension of K. Furthermore, it follows from Proposition 6 that $K[\alpha_1, \ldots, \alpha_n]$ is a finite field extension of $K[\alpha_1, \ldots, \alpha_{n-1}]$. Then $K[\alpha_1, \ldots, \alpha_n]$ is also finite over K by Proposition 2, and in particular algebraic over K by Proposition 7. Since $K[\alpha_1, \ldots, \alpha_n]$ is already a field, $K(\alpha_1, \ldots, \alpha_n)$ must coincide with $K[\alpha_1, \ldots, \alpha_n]$. □

The above proposition includes the nontrivial assertion that a simple field extension L/K that is generated by an algebraic element is algebraic itself in the sense that *every* element of L is algebraic over K. For example, using this fact we can easily see for $n \in \mathbb{N} - \{0\}$ that the real number $\cos \frac{\pi}{n}$ is algebraic over $\mathbb{Q}$. Indeed, $\cos \frac{\pi}{n} = \frac{1}{2}(e^{\pi i/n} + e^{-\pi i/n})$ is contained in $\mathbb{Q}(e^{\pi i/n})$, where $e^{\pi i/n}$ is algebraic over $\mathbb{Q}$, since it is a $2n$th root of unity. Since a finite field extension L/K is always finitely generated, for example by a K-basis of L, we obtain as a combination of Propositions 7 and 9 the following corollary:

Corollary 10. *For a field extension $K \subset L$, the following assertions are equivalent*:

(i) *L/K is finite.*
(ii) *L/K is generated by finitely many elements that are algebraic over K.*
(iii) *L/K is a finitely generated algebraic field extension.*

If $\mathfrak{A} = (\alpha_i)_{i\in I}$ is a system of generators of a field extension L/K, then L is the union of all its subfields of type $K(\alpha_{i_1}, \ldots, \alpha_{i_s})$, where $i_1, \ldots, i_s \in I$. In particular, we can conclude from Corollary 10 that L/K is algebraic as soon as all α_i are algebraic over K. Therefore, we can derive the following characterization (of not necessarily finitely generated) algebraic field extensions:

Corollary 11. *For a field extension $K \subset L$ the following assertions are equivalent*:
(i) *L/K is algebraic.*
(ii) *L/K is generated by elements that are algebraic over K.*

Next, let us show that algebraic field extensions are transitive in a natural way.

Proposition 12. *Let $K \subset L \subset M$ be field extensions. If $\alpha \in M$ is algebraic over L and if L/K is algebraic, then α is algebraic over K as well. In particular, the field extension M/K is algebraic if and only if M/L and L/K are algebraic.*

Proof. Let $f = X^n + c_1X^{n-1} + \ldots + c_n \in L[X]$ be the minimal polynomial of α over L. Then α is algebraic over the subfield $K(c_1, \ldots, c_n)$ of L, and we can conclude from Proposition 6 that

$$\big[K(c_1, \ldots, c_n, \alpha) : K(c_1, \ldots, c_n)\big] < \infty.$$

Moreover, we have

$$\big[K(c_1, \ldots, c_n) : K\big] < \infty$$

due to Proposition 9, and hence

$$\big[K(c_1, \ldots, c_n, \alpha) : K\big] < \infty$$

due to Proposition 2. But then $K(c_1, \ldots, c_n, \alpha)$ and in particular α are algebraic over K by Proposition 7.

The argument just given shows that M/K is algebraic if M/L and L/K are algebraic. The converse of this is trivial. □

Finally, let us give an example of an algebraic field extension that is not finite and hence cannot be finitely generated. We consider

$$L = \big\{\alpha \in \mathbb{C}\,;\ \alpha \text{ is algebraic over } \mathbb{Q}\big\}$$

as a subfield of $\mathbb{C}$ extending $\mathbb{Q}$. Indeed, L is a field, since for $\alpha, \beta \in L$ we get $\mathbb{Q}(\alpha, \beta) \subset L$ from Proposition 9. By its definition, $L/\mathbb{Q}$ is algebraic. Furthermore, $[L : \mathbb{Q}] = \infty$, since L contains $\mathbb{Q}(\sqrt[n]{p})$ for $n \in \mathbb{N} - \{0\}$ and p prime as

a subfield and since, as we have seen, $\mathbb{Q}(\sqrt[n]{p})$ is of degree n over $\mathbb{Q}$. We write $L = \overline{\mathbb{Q}}$ and call it the *algebraic closure* of $\mathbb{Q}$ in $\mathbb{C}$.

Exercises

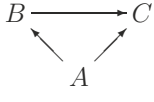
1. *Let L/K be a field extension. Discuss the problem of showing for two elements $a, b \in L$ that are algebraic over K that their sum $a + b$ is algebraic over K as well.*
2. *Characterize algebraic field extensions in terms of finite field extensions.*
3. *Show that every element in $\mathbb{C} - \overline{\mathbb{Q}}$ is transcendental over $\overline{\mathbb{Q}}$.*
4. Let L/K be a finite field extension such that $p = [L : K]$ is prime. Show: There exists an element $\alpha \in L$ such that $L = K(\alpha)$.
5. Let L/K be a finite field extension of degree $[L : K] = 2^k$. Let $f \in K[X]$ be a polynomial of degree 3 having a zero in L. Show that f admits a zero already in K.
6. Show that a field extension L/K is algebraic if and only if every subring R satisfying $K \subset R \subset L$ is a field.
7. Let L/K be a finite field extension. Show:
   (i) For $a \in L$, the minimal polynomial of a over K coincides with the minimal polynomial of the K-vector space homomorphism $\varphi_a \colon L \longrightarrow L$, $x \longmapsto ax$.
   (ii) If $L = K(a)$, the minimal polynomial of a over K coincides with the characteristic polynomial of φ_a.
   (iii) For $a \in L$, the characteristic polynomial of φ_a is also referred to as the *field polynomial* of a, relative to the field extension L/K. It is always a power of the minimal polynomial of a over K.
8. Let $\alpha \in \mathbb{C}$ satisfy $\alpha^3 + 2\alpha - 1 = 0$, so that it is algebraic over $\mathbb{Q}$. Determine the minimal polynomial of α as well as that of $\alpha^2 + \alpha$, in each case over $\mathbb{Q}$.
9. Let K be a field and x an element of an extension field of K, and assume that x is transcendental over K. Show for $n \in \mathbb{N} - \{0\}$ that x^n is transcendental over K and that $[K(x) : K(x^n)] = n$.
10. Let L/K be a field extension and let $\alpha \in L$ be algebraic over K. Show for $n \in \mathbb{N} - \{0\}$ that $[K(\alpha^n) : K] \geq \frac{1}{n}[K(\alpha) : K]$.
11. Let K be a field and $K(X)$ the function field in one variable X over K. Consider a rational function $q = f/g \in K(X) - K$, where f, g are coprime polynomials in $K[X]$. Show that q is transcendental over K and that
$$\big[K(X) : K(q)\big] = \max(\deg f, \deg g).$$
Determine the minimal polynomial of X over $K(q)$. *Hint*: Use Exercise 3 in Section 2.7.
12. Let L/K be a field extension. Show that two elements $\alpha, \beta \in L$ are algebraic over K if and only if $\alpha + \beta$ and $\alpha \cdot \beta$ are algebraic over K.
13. Consider two complex numbers $\alpha, \beta \in \mathbb{C}$ as well as exponents $m, n \in \mathbb{N}$ such that $\gcd(m, n) = 1$ and $\alpha^m = 2$, $\beta^n = 3$. Show that $\mathbb{Q}(\alpha, \beta) = \mathbb{Q}(\alpha \cdot \beta)$ and determine the minimal polynomial of $\alpha \cdot \beta$ over $\mathbb{Q}$.

3.3 Integral Ring Extensions*

In many respects, the theory of integral ring extensions to be dealt with in the present section can be viewed as a generalization of the theory of finite and algebraic field extensions, which was presented in Section 3.2. Even if one is primarily interested in the case of fields, the more general scope of ring theory leads to new insight, as we will see, for example, below in Corollary 8. We used vector spaces over fields as a technical tool in 3.2. In a similar way, we will rely on modules when dealing with ring extensions. See Section 2.9 for the definition of modules over rings.

Given a ring extension $R \subset R'$, the inclusion map $R \hookrightarrow R'$ can always be viewed as a ring homomorphism. Instead of restricting ourselves to ring extensions, we will look in the following at the more general case of ring homomorphisms. For every ring homomorphism $\varphi\colon A \longrightarrow B$, we can view B in a natural way as an A-module; just multiply elements $a \in A$ by elements $b \in B$ by carrying out the product $\varphi(a)b$ in B. We say that φ is *finite* if B is a finite A-module under φ; we will also say in this case that B is finite over A or, if φ is an inclusion homomorphism, that the extension $A \hookrightarrow B$ is finite. Furthermore, φ as well as B over A, resp. the ring extension $A \hookrightarrow B$, are said to be *of finite type* if there exists an epimorphism $\Phi\colon A[X_1, \ldots, X_n] \longrightarrow B$ from a polynomial ring in finitely many variables over A to B that extends φ. Note that every finite ring homomorphism is, in particular, of finite type. That a ring homomorphism $\varphi\colon A \longrightarrow B$ is of finite type can also be characterized by the fact that there exist elements $x_1, \ldots, x_n \in B$ satisfying $B = \varphi(A)[x_1, \ldots, x_n]$. Here $\varphi(A)[x_1, \ldots, x_n] \subset B$ indicates the subring of all expressions $f(x_1, \ldots, x_n)$ for polynomials $f \in \varphi(A)[X_1, \ldots, X_n]$, as explained in 2.5. By abuse of notation, this ring will also be denoted by $A[x_1, \ldots, x_n]$.

In the preceding situation it is convenient to employ the terminology of algebras. An *algebra* B over a ring A is just a ring homomorphism $A \longrightarrow B$. In particular, we can talk about (module-)finite A-algebras or about A-algebras of finite type. Also note that a homomorphism between two A-algebras B and C is meant as a ring homomorphism $B \longrightarrow C$ with the additional property that it is compatible with the defining homomorphisms $A \longrightarrow B$ and $A \longrightarrow C$, in the sense that the diagram



is commutative.

By its definition, a field extension $K \subset L$ is finite if and only if it is a finite ring extension. Note that a similar assertion for finitely generated field extensions and ring extensions of finite type does not hold in general. Of course, a field extension $K \subset L$ is finitely generated if it is a ring extension of finite type. However, the converse of this fails to be true, as we will see at the end of

the present section. As a next step, we want to extend the notion of algebraic field extensions to the context of rings.

Lemma 1. *For a ring homomorphism $\varphi\colon A \longrightarrow B$ and an element $b \in B$, the following conditions are equivalent*:

(i) *There exists an* integral equation *of b over A, i.e., an equation of type $f(b) = 0$ for a monic polynomial $f \in A[X]$.*

(ii) *The subring $A[b] \subset B$, viewed as an A-module, is finitely generated.*

(iii) *Viewing B as an A-module, there exists a finitely generated submodule $M = \sum_{i=1}^{n} Am_i \subset B$ such that $1 \in M$ and $bM \subset M$.*

Proof. We start with the implication (i) $\Longrightarrow$ (ii). So assume there is an equation $f(b) = 0$ for a monic polynomial $f \in A[X]$, say

$$b^n + a_1 b^{n-1} + \ldots + a_n = 0.$$

Then b^n is an element of the A-module $M = \sum_{i=0}^{n-1} Ab^i$, and we see by induction that $b^i \in M$ for all $i \in \mathbb{N}$. Therefore, we get $A[b] \subset M$ and hence $A[b] = M$. In particular, $A[b]$ is a finitely generated A-module, and (ii) follows.

The implication (ii) $\Longrightarrow$ (iii) is trivial. Thus, it remains to verify the implication (iii) $\Longrightarrow$ (i). Let $M = \sum_{i=1}^{n} Am_i \subset B$ be a finitely generated A-submodule of B such that $1 \in M$ and $bM \subset M$. The latter inclusion shows that there is a set of equations

$$\begin{aligned} bm_1 &= a_{11}m_1 + \ldots + a_{1n}m_n, \\ &\cdots \\ &\cdots \\ &\cdots \\ bm_n &= a_{n1}m_1 + \ldots + a_{nn}m_n \end{aligned}$$

with coefficients $a_{ij} \in A$ that in terms of matrices can be rewritten as

$$\Delta \cdot \begin{pmatrix} m_1 \\ \vdots \\ m_n \end{pmatrix} = 0 \ ,$$

using the matrix $\Delta = (\delta_{ij} b - a_{ij})_{i,j=1,\ldots,n} \in B^{n\times n}$; here δ_{ij} is Kronecker's symbol given by $\delta_{ij} = 1$ for $i = j$, and $\delta_{ij} = 0$ for $i \neq j$. Now we apply *Cramer's rule*, i.e., the relation

$$(*) \qquad \Delta^* \cdot \Delta = (\det \Delta) \cdot E$$

involving the adjoint matrix $\Delta^* \in B^{n\times n}$ of Δ, the determinant of Δ, as well as the unit matrix $E \in B^{n\times n}$; see, for example, [4], Satz 4.4/3. This equation is established in linear algebra for matrices with coefficients in a field, but we claim that it naturally extends to the more general setting of coefficients in rings, as is needed here. Indeed, comparing the coefficients of the matrices occurring in

$(*)$ on the left-hand and the right-hand sides, the equation $(*)$ consists of a system of polynomial identities between the coefficients of Δ. More generally, viewing the coefficients c_{ij} of Δ as *variables*, these identities can be formulated over the polynomial ring $\mathbb{Z}[c_{ij}]$. They can then be derived from the classical case of fields by embedding $\mathbb{Z}[c_{ij}]$ into its field of fractions $\mathbb{Q}(c_{ij})$.

Now using Cramer's rule $(*)$, we get

$$(\det \Delta) \cdot \begin{pmatrix} m_1 \\ \vdots \\ m_n \end{pmatrix} = \Delta^* \cdot \Delta \cdot \begin{pmatrix} m_1 \\ \vdots \\ m_n \end{pmatrix} = 0$$

and hence $(\det \Delta) \cdot m_i = 0$ for $i = 1, \ldots, n$. Since the identity element $1 \in B$ is a linear combination of the elements m_i with coefficients in A, we conclude that $\det \Delta = (\det \Delta) \cdot 1 = 0$. Therefore,

$$\det(\delta_{ij} X - a_{ij}) \in A[X]$$

is a monic polynomial admitting b as a zero, as desired. □

Definition 2. *Let $\varphi\colon A \longrightarrow B$ be a ring homomorphism. An element $b \in B$ is called* integral *over A with respect to φ if b and φ satisfy the equivalent conditions of Lemma* 1. *Moreover, we say that B is* integral *over A or that φ is* integral *if every element $b \in B$ is integral over A in the way just described.*

It is obvious that the notions "integral" and "algebraic" coincide if we restrict ourselves to field extensions. Furthermore, by establishing the equivalences of Lemma 1, we have already exhibited the crucial relations between integral and finite ring extensions. Below we formulate some special consequences of these that may be viewed as generalizations of the results 3.2/7, 3.2/9, and 3.2/12.

Corollary 3. *Every finite ring homomorphism $A \longrightarrow B$ is integral.*

Proof. Use condition (iii) of Lemma 1 for $M = B$ to see that $A \longrightarrow B$ is integral. □

Corollary 4. *Let $\varphi\colon A \longrightarrow B$ be a ring homomorphism of finite type and assume $B = A[b_1, \ldots, b_r]$ for elements $b_1, \ldots, b_r \in B$ that are integral over A. Then $A \longrightarrow B$ is finite and, in particular, integral.*

Proof. Consider the chain of "simple" ring extensions

$$\varphi(A) \subset \varphi(A)[b_1] \subset \ldots \subset \varphi(A)[b_1, \ldots, b_r] = B.$$

Each of these is finite by Lemma 1, and we easily conclude by induction that B is finite over A. Indeed, to carry out the induction step, just multiply the elements of a module generating system of B over $\varphi(A)[b_1, \ldots, b_{r-1}]$ by those

of a corresponding system of $\varphi(A)[b_1, \ldots, b_{r-1}]$ over A that is provided by the induction hypothesis. Thereby one obtains a module generating system of B over A, as is seen using a similar argumentation to that given in the proof of 3.2/2. □

Corollary 5. *Let $A \longrightarrow B$ and $B \longrightarrow C$ be two finite (resp. integral) ring homomorphisms. Then their composition $A \longrightarrow C$ is also finite (resp. integral).*

Proof. To settle the assertion for finite homomorphisms, we use the same argument as the one applied in the induction step of the proof of Corollary 4. Hence, it remains to consider the case of integral homomorphisms. Therefore, assume that $A \longrightarrow B$ and $B \longrightarrow C$ are integral and consider an element $c \in C$. Then c satisfies an integral equation over B, say

$$c^n + b_1 c^{n-1} + \ldots + b_n = 0, \qquad b_1, \ldots, b_n \in B,$$

and we can conclude that $c \in C$ is integral over $A[b_1, \ldots, b_n]$. In particular, the extension $A[b_1, \ldots, b_n] \longrightarrow A[b_1, \ldots, b_n, c]$ is finite, as we see from Corollary 4. The same result also shows that $A \longrightarrow A[b_1, \ldots, b_n]$ is finite, so that altogether $A \longrightarrow A[b_1, \ldots, b_n, c]$ is finite. But then this homomorphism is integral as well, see Corollary 3, and it follows that c is integral over A. Finally, letting c vary over C, it follows that $A \longrightarrow C$ is integral. □

Next we want to prove a fundamental theorem that significantly clarifies the structure of algebras of finite type over fields. An analogue for field extensions, namely the decomposition of an arbitrary field extension into a purely transcendental and an algebraic one, will be dealt with in 7.1.

Theorem 6 (Noether normalization). *Let K be a field and $K \hookrightarrow B$ a nonzero K-algebra of finite type. Then there exist elements $x_1, \ldots, x_r \in B$ such that B is finite over the subring $K[x_1, \ldots, x_r] \subset B$, while the system $x_1, \ldots, x_r \in B$ is algebraically independent over K (cf. 2.5/6).*

In other words, there exists a finite monomorphism $K[X_1, \ldots, X_r] \hookrightarrow B$ of K-algebras, where $K[X_1, \ldots, X_r]$ is a polynomial ring in finitely many variables over K.

Proof. Let $B = K[b_1, \ldots, b_n]$ for certain elements $b_1, \ldots, b_n \in B$. If $b_1, \ldots, b_n$ are algebraically independent over K, nothing has to be shown. Assuming the contrary, there exists a nontrivial relation of type

$$(*) \qquad \sum_{(\nu_1 \ldots \nu_n) \in I} a_{\nu_1 \ldots \nu_n} b_1^{\nu_1} \ldots b_n^{\nu_n} = 0$$

with coefficients $a_{\nu_1 \ldots \nu_n} \in K^*$, where the summation extends over a finite set I of n-tuples $(\nu_1, \ldots, \nu_n) \in \mathbb{N}^n$. Now introduce new elements $x_1, \ldots, x_{n-1} \in B$, say

$$x_1 = b_1 - b_n^{s_1}, \quad \ldots, \quad x_{n-1} = b_{n-1} - b_n^{s_{n-1}},$$

for certain exponents $s_1, \ldots, s_{n-1} \in \mathbb{N} - \{0\}$ whose choice has still to be made precise. Then, in any case, we get

$$B = K[b_1, \ldots, b_n] = K[x_1, \ldots, x_{n-1}, b_n].$$

Substituting $b_i = x_i + b_n^{s_i}$ for $i = 1, \ldots, n-1$ in the relation $(*)$ and decomposing powers $b_i^{\nu_i} = (x_i + b_n^{s_i})^{\nu_i}$ into the sum of $b_n^{s_i \nu_i}$ and terms of lower degree in b_n, we get a new relation of type

$$(**) \qquad \sum_{(\nu_1 \ldots \nu_n) \in I} a_{\nu_1 \ldots \nu_n} b_n^{s_1 \nu_1 + \ldots + s_{n-1} \nu_{n-1} + \nu_n} + f(x_1, \ldots, x_{n-1}, b_n) = 0.$$

Here $f(x_1, \ldots, x_{n-1}, b_n)$ is a polynomial expression in b_n with coefficients in $K[x_1, \ldots, x_{n-1}]$ such that the corresponding degree in b_n is strictly less than the maximum of all sums $s_1 \nu_1 + \ldots + s_{n-1} \nu_{n-1} + \nu_n$ for $(\nu_1, \ldots, \nu_n) \in I$. As is easily checked, the integers $s_1, \ldots, s_{n-1} \in \mathbb{N}$ can be chosen in such a way that the exponents $s_1 \nu_1 + \ldots + s_{n-1} \nu_{n-1} + \nu_n$ for index tuples $(\nu_1, \ldots, \nu_n) \in I$ occurring in $(**)$ are all different. Indeed, choose $t \in \mathbb{N}$ bigger than the maximum of all indices $\nu_1, \ldots, \nu_n$ for $(\nu_1, \ldots, \nu_n) \in I$ and let

$$s_1 = t^{n-1}, \ldots, s_{n-1} = t^1.$$

Now view the relation $(**)$ as a polynomial equation in b_n with coefficients in $K[x_1, \ldots, x_{n-1}]$. Then it follows that there is a term of type $a b_n^N$ with a coefficient $a \in K^*$ whose degree N strictly dominates the degrees of all other terms. In particular, multiplying it by a^{-1}, the equation $(**)$ can be read as an integral equation of b_n over $K[x_1, \ldots, x_{n-1}]$, and we see from Corollary 4 that the extension $K[x_1, \ldots, x_{n-1}] \hookrightarrow B$ is finite. If $x_1, \ldots, x_{n-1}$ happen to be algebraically independent over K, we are done. Otherwise, we can apply the just described process anew to the ring $K[x_1, \ldots, x_{n-1}]$. Continuing in this way, we finally arrive at a system $x_1, \ldots, x_r$ that is algebraically independent over K. That the inclusion $K[x_1, \ldots, x_r] \hookrightarrow B$ is finite follows from Corollary 5. $\square$

One can show that the integer r occurring in the theorem on Noether normalization is unique; it is the so-called *dimension* of the ring B. For an integral domain B, the uniqueness of r can easily be deduced from a corresponding uniqueness assertion on the transcendence degree of field extensions; cf. 7.1/5. Looking ahead to Section 7.1, we want to briefly explain this. Indeed, if $x_1, \ldots, x_r \in B$ are algebraically independent over K and the extension $K[x_1, \ldots, x_r] \hookrightarrow B$ is finite, then the field of fractions $Q(B)$ is algebraic over the purely transcendental extension $K(x_1, \ldots, x_r)$ of K. Therefore, $x_1, \ldots, x_r$ is a transcendence basis of $Q(B)/K$, cf. 7.1/2, and we have $\mathrm{transdeg}_K Q(B) = r$.

As an application of Noether normalization, we want to justify the fact mentioned already before that a finitely generated field extension is not necessarily a ring extension of finite type. To do this we need an auxiliary result.

Lemma 7. *Let $A \hookrightarrow B$ be an integral extension of integral domains. If one of the rings A and B is a field, then the same is true for the other as well.*

Proof. Let A be a field and $b \neq 0$ an element of B. Then b satisfies an integral equation over A, say

$$b^n + a_1 b^{n-1} + \ldots + a_n = 0, \qquad a_1, \ldots, a_n \in A,$$

and we may assume $a_n \neq 0$. Indeed, pass to the field of fractions of B and multiply the equation by a suitable power of b^{-1}. Now, if $a_n \neq 0$, its inverse a_n^{-1} exists in A and we get

$$b^{-1} = -a_n^{-1}(b^{n-1} + a_1 b^{n-2} + \ldots + a_{n-1}) \in B.$$

Thereby we see that B is a field.

Conversely, if B is a field, consider an element $a \in A$, $a \neq 0$. Then its inverse a^{-1} exists in B and satisfies an integral equation over A, say

$$a^{-n} + a_1 a^{-n+1} + \ldots + a_n = 0, \qquad a_1, \ldots, a_n \in A.$$

But then

$$a^{-1} = -a_1 - a_2 a - \ldots - a_n a^{n-1} \in A,$$

and it follows that A is a field. □

Corollary 8. *Let $K \subset L$ be a field extension satisfying $L = K[x_1, \ldots, x_n]$ for some elements $x_1, \ldots, x_n \in L$, i.e., assume that $K \subset L$, as a ring extension, is of finite type. Then the extension $K \subset L$ is finite.*

Proof. Due to Theorem 6 on Noether normalization, there are elements $y_1, \ldots, y_r$ in L such that the ring extension $K[y_1, \ldots, y_r] \hookrightarrow L$ is finite and the elements $y_1, \ldots, y_r$ are algebraically independent over K. Since L is a field, the same is true for $K[y_1, \ldots, y_r]$ by Lemma 7. However, a polynomial ring over a field K in r variables cannot be a field if $r > 0$. Therefore, we have necessarily $r = 0$, and the extension $K \hookrightarrow L$ is finite. □

A situation as considered in Corollary 8 can easily be set up by looking at polynomial rings modulo maximal ideals.

Corollary 9. *Let $K[X_1, \ldots, X_n]$ be the polynomial ring in n variables over a field K and let $\mathfrak{m} \subset K[X_1, \ldots, X_n]$ be a maximal ideal. Then the canonical map $K \longrightarrow K[X_1, \ldots, X_n]/\mathfrak{m} = L$ is finite, so that L/K is a finite field extension.*

Proof. Since $\mathfrak{m}$ is a maximal ideal in $K[X_1, \ldots, X_n]$, we see that L is a field. Furthermore, if $x_i \in L$ is the residue class of the variable X_i for each i, we get $L = K[x_1, \ldots, x_n]$. Therefore, L/K is of finite type, and thus a finite field extension by Corollary 8. □

Now consider the field of rational functions $K(X)$ in one variable X over a field K. Then the field extension $K(X)/K$ is finitely generated, namely by the variable X. However, as a ring extension it cannot be of finite type following Corollary 8, since the degree $[K(X) : K]$ is infinite. Thereby we see that as mentioned before, the properties "finitely generated" and "of finite type" are not equivalent in the setting of field extensions.

Exercises

1. *Let $A \subset B$ be an integral ring extension. Discuss the problem of whether one can define for an element $b \in B$ its minimal polynomial over A. As an example, consider the extension $A = \{\sum c_i X^i \in K[X]\,;\, c_1 = 0\} \subset K[X] = B$, where $K[X]$ is the polynomial ring in one variable over a field K.*

2. For a ring homomorphism $A \longrightarrow B$, let $\overline{A}$ denote the set of all elements in B that are integral over A. Show that $\overline{A}$ is a subring of B and that $A \longrightarrow B$ restricts to an integral homomorphism $A \longrightarrow \overline{A}$. The ring $\overline{A}$ is called the *integral closure* of A in B.

3. Let A be a unique factorization domain. Show that A is integrally closed in its field of fractions, i.e., that the integral closure of A in $Q(A)$ in the sense of Exercise 2 coincides with A.

4. Let $\varphi\colon A \hookrightarrow A'$ be an integral ring extension. Show for every maximal ideal $\mathfrak{m}' \subset A'$ that the ideal $\varphi^{-1}(\mathfrak{m}') \subset A$ is maximal as well, and conversely for every maximal ideal $\mathfrak{m} \subset A$, that there exists a maximal ideal $\mathfrak{m}' \subset A'$ satisfying $\varphi^{-1}(\mathfrak{m}') = \mathfrak{m}$. *Hint*: For a maximal ideal $\mathfrak{m} \subset A$, consider the multiplicative system $S = A - \mathfrak{m}$, as well as the associated rings of fractions $S^{-1}A$ and $S^{-1}A'$ introduced in Section 2.7. In addition, one may use the fact that every nonzero ring admits a maximal ideal; cf. 3.4/6.

3.4 Algebraic Closure

The objective of the present section is to construct for a given field K a so-called algebraic closure, i.e., a minimal algebraic extension field $\overline{K}$ such that every nonconstant polynomial in $\overline{K}[X]$ admits at least one zero in $\overline{K}$. We start with *Kronecker's construction* that was already mentioned before. It allows for a single nonconstant polynomial $f \in K[X]$ to set up a finite extension field L/K such that f acquires a zero in L.

Proposition 1. *Let K be a field and $f \in K[X]$ a polynomial of degree ≥ 1. Then there exists a finite algebraic field extension $K \subset L$ such that f admits a zero in L. If f is irreducible, we can set $L := K[X]/(f)$.*

Proof. We may assume that f is irreducible; otherwise, decompose f into its prime factors and replace it by one of these. Then (f) is a maximal ideal in

$K[X]$ by 2.4/6, and it follows that $L := K[X]/(f)$ is a field. Now consider the composition

$$K \hookrightarrow K[X] \xrightarrow{\pi} K[X]/(f) = L,$$

where π is the canonical epimorphism. Since the homomorphism $K \longrightarrow L$ is injective as a homomorphism between fields, we can view L as a field extending K. Then, writing $x := \pi(X)$ and $f = \sum_{i=0}^{n} c_i X^i$, we get

$$f(x) = \sum_{i=0}^{n} c_i x^i = \sum_{i=0}^{n} c_i \pi(X)^i = \pi\Big(\sum_{i=0}^{n} c_i X^i\Big) = \pi(f) = 0,$$

which shows that x is a zero of f. In particular, x is algebraic over K and satisfies $L = K(x)$. Assuming f to be monic, it is the minimal polynomial of x over K, and we see from 3.2/6 that L/K is a finite field extension of degree $n = \deg f$. □

Using Kronecker's construction, we say that L is obtained from K via *adjunction of a zero* x of f. In more precise terms, if f is irreducible, the zero x is defined as a residue class of the variable X and is forced to become a zero of f by passing from $K[X]$ to its residue class ring $L = K[X]/(f)$, which is a field. Then, over L, the linear factor $X - x$ splits off from f, say $f = (X - x) \cdot g$, and Kronecker's construction can again be applied to g, unless the latter is already constant. Thus, after finitely many steps of this type, we arrive at an extension field K' of K over which f admits a factorization into linear factors. In principle, such a process has to be applied simultaneously to all nonconstant polynomials in $K[X]$ in order to construct an algebraic closure of K.

Definition 2. *A field K is called* algebraically closed *if every nonconstant polynomial f of $K[X]$ admits a zero in K, or in other words, if f decomposes in $K[X]$ into a product of linear factors. This means that f admits a product decomposition $f = c\prod_i (X - \alpha_i)$ with a constant $c \in K^*$ as well as zeros $\alpha_i \in K$.*

Remark 3. *A field K is algebraically closed if and only if every algebraic field extension L/K is trivial.*

Proof. Assume that K is algebraically closed and that $K \subset L$ is an algebraic field extension. Furthermore, consider an element $\alpha \in L$ together with its minimal polynomial $f \in K[X]$. Then f decomposes over K into a product of linear factors and hence is linear, since it is irreducible. In particular, this shows that $\alpha \in K$ and therefore $L = K$. Conversely, assume that K does not admit any nontrivial algebraic field extension and consider a polynomial $f \in K[X]$ of degree ≥ 1. Using Kronecker's construction, there exists an algebraic field extension L/K such that f admits a zero in L. However, by our assumption we must have $L = K$, so that f has a zero in K. Consequently, K is algebraically closed. □

Theorem 4. *Every field K admits an extension field L that is algebraically closed.*

For the proof of the theorem we need to know that every ring $R \neq 0$ contains a maximal ideal. The latter result is a consequence of Zorn's lemma, whose assertion we will explain next.

Let M be a set. A (*partial*) *order* on M is a relation[1] $\leq$ such that the following conditions are satisfied:

$$\begin{aligned} x \leq x \text{ for all } x \in M \qquad & \text{(reflexivity)} \\ x \leq y,\ y \leq z \implies x \leq z \qquad & \text{(transitivity)} \\ x \leq y,\ y \leq x \implies x = y \qquad & \text{(antisymmetry)} \end{aligned}$$

The order is called *total* if every two elements $x, y \in M$ are comparable, i.e., if we have $x \leq y$ or $y \leq x$.

For example, the standard less-than-or-equal relation $\leq$ between real numbers constitutes a total order on $\mathbb{R}$. But we can also look at a set X and define M as its power set consisting of all subsets of X. Then the inclusion of subsets in X gives rise to a partial order on M. In general, it is not a total order, since for $U, U' \subset X$ we do not necessarily have $U \subset U'$ or $U' \subset U$. In a similar way, we can consider for a ring R the set M of its proper ideals $\mathfrak{a} \subsetneq R$, together with the inclusion as partial order. Then $\mathfrak{a}$ is a maximal ideal in R if and only if $\mathfrak{a}$ is a maximal element of M. To be more specific on maximality, let us introduce for a set M with partial order $\leq$ and an element $a \in M$ the following terminology:

a is called the *greatest* element of M if $x \leq a$ holds for all $x \in M$; such an element a is unique if it exists.

a is called a *maximal* element of M if $a \leq x$ for some $x \in M$ always implies $a = x$.

a is called an *upper bound* for a subset $N \subset M$ if $x \leq a$ for all $x \in N$.

If there exists a greatest element in M, it is the unique maximal element in M. However, note that a partially ordered set M can contain several different maximal elements. If that is the case, there cannot exist a greatest element in M.

Lemma 5 (Zorn). *Let M be a partially ordered set such that every subset of M that is totally ordered with respect to the order induced from M admits an upper bound in M. Then there exists a maximal element in M.*[2]

For an elementary justification of the above result we refer to [12], Appendix 2, §2. However, it should be pointed out that Zorn's lemma is of axiomatic character. It is equivalent to the so-called axiom of choice, asserting

[1] A relation on M is a subset $R \subset M \times M$, where in the present case, we will write $x \leq y$ if $(x, y) \in R$.

[2] Note that the empty subset in M is totally ordered and therefore admits an upper bound in M if the assumptions of the lemma are met. In particular, we must have $M \neq \emptyset$ then.

that the Cartesian product of a nonempty family of nonempty sets is nonempty. As an application of Zorn's lemma we prove the following result:

Proposition 6. *Let R be a ring and $\mathfrak{a} \subsetneq R$ a proper ideal. Then R admits a maximal ideal $\mathfrak{m}$ containing $\mathfrak{a}$. In particular, every ring $R \neq 0$ admits a maximal ideal.*

Proof. Let M be the set of all proper ideals $\mathfrak{b} \subsetneq R$ containing $\mathfrak{a}$. Then M is partially ordered under the inclusion of ideals. Furthermore, since $\mathfrak{a} \in M$, we see that $M \neq \emptyset$. We claim that every totally ordered subset $N \subset M$ admits an upper bound in M. Indeed, let N be such a subset, where we may assume $N \neq \emptyset$. Then $\mathfrak{c} = \bigcup_{\mathfrak{b} \in N} \mathfrak{b}$ is a proper ideal in R containing $\mathfrak{a}$, as is easily checked using the total order on N, and it follows that $\mathfrak{c} \in M$ is an upper bound of N. Therefore, the assumptions of Zorn's lemma are met, and M admits a maximal element, namely a maximal ideal $\mathfrak{m} \subsetneq R$ such that $\mathfrak{a} \subset \mathfrak{m}$. □

Proof of Theorem 4. We are now able to construct for a given field K an extension field L that is algebraically closed. The construction process we will use is based on polynomial rings in infinitely many variables over K and goes back to E. Artin. In a first step we set up a field L_1 extending K such that every polynomial $f \in K[X]$ of degree $\deg f \geq 1$ admits a zero in L_1. To do this we consider the system of variables $\mathfrak{X} = (X_f)_{f \in I}$ that is indexed by the set

$$I = \big\{ f \in K[X] \,;\, \deg f \geq 1 \big\},$$

and work with the polynomial ring $K[\mathfrak{X}]$. More specifically, we look at the ideal

$$\mathfrak{a} = \big(f(X_f) \,;\, f \in I \big) \subset K[\mathfrak{X}]$$

that is generated by the family of polynomials $f(X_f)$, where the variable X of f is replaced by X_f, for each $f \in I$. We claim that $\mathfrak{a}$ is a proper ideal in $K[\mathfrak{X}]$. Indeed, suppose that is not the case. Then we have $1 \in \mathfrak{a}$, and there is an equation

$$\sum_{i=1}^{n} g_i f_i(X_{f_i}) = 1$$

for suitable polynomials $f_1, \ldots, f_n \in I$ and $g_1, \ldots, g_n \in K[\mathfrak{X}]$. Applying Kronecker's construction to the finitely many polynomials f_i, there exists a field K' extending K such that each f_i admits a zero α_i in K'. Now, going back to the above equation, we may substitute X_{f_i} by α_i for $i = 1, \ldots, n$, using the substitution homomorphism $K[\mathfrak{X}] \longrightarrow K'$ that substitutes X_{f_i} by α_i and the remaining variables by arbitrary values in K', for example by 0. Hence, the left-hand side of the above equation vanishes, contradicting 1 on the right-hand side. Therefore $\mathfrak{a}$ must be a proper ideal in $K[\mathfrak{X}]$, as claimed.

Next, apply Proposition 6 and choose a maximal ideal $\mathfrak{m} \subset K[\mathfrak{X}]$ containing $\mathfrak{a}$. Then $L_1 = K[\mathfrak{X}]/\mathfrak{m}$ is a field, which we will view as an extension of K via the composition of canonical maps

$$K \hookrightarrow K[\mathfrak{X}] \longrightarrow K[\mathfrak{X}]/\mathfrak{m} = L_1.$$

Similarly as for Kronecker's construction, we conclude for $f \in I$ that the residue class $\overline{X}_f$ of $X_f \in K[\mathfrak{X}]$ in $K[\mathfrak{X}]/\mathfrak{m}$ is a zero of $f \in K[X]$. Note that again, the zeros of the polynomials $f \in I$ come into existence by passing to residue classes modulo $\mathfrak{a}$, resp. $\mathfrak{m}$.

To end the proof of Theorem 4 we proceed as follows. Iterating the just described construction, we arrive at a chain of fields

$$K = L_0 \subset L_1 \subset L_2 \subset \dots$$

such that each nonconstant polynomial $f \in L_n[X]$, $n \in \mathbb{N}$, admits a zero in L_{n+1}. The union

$$L = \bigcup_{n=0}^{\infty} L_n$$

of this ascending chain of fields is itself a field, and we claim that L is algebraically closed. Indeed, consider a nonconstant polynomial $f \in L[X]$. There exists an index $n \in \mathbb{N}$ satisfying $f \in L_n[X]$, since f has only finitely many nonzero coefficients. Then f admits a zero in L_{n+1} by our construction and thereby in L. This shows that L is algebraically closed, which completes our proof of Theorem 4. □

Let us point out that we actually have $L = L_1$ in the setting of the above proof; cf. Exercise 10 in Section 3.7. However, to justify this fact we need some additional information that is not yet available at the present stage.

Corollary 7. *Let K be a field. Then there exists an algebraically closed field $\overline{K}$ extending K, where $\overline{K}$ is algebraic over K; such a field $\overline{K}$ is called an* algebraic closure *of K.*

Proof. If we look a bit closer at the construction of an algebraically closed field L extending a field K, as exercised in the proof of Theorem 4 above, we can easily realize that L is algebraic over K and therefore admits the properties of an algebraic closure of K. Indeed, by its construction, the extension L_n/L_{n-1} is generated by a family of algebraic elements so that L_n/L_{n-1} is algebraic due to 3.2/11. Then, using induction, it follows from 3.2/12 that all L_n are algebraic over K. Since L is the union of the L_n, we see that L is algebraic over K.

Alternatively, if L is an arbitrary algebraically closed field extending K, we can set

$$\overline{K} = \big\{\alpha \in L\,;\, \alpha \text{ is algebraic over } K\big\}.$$

Then $\overline{K}$ is a field and hence an algebraic extension field of K, since $\alpha, \beta \in \overline{K}$ implies $K(\alpha, \beta) \subset \overline{K}$. Furthermore, $\overline{K}$ is algebraically closed. Indeed, consider a nonconstant polynomial $f \in \overline{K}[X]$. Since L is algebraically closed, f admits a zero γ in L. This zero is algebraic over $\overline{K}$, and by 3.2/12, algebraic over K, so that we get $\gamma \in \overline{K}$. □

As an example, we refer to the (not yet established) fact that $\mathbb{C}$ is an algebraic closure of $\mathbb{R}$. Moreover, we can define an algebraic closure $\overline{\mathbb{Q}}$ of $\mathbb{Q}$ by setting

$$\overline{\mathbb{Q}} = \big\{\alpha \in \mathbb{C}\,;\, \alpha \text{ is algebraic over } \mathbb{Q}\big\}.$$

Note that $\overline{\mathbb{Q}}$ is different from $\mathbb{C}$, since $\mathbb{C}$ contains elements such as e and π that are transcendental and therefore not algebraic over $\mathbb{Q}$. Alternatively, the inequality $\overline{\mathbb{Q}} \neq \mathbb{C}$ can be justified by means of a cardinality argument if we use the fact that the algebraic closure of a field is unique up to (noncanonical) isomorphism (cf. Corollary 10 below). Indeed, $\mathbb{C}$ is of uncountable cardinality, while the explicit construction of an algebraic closure of $\mathbb{Q}$ via the proof of Theorem 4 implies that $\overline{\mathbb{Q}}$ is countable.

Finally, we want to show that every two algebraic closures of a given field K are isomorphic over K, although in general, there will exist several isomorphisms of this type, none of them canonical. To settle this question we must study the problem of extending field homomorphisms $K \longrightarrow L$ to algebraic field extensions K'/K. Let us add that the corresponding results of Lemma 8 and Proposition 9 below not only are of interest for the question on the uniqueness of algebraic closures, but also play an important role for the characterization of separable field extensions in 3.6, as well as for setting up Galois theory in 4.1.

We still need a convenient notation for the transport of polynomials with respect to homomorphisms. If $\sigma\colon K \longrightarrow L$ is a field homomorphism and $K[X] \longrightarrow L[X]$ the induced homomorphism on polynomial rings, we denote by $f^\sigma \in L[X]$ the image of a polynomial $f \in K[X]$. For every zero $\alpha \in K$ of f, it is immediately clear that its image $\sigma(\alpha)$ is a zero of f^σ.

Lemma 8. *Let K be a field and $K' = K(\alpha)$ a simple algebraic field extension of K with attached minimal polynomial $f \in K[X]$ of α. Furthermore, let $\sigma\colon K \longrightarrow L$ be a field homomorphism.*

(i) *If $\sigma'\colon K' \longrightarrow L$ is a field homomorphism extending σ, then $\sigma'(\alpha)$ is a zero of f^σ.*

(ii) *Conversely, for every zero $\beta \in L$ of $f^\sigma \in L[X]$, there is precisely one extension $\sigma'\colon K' \longrightarrow L$ of σ such that $\sigma'(\alpha) = \beta$.*

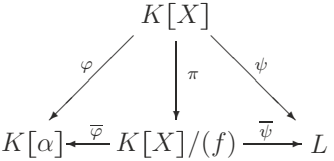
In particular, the different extensions σ' of σ are in one-to-one correspondence with the distinct zeros of f^σ in L, and the number of these is $\leq \deg f$.

Proof. For every extension $\sigma'\colon K' \longrightarrow L$ of σ we get from $f(\alpha) = 0$ necessarily $f^\sigma(\sigma'(\alpha)) = \sigma'(f(\alpha)) = 0$. Moreover, since $K' = K[\alpha]$ by 3.2/9, every extension $\sigma'\colon K' \longrightarrow L$ of σ is uniquely determined by the image $\sigma'(\alpha)$ of α.

It remains to show for each zero $\beta \in L$ of f^σ that there is an extension $\sigma'\colon K' \longrightarrow L$ of σ satisfying $\sigma'(\alpha) = \beta$. To do this, consider the substitution homomorphisms

$$\begin{aligned} &\varphi\colon K[X] \longrightarrow K[\alpha], && g \longmapsto g(\alpha),\\ &\psi\colon K[X] \longrightarrow L, && g \longmapsto g^\sigma(\beta). \end{aligned}$$

We have $(f) = \ker\varphi$ by 3.2/5, as well as $(f) \subset \ker\psi$, since $f^\sigma(\beta) = 0$. If $\pi\colon K[X] \longrightarrow K[X]/(f)$ denotes the canonical projection, we obtain via the fundamental theorem on homomorphisms 2.3/4 a commutative diagram



with homomorphisms $\overline{\varphi}$ and $\overline{\psi}$ that are unique. Since $\overline{\varphi}$ is an isomorphism, we recognize $\sigma' := \overline{\psi} \circ \overline{\varphi}^{-1}$ as an extension of σ satisfying $\sigma'(\alpha) = \beta$. □

Proposition 9. *Let $K \subset K'$ be an algebraic field extension and $\sigma\colon K \longrightarrow L$ a field homomorphism with image in an algebraically closed field L. Then σ admits an extension $\sigma'\colon K' \longrightarrow L$. In addition, if K' is algebraically closed and L algebraic over $\sigma(K)$, then every extension σ' of σ is an isomorphism.*

Proof. The main work was already done in Lemma 8, and it remains only to apply Zorn's lemma. Let M be the set of all pairs (F, τ) consisting of an intermediate field F, $K \subset F \subset K'$, as well as an extension $\tau\colon F \longrightarrow L$ of σ. Then M is partially ordered under the relation $\leq$ if we write $(F, \tau) \leq (F', \tau')$ when $F \subset F'$ and $\tau'|_F = \tau$ hold. Since (K, σ) belongs to M, we see that M is not empty. Furthermore, using the standard union argument, we see that every totally ordered subset of M admits an upper bound. Thus, the assumptions of Zorn's lemma are met, and it follows that M contains a maximal element (F, τ). But then $F = K'$, since otherwise, we could fix an element $\alpha \in K' - F$ and extend τ with the help of Lemma 8 to a homomorphism $\tau'\colon F(\alpha) \longrightarrow L$, contradicting the maximality of (F, τ). In particular, the existence of the desired extension $\sigma'\colon K' \longrightarrow L$ of σ is clear.

If, in addition, K' is algebraically closed, then the same is true for $\sigma'(K')$. Furthermore, if L is algebraic over $\sigma(K)$, it is algebraic over $\sigma'(K')$ as well, and we conclude that $\sigma'(K') = L$ using Remark 3. Since field homomorphisms are always injective, σ' is an isomorphism. □

Corollary 10. *Let $\overline{K}_1$ and $\overline{K}_2$ be two algebraic closures of a field K. Then there exists an isomorphism $\overline{K}_1 \xrightarrow{\sim} \overline{K}_2$, noncanonical in general, that extends the identity map on K.*

Exercises

1. *Let $f \in \mathbb{Q}[X]$ be a polynomial of degree > 1. Why is it less complicated to construct a zero of f within the "setting of algebra" than within the "setting of analysis"?*
2. *Why is it not possible to prove the existence of an algebraic closure $\overline{K}$ of a field K along the following lines: Consider all algebraic extensions of K and observe for*

a family $(K_i)_{i\in I}$ of such fields that is totally ordered with respect to the inclusion relation that also the union $\bigcup_{i\in I} K_i$ is an algebraic extension of K. Therefore, Zorn's lemma yields the existence of a maximal algebraic extension and thereby of an algebraic closure of K.

3. *Why should one make a difference between particular algebraic closures of a field K and avoid talking about "the" algebraic closure of K?*

4. Let K be a field and $f \in K[X]$ a polynomial of degree ≥ 1. Show for a maximal ideal $\mathfrak{m} \subset K[X]$ containing f that $L = K[X]/\mathfrak{m}$ can be viewed as an algebraic extension field of K in which f admits a zero. Furthermore, show that L coincides with the extension field one obtains by applying Kronecker's construction to a suitable irreducible factor of f.

5. Let $\overline{K}$ be an algebraic closure of a field K. Show that $\overline{K}$ is countable if the same is true for K.

6. Show that every algebraically closed field consists of infinitely many elements.

7. Let L/K be a finite field extension of degree $[L : K] = n$. Assume there is an element $\alpha \in L$ together with isomorphisms $\sigma_i \colon L \longrightarrow L$, $i = 1, \ldots, n$, satisfying $\sigma_i|_K = \mathrm{id}_K$ and $\sigma_i(\alpha) \neq \sigma_j(\alpha)$ for $i \neq j$. Show that $L = K(\alpha)$.

8. Let $\overline{\mathbb{Q}}$ be an algebraic closure of $\mathbb{Q}$. Determine all homomorphisms $\mathbb{Q}(\sqrt[4]{2}, i) \longrightarrow \overline{\mathbb{Q}}$ as well as their images.

3.5 Splitting Fields

In the present section we start with some preparations that will be of particular interest later when we study Galois theory. As prerequisites we use the existence of algebraically closed fields, as established in Section 3.4, as well as the results 3.4/8 and 3.4/9 on the extension of field homomorphisms. If L/K and L'/K are two field extensions and $\sigma \colon L \longrightarrow L'$ is a field homomorphism, we call σ a *K-homomorphism* if it restricts to the identity map on K.

Definition 1. *Let $\mathfrak{F} = (f_i)_{i\in I}$, $f_i \in K[X]$, be a family of nonconstant polynomials with coefficients in a field K. A* splitting field *(over K) of the family $\mathfrak{F}$ is a field L extending K such that:*

(i) *Every polynomial f_i decomposes into a product of linear factors over L.*

(ii) *The extension L/K is generated by the zeros of the polynomials f_i.*

In its simplest case, the family $\mathfrak{F}$ consists of a single polynomial $f \in K[X]$. Then, if $\overline{K}$ is an algebraic closure of K and $a_1, \ldots, a_n$ are the zeros of f in $\overline{K}$, it follows that $L = K(a_1, \ldots, a_n)$ is a splitting field of f over K. In a similar way, one shows that splitting fields exist for arbitrary families $\mathfrak{F}$ of nonconstant polynomials $f_i \in K[X]$. Just choose an algebraic closure $\overline{K}$ of K and define L as the subfield of $\overline{K}$ that is generated over K by all zeros of the polynomials f_i. If the family $\mathfrak{F}$ consists of only finitely many polynomials $f_1, \ldots, f_n$, then every splitting field of the product $f_1 \cdot \ldots \cdot f_n$ is a splitting field of $\mathfrak{F}$ and vice versa.

Proposition 2. *Let L_1, L_2 be two splitting fields of a family $\mathfrak{F}$ of nonconstant polynomials in $K[X]$, for a field K, and let $\overline{L}_2$ be an algebraic closure of L_2. Then every K-homomorphism $\overline{\sigma}\colon L_1 \longrightarrow \overline{L}_2$ restricts to a K-isomorphism $\sigma\colon L_1 \xrightarrow{\sim} L_2$.*

In particular, since the inclusion $K \hookrightarrow \overline{L}_2$ extends to a K-homomorphism $\overline{\sigma}\colon L_1 \longrightarrow \overline{L}_2$ by 3.4/9, it follows that L_1 is K-isomorphic to L_2. Therefore, we can state the following corollary:

Corollary 3. *Let L_1, L_2 be two splitting fields of a family of nonconstant polynomials in $K[X]$. Then there exists a K-isomorphism $L_1 \xrightarrow{\sim} L_2$.*

Proof of Proposition 2. First we consider the case in which $\mathfrak{F}$ consists of a single polynomial f, which we may assume to be monic. Let $a_1, \ldots, a_n$ be the zeros of f in L_1 and $b_1, \ldots, b_n$ the zeros of f in $L_2 \subset \overline{L}_2$. Then we get

$$f^{\overline{\sigma}} = \prod \big(X - \overline{\sigma}(a_i)\big) = \prod (X - b_i),$$

and $\overline{\sigma}$ maps the set of the a_i bijectively onto the set of the b_i. Hence, we see that

$$L_2 = K(b_1, \ldots, b_n) = K\big(\overline{\sigma}(a_1), \ldots, \overline{\sigma}(a_n)\big) = \overline{\sigma}(L_1),$$

i.e., $\overline{\sigma}$ restricts to a K-isomorphism $\sigma\colon L_1 \longrightarrow L_2$, as claimed.

The special case just dealt with settles the assertion of the proposition for finite families $\mathfrak{F}$ as well; just view L_1 and L_2 as splitting fields of the product of all polynomials in $\mathfrak{F}$. Finally, the general case is derived by viewing L_1 and L_2 as unions of splitting fields corresponding to finite subfamilies of $\mathfrak{F}$. □

We want to characterize the property of a field L being a splitting field of a family of polynomials in $K[X]$ by equivalent conditions and then introduce normal field extensions.

Theorem 4. *The following conditions are equivalent for a field K and an algebraic extension $K \subset L$:*

(i) *Every K-homomorphism $L \longrightarrow \overline{L}$ into an algebraic closure $\overline{L}$ of L restricts to an automorphism of L.*

(ii) *L is a splitting field of a family of polynomials in $K[X]$.*

(iii) *Every irreducible polynomial in $K[X]$ that admits a zero in L decomposes over L completely into linear factors.*

Definition 5. *An algebraic field extension $K \subset L$ is called* normal *if it satisfies the equivalent conditions of Theorem* 4.

Proof of Theorem 4. We start with the implication from (i) to (iii) and consider an irreducible polynomial $f \in K[X]$ admitting a zero $a \in L$. If $b \in \overline{L}$ is another zero of f, we can conclude from 3.4/8 that there is a K-homomorphism

$\sigma\colon K(a) \longrightarrow \overline{L}$ satisfying $\sigma(a) = b$. Furthermore, using 3.4/9, we can extend σ to a K-homomorphism $\sigma'\colon L \longrightarrow \overline{L}$. Now, if condition (i) is given, we obtain $\sigma'(L) = L$ and thereby see that $b = \sigma'(a) \in L$. Hence, all zeros of f are contained in L, and f decomposes over L into a product of linear factors.

Next we show that (iii) implies condition (ii). Let $(a_i)_{i\in I}$ be a family of elements in L such that the field extension L/K is generated by the elements a_i. Furthermore, let f_i be the minimal polynomial of a_i over K. Then, according to (iii), all f_i decompose over L into a product of linear factors, and we see that L is a splitting field of the family $\mathfrak{F} = (f_i)_{i\in I}$.

Finally, assume condition (ii), i.e., that L is a splitting field of a family $\mathfrak{F}$ of polynomials in $K[X]$. If $\sigma\colon L \longrightarrow \overline{L}$ is a K-homomorphism, it follows that $\sigma(L)$, just like L, is a splitting field of $\mathfrak{F}$. However, then we must have $\sigma(L) = L$, since both fields are subfields of $\overline{L}$; see also Proposition 2. □

Every polynomial of degree 2 decomposes over a given field L into a product of linear factors, provided it admits a zero in L. Thus, we see from condition (ii) (or even (iii)) of Theorem 4 that field extensions of degree 2 are always normal. Furthermore, we can draw the following conclusion:

Remark 6. *If $K \subset L \subset M$ is a chain of algebraic field extensions and if M/K is normal, then the same is true for M/L.*

Let us add that the property of a field extension being normal is *not* transitive, i.e., for a chain of fields $K \subset L \subset M$ such that L/K and M/L are normal, the extension M/K does not need to be normal. For example, $\mathbb{Q}(\sqrt{2})/\mathbb{Q}$ and $\mathbb{Q}(\sqrt[4]{2})/\mathbb{Q}(\sqrt{2})$ are normal extensions, since they are of degree 2, while the extension $\mathbb{Q}(\sqrt[4]{2})/\mathbb{Q}$ is not normal. Indeed, the polynomial $X^4 - 2$ is irreducible over $\mathbb{Q}$ and admits in $\mathbb{Q}(\sqrt[4]{2})$ the zero $\sqrt[4]{2}$. On the other hand, $X^4 - 2$ is not a product of linear factors over $\mathbb{Q}(\sqrt[4]{2})$, since the complex zero $i \cdot \sqrt[4]{2}$ does not belong to $\mathbb{Q}(\sqrt[4]{2}) \subset \mathbb{R}$.

It is often useful to know that every algebraic field extension L/K admits a *normal closure*, by which we mean an extension field L' of L such that L'/L is algebraic and L'/K is normal with the property that there is no proper subfield of L' satisfying these conditions. Thus, L' is, so to speak, a minimal extension of L such that L'/K is normal.

Proposition 7. *Let L/K be an algebraic field extension.*

(i) *L/K admits a normal closure L'/K, where L' is unique up to (non-canonical) isomorphism over L.*

(ii) *L'/K is finite if L/K is finite.*

(iii) *If M/L is an algebraic field extension such that M/K is normal, we can choose L' as an intermediate field of M/L. In this case, L' is unique. More precisely, if $(\sigma_i)_{i\in I}$ is the family of all K-homomorphisms from L to M, then $L' = K(\sigma_i(L); i \in I)$. We call L' the* normal closure *of L in M.*

Proof. Assume $L = K(\mathfrak{A})$, where $\mathfrak{A} = (a_j)_{j\in J}$ is a family of elements in L. Let f_j be the minimal polynomial of a_j over K. If M is an algebraic extension field of L such that M/K is normal (for example, choose for M an algebraic closure of L), then it follows from Theorem 4 (iii) that the polynomials f_j decompose in $M[X]$ into a product of linear factors. Now let L' be the subfield of M that is generated over K by the zeros of the f_j. Then L' is defined as a splitting field of the f_j. Furthermore, we have $L \subset L' \subset M$, and it is clear that L'/K is a normal closure of L/K. Also we see that L'/K is finite if L/K is finite. On the other hand, if L'/K is a normal closure of L/K, then the field L' contains necessarily a splitting field of the f_j and thus, due to the minimality condition, is a splitting field of the f_j over K.

To establish the uniqueness assertion, consider two normal closures L'_1/K and L'_2/K of L/K. As we have just seen, L'_1 and L'_2 are splitting fields of the polynomials f_j over K and hence also splitting fields of the f_j over L. Then Corollary 3 yields the existence of an L-isomorphism $L'_1 \longrightarrow L'_2$, which implies the uniqueness assertion of (i) and, in conjunction with Theorem 4 (i), also the uniqueness assertion of (iii).

Finally, to derive the explicit characterization of L' as given in (iii), consider a K-homomorphism $\sigma\colon L \longrightarrow M$. By 3.4/8 it maps the zeros of the f_j to zeros of the same type. Since L' is generated over K by these zeros, we see that $K(\sigma_i(L); i \in I) \subset L'$. Conversely, for every zero $a \in L'$ of one of the polynomials f_j we can define, due to 3.4/8 again, a K-homomorphism $K(a_j) \longrightarrow L'$ such that $a_j \longmapsto a$. This can be extended via 3.4/9 to a K-automorphism of an algebraic closure of L' and subsequently be restricted to a K-homomorphism $\sigma\colon L \longrightarrow L'$, using the normality of L'/K. Thus, $a \in K(\sigma_i(L); i \in I)$, and the equality $L' = K(\sigma_i(L); i \in I)$ is clear. □

Exercises

1. *Give a detailed justification for the fact that field extensions of degree* 2 *are always normal.*
2. *Let L/K be a field extension, where L is a splitting field of a nonconstant polynomial $f \in K[X]$. Explain once again why every irreducible polynomial of $K[X]$ admitting a zero in L will decompose over L into a product of linear factors.*
3. *Let K be a field and L a splitting field of the family of all nonconstant polynomials in $K[X]$. Show that L is an algebraic closure of K.*
4. Show for a finite field extension L/K that condition (i) in Theorem 4 is equivalent to the following one:

   (i′) Every K-homomorphism $L \longrightarrow L'$ into a finite field extension L' of L restricts to an automorphism of L.

   Replace condition (i) of Theorem 4 by (i′) and sketch a proof of this theorem for finite extensions L/K that avoids the existence of an algebraic closure of K.
5. Consider $L = \mathbb{Q}(\sqrt[4]{2}, i)$ as a field extending $\mathbb{Q}$.

   (i) Show that L is a splitting field of the polynomial $X^4 - 2 \in \mathbb{Q}[X]$.

(ii) Determine the degree of L over $\mathbb{Q}$, as well as all $\mathbb{Q}$-automorphisms of L.

(iii) Show that $L = \mathbb{Q}(\sqrt[4]{2} + i)$ using Exercise 7 of Section 3.4.

6. Determine a splitting field L of the polynomial $X^4 + 2X^2 - 2$ over $\mathbb{Q}$, as well as the degree $[L : \mathbb{Q}]$.

7. Check whether the field extension $\mathbb{Q}\left(\sqrt{2 + \sqrt{2}}\right)/\mathbb{Q}$ is normal.

8. Let K be a field and $f \in K[X]$ a polynomial of degree $n > 0$. Let L be a splitting field of f over K. Show:

(i) $[L : K]$ divides $n!$.

(ii) If $[L : K] = n!$, then f is irreducible.

9. Determine a splitting field L over $\mathbb{Q}$ of the family $\{X^4 + 1,\ X^5 + 2\}$ and compute the degree $[L : \mathbb{Q}]$.

10. Consider $f = X^6 - 7X^4 + 3X^2 + 3$ as a polynomial in $\mathbb{Q}[X]$, as well as in $\mathbb{F}_{13}[X]$. In either case, decompose f into its irreducible factors and determine a splitting field of f over $\mathbb{Q}$, resp. $\mathbb{F}_{13}$.

11. Let $K(\alpha)/K$ and $K(\beta)/K$ be simple algebraic field extensions with minimal polynomials f of α, resp. g of β, over K. Show that f is irreducible over $K(\beta)$ if and only if g is irreducible over $K(\alpha)$. Furthermore, show that f and g are irreducible if $\deg f$ and $\deg g$ are prime to each other.

12. Let L/K be a normal algebraic field extension and $f \in K[X]$ a monic irreducible polynomial. Let $f = f_1 \dots f_r$ be a prime factorization of f in $L[X]$ with monic factors f_i. Show for every two factors f_i, f_j, $i \neq j$, that there is a K-automorphism $\sigma\colon L \longrightarrow L$ satisfying $f_j = f_i^{\sigma}$.

13. Let L/K and L'/K be normal algebraic field extensions and let L'' be a field containing L and L' as subfields.

(i) Show that $(L \cap L')/K$ is a normal algebraic field extension.

(ii) Use (i) to give an alternative proof of Proposition 7.

3.6 Separable Field Extensions

When looking at polynomials $f \in K[X]$ over a field K, it is convenient to consider their zeros in an algebraic closure $\overline{K}$ of K. Assertions on the zeros of polynomials f are quite often independent of the choice of $\overline{K}$, since such an algebraic closure is unique up to K-isomorphism. For example, it is meaningful to say that f has only simple zeros, or that f has multiple zeros. Nonconstant polynomials with only simple zeros, i.e., zeros of multiplicity 1, are called *separable*.

Lemma 1. *Let K be a field and $f \in K[X]$ a nonconstant polynomial.*

(i) *The multiple zeros of f (in an algebraic closure $\overline{K}$ of K) coincide with the common zeros of f and its derivative f', or equivalently, with the zeros of* $\gcd(f, f')$.

(ii) *If f is irreducible, it has multiple zeros if and only if its derivative f' is identically zero.*

Proof. Assertion (i) is a consequence of 2.6/3, at least when K is algebraically closed. (In Section 2.6, zeros of polynomials $f \in K[X]$ were always considered in K, not yet in an algebraic closure of K.) To settle the general case in which K is not necessarily algebraically closed, it is enough to point out that a greatest common divisor $d = \gcd(f, f')$ in $K[X]$ is at the same time also a greatest common divisor of f and f' in $\overline{K}[X]$. To justify this, use the ideal-theoretic characterization of the greatest common divisor in principal ideal domains 2.4/13. Indeed, from the equation $d \cdot K[X] = f \cdot K[X] + f' \cdot K[X]$ we conclude that $d \cdot \overline{K}[X] = f \cdot \overline{K}[X] + f' \cdot \overline{K}[X]$, i.e., we have $d = \gcd(f, f')$ in $K[X]$ as well as in $\overline{K}[X]$.

To verify (ii), assume that f is irreducible and, in addition, monic. Then, if $a \in \overline{K}$ is a zero of f, we recognize f as the minimal polynomial of a over K. Furthermore, we know from (i) that a is a multiple zero of f if and only if a is a zero of f' as well. However, since $\deg f' < \deg f$ and f is the minimal polynomial of a over K, it is clear that f' can vanish at a only if it equals the zero polynomial. □

A nonconstant polynomial $f \in K[X]$ has always a nontrivial derivative $f' \neq 0$ if $\operatorname{char} K = 0$. Therefore, assertion (ii) of the lemma implies that irreducible polynomials are separable in characteristic 0. On the other hand, there are irreducible polynomials in characteristic > 0 that are not separable. For example, let p be a prime number, t a variable, and consider the function field $K = \mathbb{F}_p(t) = Q(\mathbb{F}_p[t])$. Then $X^p - t$, as a polynomial in $K[X]$, is irreducible by Eisenstein's criterion 2.8/1, but not separable, since $f' = pX^{p-1} = 0$. In the following, let us look more closely at the case of positive characteristic.

Proposition 2. *Let K be a field and $f \in K[X]$ an irreducible polynomial.*

(i) *If $\operatorname{char} K = 0$, then f is separable.*

(ii) *If $\operatorname{char} K = p > 0$, choose $r \in \mathbb{N}$ maximal such that f is a polynomial in X^{p^r}, i.e., such that there is a polynomial $g \in K[X]$ satisfying $f(X) = g(X^{p^r})$. Then every zero of f is of multiplicity p^r and g is an irreducible polynomial that is separable. The zeros of f equal the p^rth roots of the zeros of g.*

Proof. The case $\operatorname{char} K = 0$ was discussed before, so assume $\operatorname{char} K = p > 0$. Furthermore, write

$$f = \sum_{i=0}^{n} c_i X^i, \qquad f' = \sum_{i=1}^{n} i c_i X^{i-1}.$$

Then $f' = 0$ is equivalent to $ic_i = 0$ for $i = 1, \ldots, n$. Since ic_i vanishes precisely when we have $p \,|\, i$ or $c_i = 0$, we conclude that f' is the zero polynomial precisely when there exists some $h \in K[X]$ such that $f(X) = h(X^p)$.

Now assume $f(X) = g(X^{p^r})$, as specified in assertion (ii). If we apply the above reasoning to g instead of f, we get $g' \neq 0$ due to the maximality of r. Furthermore, g is irreducible, since the same is true for f. Therefore, g is separable by Lemma 1 (ii). Now choose an algebraic closure $\overline{K}$ of K and consider a factorization

$$g = \prod_i (X - a_i), \qquad a_i \in \overline{K},$$

where we have assumed f and therefore also g to be monic. Introducing p^rth roots $c_i \in \overline{K}$ such that $c_i^{p^r} = a_i$, we obtain

$$f = \prod_i (X^{p^r} - c_i^{p^r}) = \prod_i (X - c_i)^{p^r}$$

using 3.1/3, and thereby see that all zeros of f are of multiplicity p^r. □

Next we want to define the notion of separability for algebraic field extensions.

Definition 3. *Let $K \subset L$ be an algebraic field extension. An element $\alpha \in L$ is called* separable *over K if α is a zero of a separable polynomial of $K[X]$, or equivalently, if the minimal polynomial of α over K is separable. The field L is called* separable *over K if every element $\alpha \in L$ is separable over K.*

A field K is called *perfect* if every algebraic extension of K is separable. For example, we can deduce from Proposition 2 (i) the following:

Remark 4. *Every algebraic field extension in characteristic 0 is separable. In particular, all fields of characteristic 0 are perfect.*

For a prime number p and a variable t we have already seen that the polynomial $X^p - t \in \mathbb{F}_p(t)[X]$ is irreducible, but not separable. Therefore, the field $\mathbb{F}_p(t)[X]/(X^p - t)$ is not separable over $\mathbb{F}_p(t)$. Equivalently, we can state that the algebraic field extension $\mathbb{F}_p(t)/\mathbb{F}_p(t^p)$ fails to be separable, since $X^p - t^p$, as an irreducible polynomial in $\mathbb{F}_p(t^p)[X]$, is the minimal polynomial of t over $\mathbb{F}_p(t^p)$.

In the following we want to characterize separable algebraic field extensions more thoroughly. In particular, we want to show that an algebraic field extension is separable as soon as it is generated by separable elements. To achieve this, we need the notion of the separable degree as a replacement of the usual degree, which was used in studying general algebraic field extensions.

Definition 5. *For an algebraic field extension $K \subset L$, let us denote by $\mathrm{Hom}_K(L, \overline{K})$ the set of all K-homomorphisms from L into an algebraic closure $\overline{K}$ of K. Then the* separable degree *of L over K, denoted by $[L\colon K]_s$, is defined as the number of elements in $\mathrm{Hom}_K(L, \overline{K})$, i.e.,*

$$[L : K]_s := \#\,\mathrm{Hom}_K(L, \overline{K}).$$

It follows from 3.4/10 that the separable degree of an algebraic extension L/K is independent of the choice of an algebraic closure $\overline{K}$ of K. Let us compute the separable degree for simple algebraic field extensions.

Lemma 6. *Let $K \subset L = K(\alpha)$ be a simple algebraic field extension with minimal polynomial $f \in K[X]$ of α over K.*

(i) *The separable degree $[L:K]_s$ equals the number of different zeros of f in an algebraic closure of K.*

(ii) *The element α is separable over K if and only if $[L:K] = [L:K]_s$.*

(iii) *Assume* $\operatorname{char} K = p > 0$ *and let p^r be the multiplicity of the zero α of f (cf. Proposition 2 (ii)). Then $[L:K] = p^r[L:K]_s$.*

Proof. Assertion (i) is a reformulation of 3.4/8. To justify (ii), let $n = \deg f$. Then α is separable if and only if f does not admit multiple zeros and hence has n distinct zeros, or according to (i), if and only if $n = [L:K]_s$. However, due to 3.2/6 we have $[L:K] = \deg f = n$, and it follows that α is separable precisely when $[L:K] = [L:K]_s$. Finally, assertion (iii) is a direct consequence of Proposition 2 (ii). □

To handle the separable degree of more general algebraic field extensions we need an analogue of the multiplicativity formula 3.2/2.

Proposition 7 (Multiplicativity formula). *Let $K \subset L \subset M$ be algebraic field extensions. Then*

$$[M:K]_s = [M:L]_s \cdot [L:K]_s.$$

Proof. Fix an algebraic closure $\overline{K}$ of M. Then $K \subset L \subset M \subset \overline{K}$, and we may view $\overline{K}$ also as an algebraic closure of K and of L. Furthermore, let

$$\operatorname{Hom}_K(L, \overline{K}) = \{\sigma_i \,;\, i \in I\}, \qquad \operatorname{Hom}_L(M, \overline{K}) = \{\tau_j \,;\, j \in J\},$$

where in each case, the σ_i as well as the τ_j are distinct. Now extend the K-homomorphisms $\sigma_i \colon L \longrightarrow \overline{K}$ via 3.4/9 to K-automorphisms $\overline{\sigma}_i \colon \overline{K} \longrightarrow \overline{K}$. The desired multiplicativity formula will then be a consequence of the following two assertions:

(1) The maps $\overline{\sigma}_i \circ \tau_j \colon M \longrightarrow \overline{K}$, $i \in I$, $j \in J$, are distinct.

(2) $\operatorname{Hom}_K(M, \overline{K}) = \{\overline{\sigma}_i \circ \tau_j \,;\, i \in I, j \in J\}$.

To verify assertion (1), consider an equation of type $\overline{\sigma}_i \circ \tau_j = \overline{\sigma}_{i'} \circ \tau_{j'}$. Since τ_j and $\tau_{j'}$ restrict to the identity on L, we can conclude that $\sigma_i = \sigma_{i'}$ and hence $i = i'$. The latter implies $\tau_j = \tau_{j'}$, and thus $j = j'$. It follows that the maps specified in (1) are distinct. Since they are K-homomorphisms, it remains to show for (2) that every K-homomorphism $\tau \colon M \longrightarrow \overline{K}$ is as specified in (1). For $\tau \in \operatorname{Hom}_K(M, \overline{K})$ we have $\tau|_L \in \operatorname{Hom}_K(L, \overline{K})$. Hence, there exists an index $i \in I$ such that $\tau|_L = \sigma_i$. Then we obtain $\overline{\sigma}_i^{-1} \circ \tau \in \operatorname{Hom}_L(M, \overline{K})$, and there

exists an index $j \in J$ such that $\overline{\sigma}_i^{-1} \circ \tau = \tau_j$. Therefore, $\tau = \overline{\sigma}_i \circ \tau_j$ and (2) is clear. □

If we take into account that algebraic field extensions are separable in characteristic 0 (Remark 4), we can apply the multiplicativity formulas of 3.2/2 and Proposition 7 inductively and thereby derive from Lemma 6 the following result:

Proposition 8. *Let $K \subset L$ be a finite field extension.*

(i) *If* $\operatorname{char} K = 0$, *then* $[L : K] = [L : K]_s$.

(ii) *If* $\operatorname{char} K = p > 0$, *then* $[L : K] = p^r [L : K]_s$ *for some exponent* $r \in \mathbb{N}$. *In particular*, $1 \leq [L : K]_s \leq [L : K]$, *and* $[L : K]_s$ *divides* $[L : K]$.

We are now able to characterize finite separable field extensions in terms of the separable degree.

Theorem 9. *For a finite field extension $K \subset L$ the following conditions are equivalent*:

(i) *L/K is separable.*

(ii) *There exist elements $a_1, \ldots, a_n \in L$ that are separable over K and satisfy $L = K(a_1, \ldots, a_n)$.*

(iii) $[L : K]_s = [L : K]$.

Proof. The implication from (i) to (ii) is trivial. If $a \in L$ is separable over K, then the same is true over every intermediate field of L/K. Therefore, using the multiplicativity formulas in 3.2/2 and in Proposition 7, the implication from (ii) to (iii) can be reduced to the case of a simple field extension. However, that case was already dealt with in Lemma 6 (ii).

It remains to show that (iii) implies (i). Consider an element $a \in L$ with its minimal polynomial $f \in K[X]$ over K. To show that a is separable over K, which amounts to showing that f admits only simple zeros, we are reduced to the case $\operatorname{char} K = p > 0$, due to Remark 4. Then, by Proposition 2 (ii), there is an exponent $r \in \mathbb{N}$ such that every zero of f is of multiplicity p^r. Hence, we get

$$[K(a) : K] = p^r \cdot [K(a) : K]_s$$

from Lemma 6. Using the multiplicativity formulas of 3.2/2 and Proposition 7 in conjunction with the estimate between the degree and the separable degree in Proposition 8, we obtain

$$\begin{aligned} [L : K] &= [L : K(a)] \cdot [K(a) : K] \\ &\geq [L : K(a)]_s \cdot p^r \cdot [K(a) : K]_s = p^r \cdot [L : K]_s. \end{aligned}$$

Now, if $[L : K]_s = [L : K]$, we must have $r = 0$. Then all zeros of f are simple and a is separable over K, which shows that (iii) implies condition (i). □

Corollary 10. *Let $K \subset L$ be an algebraic field extension and $\mathfrak{A}$ a family of elements in L such that L/K is generated by $\mathfrak{A}$. Then the following conditions are equivalent*:

(i) *L/K is separable.*

(ii) *Every $a \in \mathfrak{A}$ is separable over K.*

Each of these conditions implies $[L : K] = [L : K]_s$.

Proof. Every $a \in L$ is contained in a subfield of type $K(a_1, \ldots, a_n)$, where $a_1, \ldots, a_n \in \mathfrak{A}$. In this way, the equivalence between (i) and (ii) is a direct consequence of Theorem 9. Furthermore, for L/K finite separable, we conclude that $[L : K] = [L : K]_s$, again by Theorem 9. Now let L/K be separable while $[L : K] = \infty$. Then every intermediate field E of L/K is separable over K as well, so that we obtain $[E : K] = [E : K]_s$ for $[E : K] < \infty$ and hence $[L : K]_s \geq [E : K]$ using the multiplicativity formula of Proposition 7. Since there exist intermediate fields E of L/K of arbitrarily large degree, we see that $[L : K]_s = \infty = [L : K]$. □

Corollary 11. *Let $K \subset L \subset M$ be algebraic field extensions. Then M/K is separable if and only if M/L and L/K are separable.*

Proof. We have only to show that the separability of M/L and L/K implies the separability of M/K. Fix an element $a \in M$ with minimal polynomial $f \in L[X]$ over L. Furthermore, let L' be the intermediate field of L/K that is generated over K by the coefficients of f. Since M/L is separable, we conclude that f is separable. Therefore, $L'(a)/L'$ is separable, and the same is true for L'/K, since L/K is separable. Moreover, $L'(a)/L'$ and L'/K are finite, and we see using the multiplicativity formulas that

$$\begin{aligned}\big[L'(a) : K\big]_s &= \big[L'(a) : L'\big]_s \cdot \big[L' : K\big]_s \\ &= \big[L'(a) : L'\big] \cdot \big[L' : K\big] = \big[L'(a) : K\big].\end{aligned}$$

This shows that $L'(a)$, and in particular a, are separable over K. □

Finally, we want to prove the *primitive element theorem*, which asserts that finite separable field extensions are simple.

Proposition 12. *Let L/K be a finite field extension, say $L = K(a_1, \ldots, a_r)$, and assume that the elements $a_2, \ldots, a_r$ are separable over K. Then the extension L/K admits a* primitive element, *i.e., an element $a \in L$ such that $L = K(a)$. In particular, every finite separable field extension admits a primitive element.*

Proof. Let us first consider the case that K consists of only finitely many elements. Then, since $[L : K] < \infty$, also L is finite. In particular, the multiplicative group L^* is finite and hence cyclic, as we will show below in Proposition 14. Any element $a \in L$ generating L^* as a cyclic group will also generate the field

extension L/K, and we see that L/K is simple. Note that this argument does not use any separability assumption on the extension L/K. However, L/K is automatically separable if K is a finite field, as we will see in 3.8/4.

It remains to look at the case that K consists of infinitely many elements. Using an induction argument we may assume $r = 2$, i.e., that L is generated over K by two elements a and b with b being separable over K. Write $n = [L : K]_s$ and let $\sigma_1, \ldots, \sigma_n$ be the distinct elements of $\mathrm{Hom}_K(L, \overline{K})$, where as usual, $\overline{K}$ is an algebraic closure of K. Then consider the polynomial

$$P = \prod_{i \neq j} \Big[\big(\sigma_i(a) - \sigma_j(a)\big) \ + \ \big(\sigma_i(b) - \sigma_j(b)\big) \cdot X \Big].$$

We claim that $P \in \overline{K}[X]$ is not the zero polynomial. Indeed, for $i \neq j$ we must have $\sigma_i(a) \neq \sigma_j(a)$ or $\sigma_i(b) \neq \sigma_j(b)$, since otherwise σ_i would coincide with σ_j on $L = K[a, b]$. Since K contains infinitely many elements, but P can admit only finitely many zeros, there must exist an element $c \in K$ satisfying $P(c) \neq 0$. This implies that the elements

$$\sigma_i(a) + c\sigma_i(b) = \sigma_i(a + cb) \in \overline{K}, \qquad i = 1, \ldots, n,$$

are distinct and hence that $L' = K(a + cb)$ is a simple algebraic field extension of K of separable degree

$$[L' : K]_s \geq n = [L : K]_s.$$

Now $L' \subset L$ implies $[L' : K]_s \leq [L : K]_s$ and therefore $[L' : K]_s = [L : K]_s$. Let us show that we have, in fact, $L' = L$ and hence that L/K is simple. By our assumption, $b \in L$ is separable over K and hence also over L', so that $[L'(b) : L']_s = [L'(b) : L']$. Furthermore, the multiplicativity formula of Proposition 7 yields

$$[L : K]_s \geq [L'(b) : K]_s = [L'(b) : L']_s \cdot [L' : K]_s = [L'(b) : L'] \cdot [L : K]_s.$$

This shows that $[L'(b) : L'] = 1$ and hence $L'(b) = L'$. Therefore, b belongs to $L' = K(a+cb)$, and the same is true for a. In particular, we get $L = K(a, b) = L'$, and L is a simple field extension of K. □

It remains to verify that the multiplicative group of a finite field is cyclic. To do this we need an auxiliary result from group theory.

Lemma 13. *Let a and b be two elements of finite order in an abelian group G, say* $\operatorname{ord} a = m$ *and* $\operatorname{ord} b = n$. *Then there exists an element of order* $\operatorname{lcm}(m, n)$ *in G.*

More precisely, choose integer decompositions $m = m_0 m'$, $n = n_0 n'$, where $\operatorname{lcm}(m, n) = m_0 n_0$ *and* $\gcd(m_0, n_0) = 1$. *Then $a^{m'} b^{n'}$ is an element of order* $\operatorname{lcm}(m, n)$. *In particular, ab is of order mn if m and n are prime to each other.*

Proof. Let us first assume that m and n are prime to each other, and show that ab is of order mn. Clearly, we have $(ab)^{mn} = (a^m)^n(b^n)^m = 1$. On the other hand, the equation $(ab)^t = 1$ implies $a^{nt} = a^{nt}b^{nt} = 1$, and we get $m \mid t$, since $\gcd(m,n) = 1$. Similarly, it follows that $n \mid t$ and hence $mn \mid t$, so that $\operatorname{ord}(ab) = mn$.

In the general case, choose decompositions $m = m_0m'$, $n = n_0n'$ such that $\operatorname{lcm}(m,n) = m_0n_0$ and $\gcd(m_0,n_0) = 1$. To achieve this, consider a prime factorization $p_1^{\nu_1} \cdot \ldots \cdot p_r^{\nu_r}$ of $\operatorname{lcm}(m,n)$ and define m_0 as the product of all prime powers $p_i^{\nu_i}$ dividing m, as well as n_0 as the product of all prime powers $p_i^{\nu_i}$ not dividing m. Then we get $m_0 \mid m$ as well as $n_0 \mid n$, and the resulting decompositions $m = m_0m'$, $n = n_0n'$ clearly satisfy $\operatorname{lcm}(m,n) = m_0n_0$, as well as $\gcd(m_0,n_0) = 1$.

Now, since $a^{m'}$ is of order m_0 and $b^{n'}$ is of order n_0, the order of $a^{m'}b^{n'}$ equals m_0n_0 by the special case considered before. □

Proposition 14. *Let K be a field and H a finite subgroup of the multiplicative group K^*. Then H is cyclic.*

Proof. Fix an element $a \in H$ of maximal order m and let H_m be the subgroup of all elements in H whose order divides m. Then all elements of H_m are zeros of the polynomial $X^m - 1$, so that H_m can contain at most m elements. On the other hand, H_m contains the cyclic group $\langle a \rangle$ generated by a, whose order is m. Therefore, we must have $H_m = \langle a \rangle$, and H_m is cyclic. We claim that in fact, $H = H_m$. Indeed, if there existed an element $b \in H$ not belonging to H_m and hence whose order n did not divide m, then H would contain an element of order $\operatorname{lcm}(m,n) > m$, due to Lemma 13. However, this is in contradiction to the choice of a. □

Exercises

1. *Recall the proof of the following fact*: *For an algebraic field extension L/K and two elements $a, b \in L$ that are separable over K, their sum $a + b$ is separable over K as well. More thoroughly, show that the set of all elements in L that are separable over K yields an intermediate field of L/K. It is called the* separable closure *of K in L.*

2. *Let K be a field and $f \in K[X]$ a nonconstant polynomial. Why is the assertion that f has multiple zeros in an algebraic closure $\overline{K}$ of K independent of the choice of $\overline{K}$?*

3. *The proof of Proposition* 12 *yields a practical method for determining primitive elements of finite separable field extensions. Give a sketch of this process.*

4. Let $K \subset L \subset M$ be algebraic field extensions such that M/K is normal. Show that $[L : K]_s = \#\operatorname{Hom}_K(L, M)$.

5. For a prime number p consider the function field $L = \mathbb{F}_p(X, Y)$ in two variables over $\mathbb{F}_p$, as well as the Frobenius homomorphism $\sigma\colon L \longrightarrow L$, $a \longmapsto a^p$. Let

$K = \sigma(L)$ be the image of σ. Determine the degrees $[L : K]$ and $[L : K]_s$, and show that the field extension L/K is not simple.

6. Let L/K be a field extension in characteristic $p > 0$ and consider an element $\alpha \in L$ that is algebraic over K. Show that α is separable over K if and only if $K(\alpha) = K(\alpha^p)$.

7. An algebraic field extension L/K is simple if and only if it admits only finitely many intermediate fields. Prove this assertion using the following steps:
   (i) First discuss the case that K is a finite field, so that after this, K can be assumed to be infinite.
   (ii) Assume $L = K(\alpha)$ and let $f \in K[X]$ be the minimal polynomial of α over K. Show that the set of intermediate fields of L/K can be identified with a subset of the set of divisors of f in $L[X]$.
   (iii) Assume that L/K admits only finitely many intermediate fields. To show that L/K is simple, reduce to the case that L is generated over K by two elements α, β. Finally, if $L = K(\alpha, \beta)$, consider fields of type $K(\alpha + c\beta)$ for constants $c \in K$.

8. Let K be a finite field. Show that the product of all elements in K^* equals -1. As an application, deduce for prime numbers p the divisibility relation $p \mid ((p-1)!+1)$.

3.7 Purely Inseparable Field Extensions

In the last section we introduced the separable degree $[L : K]_s$ for algebraic field extensions L/K and saw that $1 \leq [L : K]_s \leq [L : K]$; cf. 3.6/8. In particular, we proved that a finite extension L/K is separable if and only if $[L : K]_s = [L : K]$. Next we turn to the opposite situation and consider algebraic field extensions L/K satisfying $[L : K]_s = 1$. Since algebraic field extensions are separable in characteristic 0, we assume in the following that K is a *field of characteristic* $p > 0$.

A polynomial $f \in K[X]$ is called *purely inseparable* if it admits precisely one zero α (in an algebraic closure $\overline{K}$ of K). Since the attached minimal polynomial $m_\alpha \in K[X]$ divides f, we conclude by induction on the degree of f that f, assuming it is monic, is a power of m_α and thereby a power of an irreducible purely inseparable monic polynomial. Furthermore, if $h \in K[X]$ is such a monic irreducible polynomial that is purely inseparable, we see using 3.1/3 and 3.6/2 (ii) that h is of type $X^{p^n} - c$ for some $n \in \mathbb{N}$ and some $c \in K$. Conversely, it is clear that all polynomials of this type are purely inseparable. Hence, the monic purely inseparable polynomials in $K[X]$ consist of precisely all powers of polynomials of type $X^{p^n} - c$.

Definition 1. *Let $K \subset L$ be an algebraic field extension. An element $\alpha \in L$ is called* purely inseparable *over K if α is a zero of a purely inseparable polynomial in $K[X]$, or equivalently, if the minimal polynomial of α over K is of type $X^{p^n} - c$ for $n \in \mathbb{N}$ and $c \in K$. Furthermore, L is called* purely inseparable *over K if every element $\alpha \in L$ is purely inseparable over K in the sense just defined.*

It is immediately clear from the definition that purely inseparable field extensions are normal. The trivial extension K/K is the only algebraic field extension that is separable and purely inseparable at the same time. Also note that the extension $\mathbb{F}_p(t)/\mathbb{F}_p(t^p)$ of the preceding section is an example of a nontrivial purely inseparable field extension.

Proposition 2. *For an algebraic field extension $K \subset L$, the following conditions are equivalent*:

(i) *L is purely inseparable over K.*

(ii) *There exists a family $\mathfrak{A} = (a_i)_{i \in I}$ of elements in L such that $L = K(\mathfrak{A})$ and each a_i is purely inseparable over K.*

(iii) $[L : K]_s = 1$.

(iv) *For every element $a \in L$, there is an integer $n \in \mathbb{N}$ such that $a^{p^n} \in K$.*

Proof. The implication from (i) to (ii) is trivial. Next, to derive (iii) from (ii) it is enough to show that $[K(a_i) : K]_s = 1$ for all $i \in I$. The reason is that any K-homomorphism $L \longrightarrow \overline{K}$ into an algebraic closure $\overline{K}$ of K is uniquely determined by the images of the elements a_i. On the other hand, the minimal polynomial of each element a_i is of type $X^{p^n} - c$ and therefore admits only a single zero in $\overline{K}$. Hence, we get $[K(a_i) : K]_s = 1$ by 3.4/8, as desired.

Now let us assume condition (iii) and derive (iv) from it. Choosing an element $a \in L$, we get

$$[L : K(a)]_s \cdot [K(a) : K]_s = [L : K]_s = 1$$

and thereby $[K(a) : K]_s = 1$. This means that the minimal polynomial of a over K admits only a single zero and hence by 3.6/2, that it is of type $X^{p^n} - c$. But then we conclude that $a^{p^n} \in K$, as required in (iv). Finally, assume $a^{p^n} \in K$ for some $a \in L$ and hence that a is a zero of a polynomial of type $X^{p^n} - c \in K[X]$, which is a polynomial admitting only a single zero. Then the minimal polynomial of a over K is of the same type, and we see that a is purely inseparable over K. This shows that (iv) implies (i). □

Corollary 3. *Let $K \subset L \subset M$ be algebraic field extensions. Then M/K is purely inseparable if and only if M/L and L/K are purely inseparable.*

Proof. $[M : K]_s = [M : L]_s \cdot [L : K]_s$; cf. 3.6/7. □

Next, we want to show that every algebraic field extension can be decomposed into a separable extension, followed by a purely inseparable extension. For normal extensions, this is possible in reverse order as well.

Proposition 4. *Let L/K be an algebraic field extension. Then there exists a unique intermediate field K_s of L/K such that L/K_s is purely inseparable and K_s/K is separable. The field K_s is called the* separable closure *of K in L, i.e.,*

$$K_s = \{a \in L \,;\, a \text{ separable over } K\},$$

and we have $[L : K]_s = [K_s : K]$. *If* L/K *is normal, the extension* K_s/K *is normal, too.*

Proposition 5. *Let* L/K *be a* normal *algebraic field extension. Then there exists a unique intermediate field* K_i *of* L/K *such that* L/K_i *is separable and* K_i/K *is purely inseparable.*

Proof of Proposition 4. We write

$$K_s = \{a \in L\,;\, a \text{ separable over } K\}.$$

Then K_s is a field. Indeed, for $a, b \in K_s$ we see from 3.6/9 that $K(a, b)$ is a separable extension of K, so that $K(a, b) \subset K_s$. Therefore, K_s is the biggest separable extension of K that is contained in L. Now consider an element $a \in L$ and let $f \in K_s[X]$ be the minimal polynomial of a over K_s. Then, by 3.6/2, there exists a separable polynomial $g \in K_s[X]$ such that $f(X) = g(X^{p^r})$ for some exponent $r \in \mathbb{N}$. Moreover, g is irreducible, since f is irreducible. It follows that g is the minimal polynomial of $c = a^{p^r}$ over K_s and that c is separable over K_s, hence by 3.6/11 also separable over K. However, then we must have $c \in K_s$ and therefore $g = X - c$, as well as $f = X^{p^r} - c$. Thus, a is purely inseparable over K_s, and the same is true for L over K_s.

Since L/K_s is purely inseparable and K_s/K is separable, we get the stated relation on degrees

$$[L : K]_s = [L : K_s]_s \cdot [K_s : K]_s = [K_s : K].$$

To justify the uniqueness of K_s, consider an intermediate field K' of L/K such that L/K' is purely inseparable and K'/K is separable. Then we have $K' \subset K_s$ by the definition of K_s, and the extension K_s/K' is separable. On the other hand, the latter extension is purely inseparable, since L/K' is purely inseparable. This shows that K_s/K' is trivial and hence that K_s is unique, as claimed.

It remains to show that K_s/K is normal if the same is true for L/K. To do this, consider a K-homomorphism $\sigma\colon K_s \longrightarrow \overline{L}$ into an algebraic closure $\overline{L}$ of L. Since we can view $\overline{L}$ as an algebraic closure of K as well, we can extend σ due to 3.4/9 to a K-homomorphism $\sigma'\colon L \longrightarrow \overline{L}$. Now, assuming L/K to be normal, σ' restricts to a K-automorphism of L. Furthermore, the uniqueness of K_s implies that σ restricts to a K-automorphism of K_s, and it follows that K_s/K is normal. □

Proof of Proposition 5. Since the extension L/K is assumed to be normal, we can identify the set of K-homomorphisms of L into an algebraic closure $\overline{L}$ of L with the set of K-automorphisms of L, the latter forming a group G. Let

$$K_i = \{a \in L\,;\, \sigma(a) = a \text{ for all } \sigma \in G\}$$

be the subset in L that is left invariant under the members of G; it is easily verified that K_i is a field, the so-called *fixed field* attached to G. Using 3.4/9,

every K-homomorphism $K_i \longrightarrow \overline{L}$ extends to a K-homomorphism $L \longrightarrow \overline{L}$. Since such a homomorphism leaves K_i fixed due to its definition, we can conclude that $\#\operatorname{Hom}_K(K_i, \overline{L}) = 1$. Hence, viewing $\overline{L}$ as an algebraic closure of K, we see that K_i/K is purely inseparable. Indeed, the equivalence between (i) and (iii) in Proposition 2 shows that K_i is the biggest purely inseparable extension of K contained in L. To see that L/K_i is separable, consider an element $a \in L$, as well as a maximal system of elements $\sigma_1, \ldots, \sigma_r \in G$ such that $\sigma_1(a), \ldots, \sigma_r(a)$ are distinct. Such a finite system will always exist, even if G is not finite, since $\sigma(a)$, for $\sigma \in G$, is a zero of the minimal polynomial of a over K. Also note that a will necessarily appear among the elements $\sigma_i(a)$. Since every $\sigma \in G$ induces a bijective self-map on the set $\{\sigma_1(a), \ldots, \sigma_r(a)\}$, we can conclude that the polynomial

$$f = \prod_{i=1}^{r} \big(X - \sigma_i(a)\big)$$

has coefficients in K_i, since these are fixed by the elements of G. In particular, a is a zero of a separable polynomial in $K_i[X]$, and we see that a is separable over K_i. Letting a vary over all of L, it follows that L/K_i is separable. Finally, the uniqueness of K_i is deduced, similarly as in Proposition 4, from the fact that K_i is the biggest intermediate field of L/K that is purely inseparable over K. □

Exercises

1. *Let L/K be a field extension, and let $a, b \in L$ be elements that are purely inseparable over K. Give an explicit argument showing that $a + b$ and $a \cdot b$ are purely inseparable over K as well.*

2. *Looking at a finite field extension L/K, its inseparable degree could be defined by $[L : K]_i = [L : K] \cdot [L : K]_s^{-1}$. Discuss the drawbacks of this notion in comparison to the separable degree when formulating and proving the results of the present section on purely inseparable field extensions.*

3. *Consider a simple algebraic field extension L/K and give a direct argument for the fact known from Proposition 4 that there exists an intermediate field K_s such that L/K_s is purely inseparable and K_s/K is separable.*

4. Let K be a field of characteristic $p > 0$. Show that the Frobenius homomorphism $\sigma\colon K \longrightarrow K$, $a \longmapsto a^p$, is surjective if and only if K is perfect.

5. Let L/K be a field extension and let $\alpha \in L$ be separable over K, as well as $\beta \in L$ purely inseparable over K. Show:
   (i) $K(\alpha, \beta) = K(\alpha + \beta)$,
   (ii) $K(\alpha, \beta) = K(\alpha \cdot \beta)$ if $\alpha \neq 0 \neq \beta$.

6. Let K be a field of characteristic $p > 0$. Show:
   (i) Given $n \in \mathbb{N}$ there exists a field $K^{p^{-n}}$ extending K with the following properties: If $a \in K^{p^{-n}}$, then $a^{p^n} \in K$, and for every element $b \in K$ there exists an element $a \in K^{p^{-n}}$ such that $a^{p^n} = b$.

(ii) The field $K^{p^{-n}}$ is unique up to canonical isomorphism, and there are canonical embeddings $K \subset K^{p^{-1}} \subset K^{p^{-2}} \subset \dots$.

(iii) $K^{p^{-\infty}} = \bigcup_{i=0}^{\infty} K^{p^{-i}}$ is a perfect field.

The field $K^{p^{-\infty}}$ is called the *purely inseparable closure* of K.

7. Let L/K be an algebraic field extension. Show:

   (i) If K is perfect, the same is true for L.

   (ii) If L is perfect and L/K is finite, then K is perfect.

   Give an example showing that assertion (ii) will fail to be true in general if the finiteness of L/K is not assumed.

8. Let L/K be a separable algebraic field extension. Show that the following conditions are equivalent:

   (i) Every nonconstant separable polynomial in $L[X]$ admits a factorization into linear factors.

   (ii) Choosing an algebraic closure $\overline{K}$ of K and a K-embedding $L \hookrightarrow \overline{K}$, the extension $\overline{K}/L$ is purely inseparable.

   Show for a field K that there always exists an extension field $L = K_{\text{sep}}$ satisfying the preceding conditions and that it is unique up to (noncanonical) isomorphism over K. The field K_{sep} is called a *separable algebraic closure* of K.

9. Let L/K be a normal algebraic field extension in characteristic > 0. Consider the intermediate fields K_s and K_i as specified in Propositions 4 and 5. Show that $L = K_s(K_i) = K_i(K_s)$.

10. Let L/K be an algebraic field extension with the property that every irreducible polynomial in $K[X]$ admits at least one zero in L. Show that L is an algebraic closure of K.

3.8 Finite Fields

We are already familiar with the finite fields of type $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ for prime numbers p; these are precisely the prime fields of characteristic > 0; cf. 3.1/2. In the following we want to construct for every nontrivial prime power q of p, say $q = p^n$ for some $n > 0$, a field $\mathbb{F}_q$ consisting of q elements. However, note that such a field will be totally different from the residue class ring $\mathbb{Z}/p^n\mathbb{Z}$ for $n > 1$, since the latter admits nontrivial zero divisors and thus cannot be a field.

Lemma 1. *Let $\mathbb{F}$ be a finite field. Then $p = \operatorname{char} \mathbb{F} > 0$, and $\mathbb{F}$ contains $\mathbb{F}_p$ as its prime subfield. Moreover, $\mathbb{F}$ consists of precisely $q = p^n$ elements, where $n = [\mathbb{F} : \mathbb{F}_p]$. In addition, $\mathbb{F}$ is a splitting field of the polynomial $X^q - X$ over $\mathbb{F}_p$, and it follows that the extension $\mathbb{F}/\mathbb{F}_p$ is normal.*

Proof. Since $\mathbb{F}$ is finite, the same is true for its prime subfield. Hence, the latter is of type $\mathbb{F}_p$, where $p = \operatorname{char} \mathbb{F} > 0$. Furthermore, the finiteness of $\mathbb{F}$ shows that the degree $n = [\mathbb{F} : \mathbb{F}_p]$ is finite, and we see, for example using an $\mathbb{F}_p$-vector space

isomorphism $\mathbb{F} \xrightarrow{\sim} (\mathbb{F}_p)^n$, that $\mathbb{F}$ consists of $q = p^n$ elements. In particular, the multiplicative group $\mathbb{F}^*$ is of order $q-1$. Therefore, every element in $\mathbb{F}^*$ is a zero of the polynomial $X^{q-1}-1$, and it follows that every element of $\mathbb{F}$ is a zero of the polynomial $X^q - X$. Thus, all in all, $\mathbb{F}$ consists of $q = p^n$ zeros of $X^q - X$ and thereby of all zeros of this polynomial. It follows that $X^q - X$ factorizes over $\mathbb{F}$ into a product of linear factors, and we can conclude that $\mathbb{F}$ is a splitting field of the polynomial $X^q - X \in \mathbb{F}_p[X]$. □

Theorem 2. *Let p be a prime number. For every integer $n \in \mathbb{N} - \{0\}$ there exists an extension field $\mathbb{F}_q/\mathbb{F}_p$ consisting of $q = p^n$ elements. Furthermore, up to isomorphism, $\mathbb{F}_q$ is uniquely characterized as a splitting field of the polynomial $X^q - X$ over $\mathbb{F}_p$. In fact, the elements of $\mathbb{F}_q$ are recognized as the q different zeros of $X^q - X$.*

Every finite field of characteristic p is isomorphic to precisely one of the finite fields of type $\mathbb{F}_q$.

Proof. Write $f = X^q - X$. Since $f' = -1$, the polynomial f does not admit multiple zeros and therefore has q simple zeros in an algebraic closure $\overline{\mathbb{F}}_p$ of $\mathbb{F}_p$. If $a, b \in \overline{\mathbb{F}}_p$ are two zeros of f, the binomial formula 3.1/3,

$$(a \pm b)^q = a^q \pm b^q = a \pm b,$$

implies that $a \pm b$ is a zero of f. Moreover, for $b \neq 0$ the equation

$$(ab^{-1})^q = a^q (b^q)^{-1} = ab^{-1}$$

shows that ab^{-1} is a zero of f as well. Thereby we see that the q zeros of f in $\overline{\mathbb{F}}_p$ form a field consisting of q elements. In fact, it is a splitting field of f over $\mathbb{F}_p$ (constructed as a subfield of $\overline{\mathbb{F}}_p$). This verifies the existence of a field of characteristic p consisting of $q = p^n$ elements. Finally, the stated uniqueness assertions can be derived from Lemma 1. □

In the following we fix a prime number p. In dealing with finite fields of characteristic $p > 0$, it is common practice to choose an algebraic closure $\overline{\mathbb{F}}_p$ of $\mathbb{F}_p$ and to view the fields $\mathbb{F}_{p^n}$ for $n \in \mathbb{N} - \{0\}$ as subfields of $\overline{\mathbb{F}}_p$, using 3.4/9. Since $\mathbb{F}_{p^n}$ is normal over $\mathbb{F}_p$, we conclude from 3.5/4 (i) that $\mathbb{F}_{p^n}$, as a subfield of $\overline{\mathbb{F}}_p$, is unique.

Corollary 3. *Embed the fields $\mathbb{F}_q$ for $q = p^n$, $n \in \mathbb{N} - \{0\}$, into an algebraic closure $\overline{\mathbb{F}}_p$ of $\mathbb{F}_p$. Then an inclusion $\mathbb{F}_q \subset \mathbb{F}_{q'}$ holds for $q = p^n$ and $q' = p^{n'}$ if and only if $n \,|\, n'$. Furthermore, the extensions of type $\mathbb{F}_q \subset \mathbb{F}_{q'}$ are the only extensions between finite fields of characteristic p, up to isomorphism.*

Proof. Assume $\mathbb{F}_q \subset \mathbb{F}_{q'}$ and let $m = [\mathbb{F}_{q'} : \mathbb{F}_q]$. Then

$$p^{n'} = \#\mathbb{F}_{q'} = (\#\mathbb{F}_q)^m = p^{mn},$$

and we see that $n \mid n'$. Conversely, if $n' = mn$, consider an element $a \in \mathbb{F}_q$, i.e., an element $a \in \overline{\mathbb{F}}_p$ satisfying $a^q = a$. Using a recursive argument, we obtain $a^{q'} = a^{(q^m)} = (a^q)^{(q^{m-1})} = a^{(q^{m-1})} = a$ and hence $\mathbb{F}_q \subset \mathbb{F}_{q'}$. That there cannot exist any further extensions between finite fields of characteristic p, up to isomorphism, follows from the extension result 3.4/9. Indeed, if $\mathbb{F} \subset \mathbb{F}'$ is an extension of finite fields in characteristic p, we can extend the inclusion $\mathbb{F}_p \subset \overline{\mathbb{F}}_p$ to a homomorphism $\mathbb{F} \longrightarrow \overline{\mathbb{F}}_p$ and that to a homomorphism $\mathbb{F}' \longrightarrow \overline{\mathbb{F}}_p$, so that up to isomorphism, we are reduced to the case $\mathbb{F} \subset \mathbb{F}' \subset \overline{\mathbb{F}}_p$. □

Corollary 4. *Every algebraic extension of a finite field is normal and separable. In particular, finite fields are perfect.*

Proof. Let $\mathbb{F} \subset K$ be an algebraic field extension, where $\mathbb{F}$ is finite. If K is finite as well, say $K = \mathbb{F}_q$ for $q = p^n$, then K is a splitting field of the separable polynomial $X^q - X$ and therefore is normal and separable over $\mathbb{F}_p$, resp. $\mathbb{F}$. If K is infinite, we can view it as a union of finite extensions of $\mathbb{F}$. □

We have already seen in 3.6/14 that the multiplicative group of a finite field is cyclic. Therefore we can state the following:

Proposition 5. *Let q be a power of a prime number. Then the multiplicative group of $\mathbb{F}_q$ is cyclic of order $q - 1$.*

Finally, let us look at a finite extension $\mathbb{F}_{q'}/\mathbb{F}_q$ of degree n and determine the automorphism group $\mathrm{Aut}_{\mathbb{F}_q}(\mathbb{F}_{q'})$, or in other words, the corresponding Galois group, as we will say in the next chapter. To do this, assume $q = p^r$ and $q' = q^n = p^{rn}$, and fix an algebraic closure $\overline{\mathbb{F}}_p$ of $\mathbb{F}_q$. Then

$$\mathrm{Aut}_{\mathbb{F}_q}(\mathbb{F}_{q'}) = \mathrm{Hom}_{\mathbb{F}_q}(\mathbb{F}_{q'}, \overline{\mathbb{F}}_p),$$

since $\mathbb{F}_{q'}/\mathbb{F}_q$ is normal, and furthermore,

$$\#\,\mathrm{Aut}_{\mathbb{F}_q}(\mathbb{F}_{q'}) = \big[\mathbb{F}_{q'} : \mathbb{F}_q\big]_s = \big[\mathbb{F}_{q'} : \mathbb{F}_q\big] = n,$$

since $\mathbb{F}_{q'}/\mathbb{F}_q$ is separable. Now consider the *Frobenius homomorphism*

$$\sigma \colon \mathbb{F}_{q'} \longrightarrow \mathbb{F}_{q'}, \qquad a \longmapsto a^p,$$

of $\mathbb{F}_{q'}$ that was introduced in 3.1; concerning the additivity of σ see 3.1/3. The rth power σ^r leaves $\mathbb{F}_q$ invariant and is referred to as the *relative Frobenius homomorphism* on $\mathbb{F}_{q'}$ over $\mathbb{F}_q$. We claim that it is of order n. Indeed, $\sigma^r \in \mathrm{Aut}_{\mathbb{F}_q}(\mathbb{F}_{q'})$ is of order $\leq n$, since $a^{(p^{rn})} = a$ for all $a \in \mathbb{F}_{q'}$. On the other hand, if we had $\operatorname{ord}\sigma^r < n$ and hence $e := \operatorname{ord}\sigma < rn$, then all elements $a \in \mathbb{F}_{q'}$ would be zeros of the polynomial $X^{(p^e)} - X$, contradicting the fact that $\#\mathbb{F}_{q'} = p^{rn} > p^e$. Therefore, it follows that $\mathrm{Aut}_{\mathbb{F}_q}(\mathbb{F}_{q'})$ is cyclic of order n, generated by the relative Frobenius homomorphism σ^r. In particular, taking into account Corollary 3, we obtain the following result:

Proposition 6. *Let $\mathbb{F}_q$ be a finite field, $q = p^r$, and $\mathbb{F}/\mathbb{F}_q$ a finite field extension of degree n. Then $\mathrm{Aut}_{\mathbb{F}_q}(\mathbb{F})$ is cyclic of order n, generated by the relative Frobenius homomorphism $\mathbb{F} \longrightarrow \mathbb{F}$, $a \longmapsto a^q$.*

Exercises

1. *Explain why the extensions of type $\mathbb{F}_p(t)/\mathbb{F}_p(t^p)$, for p prime and t a variable, are, so to speak, the "simplest" examples of field extensions that are not separable.*
2. *Let $\mathbb{F}$ and $\mathbb{F}'$ be subfields of a field L. Explain why we will have $\mathbb{F} = \mathbb{F}'$ if $\mathbb{F}$ and $\mathbb{F}'$ are finite and consist of the same number of elements.*
3. Show for a prime number p and $n \in \mathbb{N} - \{0\}$:
   (i) An irreducible polynomial $f \in \mathbb{F}_p[X]$ is a divisor of $X^{p^n} - X$ if and only if $\deg f$ is a divisor of n.
   (ii) $X^{p^n} - X \in \mathbb{F}_p[X]$ equals the product over all irreducible monic polynomials $f \in \mathbb{F}_p[X]$ such that $\deg f$ divides n.
4. Show that $\mathbb{F}_{p^\infty} = \bigcup_{n=0}^{\infty} \mathbb{F}_{p^{n!}}$ is an algebraic closure of $\mathbb{F}_p$.
5. Let $\overline{\mathbb{F}}_p$ be an algebraic closure of $\mathbb{F}_p$. Show that besides the powers of the Frobenius homomorphism, there exist further automorphisms of $\overline{\mathbb{F}}_p$. *Hint*: For prime numbers ℓ study the automorphisms of $\bigcup_{\nu=0}^{\infty} \mathbb{F}_{q_\nu}$, where $q_\nu = p^{\ell^\nu}$.

3.9 Beginnings of Algebraic Geometry*

So far we have looked at zeros of polynomials in one variable. Now we want to study zeros of polynomials in several variables with coefficients in a field K, and thereby have a first look at the interesting domain of algebraic geometry; for more details, consult [3] or any other book on the subject. As the name suggests, algebraic geometry applies abstract algebraic methods in a geometric setting. This is related to the fact that zero sets of polynomials in several variables are usually not finite and that it is a truly demanding venture to discover their structures.

In the following, let $X = (X_1, \ldots, X_n)$ be a system of variables and let $\overline{K}$ be an algebraic closure of the field K under consideration. For an arbitrary subset E of the polynomial ring $K[X] = K[X_1, \ldots, X_n]$, we can look at the set

$$V(E) = \big\{x \in \overline{K}^n \,;\, f(x) = 0 \text{ for all } f \in E\big\}$$

of common zeros in $\overline{K}^n$ of the polynomials belonging to E; note that $V(E)$ is referred to as an *algebraic subset of $\overline{K}^n$ that is defined over K*. Conversely, given a subset $U \subset \overline{K}^n$, we can consider its corresponding ideal

$$I(U) = \big\{f \in K[X] \,;\, f(U) = 0\big\}$$

of all polynomials f vanishing on U. That $I(U)$ is indeed an ideal in $K[X]$ is easily verified. Furthermore, we have $V(E) = V(\mathfrak{a})$ for $\mathfrak{a}$ the ideal generated by

E in $K[X]$; use the fact that $\mathfrak{a}$ consists of all finite sums of type $\sum f_i e_i$, where $f_i \in K[X]$ and $e_i \in E$. Let us list some elementary properties of the mappings $V(\cdot)$ and $I(\cdot)$:

Lemma 1. *For ideals $\mathfrak{a}_1, \mathfrak{a}_2$, resp. a family $(\mathfrak{a}_i)_{i\in I}$ of ideals in $K[X]$, as well as for subsets $U_1, U_2 \subset \overline{K}^n$, we have*:

(i) $\mathfrak{a}_1 \subset \mathfrak{a}_2 \Longrightarrow V(\mathfrak{a}_1) \supset V(\mathfrak{a}_2)$.
(ii) $U_1 \subset U_2 \Longrightarrow I(U_1) \supset I(U_2)$.
(iii) $V(\sum_i \mathfrak{a}_i) = \bigcap_i V(\mathfrak{a}_i)$.
(iv) $V(\mathfrak{a}_1 \cdot \mathfrak{a}_2) = V(\mathfrak{a}_1 \cap \mathfrak{a}_2) = V(\mathfrak{a}_1) \cup V(\mathfrak{a}_2)$.

Proof. The assertions (i), (ii), and (iii) are easy to verify; we show only how to obtain (iv). Since

$$\mathfrak{a}_1 \cdot \mathfrak{a}_2 \subset \mathfrak{a}_1 \cap \mathfrak{a}_2 \subset \mathfrak{a}_i, \qquad i = 1, 2,$$

we conclude from (i) that

$$V(\mathfrak{a}_1 \cdot \mathfrak{a}_2) \supset V(\mathfrak{a}_1 \cap \mathfrak{a}_2) \supset V(\mathfrak{a}_1) \cup V(\mathfrak{a}_2).$$

On the other hand, consider a point $x \in \overline{K}^n - (V(\mathfrak{a}_1) \cup V(\mathfrak{a}_2))$. Since $x \notin V(\mathfrak{a}_i)$ for $i = 1, 2$, there exists in each case an element $f_i \in \mathfrak{a}_i$ such that $f_i(x) \neq 0$. Using the fact that $f_1 f_2$ belongs to $\mathfrak{a}_1 \cdot \mathfrak{a}_2$ and $(f_1 f_2)(x) = f_1(x) \cdot f_2(x)$ does not vanish, we get $x \notin V(\mathfrak{a}_1 \cdot \mathfrak{a}_2)$. This shows that

$$V(\mathfrak{a}_1 \cdot \mathfrak{a}_2) \subset V(\mathfrak{a}_1) \cup V(\mathfrak{a}_2),$$

thereby justifying assertion (iv). □

The main objective of the present section is to discuss some of the deeper results on the mappings $V(\cdot)$ and $I(\cdot)$. To begin with, we want to show for every subset $E \subset K[X]$ that there exist finitely many elements $f_1, \ldots, f_r \in E$ satisfying $V(E) = V(f_1, \ldots, f_r)$. This means that every algebraic subset of $\overline{K}^n$ that is defined over K can be viewed as the zero set of *finitely* many polynomials in $K[X]$. To justify this it is enough to prove that the ideal $\mathfrak{a}$ that is generated by E in $K[X]$ is finitely generated. A ring with the property that all its ideals are finitely generated is called a *Noetherian* ring.

Theorem 2 (Hilbert's basis theorem). *Let R be a Noetherian ring. Then the polynomial ring $R[Y]$ in a variable Y is Noetherian as well. In particular, the polynomial ring $K[X] = K[X_1, \ldots, X_n]$ in finitely many variables over a field K is Noetherian.*

In 2.4/8, a ring R was called Noetherian if every ascending chain of ideals $\mathfrak{a}_1 \subset \mathfrak{a}_2 \subset \ldots \subset R$ becomes stationary after finitely many steps. Let us first show that this condition is equivalent to the fact that every ideal of R is finitely generated. Indeed, given a chain as specified before, we can consider the union $\mathfrak{a} = \bigcup_i \mathfrak{a}_i$ as an ideal in R. If $\mathfrak{a}$ admits a finite system of generators $f_1, \ldots, f_r$,

then all f_ρ and hence also $\mathfrak{a}$ are contained in one of the ideals $\mathfrak{a}_i$. Consequently, the chain of ideals becomes stationary at this position. Conversely, consider an ideal $\mathfrak{a} \subset R$ that is not finitely generated. Then, for finitely many elements $f_1, \ldots, f_r \in \mathfrak{a}$ we always have $(f_1, \ldots, f_r) \neq \mathfrak{a}$. Thus, using an inductive construction, there exists in $\mathfrak{a}$ an infinite strictly ascending chain of ideals.

Proof of Theorem 2. Let R be a Noetherian ring and $\mathfrak{a} \subset R[Y]$ an ideal. For $i \in \mathbb{N}$ define $\mathfrak{a}_i \subset R$ as the set of all elements $a \in R$ such that there exists a polynomial of type

$$aY^i + \text{ terms of lower degree in } Y$$

in $\mathfrak{a}$. It is verified without difficulty that every $\mathfrak{a}_i$ is an ideal in R and that we obtain an ascending chain

$$\mathfrak{a}_0 \subset \mathfrak{a}_1 \subset \ldots \subset R,$$

since $f \in \mathfrak{a}$ implies $Yf \in \mathfrak{a}$. Now, using the fact that R is Noetherian, the chain becomes stationary, say at the position of the ideal $\mathfrak{a}_{i_0}$. Then, for $i = 0, \ldots, i_0$, choose polynomials $f_{ij} \in \mathfrak{a}$ of degree $\deg f_{ij} = i$ such that for fixed i the leading coefficients a_{ij} of the f_{ij} generate the ideal $\mathfrak{a}_i$. We claim that the polynomials f_{ij} will generate the ideal $\mathfrak{a}$. To justify this, consider a polynomial $g \in \mathfrak{a}$, where we may assume $g \neq 0$. Furthermore, let $a \in R$ be the leading coefficient of g and write $d = \deg g$ as well as $i = \min\{d, i_0\}$. Then we get $a \in \mathfrak{a}_i$, and hence there is an equation

$$a = \sum_j c_j a_{ij}, \qquad c_j \in R.$$

Next, observe that the polynomial

$$g_1 = g - Y^{d-i} \cdot \sum_j c_j f_{ij}$$

belongs to $\mathfrak{a}$ again and that its degree is strictly less than the degree d of g, since the coefficient of Y^d has become trivial now. If $g_1 \neq 0$, we can apply the same process to g_1 in place of g again, and so forth. After finitely many steps we arrive at a polynomial g_s that is zero, and the process stops. It follows that g is a linear combination of the f_{ij} with coefficients in $R[Y]$, and we see that the f_{ij} generate $\mathfrak{a}$. □

Given an ideal $\mathfrak{a}$ of a ring R, we can always consider its *radical*

$$\operatorname{rad} \mathfrak{a} = \big\{a \in R;\ \text{there exists some } n \in \mathbb{N} \text{ such that } a^n \in \mathfrak{a}\big\}.$$

Using the binomial formula, it is easily seen that the radical of $\mathfrak{a}$ is itself an ideal in R. Ideals with the property that $\mathfrak{a} = \operatorname{rad} \mathfrak{a}$ are called *reduced*. For every subset $U \subset \overline{K}^n$, its corresponding ideal $I(U) \subset K[X]$ is reduced. Indeed, a polynomial $f \in K[X]$ vanishes at a point $x \in \overline{K}^n$ if and only if some power f^r,

where $r > 0$, vanishes at x. We want to have a closer look at the relationship between ideals in $K[X]$ and algebraic sets in $\overline{K}^n$.

Proposition 3. *The mappings $I(\cdot)$ and $V(\cdot)$ define mutually inverse and inclusion-inverting bijections*

$$\big\{\text{algebraic subsets} \subset \overline{K}^n\big\} \underset{V}{\overset{I}{\rightleftarrows}} \big\{\text{reduced ideals} \subset K[X]\big\},$$

where, in more precise terms, on the left-hand side we mean algebraic subsets of $\overline{K}^n$ that are defined over K.

To carry out the *proof* we have to establish the equations

$$V\big(I(U)\big) = U, \qquad I\big(V(\mathfrak{a})\big) = \mathfrak{a},$$

for algebraic subsets $U \subset \overline{K}^n$ and reduced ideals $\mathfrak{a} \subset K[X]$. The first of these is of elementary nature. To justify it, assume $U = V(\mathfrak{a})$ for some ideal $\mathfrak{a} \subset K[X]$. We then have to show that $V(I(V(\mathfrak{a}))) = V(\mathfrak{a})$. Since all polynomials in $\mathfrak{a}$ vanish on $V(\mathfrak{a})$, we see that $\mathfrak{a} \subset I(V(\mathfrak{a}))$ and therefore $V(\mathfrak{a}) \supset V(I(V(\mathfrak{a})))$. On the other hand, all polynomials in $I(V(\mathfrak{a}))$ vanish on $V(\mathfrak{a})$, so that $V(\mathfrak{a}) \subset V(I(V(\mathfrak{a})))$ and hence $V(I(V(\mathfrak{a}))) = V(\mathfrak{a})$. The second equation $I(V(\mathfrak{a})) = \mathfrak{a}$ is more involved; it is a special case of a result referred to as *Hilbert's Nullstellensatz* (German for Hilbert's theorem on the zero locus of polynomials):

Theorem 4 (Hilbert's Nullstellensatz). *Let $\mathfrak{a}$ be an ideal of the polynomial ring $K[X] = K[X_1, \dots, X_n]$ and let $V(\mathfrak{a})$ be the corresponding zero set in $\overline{K}^n$. Then $I(V(\mathfrak{a})) = \operatorname{rad}\mathfrak{a}$. In other words, a polynomial $f \in K[X]$ vanishes on $V(\mathfrak{a})$ precisely when a power f^r belongs to $\mathfrak{a}$.*

We start by proving a lemma known as a weak version of Hilbert's Nullstellensatz.

Lemma 5. *Let $A = K[x_1, \dots, x_n] \neq 0$ be a ring of finite type over a field K. Then the inclusion $K \hookrightarrow \overline{K}$ can be extended to a K-homomorphism $A \longrightarrow \overline{K}$.*

Proof. For a maximal ideal $\mathfrak{m} \subset A$, consider the canonical map $K \longrightarrow A/\mathfrak{m}$. Since $A/\mathfrak{m}$ is a field that is of finite type over K in the ring-theoretic sense, we conclude from 3.3/8 that $A/\mathfrak{m}$ is finite over K. Then 3.4/9 shows that there exists a K-homomorphism $A/\mathfrak{m} \longrightarrow \overline{K}$, and the composition of the latter with the projection $A \longrightarrow A/\mathfrak{m}$ yields the desired K-homomorphism from A to $\overline{K}$. □

Now we come to the *proof* of Theorem 4. Since all polynomials of $\mathfrak{a}$ vanish on $V(\mathfrak{a})$, we obtain $\mathfrak{a} \subset I(V(\mathfrak{a}))$ and even $\operatorname{rad}\mathfrak{a} \subset I(V(\mathfrak{a}))$, since all ideals of type $I(U)$ are reduced. To derive the opposite inclusion we proceed indirectly.

Assume that there is a polynomial $f \in I(V(\mathfrak{a}))$ such that $f^r \notin \mathfrak{a}$ for all $r \in \mathbb{N}$. Then the multiplicative system $S = \{1, f, f^2, \ldots\}$ is disjoint from $\mathfrak{a}$. Using Zorn's lemma 3.4/5 (or alternatively, the fact that $K[X]$ is Noetherian), there exists an ideal $\mathfrak{p} \subset K[X]$ that is maximal among all ideals $\mathfrak{q} \subset K[X]$ such that $\mathfrak{a} \subset \mathfrak{q}$ and $\mathfrak{q} \cap S = \emptyset$. We claim that $\mathfrak{p}$ is a prime ideal. Indeed, choose $a, b \in K[X] - \mathfrak{p}$. By the choice of $\mathfrak{p}$, the ideals $(a, \mathfrak{p})$ and $(b, \mathfrak{p})$ that are generated by a and $\mathfrak{p}$, resp. b and $\mathfrak{p}$, in $K[X]$ must have a nonempty intersection with S, so that

$$S \cap (ab, \mathfrak{p}) \supset S \cap \big((a, \mathfrak{p}) \cdot (b, \mathfrak{p})\big) \neq \emptyset.$$

In particular, $ab \notin \mathfrak{p}$ and $\mathfrak{p}$ is a prime ideal.

Next, consider the residue class ring $A = K[X]/\mathfrak{p}$ as a ring extension of finite type of K. Let $\tilde{f} \in A$ be the residue class of f. Since $f \notin \mathfrak{p}$ by the choice of $\mathfrak{p}$ and since A is an integral domain, we can consider the subring $A[\tilde{f}^{-1}]$ of the field of fractions $Q(A)$. Using Lemma 5, there is a K-homomorphism $A[\tilde{f}^{-1}] \longrightarrow \overline{K}$, and composing it with canonical maps, we arrive at a K-homomorphism

$$\varphi\colon K[X] \longrightarrow A \hookrightarrow A[\tilde{f}^{-1}] \longrightarrow \overline{K}.$$

Now look at the point $x = (\varphi(X_1), \ldots, \varphi(X_n)) \in \overline{K}^n$. We may view the map φ as the substitution homomorphism evaluating polynomials of $K[X]$ at x. Since $\mathfrak{a} \subset \mathfrak{p} \subset \ker \varphi$, it follows that $x \in V(\mathfrak{a})$. On the other hand, $f(x) = \varphi(f)$ is nonzero, since it is the image of the unit $\tilde{f} \in A[\tilde{f}^{-1}]$. However, $f(x) \neq 0$ for a point $x \in V(\mathfrak{a})$ contradicts the fact that $f \in I(V(\mathfrak{a}))$ by our choice of f. In particular, the assumption that there is no power of f belonging to $\mathfrak{a}$ cannot be maintained, and we see that $I(V(\mathfrak{a})) \subset \operatorname{rad} \mathfrak{a}$. □

For an algebraically closed field K, the algebraic subsets of K^n corresponding to maximal ideals $\mathfrak{m} \subset K[X]$ are particularly simple to describe, since they are reduced to the one-point sets in K^n:

Corollary 6. *Let K be an algebraically closed field. An ideal $\mathfrak{m}$ of the polynomial ring $K[X] = K[X_1, \ldots, X_n]$ is maximal if and only if there exists a point $x = (x_1, \ldots, x_n) \in K^n$ such that $\mathfrak{m} = (X_1 - x_1, \ldots, X_n - x_n)$. In particular, $V(\mathfrak{m}) = \{x\}$ and $I(x) = \mathfrak{m}$ in this case.*

Therefore, if K is algebraically closed, the maximal ideals in $K[X]$ correspond bijectively to the points of K^n under the bijection mentioned in Proposition 3.

Proof. First note that $(X_1, \ldots, X_n) \subset K[X]$ is a maximal ideal, since the residue class ring $K[X]/(X_1, \ldots, X_n)$ is isomorphic to K. In the same way, we see that ideals of type $(X_1 - x_1, \ldots, X_n - x_n) \subset K[X]$ for points $x = (x_1, \ldots, x_n)$ in K^n are maximal as well; just use a K-automorphism on $K[X]$ transforming the variables $X_1, \ldots, X_n$ to $X_1 - x_1, \ldots, X_n - x_n$. Now consider an arbitrary maximal ideal $\mathfrak{m} \subset K[X]$. Due to Lemma 5, there is a K-homomorphism

$K[X]/\mathfrak{m} \longrightarrow K$, where the latter is necessarily an isomorphism, since $K[X]/\mathfrak{m}$ is a field extending K. Composing this isomorphism with the canonical projection $K[X] \longrightarrow K[X]/\mathfrak{m}$, we arrive at an epimorphism $K[X] \longrightarrow K$ admitting $\mathfrak{m}$ as its kernel. For $i = 1, \ldots, n$, let $x_i \in K$ be the corresponding image of X_i. Then we get $X_i - x_i \in \mathfrak{m}$ for all i, and we see that $\mathfrak{m}$ must coincide with the ideal $(X_1 - x_1, \ldots, X_n - x_n)$, since that ideal is already maximal in $K[X]$. This establishes the desired characterization of maximal ideals in $K[X]$, while the remaining assertions are easily deduced from this one. □

More generally, one can show for a not necessarily algebraically closed field K that an ideal in $K[X]$ is maximal if and only if it is of type $I(\{x\})$ for a point $x \in \overline{K}^n$; cf. Exercise 2. However, the set $\{x\}$ does not necessarily form an algebraic subset of $\overline{K}^n$ that is defined over K, and furthermore, x is in general not uniquely determined by the corresponding maximal ideal $I(\{x\}) \subset K[X]$. For example, every K-automorphism $\sigma\colon \overline{K} \longrightarrow \overline{K}$ maps the point $x = (x_1, \ldots, x_n)$ to a point $\sigma(x) := (\sigma(x_1), \ldots, \sigma(x_n))$ satisfying $I(\{x\}) = I(\{\sigma(x)\})$. The smallest algebraic subset of $\overline{K}^n$ containing x and defined over K is $V(I\{x\})$, which one can show consists of all points $\sigma(x)$ for σ varying over the K-automorphisms of $\overline{K}$.

Now consider an ideal $\mathfrak{a} \subset K[X]$ and its associated algebraic set $V(\mathfrak{a}) \subset \overline{K}^n$. Viewing polynomials of $K[X]$ as $\overline{K}$-valued functions on $\overline{K}^n$, we can restrict this domain to the algebraic set $V(\mathfrak{a})$. This restriction process gives rise to a ring homomorphism $K[X] \longrightarrow \mathrm{Map}(V(\mathfrak{a}), \overline{K})$ whose kernel contains $\mathfrak{a}$. In particular, the elements of the residue class ring $K[X]/\mathfrak{a}$ may canonically be viewed as "functions" on $V(\mathfrak{a})$; the ring $K[X]/\mathfrak{a}$ is referred to as the ring of *polynomial functions* (modulo $\mathfrak{a}$) on the algebraic set $V(\mathfrak{a})$. Proceeding like this, a little bit of care is necessary, since the map $K[X]/\mathfrak{a} \longrightarrow \mathrm{Map}(V(\mathfrak{a}), \overline{K})$ will not be injective in general. For example, nilpotent elements in $K[X]/\mathfrak{a}$ give rise to the zero function on $V(\mathfrak{a})$, and one can deduce from Hilbert's Nullstellensatz that these are the only elements in $K[X]/\mathfrak{a}$ with this property. Indeed, the kernel of the map $K[X] \longrightarrow \mathrm{Map}(V(\mathfrak{a}), \overline{K})$ is given by the ideal $\operatorname{rad}\mathfrak{a}$, which implies that the kernel of the induced homomorphism $K[X]/\mathfrak{a} \longrightarrow \mathrm{Map}(V(\mathfrak{a}), \overline{K})$ equals the radical of the zero ideal in $K[X]/\mathfrak{a}$, which consists of all nilpotent elements in $K[X]/\mathfrak{a}$.

Exercises

Let K be a field, $\overline{K}$ an algebraic closure of K, and $X = (X_1, \ldots, X_n)$ a system of variables.

1. *For subsets $E \subset K[X]$ and $U \subset K^n$ set*

$$V_K(E) = \{x \in K^n\,;\, f(x) = 0 \text{ for all } f \in E\},$$
$$I(U) = \{f \in K[X]\,;\, f(U) = 0\}.$$

Review the results of the present section and examine which of them remain valid, and which not, if we consider zeros of polynomials $f \in K[X]$ merely in K^n and not in $\overline{K}^n$, in other words, if we use the mapping $V_K(\cdot)$ instead of $V(\cdot)$.

2. Consider the substitution homomorphism $h_x \colon K[X] \longrightarrow \overline{K},\ f \longmapsto f(x)$, for elements $x \in \overline{K}^n$. Show that the ideals of type $\ker h_x$ are precisely the maximal ideals of $K[X]$.

3. Let $\mathfrak{m} \subset K[X]$ be a maximal ideal. Show that $\mathfrak{m} = (f_1, \ldots, f_n)$ for polynomials $f_1, \ldots, f_n$, where f_i for each i is a monic polynomial in X_i with coefficients in $K[X_1, \ldots, X_{i-1}]$.

4. Let $U \subset \overline{K}^n$ be an algebraic subset defined over K. Then U is said to be *irreducible* over K if there does not exist a decomposition $U = U_1 \cup U_2$ for algebraic subsets $U_1, U_2 \subsetneq U$ that are defined over K. Show:
   (i) $U \subset \overline{K}^n$ is irreducible over K if and only if the corresponding ideal $I(U)$ is prime in $K[X]$.
   (ii) There exists a decomposition $U = U_1 \cup \ldots \cup U_r$ of U into algebraic subsets that are defined and irreducible over K. For decompositions that cannot be shortened, the $U_1, \ldots, U_r$ are unique, up to numeration.

5. Let A be a K-algebra of finite type. Show that A is a *Jacobson ring*, i.e., that every reduced ideal $\mathfrak{a} \subsetneq A$ is an intersection of maximal ideals.

4. Galois Theory

Background and Overview

In Chapter 3 we saw that every field K admits an algebraic closure $\overline{K}$ and that this closure is unique up to K-isomorphism. Hence, given an algebraic equation $f(x) = 0$ for a nonconstant polynomial $f \in K[X]$, we know that f factorizes over $\overline{K}$ into linear factors. In particular, $\overline{K}$ contains "all" solutions of the algebraic equation $f(x) = 0$. The subfield $L \subset \overline{K}$ generated over K by these solutions is a splitting field of f, and the extension L/K is finite, as well as normal; see 3.5/5. Alternatively, a splitting field L of f can be obtained in terms of Kronecker's construction, by successively adjoining all solutions of the equation $f(x) = 0$ to K. If we want to clarify the "nature" of the solutions, for example, if we want to see whether we can solve the equation by radicals, the structure of the extension L/K has to be studied.

At this point Galois theory comes into play, with its group-theoretic methods. At the center of interest is the group $\mathrm{Aut}_K(L)$ of all K-automorphisms of L. If L/K is separable and thus a *Galois extension*, the group $\mathrm{Aut}_K(L)$ is referred to as the *Galois group* of L/K and is denoted by $\mathrm{Gal}(L/K)$. Every K-automorphism $L \longrightarrow L$ restricts to a bijective self-map on the zero set of f and, in fact, is uniquely determined by the images of these zeros. Therefore, the elements of $\mathrm{Aut}_K(L)$ can readily be identified with the corresponding permutations on the zero set of f. Interpreting $\overline{K}$ as an algebraic closure of L, we can just as well view $\mathrm{Aut}_K(L)$ as the set of all K-homomorphisms $L \longrightarrow \overline{K}$; see 3.5/4. Also note that the results 3.4/8 and 3.4/9 provide an explicit description of those homomorphisms. For example, let us assume that f does not admit multiple zeros, or more generally that L, as a splitting field of f, is separable over K. Then the extension L/K is simple by the primitive element theorem 3.6/12, say $L = K(\alpha)$, and the minimal polynomial $g \in K[X]$ of α factorizes over L into linear factors, due to 3.5/4. The corresponding zeros $\alpha_1, \ldots, \alpha_n \in L$ satisfy $L = K(\alpha_i)$, and there is a unique automorphism $\sigma_i \in \mathrm{Aut}_K(L)$ such that $\sigma_i(\alpha) = \alpha_i$ for each i; see 3.4/8. This automorphism is characterized by $h(\alpha) \longmapsto h(\alpha_i)$ for polynomials $h \in K[X]$. It follows that the Galois group $\mathrm{Gal}(L/K)$ consists of the elements $\sigma_1, \ldots, \sigma_n$, where their number n equals the degree of g, resp. the degree of the extension L/K. Such an explicit description of automorphism groups $\mathrm{Aut}_K(L)$ was discovered by Galois himself, and it is for this reason that these groups are referred to as Galois groups.

The first basic result we will prove in the present chapter is the so-called *fundamental theorem of Galois theory*; see Section 4.1. It asserts for a finite Galois extension L/K that the subgroups H of the corresponding Galois group $\mathrm{Gal}(L/K)$ correspond bijectively to the intermediate fields E of L/K via the mappings $H \longmapsto L^H$, resp. $E \longmapsto \mathrm{Aut}_E(L)$; here L^H denotes the subfield consisting of all elements in L that are invariant under the automorphisms belonging to H. Furthermore, an intermediate field E of L/K is normal over K if and only if $\mathrm{Aut}_E(L)$, as a subgroup of $\mathrm{Gal}(L/K)$, is normal. This is only one facet of the quite general fact that the Galois group $\mathrm{Gal}(L/K)$ encodes several properties of the extension L/K. In particular, the problem to determine all intermediate fields of L/K simplifies to that of determining all subgroups of $\mathrm{Gal}(L/K)$.

In Section 4.2 we generalize the fundamental theorem of Galois theory to Galois extensions that are not necessarily finite. To do this we interpret Galois groups as topological groups following W. Krull, and look especially at their *closed* subgroups. As an example, we determine the absolute Galois group $\mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F})$ of a finite field $\mathbb{F}$, where $\overline{\mathbb{F}}$ is an algebraic closure of $\mathbb{F}$. Next, in 4.3, some examples are considered showing how the Galois group of algebraic equations can be worked out in special cases. Also we prove at this place that the generic equation of degree n admits the full permutation group $\mathfrak{S}_n$ as its Galois group. Related considerations lead us to the fundamental theorem on symmetric polynomials, whose more advanced version is derived in 4.4. As an application we study the discriminant of a polynomial f, which is nonzero precisely when f does not admit multiple zeros. Also we consider the resultant of two polynomials as a possible means for computing discriminants.

The purpose of Sections 4.5–4.8 is essentially to prepare the characterization of the solvability of algebraic equations by radicals, although a final treatment of this subject has to be postponed until Chapter 6. In 4.5 and 4.8 we study so-called *radical extensions*, i.e., extensions that occur by adjoining solutions of pure equations of type $x^n - c = 0$. For $c = 1$ this concerns nth *roots of unity*, i.e., nth roots of 1. In the remaining cases, we are dealing with *cyclic extensions*, i.e., Galois extensions with cyclic Galois group, provided we assume that the coefficient field K contains all nth roots of unity. Certain modifications are necessary if the characteristic of the field K under consideration divides n. As an auxiliary tool, we prove the theorem on the *linear independence of characters* in 4.6 and study subsequently, in 4.7, the *norm* and *trace* for finite field extensions. E. Artin based his set-up of Galois theory on techniques of this kind, see [1] and [2], while we have preferred to follow a more conventional approach in Section 4.1.

In Sections 4.9 and 4.10 we generalize the characterization of cyclic extensions to certain classes of abelian extensions, called *Kummer extensions* of some given exponent n, named after E. Kummer. First, in 4.9, we assume that the characteristic of the field under consideration does not divide n; this is the easiest case. Then, in 4.10, we explain Kummer theory from a more axiomatic point of view, applying it to study Kummer extensions of exponents p^r over fields of

characteristic $p > 0$. As a necessary technical tool, we explain the formalism of *Witt vectors*, which was introduced by E. Witt.

The chapter ends in 4.11 with an example of descent theory. Considering a finite Galois extension L/K, its aim is to describe K-vector spaces in the style of the fundamental theorem of Galois theory as fixed spaces of L-vector spaces that are equipped with a certain action of the Galois group $\mathrm{Gal}(L/K)$.

4.1 Galois Extensions

An algebraic field extension L/K is called normal, see 3.5, if L is a splitting field of a family of polynomials in $K[X]$, or equivalently, if every irreducible polynomial in $K[X]$ having a zero in L decomposes over L into a product of linear factors; cf. 3.5/4 (ii) and (iii). In the sequel the remaining characterizing property of normal extensions 3.5/4 (i) will play a fundamental role, namely that on choosing an algebraic closure $\overline{L}$ of L, every K-homomorphism $L \longrightarrow \overline{L}$ restricts to an automorphism of L. Then, viewing $\overline{L}$ as an algebraic closure $\overline{K}$ of K, the set $\mathrm{Hom}_K(L, \overline{K})$ of all K-homomorphisms of L to $\overline{K}$ can be identified with the group $\mathrm{Aut}_K(L)$ of all K-automorphisms of L. Let us add along the way that two elements $a, b \in L$ are said to be *conjugate* (*over* K) if there exists an automorphism $\sigma \in \mathrm{Aut}_K(L)$ such that $\sigma(a) = b$; however, we will use such terminology only on rare occasions.

Definition 1. *An algebraic field extension L/K is called* Galois *if it is normal and separable. Then* $\mathrm{Gal}(L/K) := \mathrm{Aut}_K(L)$ *is called the* Galois group *of the Galois extension L/K.*

In the literature, normal field extensions are also referred to as *quasi-Galois* extensions. Looking at a splitting field of a separable polynomial over a field K, we obtain an example of a finite Galois extension. Furthermore, as we saw in 3.8/4, every algebraic extension $\mathbb{F}/\mathbb{F}_q$ of a finite field $\mathbb{F}_q$, where q is a prime power, is Galois. If $\mathbb{F}/\mathbb{F}_q$ is *finite*, the corresponding Galois group $\mathrm{Gal}(\mathbb{F}/\mathbb{F}_q)$ is cyclic of order $n = [\mathbb{F} : \mathbb{F}_q]$ and is generated by the relative Frobenius homomorphism $\mathbb{F} \longrightarrow \mathbb{F}$, $a \longmapsto a^q$; cf. 3.8/6.

Remark 2. *Let L/K be a Galois extension and E an intermediate field. Then*:

(i) *The extension L/E is Galois and the Galois group* $\mathrm{Gal}(L/E)$ *is naturally a subgroup of* $\mathrm{Gal}(L/K)$.

(ii) *If E/K is Galois as well, then every K-automorphism of L restricts to a K-automorphism of E and* $\mathrm{Gal}(L/K) \longrightarrow \mathrm{Gal}(E/K)$, $\sigma \longmapsto \sigma|_E$, *is a surjective group homomorphism.*

Proof. It follows from 3.5/6 and 3.6/11 that the extension L/E is Galois. Since each E-automorphism of L is at the same time a K-automorphism, we recognize $\mathrm{Gal}(L/E)$ as a subgroup of $\mathrm{Gal}(L/K)$. Furthermore, if E/K is Galois, then ev-

ery K-automorphism of L restricts to a K-automorphism of E, due to 3.5/4 (i). Thereby one obtains a group homomorphism $\mathrm{Gal}(L/K) \longrightarrow \mathrm{Gal}(E/K)$, which is surjective by 3.4/9 due to the fact that L/K is normal. □

Combining the defining properties of the separable degree with the results 3.6/8 and 3.6/9, we have the following:

Remark 3. *Let L/K be a finite normal field extension. Then*

$$\operatorname{ord} \mathrm{Aut}_K(L) = [L : K]_s \leq [L : K].$$

In particular, $\operatorname{ord} \mathrm{Aut}_K(L) = [L : K]$ *is equivalent to the fact that L/K is separable.*

As it will turn out, a fundamental property of Galois extensions L/K consists in the fact that K is the invariant or fixed field of the Galois group $\mathrm{Gal}(L/K)$, i.e., K equals the set of all elements of L that are left invariant under all automorphisms in $\mathrm{Gal}(L/K)$. To prove this assertion, which is part of the fundamental theorem of Galois theory, we start by studying fixed fields that are constructed with respect to the action of automorphism groups.

Proposition 4. *Let L be a field and G a subgroup of* $\mathrm{Aut}(L)$*, the group of automorphisms of L. Furthermore, consider*

$$K = L^G = \big\{a \in L\,;\, \sigma(a) = a \text{ for all } \sigma \in G\big\},$$

the fixed field *attached to G.*

(i) *If G is finite, then L/K is a finite Galois extension, in fact, of degree* $[L : K] = \operatorname{ord} G$ *and with Galois group* $\mathrm{Gal}(L/K) = G$.

(ii) *If G is infinite, but L/K is known to be algebraic, then L/K is an infinite Galois extension with Galois group* $\mathrm{Gal}(L/K)$ *containing G as a subgroup.*

Proof. First, it is easily checked that $K = L^G$ is indeed a subfield of L. Now assume that G is finite, or if not, that L/K is algebraic. To see that L/K is separable algebraic, consider an element $a \in L$, as well as a maximal system of elements $\sigma_1, \ldots, \sigma_r \in G$ such that $\sigma_1(a), \ldots, \sigma_r(a)$ are distinct. Such a finite system exists always, also in the case that G is infinite and the extension L/K is known to be algebraic. Indeed, in the latter case we see for each $\sigma \in G$ that $\sigma(a)$ is a zero of the minimal polynomial of a over K. Also observe that the element a itself will occur among the $\sigma_i(a)$. Every $\sigma \in G$ gives rise to a self-map on the set $\{\sigma_1(a), \ldots, \sigma_r(a)\}$ that is necessarily bijective, and it follows that the polynomial

$$f = \prod_{i=1}^{r} \big(X - \sigma_i(a)\big)$$

has coefficients in K, since these are left invariant by G. In particular, a is a zero of a separable polynomial in $K[X]$ and hence is separable algebraic over

K. The same is then true for the extension L/K. Furthermore, L/K is normal, since L is a splitting field over K of all polynomials f of the type just considered. Thereby we see that L/K is a Galois extension.

Now let $n = \operatorname{ord} G$, where $n = \infty$ is not excluded. Then the argumentation given before shows that $[K(a) : K] \leq n$ for every $a \in L$. This implies $[L : K] \leq n$ if we apply the primitive element theorem 3.6/12 to subfields of L that are finite over K. Since G is a subgroup of $\operatorname{Aut}_K(L) = \operatorname{Gal}(L/K)$, we deduce from Remark 3 that

$$n = \operatorname{ord} G \leq \operatorname{ord} \operatorname{Gal}\big(L/K\big) \leq [L : K] \leq n$$

and therefore that $\operatorname{ord} G = [L : K]$. In addition, for $n < \infty$ we can conclude that $G = \operatorname{Gal}(L/K)$. □

Corollary 5. *Let L/K be a normal algebraic field extension with automorphism group $G = \operatorname{Aut}_K(L)$. Then:*

(i) *L/L^G is a Galois extension with Galois group G.*

(ii) *If L/K is separable and therefore Galois, then $L^G = K$.*

(iii) *Assume* $\operatorname{char} K > 0$. *Then L^G is purely inseparable over K, and the chain $K \subset L^G \subset L$ coincides with the chain $K \subset K_i \subset L$ of* 3.7/5.

Proof. We know by Proposition 4 that L/L^G is a Galois extension. The corresponding Galois group coincides with G in this case, since $\operatorname{Aut}_{L^G}(L) = \operatorname{Aut}_K(L)$. Furthermore, the definition of L^G shows that $[L^G : K]_s = 1$. Indeed, if $\overline{K}$ is an algebraic closure of K containing L, then using 3.4/9, every K-homomorphism $L^G \longrightarrow \overline{K}$ extends to a K-homomorphism $L \longrightarrow \overline{K}$ or, since L/K is normal, to a K-automorphism of L. However, all K-automorphisms of L are trivial on L^G. Now if L/K is separable, the same is true for L^G/K, and we get $L^G = K$, since $[L^G : K] = [L^G : K]_s = 1$. On the other hand, if L/K is not separable (for $\operatorname{char} K > 0$), we see from 3.7/2 that L^G/K is purely inseparable. That the chain $K \subset L^G \subset L$ coincides with the one of 3.7/5 follows from the uniqueness assertion in 3.7/5. □

Theorem 6 (Fundamental theorem of Galois theory). *Let L/K be a finite Galois extension with Galois group $G = \operatorname{Gal}(L/K)$. Then the maps*

$$\begin{aligned} \big\{\text{subgroups of } G\big\} &\underset{\Psi}{\overset{\Phi}{\rightleftarrows}} \big\{\text{intermediate fields of } L/K\big\}, \\ H &\longmapsto L^H, \\ \operatorname{Gal}\big(L/E\big) &\longmapsfrom E, \end{aligned}$$

that assign to a subgroup $H \subset G$ the fixed field L^H, resp. to an intermediate field E of L/K the Galois group of the Galois extension L/E, are bijective and mutually inverse.

The fixed field L^H is normal and therefore Galois over K if and only if H is a normal subgroup of G. In the latter case, the surjective group homomorphism

$$G \longrightarrow \mathrm{Gal}\big(L^H/K\big),$$
$$\sigma \longmapsto \sigma|_{L^H},$$

admits H as its kernel and hence induces an isomorphism

$$G/H \xrightarrow{\sim} \mathrm{Gal}\big(L^H/K\big).$$

Remark 7. *If in Theorem* 6 *we do not require that the Galois extension L/K be finite, we still get $\Phi \circ \Psi = \mathrm{id}$; in particular, Φ is surjective and Ψ injective. However, for a subgroup $H \subset G$, its image $(\Psi \circ \Phi)(H)$ will in general be different from H itself*; *cf.* 4.2/3 *or* 4.2/4.

The second part of Theorem 6 *remains valid for arbitrary Galois extensions if we restrict ourselves to subgroups $H \subset \mathrm{Gal}(L/K)$ satisfying the condition that $(\Psi \circ \Phi)(H) = H$, or equivalently, that $H = \mathrm{Gal}(L/L^H)$. These are the* closed *subgroups of* $\mathrm{Gal}(L/K)$; *cf.* 4.2.

Proof of Theorem 6 *and Remark* 7. Let L/K be a Galois extension that is not necessarily finite. If E is an intermediate field of L/K, then L/E is Galois and the Galois group $H = \mathrm{Gal}(L/E)$ is a subgroup of $G = \mathrm{Gal}(L/K)$; cf. Remark 2. Using Corollary 5 (ii), we get $E = L^H$, so that $\Phi \circ \Psi = \mathrm{id}$, as claimed. No finiteness condition on L/K is used for this argument. Now consider a subgroup $H \subset G$ and look at the intermediate field $E = L^H$ of L/K. If G is finite, the same is true for H, and we obtain $H = \mathrm{Gal}(L/E)$ from Proposition 4, or in other words, $\Psi \circ \Phi = \mathrm{id}$. In particular, Φ and Ψ are bijective and mutually inverse to each other.

Next consider a subgroup $H \subset G$ and assume in view of Remark 7 that $H = \mathrm{Gal}(L/L^H)$; this is automatically the case if L/K is a finite Galois extension, as we just have seen. If L^H/K is normal, there is a surjective group homomorphism

$$\varphi\colon G \longrightarrow \mathrm{Gal}\big(L^H/K\big),$$
$$\sigma \longmapsto \sigma|_{L^H},$$

by Remark 2. In particular, $\ker\varphi$ consists of all K-automorphisms of L leaving L^H fixed, i.e., $\ker\varphi = \mathrm{Gal}(L/L^H) = H$. Being the kernel of a group homomorphism, H is a normal subgroup in G and φ induces an isomorphism $G/H \xrightarrow{\sim} \mathrm{Gal}(L^H/K)$ by the fundamental theorem on homomorphisms 1.2/7.

Conversely, assume that H is a normal subgroup in G. Choosing an algebraic closure $\overline{L}$ of L, it is at the same time an algebraic closure of K and of L^H. To see that L^H/K is normal, consider a K-homomorphism $\sigma\colon L^H \longrightarrow \overline{L}$ and let us show that $\sigma(L^H) = L^H$. To do this extend σ to a K-homomorphism $\sigma'\colon L \longrightarrow \overline{L}$, using 3.4/9. Since L/K is normal, σ' restricts to an automorphism of L, and we can view σ as a K-homomorphism $L^H \longrightarrow L$. Now let $b \in \sigma(L^H)$, say $b = \sigma(a)$ for some $a \in L^H$. To check that $b \in L^H$, we have to show that b is fixed by all automorphisms of H. Therefore, let $\tau \in H$. Using $H\sigma = \sigma H$, due to the fact

that H is a normal subgroup of G, there exists an element $\tau' \in H$ such that $\tau \circ \sigma = \sigma \circ \tau'$. Hence, we get

$$\tau(b) = \tau \circ \sigma(a) = \sigma \circ \tau'(a) = \sigma(a) = b,$$

i.e., $b \in L^H$, since $a \in L^H$. It follows that $\sigma(L^H) \subset L^H$. Furthermore, extending $\sigma^{-1}\colon \sigma(L^H) \longrightarrow L^H$ to a K-homomorphism $\rho\colon L^H \longrightarrow \overline{L}$, using 3.4/9, we see in the same way that $\rho(L^H) \subset L^H$. It follows that $\sigma(L^H) = L^H$, as claimed. □

Next, let us discuss some consequences of the fundamental theorem of Galois theory.

Corollary 8. *Every finite separable field extension L/K admits only finitely many intermediate fields.*

Proof. Passing to a normal closure of L/K, see 3.5/7, we may assume that L/K is finite and Galois. Then the intermediate fields of L/K correspond bijectively to the subgroups of the finite group $\mathrm{Gal}(L/K)$. □

To be able to formulate our next result, we define for two subfields E, E' of a field L their *composite field* $E \cdot E'$. It is the smallest subfield of L containing E as well as E'. Of course, we may view $E \cdot E'$ as being obtained by adjoining all elements of E' to E, or likewise, by adjoining all elements of E to E', i.e., $E \cdot E' = E(E') = E'(E)$.

Corollary 9. *Let L/K be a finite Galois extension. For intermediate fields E and E' of L/K, consider $H = \mathrm{Gal}(L/E)$ and $H' = \mathrm{Gal}(L/E')$ as subgroups of $G = \mathrm{Gal}(L/K)$. Then*:

(i) $E \subset E' \iff H \supset H'$.

(ii) $E \cdot E' = L^{H \cap H'}$.

(iii) $E \cap E' = L^{H''}$, *where H'' is the subgroup of G generated by H and H'.*

Proof. (i) If $E \subset E'$, then every E'-automorphism of L is an E-automorphism as well, i.e., we have $H = \mathrm{Gal}(L/E) \supset \mathrm{Gal}(L/E') = H'$. On the other hand, we see that $H \supset H'$ implies $E = L^H \subset L^{H'} = E'$.

(ii) Clearly, we have $E \cdot E' \subset L^{H \cap H'}$, as well as $\mathrm{Gal}(L/E \cdot E') \subset H \cap H'$. From the latter inclusion we conclude that $E \cdot E' \supset L^{H \cap H'}$ with the help of (i).

(iii) We have $L^{H''} = L^H \cap L^{H'} = E \cap E'$. □

Definition 10. *A Galois extension L/K is called* abelian (*resp.* cyclic) *if the Galois group* $\mathrm{Gal}(L/K)$ *is abelian* (*resp. cyclic*).

Examples of cyclic, and hence abelian, Galois extensions are easy to obtain, since we can read from 3.8/4 and 3.8/6 that every extension between finite fields is cyclic.

Corollary 11. *Let L/K be a finite abelian (resp. cyclic) Galois extension. Then, for every intermediate field E of L/K, the extension E/K is a finite abelian (resp. cyclic) Galois extension.*

Proof. In each case, $\mathrm{Gal}(L/E)$ is a normal subgroup in $\mathrm{Gal}(L/K)$, since cyclic groups are abelian. It follows that the extension E/K is Galois. Furthermore, the Galois group $\mathrm{Gal}(E/K) = \mathrm{Gal}(L/K)/\mathrm{Gal}(L/E)$ is abelian, resp. cyclic, if the group $\mathrm{Gal}(L/K)$ is of this type. □

Proposition 12. *Let L/K be a field extension together with intermediate fields E and E' such that E/K and E'/K are finite Galois extensions. Then:*

(i) *$E \cdot E'$ is finite and Galois over K, and the homomorphism*

$$\varphi\colon \mathrm{Gal}\big(E \cdot E'/E\big) \longrightarrow \mathrm{Gal}\big(E'/E \cap E'\big), \qquad \sigma \longmapsto \sigma|_{E'},$$

is an isomorphism.

(ii) *The homomorphism*

$$\psi\colon \mathrm{Gal}\big(E \cdot E'/K\big) \longrightarrow \mathrm{Gal}\big(E/K\big) \times \mathrm{Gal}\big(E'/K\big), \qquad \sigma \longmapsto (\sigma|_E, \sigma|_{E'}),$$

is injective. In addition, if $E \cap E' = K$, then ψ is surjective and hence an isomorphism.

Proof. We start with assertion (i). First, using the fact that $E \cdot E' = K(E, E')$, we see that $E \cdot E'$ is normal, separable, and finite over K, since E/K and E'/K admit these properties. To show that φ is injective, observe that $\sigma|_E = \mathrm{id}$ for every $\sigma \in \mathrm{Gal}(E \cdot E'/E)$. In addition, for $\sigma \in \ker\varphi$ we conclude that $\sigma|_{E'} = \mathrm{id}$ and hence that σ is trivial on $E \cdot E'$. To derive the surjectivity of φ we apply the fundamental theorem of Galois theory and consider the equation

$$(E')^{\mathrm{im}\,\varphi} = (E \cdot E')^{\mathrm{Gal}(E \cdot E'/E)} \cap E' = E \cap E',$$

which implies $\mathrm{im}\,\varphi = \mathrm{Gal}(E'/E \cap E')$, as desired.

The injectivity of ψ in assertion (ii) is easy to obtain. Just observe that every K-automorphism $\sigma \in \ker\psi$ is trivial on E and on E', therefore also on $E \cdot E'$. Concerning the surjectivity of ψ, assume $E \cap E' = K$ and consider an element $(\sigma, \sigma') \in \mathrm{Gal}(E/K) \times \mathrm{Gal}(E'/K)$. By (i) we can extend $\sigma' \in \mathrm{Gal}(E'/K)$ to an automorphism $\tilde{\sigma}' \in \mathrm{Gal}(E \cdot E'/K)$ such that $\tilde{\sigma}'|_E = \mathrm{id}$. Likewise, we can extend σ to an automorphism $\tilde{\sigma} \in \mathrm{Gal}(E \cdot E'/K)$ such that $\tilde{\sigma}|_{E'} = \mathrm{id}$. Then $\tilde{\sigma} \circ \tilde{\sigma}'$ is a preimage of (σ, σ') with respect to ψ, since

$$(\tilde{\sigma} \circ \tilde{\sigma}')|_E = \tilde{\sigma}|_E \circ \tilde{\sigma}'|_E = \sigma$$

and

$$(\tilde{\sigma} \circ \tilde{\sigma}')|_{E'} = \tilde{\sigma}|_{E'} \circ \tilde{\sigma}'|_{E'} = \sigma'.$$

□

Exercises

1. *What sort of information can be deduced from the fundamental theorem of Galois theory for finite algebraic field extensions*?
2. *Indicate how the fundamental theorem of Galois theory could be extended to the context of finite quasi-Galois extensions.*
3. *Show that an algebraic field extension L/K is Galois if and only if K equals the fixed field of the automorphism group* $\mathrm{Aut}_K(L)$.
4. Construct a field L together with a subgroup $G \subset \mathrm{Aut}(L)$ such that L/L^G is not a Galois extension.
5. Let L/K be a finite Galois extension and $H \subset \mathrm{Gal}(L/K)$ a subgroup.
   (i) Consider an element $\alpha \in L$ such that the equation $\sigma(\alpha) = \alpha$ for an automorphism $\sigma \in \mathrm{Gal}(L/K)$ is equivalent to $\sigma \in H$. Show that $L^H = K(\alpha)$.
   (ii) Justify that associated to H, there is always an element $\alpha \in L$ as in (i).
6. Let K be a field, $f \in K[X]$ an irreducible separable polynomial, and L a splitting field of f over K, so that L/K is a finite Galois extension. If L/K is abelian, show that $L = K(\alpha)$ for every zero $\alpha \in L$ of f.
7. Consider an algebraically closed field L and an automorphism $\sigma \in \mathrm{Aut}(L)$. Let $K = L^\sigma$ be the fixed field under σ. Show that every finite field extension of K is a cyclic Galois extension.
8. For a Galois extension L/K, consider an element $\alpha \in L - K$ as well as an intermediate field K' that is maximal with respect to the condition that $\alpha \notin K'$. Show for every intermediate field E of L/K' satisfying $[E : K'] < \infty$ that E/K' is a cyclic Galois extension.
9. Let K be a field and $\overline{K}$ an algebraic closure. Show:
   (i) If E_i, $i \in I$, is a family of intermediate fields of $\overline{K}/K$ such that E_i/K is an abelian Galois extension for every i, then $K(\bigcup_{i\in I} E_i)$ is an abelian Galois extension of K as well.
   (ii) There exists a maximal abelian Galois extension K_{ab}/K. It is characterized by the following properties: (a) K_{ab}/K is an abelian Galois extension. (b) For any further abelian Galois extension L/K, the field L is isomorphic over K to an intermediate field of K_{ab}/K.
   (iii) Any two maximal abelian Galois extensions of the type discussed before are isomorphic over K.
10. For a finite Galois extension L/K, consider intermediate fields L_1, L_2 corresponding to subgroups $H_1, H_2 \subset \mathrm{Gal}(L/K)$. Show for an automorphism $\sigma \in \mathrm{Gal}(L/K)$ that $\sigma(L_1) = L_2$ is equivalent to the equation $\sigma H_1 \sigma^{-1} = H_2$.
11. Show that $L = \mathbb{Q}(\sqrt{p_1}, \ldots, \sqrt{p_n})$ for distinct prime numbers $p_1, \ldots, p_n$ is an abelian Galois extension of $\mathbb{Q}$ with Galois group $(\mathbb{Z}/2\mathbb{Z})^n$. *Hint*: Observe for $a \in \mathbb{Q}$ with $\sqrt{a} \in L$, and for $\sigma \in \mathrm{Gal}(L/\mathbb{Q})$, that $\sigma(\sqrt{a}) = \pm\sqrt{a}$. From a more general point of view, this is an example of *Kummer theory*, to be dealt with in 4.9. Considering the multiplicative subgroup $M \subset \mathbb{Q}^*$ that is generated by $p_1, \ldots, p_n$, the quotient M/M^2 can be viewed as a subgroup of the group $\mathrm{Hom}(\mathrm{Gal}(L/\mathbb{Q}), \mathbb{Z}/2\mathbb{Z})$ of all group homomorphisms from $\mathrm{Gal}(L/\mathbb{Q})$ to $\mathbb{Z}/2\mathbb{Z}$.

4.2 Profinite Galois Groups*

In the preceding section we considered Galois theory mainly for finite field extensions. We want to remove this restriction now and study some additional phenomena that occur when one is working with nonfinite Galois extensions. Given an arbitrary Galois extension L/K, we can look at the system $\mathfrak{L} = (L_i)_{i\in I}$ of all intermediate fields of L/K that are *finite* and Galois over K. Let $f_i\colon \mathrm{Gal}(L/K) \longrightarrow \mathrm{Gal}(L_i/K)$ be the restriction homomorphism according to 4.1/2. Then every $\sigma \in \mathrm{Gal}(L/K)$ determines a family of Galois automorphisms $(\sigma_i)_{i\in I}$, where $\sigma_i = \sigma|_{L_i} = f_i(\sigma)$ and $\sigma_j|_{L_i} = \sigma_i$ if $L_i \subset L_j$. Conversely, every family $(\sigma_i)_{i\in I} \in \prod_{i\in I} \mathrm{Gal}(L_i/K)$ satisfying the compatibility relation $\sigma_j|_{L_i} = \sigma_i$ for $L_i \subset L_j$ gives rise to a well-defined element $\sigma \in \mathrm{Gal}(L/K)$. Two facts are responsible for this. First, observe that L is the union of all fields $L_i \in \mathfrak{L}$, since for every $a \in L$, the normal closure of $K(a)$ in L/K is a finite Galois extension containing a; cf. 3.5/7. In particular, every $\sigma \in \mathrm{Gal}(L/K)$ is uniquely determined by its restrictions to the L_i. On the other hand, for every two Galois extensions $L_i, L_j \in \mathfrak{L}$, there is some $L_k \in \mathfrak{L}$ such that $L_i \cup L_j \subset L_k$, namely the composite field $L_i \cdot L_j = K(L_i, L_j)$. Thus, if (σ_i) is a family of Galois automorphisms satisfying $\sigma_j|_{L_i} = \sigma_i$ for $L_i \subset L_j$, then the σ_i give rise to a well-defined map $\sigma\colon L \longrightarrow L$. The latter is a K-automorphism, since for any two elements $a, b \in L$, say $a \in L_i$, $b \in L_j$, there is always an index k such that $a, b \in L_k$, and since we can use the fact that σ_k is a K-automorphism.

Next consider a subgroup $H \subset \mathrm{Gal}(L/K)$. Similarly as before, we can restrict H to each L_i and thereby look at the subgroups $H_i = f_i(H) \subset \mathrm{Gal}(L_i/K)$, $i \in I$. An element $a \in L$ is invariant under H if and only if it is invariant under a subgroup (or alternatively, all subgroups) H_i such that $a \in L_i$. However, in contrast to the situation encountered before, H will in general not be uniquely characterized by its restrictions H_i. As an example, one may consider the absolute Galois group of a finite field, which will be computed at the end of the present section. This indeterminacy of H is the actual reason for the fact that the fundamental theorem of Galois theory 4.1/6 cannot be extended to infinite Galois extensions without modifying its assertion. It is necessary to consider a certain closure of subgroups in $\mathrm{Gal}(L/K)$, and the easiest way to do this is by using concepts from topology.

Recall that a *topology* on a set X consists of a system $\mathfrak{T} = (U_i)_{i\in I}$ of subsets of X, the so-called *open* sets, such that the following conditions are satisfied:

(i) $\emptyset, X$ are open.

(ii) The union of arbitrarily many open subsets of X is open.

(iii) The intersection of finitely many open subsets of X is open.

The pair $(X, \mathfrak{T})$ (in most cases simply denoted by X) is called a *topological space*. For a point $x \in X$, open sets $U \subset X$ containing x are called *open neighborhoods* of x. Complements of open subsets of X are referred to as *closed* subsets of X. Furthermore, given an arbitrary subset $S \subset X$, we can consider its *closure* $\overline{S}$. The latter equals the intersection of all closed subsets of X containing S. In other words, it is the smallest closed subset of X containing S and hence

consists of all points $x \in X$ such that $U \cap S \neq \emptyset$ for every open neighborhood U of x. As usual, a map of topological spaces $(X', \mathfrak{T}') \longrightarrow (X, \mathfrak{T})$ is called *continuous* if the preimage of every $\mathfrak{T}$-open subset of X is $\mathfrak{T}'$-open in X', or equivalently, if the preimage of every $\mathfrak{T}$-closed subset of X is $\mathfrak{T}'$-closed in X'.

To introduce a topology on a set X, we can start out from an arbitrary system $\mathfrak{B}$ of subsets of X and look at the topology generated by it. To set up the latter, we enlarge $\mathfrak{B}$ to a system $\mathfrak{B}'$ by adding the special subset $X \subset X$, as well as all finite intersections of subsets of X that belong to $\mathfrak{B}$. Then a subset $U \subset X$ is said to be open if it is a union of sets belonging to $\mathfrak{B}'$, or in other words, if for every $x \in U$ there is some $V \in \mathfrak{B}'$ such that $x \in V \subset U$. It is easily seen that in this way, we obtain a topology $\mathfrak{T}$ on X. One calls $\mathfrak{T}$ the topology on X that is *generated* by $\mathfrak{B}$. Note that $\mathfrak{T}$ is the *weakest topology* on X such that the members of $\mathfrak{B}$ are open in X, i.e., every further topology $\mathfrak{T}'$ with this property is *finer* than $\mathfrak{T}$ in the sense that every $\mathfrak{T}$-open subset of X is $\mathfrak{T}'$-open as well. Moreover, it is easily checked that the enlargement from $\mathfrak{B}$ to $\mathfrak{B}'$ is unnecessary if X is the union of all members of $\mathfrak{B}$ and if the intersection of any two sets $U, V \in \mathfrak{B}$ is a union of subsets of X that belong to $\mathfrak{B}$.

As an example of the preceding construction, we can define the *product* of a family of topological spaces $(X_i)_{i \in I}$. To do this, consider on the Cartesian product of sets $\prod_{i \in I} X_i$ the topology generated by all subsets of type $\prod_{i \in I} U_i$, where U_i is open in X_i and $U_i = X_i$ for almost all $i \in I$. This is the weakest topology on $\prod_{i \in I} X_i$ such that all projections onto the factors X_i are continuous. Also we want to introduce the *restriction* of a topology on a set X to a subset $V \subset X$. Thereby we mean the topology on V whose open sets consist of all intersections of open sets in X with V. This topology is also referred to as the topology that is *induced* from X on V.

We now return to a Galois extension L/K as considered before, and look again at the system $\mathfrak{L} = (L_i)_{i \in I}$ of all its subextensions that are finite and Galois over K, together with the corresponding restriction homomorphisms $f_i \colon \mathrm{Gal}(L/K) \longrightarrow \mathrm{Gal}(L_i/K)$. We equip the finite group $\mathrm{Gal}(L_i/K)$ for every $i \in I$ with the discrete topology, i.e., with the topology in which every subset of $\mathrm{Gal}(L_i/K)$ is open. Moreover, we consider on $\mathrm{Gal}(L/K)$ the weakest topology such that all restrictions $f_i \colon \mathrm{Gal}(L/K) \longrightarrow \mathrm{Gal}(L_i/K)$ are continuous. Since $\mathrm{Gal}(L_i/K)$ carries the discrete topology for every i, the latter equals the topology that is generated by all fibers of the maps f_i.[1]

Remark 1. (i) *A subset $U \subset \mathrm{Gal}(L/K)$ is open if and only if for every $\sigma \in U$ there is an index $i \in I$ such that $f_i^{-1}(f_i(\sigma)) \subset U$.*

(ii) *A subset $A \subset \mathrm{Gal}(L/K)$ is closed if and only if for every $\sigma \in \mathrm{Gal}(L/K)$ not belonging to A there is an index $i \in I$ such that $f_i^{-1}(f_i(\sigma)) \cap A = \emptyset$.*

(iii) *For a subset $S \subset \mathrm{Gal}(L/K)$ its closure $\overline{S}$ consists of all $\sigma \in \mathrm{Gal}(L/K)$ such that $f_i^{-1}(f_i(\sigma)) \cap S \neq \emptyset$ for all $i \in I$.*

[1] The fibers of a map $f \colon X \longrightarrow Y$ are given by the preimages $f^{-1}(y)$ of points $y \in Y$.

Proof. We restrict ourselves to proving (i), since the remaining assertions are formal consequences of it. Let $\mathfrak{B}$ be the system of all fibers of the restriction maps f_i, $i \in I$. Due to the description of the topology generated by a system of subsets of a set X, we have only to show that it is not necessary to enlarge $\mathfrak{B}$ to a system $\mathfrak{B}'$, as explained before, by adding finite intersections of members of $\mathfrak{B}$, i.e., that for any two automorphisms $\sigma_i \in \mathrm{Gal}(L_i/K)$ and $\sigma_j \in \mathrm{Gal}(L_j/K)$, the intersection $f_i^{-1}(\sigma_i) \cap f_j^{-1}(\sigma_j)$ is a union of certain fibers of restriction maps $f_k\colon \mathrm{Gal}(L/K) \longrightarrow \mathrm{Gal}(L_k/K)$. To achieve this, choose an index $k \in I$ such that $L_i \cup L_j \subset L_k$. Since f_i is the composition of f_k with the restriction map $\mathrm{Gal}(L_k/K) \longrightarrow \mathrm{Gal}(L_i/K)$, we see that $f_i^{-1}(\sigma_i)$ is the union of fibers of $f_k\colon \mathrm{Gal}(L/K) \longrightarrow \mathrm{Gal}(L_k/K)$. The same is true for $f_j^{-1}(\sigma_j)$, and it follows that also $f_i^{-1}(\sigma_i) \cap f_j^{-1}(\sigma_j)$ is a union of fibers of f_k. □

Using Remark 1, we can easily see that $\mathrm{Gal}(L/K)$ is a *topological group*. By this one understands a group G equipped with a topology such that the group law $G \times G \longrightarrow G$ and the map of taking inverses $G \longrightarrow G$ are continuous, where, of course, $G \times G$ is considered as a product in the sense of topological spaces. To further illustrate the topology on $\mathrm{Gal}(L/K)$ let us prove the following assertion:

Remark 2. *The topological group* $\mathrm{Gal}(L/K)$ *is compact and totally disconnected.*

Prior to giving the proof, let us recall that a topological space X is called *quasicompact* if every open covering of X contains a finite subcovering. Furthermore, X is called *compact* if X is quasicompact and *Hausdorff*, which means that for $x, y \in X$ there exist disjoint open subsets $U, V \subset X$ such that $x \in U$, $y \in V$. Finally, X is called *totally disconnected* if for every subset $A \subset X$ containing at least two points, there exist two open subsets $U, V \subset X$ such that $A \subset U \cup V$ as well as $U \cap A \neq \emptyset \neq V \cap A$ and $U \cap A \cap V = \emptyset$. For example, if X carries the discrete topology, then X is Hausdorff and totally disconnected. If, in addition, X is finite, then it is compact as well.

Proof of Remark 2. The restriction maps $f_i\colon \mathrm{Gal}(L/K) \longrightarrow \mathrm{Gal}(L_i/K)$ induce an injective homomorphism

$$\mathrm{Gal}\big(L/K\big) \hookrightarrow \prod_{i \in I} \mathrm{Gal}\big(L_i/K\big),$$

which we will view as an inclusion. Furthermore, $\prod \mathrm{Gal}(L_i/K)$, as a product of finite discrete and hence compact topological spaces, is itself compact, due to Tychonoff's theorem (a fact that in the present situation can also be justified by an elementary argument). Since $\prod \mathrm{Gal}(L_i/K)$ induces on $\mathrm{Gal}(L/K)$ the given topology, the latter is seen to be compact if we can show that $\mathrm{Gal}(L/K)$ is closed in $\prod \mathrm{Gal}(L_i/K)$. To achieve this, consider a point $(\sigma_i) \in \prod \mathrm{Gal}(L_i/K)$ that does not belong to $\mathrm{Gal}(L/K)$, i.e., such that there are two indices $j, j' \in I$ such that $L_j \subset L_{j'}$, but $\sigma_{j'}|_{L_j} \neq \sigma_j$. Then the set of all $(\sigma_i') \in \prod \mathrm{Gal}(L_i/K)$

satisfying $\sigma'_j = \sigma_j$ and $\sigma'_{j'} = \sigma_{j'}$ forms an open neighborhood of (σ_i) that is disjoint from $\mathrm{Gal}(L/K)$. This shows that $\mathrm{Gal}(L/K)$ is closed in $\prod \mathrm{Gal}(L_i/K)$.

To show that $\mathrm{Gal}(L/K)$ is totally disconnected, it is enough to show that $\prod \mathrm{Gal}(L_i/K)$, as a product of discrete topological groups, is totally disconnected. Let (σ_i) and (σ'_i) be two different elements of $\prod \mathrm{Gal}(L_i/K)$. Then there exists an index $j \in I$ such that $\sigma_j \neq \sigma'_j$, and we can define open subsets $V = \prod V_i$ and $V' = \prod V'_i$ in $\prod \mathrm{Gal}(L_i/K)$ by

$$V_i = \begin{cases} \mathrm{Gal}(L_i/K) & \text{for } i \neq j, \\ \{\sigma_j\} & \text{for } i = j, \end{cases} \qquad V'_i = \begin{cases} \mathrm{Gal}(L_i/K) & \text{for } i \neq j, \\ \mathrm{Gal}(L_j/K) - \{\sigma_j\} & \text{for } i = j. \end{cases}$$

Hence, we get $(\sigma_i) \in V$, $(\sigma'_i) \in V'$, as well as $\prod \mathrm{Gal}(L_i/K) = V \cup V'$ and $V \cap V' = \emptyset$. From this we can read immediately that $\prod \mathrm{Gal}(L_i/K)$ satisfies the defining condition of a totally disconnected topological space. □

Now we are able to generalize the fundamental theorem of Galois theory 4.1/6 to arbitrary Galois extensions.

Proposition 3. *Let L/K be a (not necessarily finite) Galois extension. Then the intermediate fields of L/K correspond bijectively to the* closed *subgroups of* $\mathrm{Gal}(L/K)$. *More precisely, the assertions of the fundamental theorem* 4.1/6 *remain valid if we restrict ourselves to those subgroups $H \subset \mathrm{Gal}(L/K)$ that are* closed.

The main work for proving the proposition was already done in Section 4.1; see 4.1/7. It remains only to verify for intermediate fields E of L/K that the corresponding Galois group $\mathrm{Gal}(L/E)$ is a closed subgroup of $\mathrm{Gal}(L/K)$, and that the composition $\Psi \circ \Phi$ from 4.1/6 yields the identity on the set of all closed subgroups in $\mathrm{Gal}(L/K)$. Both facts are consequences of the following result:

Lemma 4. *Let $H \subset \mathrm{Gal}(L/K)$ be a subgroup and let $L^H \subset L$ be the corresponding fixed field. Then* $\mathrm{Gal}(L/L^H)$, *viewed as a subgroup of* $\mathrm{Gal}(L/K)$, *equals the closure of H in* $\mathrm{Gal}(L/K)$.

Proof. As before, we consider the system $(L_i)_{i \in I}$ of all intermediate fields of L/K such that L_i/K is a finite Galois extension, together with the restriction maps $f_i \colon \mathrm{Gal}(L/K) \longrightarrow \mathrm{Gal}(L_i/K)$. Let $H_i = f_i(H)$. Since an element $a \in L_i$ is invariant under H if and only if it is invariant under H_i, we get $L^H \cap L_i = L_i^{H_i}$ and hence $L^H = \bigcup_{i \in I} L_i^{H_i}$. Now consider a second subgroup $H' \subset \mathrm{Gal}(L/K)$ and set $H'_i = f_i(H')$. Then, by 4.1/4 or 4.1/6, we see that we get $L^H = L^{H'}$ if and only if $H_i = H'_i$ for all $i \in I$. However, $H' := \bigcap_{i \in I} f_i^{-1}(H_i)$ is clearly the largest subgroup in $\mathrm{Gal}(L/K)$ such that $f_i(H') = H_i$ for all $i \in I$ and hence such that $L^{H'} = L^H$. Thereby we see that $H' = \mathrm{Gal}(L/L^H)$.

On the other hand, the closure $\overline{H}$ of H in $\mathrm{Gal}(L/K)$ is computed according to Remark 1 (iii) as follows:

$$\begin{aligned}\overline{H} &= \Big\{\sigma \in \mathrm{Gal}\big(L/K\big)\,;\, f_i^{-1}\big(f_i(\sigma)\big) \cap H \neq \emptyset \text{ for all } i \in I\Big\}\\ &= \Big\{\sigma \in \mathrm{Gal}\big(L/K\big)\,;\, f_i(\sigma) \in H_i \text{ for all } i \in I\Big\}\\ &= \bigcap_{i\in I} f_i^{-1}(H_i)\\ &= H'.\end{aligned}$$

Thus, $\mathrm{Gal}(L/L^H)$ is recognized as the closure of the subgroup $H \subset \mathrm{Gal}(L/K)$. □

In the setting of Proposition 3, the open subgroups of $\mathrm{Gal}(L/K)$ can be characterized as follows:

Corollary 5. *Let L/K be a Galois extension and H a subgroup of* $\mathrm{Gal}(L/K)$. *Then the following assertions are equivalent*:

(i) *H is open in* $\mathrm{Gal}(L/K)$.

(ii) *H is closed in* $\mathrm{Gal}(L/K)$ *and the fixed field L^H is finite over K.*

Proof. First assume that H is open in $\mathrm{Gal}(L/K)$. Then H is closed in $\mathrm{Gal}(L/K)$ as well, since all its left (resp. right) cosets in $\mathrm{Gal}(L/K)$ are open, and hence the complement of H is open in $\mathrm{Gal}(L/K)$. Furthermore, using Remark 1 (i), there exists a finite Galois extension L'/K in L such that H contains the kernel of the restriction map $\mathrm{Gal}(L/K) \longrightarrow \mathrm{Gal}(L'/K)$, which equals $\mathrm{Gal}(L/L')$. Then, using Proposition 3, we get $L^H \subset L^{\mathrm{Gal}(L/L')} = L'$, and L^H is finite over K, since the same is true for L'.

Conversely, if H is closed in $\mathrm{Gal}(L/K)$ and L^H/K is finite, we can consider the normal closure $L' \subset L$ of L^H/K, which is finite over K; cf. 3.5/7. Then $\mathrm{Gal}(L/L')$ is open in $\mathrm{Gal}(L/K)$ according to Remark 1 (i), and we have $\mathrm{Gal}(L/L') \subset \mathrm{Gal}(L/L^H) = H$, again by Proposition 3. In particular, H is open in $\mathrm{Gal}(L/K)$. □

In studying infinite Galois extensions L/K, it is often convenient to view the Galois group $\mathrm{Gal}(L/K)$ as the projective limit of the finite Galois groups $\mathrm{Gal}(L_i/K)$, where $(L_i)_{i\in I}$, as usual, denotes the system of all intermediate fields of L/K that are finite and Galois over K. Let us briefly explain the formalism of projective limits.

Consider a partially ordered index set I with order relation $\leq$ as in 3.4, together with group homomorphisms $f_{ij}\colon G_j \longrightarrow G_i$ for indices $i, j \in I$, $i \leq j$, and assume the following conditions:

(i) $f_{ii} = \mathrm{id}_{G_i}$ for all $i \in I$.

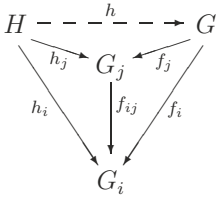
(ii) $f_{ik} = f_{ij} \circ f_{jk}$ for $i \leq j \leq k$.

Such a system $(G_i, f_{ij})_{i,j\in I}$ is called a *projective system* of groups. In a similar way one can define projective systems of sets or of sets with additional structures. For example, for a projective system of topological groups it is required that all maps f_{ij} be continuous homomorphisms. A group G together

with homomorphisms $f_i\colon G \longrightarrow G_i$ satisfying $f_i = f_{ij} \circ f_j$ for $i \leq j$ is called a *projective limit* of the system (G_i, f_{ij}) if it admits the following universal property:

If $h_i\colon H \longrightarrow G_i$, $i \in I$, are group homomorphisms satisfying $h_i = f_{ij} \circ h_j$ for $i \leq j$, then there exists a unique group homomorphism $h\colon H \longrightarrow G$ such that $h_i = f_i \circ h$ for all $i \in I$.

This condition is illustrated by the following commutative diagram:



If a projective limit G exists, it is unique up to canonical isomorphism, as is the case for any object defined in terms of a universal property. The reason is as follows. If in the above situation, together with (G, f_i), also (H, h_i) is a projective limit of (G_i, f_{ij}), then besides $h\colon H \longrightarrow G$, there is also a homomorphism $g\colon G \longrightarrow H$ satisfying the compatibilities as specified in the above diagram. Taking into account the uniqueness condition in the definition of projective limits, we see that the maps $g \circ h, \mathrm{id}_H\colon H \longrightarrow H$ coincide, and that the same is true for $h \circ g, \mathrm{id}_G\colon G \longrightarrow G$. Thus, h and g are inverse to each other. We write $G = \varprojlim_{i \in I} G_i$ for the projective limit of the system (G_i, f_{ij}), where in most cases, the homomorphisms f_i are not mentioned explicitly, certainly if they are defined in an obvious way.

If (G_i, f_{ij}) is a projective system of *topological* groups and (G, f_i) is its projective limit in the sense of ordinary groups, then we can equip G with the weakest topology such that all homomorphisms f_i are continuous. This is the topology generated by all preimages $f_i^{-1}(U)$ of open subsets $U \subset G_i$; it is referred to as the *projective limit* of the topologies on the groups G_i. Indeed, G equipped with this topology is a projective limit of (G_i, f_{ij}) in the sense of topological groups.

Let us add along the way that there is also a notion that is dual to that of projective limits, namely the notion of *inductive* (or *direct*) *limits* $\varinjlim$. The definition of an inductive system, resp. of an inductive limit, is obtained by inverting the direction of all arrows occurring in the corresponding setup of a projective system, resp. of a projective limit. In addition, for inductive systems it is required that the inherent index set I be *directed* in the sense that for arbitrary indices $i, j \in I$, there is always an index $k \in I$ such that $i, j \leq k$. Projective and inductive limits of groups (resp. sets, or rings, etc.) exist always, as is easily verified. However, here we are interested only in the projective case:

Remark 6. *Let (G_i, f_{ij}) be a projective system of groups.*

(i) *The subgroup*

$$G = \big\{(x_i)_{i\in I}\,;\, f_{ij}(x_j) = x_i \text{ for } i \le j\big\} \subset \prod_{i\in I} G_i,$$

together with the group homomorphisms $f_i\colon G \longrightarrow G_i$ *that are induced from the projections of* $\prod_{i\in I} G_i$ *onto its factors, constitutes a projective limit of* (G_i, f_{ij}).

In particular, every system $(x_i)_{i\in I} \in \prod_{i\in I} G_i$ *satisfying* $f_{ij}(x_j) = x_i$ *for* $i \le j$ *determines a unique element* $x \in \varprojlim_{i\in I} G_i$.

(ii) *If* (G_i, f_{ij}) *is a projective system of topological groups and* G *is as in* (i), *then the restriction of the product topology on* $\prod_{i\in I} G_i$ *to* G *equals the projective limit of the topologies on the* G_i.

Our main example of projective systems is inspired from Galois extensions. Indeed, let L/K be a Galois extension and let $\mathfrak{L} = (L_i)_{i\in I}$ be the system of all intermediate fields that are finite and Galois over K. We introduce a partial ordering on I by setting $i \le j$ if $L_i \subset L_j$. Furthermore, write $G_i = \mathrm{Gal}(L_i/K)$ for $i \in I$ and let $f_{ij}\colon \mathrm{Gal}(L_j/K) \longrightarrow \mathrm{Gal}(L_i/K)$ for $i \le j$ be the restriction map. Then (G_i, f_{ij}) is a projective system of groups, resp. (discrete) topological groups, and we can prove the following result:

Proposition 7. *The restriction maps* $f_i\colon \mathrm{Gal}(L/K) \longrightarrow \mathrm{Gal}(L_i/K)$ *define* $\mathrm{Gal}(L/K)$ *as the projective limit of the system* $(\mathrm{Gal}(L_i/K), f_{ij})$, *i.e.*,

$$\mathrm{Gal}\big(L/K\big) = \varprojlim_{i\in I} \mathrm{Gal}\big(L_i/K\big).$$

This holds in terms of ordinary groups, but also in terms of topological groups.

Proof. It is enough to check the defining universal property of a projective limit in terms of ordinary groups, since the topology given on $\mathrm{Gal}(L/K)$ coincides by its definition with the projective limit of the topologies on the groups $\mathrm{Gal}(L_i/K)$. Therefore, consider group homomorphisms $h_i\colon H \longrightarrow \mathrm{Gal}(L_i/K)$ that are compatible with the restriction maps f_{ij}. To verify the uniqueness part of the universal property, assume there is a group homomorphism $h\colon H \longrightarrow \mathrm{Gal}(L/K)$ satisfying $h_i = f_i \circ h$ for all $i \in I$. Fixing an element $x \in H$, write $\sigma = h(x)$, as well as $\sigma_i = h_i(x)$. Then the relation $h_i = f_i \circ h$ implies $\sigma_i = \sigma|_{L_i}$. Since L equals the union of all L_i, it follows that $\sigma = h(x)$ is uniquely determined by the elements $\sigma_i = h_i(x)$. On the other hand, we can use this observation to settle the existence part of the universal property and to construct a homomorphism $h\colon H \longrightarrow \mathrm{Gal}(L/K)$ as desired. Indeed, look at an element $x \in H$ and let $\sigma_i = h_i(x)$, $i \in I$, denote the images of x. Then the relations $h_i = f_{ij} \circ h_j$ for $i \le j$, and hence $L_i \subset L_j$, show that $\sigma_i = \sigma_j|_{L_i}$. Now use $L = \bigcup_{i\in I} L_i$ and the fact that I is directed, which means for $i, j \in I$ that there is always an index $k \in I$ satisfying $i, j \le k$, i.e., $L_i \cup L_j \subset L_k$. Then we can conclude that the σ_i determine a well-defined automorphism $\sigma \in \mathrm{Gal}(L/K)$ restricting to σ_i on each L_i. Mapping $x \in H$ in each case to the corresponding element $\sigma \in \mathrm{Gal}(L/K)$, we obtain a group homomorphism $h\colon H \longrightarrow \mathrm{Gal}(L/K)$,

as desired. Thus, all in all, $\mathrm{Gal}(L/K)$ admits the defining universal property of a projective limit of the system $(\mathrm{Gal}(L_i/K))_{i\in I}$, and we are done. □

In view of Proposition 7, one calls $\mathrm{Gal}(L/K)$ a *profinite group*, thereby indicating that it is a projective limit of finite (discrete) groups. Let us add for a projective system $(G_i, f_{ij})_{i,j\in I}$ with a *directed* index set I that in order to determine the corresponding projective limit, it is enough to execute this limit over a *cofinal subsystem*. Here a subsystem $(G_i, f_{ij})_{i,j\in I'}$ of $(G_i, f_{ij})_{i,j\in I}$ is called cofinal if for every index $i \in I$, there exists an index $i' \in I'$ such that $i \leq i'$. For example, if $(L_i)_{i\in I'}$ is a subsystem of the system $(L_i)_{i\in I}$ of all intermediate fields of L/K that are finite and Galois over K, and if for every $i \in I$ there is some $i' \in I'$ such that $L_i \subset L_{i'}$, then $\mathrm{Gal}(L/K)$ is already the projective limit of the Galois groups $\mathrm{Gal}(L_i/K)$, $i \in I'$. Note that the index set I is directed in this case, since for $i, j \in I$ there is always an index $k \in I$ such that $L_i \cup L_j \subset L_k$.

Finally, let us have a look at an example in which we compute an infinite Galois group. Let p be a prime number and $\overline{\mathbb{F}}$ an algebraic closure of the field $\mathbb{F}_p$ with p elements. Then every finite extension of $\mathbb{F}_p$ is of type $\mathbb{F}_q$ for a power $q = p^n$, see 3.8/2, and we may view all such fields $\mathbb{F}_q$ as subfields of $\overline{\mathbb{F}}$; see 3.4/9 and 3.8/3. We want to compute the Galois group $\mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F}_q)$ for a fixed power $q = p^n$, the so-called absolute *Galois group* of $\mathbb{F}_q$. To do this, we consider the system of all finite Galois extensions of $\mathbb{F}_q$, hence by 3.8/3 and 3.8/4 the system $(\mathbb{F}_{q^i})_{i\in\mathbb{N}-\{0\}}$. Then Proposition 7 says that

$$\mathrm{Gal}\big(\overline{\mathbb{F}}/\mathbb{F}_q\big) = \varprojlim_{i\in\mathbb{N}-\{0\}} \mathrm{Gal}\big(\mathbb{F}_{q^i}/\mathbb{F}_q\big).$$

Let us look more closely at the projective limit on the right-hand side. We write $\sigma\colon \overline{\mathbb{F}} \longrightarrow \overline{\mathbb{F}}$, $a \longmapsto (a^p)^n = a^q$, for the nth power of the Frobenius homomorphism on $\overline{\mathbb{F}}$; similarly as in Section 3.8, one calls σ the *relative Frobenius homomorphism* over $\mathbb{F}_q$. Recall that $\mathbb{F}_q \subset \overline{\mathbb{F}}$ is the splitting field of the polynomial $X^q - X$ over $\mathbb{F}_p$; see 3.8/2. Hence, the fixed field under the cyclic subgroup of $\mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F}_p)$ that is generated by σ equals $\mathbb{F}_q$. Furthermore, let us write σ_i for the restriction of σ to the finite extension $\mathbb{F}_{q^i}$ of $\mathbb{F}_q$. Then we can read from 3.8/3, resp. 3.8/6, the following:

Remark 8. (i) *The Galois group* $\mathrm{Gal}(\mathbb{F}_{q^i}/\mathbb{F}_q)$ *is cyclic of order* i*, generated by the restriction* σ_i *of the relative Frobenius homomorphism over* $\mathbb{F}_q$*.*

(ii) *We have* $\mathbb{F}_{q^i} \subset \mathbb{F}_{q^j}$ *if and only if* i *divides* j*. If this is the case, the restriction homomorphism* $\mathrm{Gal}(\mathbb{F}_{q^j}/\mathbb{F}_q) \longrightarrow \mathrm{Gal}(\mathbb{F}_{q^i}/\mathbb{F}_q)$ *maps the generating element* σ_j *to the generating element* σ_i*.*

Thereby we see that in order to determine the limit $\varprojlim \mathrm{Gal}(\mathbb{F}_{q^i}/\mathbb{F}_q)$, we have to execute the projective limit over the system $(\mathbb{Z}/i\mathbb{Z})_{i\in\mathbb{N}-\{0\}}$. In more detail, the order relation on $\mathbb{N}-\{0\}$ is the divisibility relation, while for $i \mid j$ the corresponding homomorphism $f_{ij}\colon \mathbb{Z}/j\mathbb{Z} \longrightarrow \mathbb{Z}/i\mathbb{Z}$ is the one mapping the residue class $\overline{1} \in \mathbb{Z}/j\mathbb{Z}$ to the residue class $\overline{1} \in \mathbb{Z}/i\mathbb{Z}$. Thus, we obtain the following result:

Proposition 9. *There exists a unique isomorphism of topological groups*

$$\mathrm{Gal}\big(\overline{\mathbb{F}}/\mathbb{F}_q\big) \simeq \varprojlim_{i\in\mathbb{N}-\{0\}} \mathbb{Z}/i\mathbb{Z}$$

such that the relative Frobenius homomorphism $\sigma \in \mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F}_q)$ *corresponds to the system of residue classes* $\overline{1} \in \mathbb{Z}/i\mathbb{Z}$, $i \in \mathbb{N}-\{0\}$.

We write $\widehat{\mathbb{Z}} = \varprojlim_{i\in\mathbb{N}-\{0\}} \mathbb{Z}/i\mathbb{Z}$, where the limit may just as well be viewed as a projective limit of *rings*, or *topological rings*.[2] In particular, we observe that $\widehat{\mathbb{Z}}$, up to canonical isomorphism, is the absolute Galois group of every *finite* field. Moreover, $\mathbb{Z}$ is canonically a subgroup of $\widehat{\mathbb{Z}}$, since the projections $\mathbb{Z} \longrightarrow \mathbb{Z}/i\mathbb{Z}$ give rise to an injective homomorphism $\mathbb{Z} \longrightarrow \widehat{\mathbb{Z}}$. In fact, $\mathbb{Z}$ corresponds to the free cyclic subgroup $\langle\sigma\rangle$ in $\mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F}_q)$ that is generated by the relative Frobenius homomorphism over $\mathbb{F}_q$. Since all projections $\mathbb{Z} \longrightarrow \mathbb{Z}/i\mathbb{Z}$ are surjective, we conclude that $\mathbb{Z}$ lies dense in $\widehat{\mathbb{Z}}$, so that σ generates a dense subgroup in $\mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F}_q)$, i.e., a subgroup whose closure equals $\mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F}_q)$. By the way, this fact is also a consequence of Lemma 4, since $\mathbb{F}_q$ can be interpreted as the fixed field $\overline{\mathbb{F}}^{\langle\sigma\rangle}$. We will see below that $\mathbb{Z}$ is a proper subgroup of $\widehat{\mathbb{Z}}$, even a subgroup that is significantly "smaller" than $\widehat{\mathbb{Z}}$ itself. In particular, it follows that the relative Frobenius homomorphism σ generates a subgroup in $\mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F}_q)$ that is not closed.

We may view $\widehat{\mathbb{Z}}$, as the notation indicates already, as a certain closure of the ring $\mathbb{Z}$, although other closures of $\mathbb{Z}$ are thinkable as well. For example, when executing the projective limit of the quotients $\mathbb{Z}/i\mathbb{Z}$, we may restrict ourselves to integers i varying only over a certain subset of $\mathbb{N}-\{0\}$. Indeed, for a prime number ℓ, the projective limit of topological rings

$$\mathbb{Z}_\ell = \varprojlim_{\nu\in\mathbb{N}} \mathbb{Z}/\ell^\nu\mathbb{Z}$$

is referred to as the ring of *integral ℓ-adic numbers*. In our situation these rings are quite useful, since their structure is easy to describe and since, on the other hand, $\widehat{\mathbb{Z}}$ decomposes into a Cartesian product of them:

Proposition 10. *There exists a canonical isomorphism of topological rings*

$$\widehat{\mathbb{Z}} = \varprojlim_{i\in\mathbb{N}-\{0\}} \mathbb{Z}/i\mathbb{Z} \simeq \prod_{\ell \text{ prime}} \mathbb{Z}_\ell.$$

Proof. We show that $P := \prod_{\ell \text{ prime}} \mathbb{Z}_\ell$, together with canonical homomorphisms $f_i\colon P \longrightarrow \mathbb{Z}/i\mathbb{Z}$ to be introduced below, admits the universal property of a projective limit of the system $(\mathbb{Z}/i\mathbb{Z})_{i\in\mathbb{N}-\{0\}}$. To do this, look at an integer

[2] For a topological ring R it is required that R be a topological group with respect to the addition, and furthermore, that the ring multiplication be continuous.

$i \in \mathbb{N}-\{0\}$ with prime factorization $i = \prod_\ell \ell^{\nu_\ell(i)}$, where, of course, almost all exponents $\nu_\ell(i)$ are zero. Applying the Chinese remainder theorem in the version of 2.4/14 yields that the canonical homomorphism

$$(*) \qquad \mathbb{Z}/i\mathbb{Z} \longrightarrow \prod_{\ell \text{ prime}} \mathbb{Z}/\ell^{\nu_\ell(i)}\mathbb{Z}$$

is an isomorphism, and we obtain a canonical homomorphism

$$f_i \colon P \longrightarrow \prod_{\ell \text{ prime}} \mathbb{Z}/\ell^{\nu_\ell(i)}\mathbb{Z} \xrightarrow{\sim} \mathbb{Z}/i\mathbb{Z}.$$

Varying i over $\mathbb{N}-\{0\}$, the homomorphisms $f_i \colon P \longrightarrow \mathbb{Z}/i\mathbb{Z}$ are compatible with the projections $f_{ij} \colon \mathbb{Z}/j\mathbb{Z} \longrightarrow \mathbb{Z}/i\mathbb{Z}$ for $i \mid j$. Furthermore, the definition of the f_i shows that the topology on P coincides with the weakest one such that all f_i are continuous. Therefore, it remains only to check that (P, f_i) is a projective limit of $(\mathbb{Z}/i\mathbb{Z}, f_{ij})$, in the sense of ordinary rings.

However, this is more or less straightforward. Fix a ring R and consider ring homomorphisms $h_i \colon R \longrightarrow \mathbb{Z}/i\mathbb{Z}$ for $i \in \mathbb{N}-\{0\}$ that are compatible with the projection maps f_{ij}. Relying on isomorphisms as in $(*)$, the h_i induce for every prime number ℓ homomorphisms $h_{i,\ell} \colon R \longrightarrow \mathbb{Z}/\ell^{\nu_\ell(i)}\mathbb{Z}$ that are compatible with the projection homomorphisms of the projective system $(\mathbb{Z}/\ell^\nu\mathbb{Z})_{\nu\in\mathbb{N}}$. Therefore, the $h_{i,\ell}$ define a ring homomorphism $h_\ell \colon R \longrightarrow \varprojlim_{\nu\in\mathbb{N}} \mathbb{Z}/\ell^\nu\mathbb{Z}$ and hence, letting ℓ vary, a ring homomorphism $h \colon R \longrightarrow P$, which satisfies $h_i = f_i \circ h$ for all i. Finally, since the $h_{i,\ell}$ are equivalent to the h_i, we can use the uniqueness part of the universal property of the limits $\mathbb{Z}_\ell$ to show that h is uniquely determined by the h_i. □

Thus, we can summarize:

Theorem 11. *Let $\mathbb{F}$ be a finite field and $\overline{\mathbb{F}}$ an algebraic closure. Then there exists a canonical isomorphism of topological groups*

$$\mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F}) \simeq \prod_{\ell \text{ prime}} \mathbb{Z}_\ell$$

such that the relative Frobenius homomorphism $\sigma \in \mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F})$ corresponds to the element $(1, 1, \ldots) \in \prod_{\ell \text{ prime}} \mathbb{Z}_\ell$. Here 1 denotes in each case the unit element in $\mathbb{Z}_\ell$, which is viewed as a ring.

In particular, we thereby see that the free cyclic subgroup $\mathbb{Z} \subset \mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F})$ that is generated by the relative Frobenius homomorphism σ is significantly "smaller" than the Galois group $\mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F})$, in fact, even significantly "smaller" than the ring of integral ℓ-adic numbers $\mathbb{Z}_\ell$. Indeed, it is not too hard to see (and this justifies the terminology of ℓ-adic numbers) that the elements of $\mathbb{Z}_\ell$ correspond bijectively to all formal infinite series $\sum_{\nu=0}^{\infty} c_\nu \ell^\nu$ with integer coefficients c_ν, $0 \le c_\nu \le \ell - 1$. In this setting, the subset $\mathbb{N}$ of natural numbers is represented by all finite sums of this type.

Exercises

1. *Make precise the basic idea that makes it possible to extend the fundamental theorem of Galois theory* 4.1/6 *in a straightforward way to infinite Galois extensions.*

2. *Explain why infinite Galois groups should be viewed rather as topological or profinite groups than as purely abstract groups.*

3. Let X be a set and $(X_i)_{i\in I}$ a system of subsets of X. For indices $i, j \in I$ such that $X_j \subset X_i$, let f_{ij} be the inclusion map $X_j \longrightarrow X_i$.
   (i) Write $i \leq j$ if $X_j \subset X_i$ and show that (X_i, f_{ij}) is a projective system of sets satisfying $\varprojlim_{i\in I} X_i = \bigcap_{i\in I} X_i$.
   (ii) Write $i \leq j$ if $X_i \subset X_j$ and assume that the index set I is directed with respect to the relation $\leq$. (However, in the present context, the condition "directed" is without significance.) Show that (X_i, f_{ji}) is an inductive system of sets satisfying $\varinjlim_{i\in I} X_i = \bigcup_{i\in I} X_i$.

4. Show that every inductive system of groups admits an (inductive) limit.

5. Let K be a field and $\overline{K}$ an algebraic closure of K. Show that the absolute Galois group $\mathrm{Gal}(\overline{K}/K)$ is independent of the choice of $\overline{K}$, up to isomorphism.

6. Let L/K be a field extension and let $(L_i)_{i\in I}$ be a system of intermediate fields such that L_i, in each case, is Galois over K and such that for $i, j \in I$ there is always an index $k \in I$ satisfying $L_i \cup L_j \subset L_k$. Furthermore, let L' be the smallest subfield of L containing all L_i. Show that L'/K is Galois and that we have $\mathrm{Gal}(L'/K) = \varprojlim \mathrm{Gal}(L_i/K)$ in the sense of topological groups.

7. Let L/K be a Galois extension and E an intermediate field such that E/K is Galois. Show:
   (i) The restriction homomorphism $\varphi\colon \mathrm{Gal}(L/K) \longrightarrow \mathrm{Gal}(E/K)$ is continuous.
   (ii) The topology on $\mathrm{Gal}(E/K)$ equals the quotient topology with respect to φ, i.e., a subset $V \subset \mathrm{Gal}(E/K)$ is open if and only if $\varphi^{-1}(V)$ is open in $\mathrm{Gal}(L/K)$.

8. Can there exist a Galois extension L/K satisfying $\mathrm{Gal}(L/K) \simeq \mathbb{Z}$?

9. Look at the situation of Theorem 11.
   (i) For a prime number ℓ, determine the fixed field of $\mathbb{Z}_\ell$, viewed as a subgroup of $\mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F})$.
   (ii) Determine all intermediate fields of $\overline{\mathbb{F}}/\mathbb{F}$.

10. Consider the ring $\mathbb{Z}_\ell = \varprojlim_\nu \mathbb{Z}/\ell^\nu\mathbb{Z}$ of integral ℓ-adic numbers associated to a prime number ℓ. For an element $a \in \mathbb{Z}_\ell$, let $v(a)$ be the maximum of all integers $\nu \in \mathbb{N}$ such that the residue class of a in $\mathbb{Z}/\ell^\nu\mathbb{Z}$ is zero; set $v(a) = \infty$ for $a = 0$. Furthermore, define the *ℓ-adic absolute value* of a by $|a|_\ell = \ell^{-v(a)}$. Show for $a, b \in \mathbb{Z}_\ell$:
    (i) $|a|_\ell = 0 \Longleftrightarrow a = 0$,
    (ii) $|a \cdot b|_\ell = |a|_\ell \cdot |b|_\ell$,
    (iii) $|a + b|_\ell \leq \max\{|a|_\ell, |b|_\ell\}$.

11. Show that the ℓ-adic absolute value $|\cdot|_\ell$ of Exercise 10 induces the topology of $\mathbb{Z}_\ell$ (in the sense that a subset $U \subset \mathbb{Z}_\ell$ is open if and only if there exists for every point of U an ℓ-adic ε-neighborhood that is contained in U). Furthermore, show that $(1-\ell)^{-1} = \sum_{i=0}^{\infty} \ell^i$, where the convergence is naturally understood in terms of the ℓ-adic absolute value. Using a similar argument, one can show that every element $a \in \mathbb{Z}_\ell$ satisfying $|a|_\ell = 1$ is a unit in $\mathbb{Z}_\ell$.

4.3 The Galois Group of an Equation

Let K be a field and $f \in K[X]$ a nonconstant polynomial. Furthermore, let L be a splitting field of f over K. Then, if f is separable, L/K is a finite Galois extension, and $\mathrm{Gal}(L/K)$ is referred to as the Galois group of f over K or, more suggestively, as the Galois group of the equation $f(x) = 0$.

Proposition 1. *Let $f \in K[X]$ be a separable polynomial of degree $n > 0$ with splitting field L over K, and let $\alpha_1, \ldots, \alpha_n \in L$ be the zeros of f. Then*

$$\begin{aligned} \varphi\colon \mathrm{Gal}\big(L/K\big) &\longrightarrow S\big(\{\alpha_1, \ldots, \alpha_n\}\big) \simeq \mathfrak{S}_n, \\ \sigma &\longmapsto \sigma|_{\{\alpha_1,\ldots,\alpha_n\}}, \end{aligned}$$

defines an injective group homomorphism from the Galois group of L/K to the group of permutations of $\alpha_1, \ldots, \alpha_n$, resp. to the group $\mathfrak{S}_n$ of permutations of n elements. In particular, $\mathrm{Gal}(L/K)$ *can be viewed as a subgroup of $\mathfrak{S}_n$, and it follows that $[L : K] = \operatorname{ord} \mathrm{Gal}(L/K)$ divides $\operatorname{ord} \mathfrak{S}_n = n!$.*

Furthermore, f is irreducible if and only if $\mathrm{Gal}(L/K)$ *operates* transitively *on the zero set $\{\alpha_1, \ldots, \alpha_n\}$, i.e., if and only if for every two of these zeros α_i, α_j, there exists an automorphism $\sigma \in \mathrm{Gal}(L/K)$ such that $\sigma(\alpha_i) = \alpha_j$. Such is always the case for $[L : K] = n!$, resp.* $\mathrm{Gal}(L/K) \simeq \mathfrak{S}_n$.

Proof. Consider an automorphism $\sigma \in \mathrm{Gal}(L/K)$. Since σ leaves the coefficients of f invariant, it maps zeros of f to zeros of f. Furthermore, σ is injective and hence induces on $\{\alpha_1, \ldots, \alpha_n\}$ an injective and therefore bijective self-map, in other words, a permutation. This shows that the map φ is well defined. It is injective as well, due to $L = K(\alpha_1, \ldots, \alpha_n)$, since every K-homomorphism of $\mathrm{Gal}(L/K)$ is uniquely determined by its values on the elements $\alpha_1, \ldots, \alpha_n$.

Now assume that f is irreducible, and consider two zeros α_i, α_j of f. Using 3.4/8, there is a K-homomorphism $\sigma\colon K(\alpha_i) \longrightarrow K(\alpha_j)$ such that $\sigma(\alpha_i) = \alpha_j$. This extends by 3.4/9 to a K-homomorphism $\sigma'\colon L \longrightarrow \overline{L}$, where $\overline{L}$ is an algebraic closure of L. Since the extension L/K is normal, it follows that σ' restricts to a K-automorphism of L and thereby to an element $\sigma'' \in \mathrm{Gal}(L/K)$ satisfying $\sigma''(\alpha_i) = \alpha_j$.

On the other hand, if f is reducible, consider a factorization $f = gh$ into nonconstant polynomials $g, h \in K[X]$. Then every $\sigma \in \mathrm{Gal}(L/K)$ induces a self-map on the zeros of g and, in the same way, on the zeros of h. However,

since f is separable by our assumption, the zeros of g are disjoint from those of h, and we thereby see that σ cannot act transitively on the zeros of f. □

Since every finite Galois extension L/K is simple by the primitive element theorem 3.6/12, it follows that L is a splitting field of a polynomial of degree $n = [L : K]$ in $K[X]$. Thus, we can conclude the following corollary:

Corollary 2. *For a finite Galois extension L/K of degree n, the Galois group* $\mathrm{Gal}(L/K)$ *may be viewed as a subgroup of the permutation group $\mathfrak{S}_n$.*

Also note in the setting of Proposition 1 that the Galois group $\mathrm{Gal}(L/K)$ will in general be a proper subgroup of $\mathfrak{S}_n$. For example, if $f \in K[X]$ is the minimal polynomial of a primitive element of L/K and n is its degree, we get $\mathrm{ord}(\mathrm{Gal}(L/K)) = n < n! = \mathrm{ord}\,\mathfrak{S}_n$ for $n > 2$. Hence, as a general rule, not every permutation of the zeros of f as in Proposition 1 is induced from a Galois automorphism.

Let us compute the Galois group of polynomials $f \in K[X]$ in some special cases.

(1) Consider a polynomial $f = X^2 + aX + b \in K[X]$, where we assume that f does not admit a zero in K. Then f is irreducible in $K[X]$ and, furthermore, separable if $\mathrm{char}\,K \neq 2$ or $a \neq 0$. Adjoining a zero α of f to K, the resulting field $L = K(\alpha)$ is a splitting field of f over K, so that L/K is a Galois extension of degree 2. In particular, the Galois group $\mathrm{Gal}(L/K)$ is of order 2 and thus necessarily cyclic.

(2) Next consider a polynomial of type $f = X^3 + aX + b \in K[X]$, where $\mathrm{char}\,K \neq 2, 3$. Let us mention along the way that every monic polynomial $X^3 + c_1X^2 + \ldots \in K[X]$ of degree 3 can be assumed to be of the latter form, using the substitution $X \longmapsto X - c$ for $c = \frac{1}{3}c_1$; the splitting field and Galois group of f remain unchanged under such a transformation. We assume that f does not have a zero in K. Then f is irreducible in $K[X]$ and, by our assumption on $\mathrm{char}\,K$, also separable. Let L be a splitting field of f over K and $\alpha \in L$ a zero of f. Then $K(\alpha)/K$ is an extension of degree 3, and the degree $[L : K]$ may be 3 or 6, depending on whether or not $K(\alpha)$ is already a splitting field of f. Likewise, $\mathrm{Gal}(L/K)$ will be of order 3 or 6, where in either case, we can view this group as a subgroup of $\mathfrak{S}_3$, according to Proposition 1. In the first case, $\mathrm{Gal}(L/K)$ is cyclic of order 3; every element $\sigma \in \mathrm{Gal}(L/K)$ different from the identity is a generating element, since $\mathrm{ord}\,\sigma > 1$ and $(\mathrm{ord}\,\sigma)|3$ imply $\mathrm{ord}\,\sigma = 3$. In the second case we get $\mathrm{Gal}(L/K) = \mathfrak{S}_3$, since $\mathrm{ord}\,\mathrm{Gal}(L/K) = 6 = \mathrm{ord}\,\mathfrak{S}_3$.

We want to explain a general principle allowing one to find out which of the two cases we are facing when considering special examples. Let $\alpha_1, \alpha_2, \alpha_3 \in L$ be the zeros of f and write

$$\delta = (\alpha_1 - \alpha_2)(\alpha_2 - \alpha_3)(\alpha_1 - \alpha_3).$$

Then $\Delta = \delta^2$ is called the *discriminant* of the polynomial f; see also Section 4.4 for this notion. Since Δ is invariant under the automorphisms in $\mathrm{Gal}(L/K)$, we

get $\Delta \in K$, where in our special case, an easy calculation shows that

$$\Delta = -4a^3 - 27b^2.$$

Applying an automorphism $\sigma \in \mathrm{Gal}(L/K)$ to δ, it is possible that the factors of δ change signs. Therefore, we get $\sigma(\delta) = \pm\delta$, depending on whether σ corresponds to an even or an odd permutation in $\mathfrak{S}_3$. (A permutation $\pi \in \mathfrak{S}_n$ is called even, resp. odd, if

$$\mathrm{sgn}(\pi) = \prod_{i<j} \frac{\pi(i) - \pi(j)}{i - j},$$

the sign of π, assumes the value 1, resp. -1; note that the function sgn is multiplicative in the sense that $\mathrm{sgn}(\pi \circ \pi') = \mathrm{sgn}(\pi) \cdot \mathrm{sgn}(\pi')$ for π, $\pi' \in \mathfrak{S}_n$; see also 5.3.)

The even permutations in $\mathfrak{S}_n$ form a subgroup, the so-called alternating group $\mathfrak{A}_n$. This group coincides with the kernel of the group homomorphism

$$\mathfrak{S}_n \longrightarrow \{1, -1\}, \qquad \pi \longmapsto \mathrm{sgn}(\pi),$$

which is surjective for $n > 1$. Thus, $\mathfrak{A}_n$ is a normal subgroup of index 2 in $\mathfrak{S}_n$ for $n > 1$. In addition, we see that all permutations $\pi \in \mathfrak{S}_n$ whose order is odd must belong to $\mathfrak{A}_n$. In particular, $\mathfrak{A}_3$ is the only subgroup in $\mathfrak{S}_3$ of order 3. Hence, the following equivalences hold:

$$\begin{aligned} & \operatorname{ord} \mathrm{Gal}(L/K) = 3 \\ \iff\ & \mathrm{Gal}(L/K) \subset \mathfrak{S}_3 \text{ contains only even permutations} \\ \iff\ & \delta \in K \\ \iff\ & \Delta \text{ admits a square root in } K. \end{aligned}$$

Therefore, we can figure out whether $\mathrm{Gal}(L/K)$ is of order 3 or 6 by checking whether the discriminant Δ admits a square root in K or not.

To look at an explicit example, let L be the splitting field of the polynomial $f = X^3 - X + 1 \in \mathbb{Q}[X]$; it is irreducible, since f does not split off a linear factor in $\mathbb{Z}[X]$. Then we get $\mathrm{Gal}(L/\mathbb{Q}) = \mathfrak{S}_3$, since $\sqrt{\Delta} = \sqrt{-23} \notin \mathbb{Q}$.

(3) To give another example, let us look at special irreducible polynomials of degree 4, more precisely, irreducible monic polynomials $f \in \mathbb{Q}[X]$ that are biquadratic in the sense that their linear and cubic terms are trivial. Such polynomials are of type $f = (X^2 - a)^2 - b$, where we will assume $b > a^2$ in the following. For instance, the polynomials $X^4 - 2$ and $X^4 - 4X^2 - 6$ are of this type. The zeros of f in $\mathbb{C}$ are given by

$$\alpha = \sqrt{a + \sqrt{b}}, \qquad -\alpha, \qquad \beta = \sqrt{a - \sqrt{b}}, \qquad -\beta,$$

where $\sqrt{b} > |a|$. Therefore, α is real, in contrast to β, which is a square root of a negative real number. Furthermore, the splitting field of f in $\mathbb{C}$ is given by $L = \mathbb{Q}(\alpha, \beta)$. To determine its degree over $\mathbb{Q}$, observe that α is a zero of f

and hence has degree 4 over $\mathbb{Q}$, which means that $[\mathbb{Q}(\alpha):\mathbb{Q}]=4$. Further, β, as a square root of $a-\sqrt{b}\in\mathbb{Q}(\alpha)$, is of degree ≤ 2 over $\mathbb{Q}(\alpha)$. Since $\mathbb{Q}(\alpha)$ is contained in $\mathbb{R}$ but β is not, it follows that in fact, β is of degree 2 over $\mathbb{Q}(\alpha)$ and that we can conclude that $[L:\mathbb{Q}]=[\mathbb{Q}(\alpha,\beta):\mathbb{Q}]=8$.

Now let us determine the Galois group $\mathrm{Gal}(L/\mathbb{Q})$. To do this, we use Proposition 1 and view $\mathrm{Gal}(L/\mathbb{Q})$ as a subgroup of the permutation group $S(\{\alpha,-\alpha,\beta,-\beta\})$ of the zeros of f. In doing so, we know already that $L/\mathbb{Q}$ is of degree 8 and hence that $\mathrm{Gal}(L/\mathbb{Q})$ is of order 8. Furthermore, every $\sigma\in\mathrm{Gal}(L/\mathbb{Q})$ is a field homomorphism and as such satisfies the relations $\sigma(-\alpha)=-\sigma(\alpha)$, $\sigma(-\beta)=-\sigma(\beta)$. However, there are precisely eight permutations in $S(\{\alpha,-\alpha,\beta,-\beta\})$ satisfying these conditions. Indeed, if we want to set up such a permutation, there are precisely four possibilities to select $\sigma(\alpha)$, where $\sigma(-\alpha)$ is determined by $\sigma(-\alpha)=-\sigma(\alpha)$. Then there remain two possibilities to select $\sigma(\beta)$, where again, $\sigma(-\beta)$ is determined by the relation $\sigma(-\beta)=-\sigma(\beta)$. As a result, there are precisely eight permutations in $S(\{\alpha,-\alpha,\beta,-\beta\})$ satisfying the relations $\sigma(-\alpha)=-\sigma(\alpha)$, $\sigma(-\beta)=-\sigma(\beta)$, and it follows that these must coincide with the elements of $\mathrm{Gal}(L/\mathbb{Q})$. To explicitly describe this group, consider the two elements $\sigma,\tau\in\mathrm{Gal}(L/\mathbb{Q})$ that are given by

$$\begin{aligned}\sigma:&\quad \alpha\longmapsto\beta,\qquad \beta\longmapsto-\alpha,\\ \tau:&\quad \alpha\longmapsto-\alpha,\quad \beta\longmapsto\beta.\end{aligned}$$

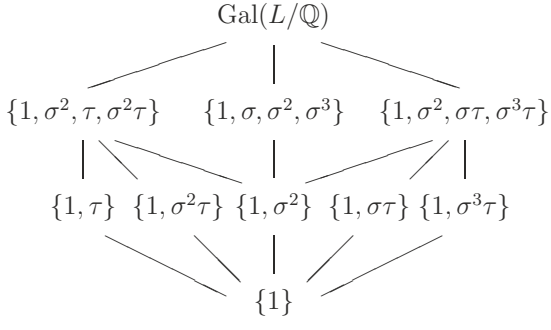
The subgroup $\langle\sigma\rangle\subset\mathrm{Gal}(L/\mathbb{Q})$ generated by σ is cyclic of order 4 and therefore normal in $\mathrm{Gal}(L/\mathbb{Q})$, since it is of index 2. Furthermore, τ is of order 2. Since $\tau\notin\langle\sigma\rangle$, we obtain

$$\mathrm{Gal}(L/\mathbb{Q})=\langle\sigma,\tau\rangle=\langle\sigma\rangle\cup\tau\langle\sigma\rangle=\langle\sigma\rangle\cup\langle\sigma\rangle\tau,$$

or more explicitly,

$$\mathrm{Gal}(L/\mathbb{Q})=\{1,\sigma,\sigma^2,\sigma^3,\tau,\sigma\tau,\sigma^2\tau,\sigma^3\tau\}.$$

To describe the group law on $\mathrm{Gal}(L/\mathbb{Q})$ it is enough to check that σ and τ satisfy the relation $\tau\sigma=\sigma^3\tau$. Now it is easy to specify all subgroups of $\mathrm{Gal}(L/\mathbb{Q})$ by way of the following scheme:



The subgroups of $\mathrm{Gal}(L/\mathbb{Q})$ are in one-to-one correspondence with the intermediate fields of $L/\mathbb{Q}$, due to the fundamental theorem of Galois theory 4.1/6. These fields can be determined by considering suitable elements of degree 2 or 4 in L that are invariant under the above groups.

As a counterpart to the preceding class of biquadratic polynomials, we want to determine the Galois group of the polynomial $f = X^4 - 4X^2 + 16 \in \mathbb{Q}[X]$. Also in this case, f is of type $(X^2-a)^2-b$, where, however, $a = 2$ and $b = -12$ do not satisfy the above condition $b > a^2$. The zeros of f in $\mathbb{C}$ are given by

$$\alpha = 2e^{2\pi i/12}, \qquad -\alpha, \qquad \beta = 2e^{-2\pi i/12}, \qquad -\beta,$$

resp.

$$2\zeta, \qquad 2\zeta^7, \qquad 2\zeta^{11}, \qquad 2\zeta^5,$$

where $\zeta = e^{2\pi i/12}$ has to be viewed as a square root of $\frac{1}{2} + \frac{1}{2}i\sqrt{3}$, and likewise, $e^{-2\pi i/12}$ as a square root of $\frac{1}{2} - \frac{1}{2}i\sqrt{3}$. In particular, adjoining a zero of f to $\mathbb{Q}$, say α, we see that $L = \mathbb{Q}(\alpha) = \mathbb{Q}(\zeta)$ is a splitting field of f over $\mathbb{Q}$. Thus, the Galois group of $L/\mathbb{Q}$ is of order 4, and its members are characterized by

$$\begin{aligned}
\sigma_1 &: \quad \zeta \longmapsto \zeta, \\
\sigma_2 &: \quad \zeta \longmapsto \zeta^5, \\
\sigma_3 &: \quad \zeta \longmapsto \zeta^7, \\
\sigma_4 &: \quad \zeta \longmapsto \zeta^{11}.
\end{aligned}$$

From this we can read the relations $\sigma_1 = \mathrm{id}$, $\sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \mathrm{id}$, as well as $\sigma_2 \circ \sigma_3 = \sigma_4$. Furthermore, we observe that $\mathrm{Gal}(L/\mathbb{Q})$ is commutative and hence that $\mathrm{Gal}(L/\mathbb{Q}) \simeq \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$. Besides the trivial subgroups, $\mathrm{Gal}(L/\mathbb{Q})$ contains only the subgroups $\langle\sigma_2\rangle$, $\langle\sigma_3\rangle$, $\langle\sigma_4\rangle$, where in terms of the fundamental theorem of Galois theory 4.1/6, these correspond to the intermediate fields $\mathbb{Q}(\zeta^3)$, $\mathbb{Q}(\zeta^2)$, $\mathbb{Q}(\sqrt{3})$; use the fact that $\sqrt{3} = \zeta + \zeta^{11}$. Thus, up to the trivial intermediate fields $\mathbb{Q}$ and L, these are the only intermediate fields of $L/\mathbb{Q}$. Extensions of type $L/\mathbb{Q}$ will be studied in more detail in Section 4.5. Indeed, L is constructed from $\mathbb{Q}$ by adjoining a so-called *primitive* 12th root of unity ζ, and it belongs to the class of *cyclotomic fields*.

(4) As a last example, let us study the *generic equation* of degree n. To do this, we fix a field k and consider over it the field L of rational functions in finitely many variables $T_1, \ldots, T_n$, namely

$$L = k(T_1, \ldots, T_n) = Q\big(k[T_1, \ldots, T_n]\big).$$

Every permutation $\pi \in \mathfrak{S}_n$ defines an automorphism of L by applying π to the variables $T_1, \ldots, T_n$:

$$\begin{aligned}
k(T_1, \ldots, T_n) &\longrightarrow k(T_1, \ldots, T_n), \\
\frac{g(T_1, \ldots, T_n)}{h(T_1, \ldots, T_n)} &\longmapsto \frac{g(T_{\pi(1)}, \ldots, T_{\pi(n)})}{h(T_{\pi(1)}, \ldots, T_{\pi(n)})}.
\end{aligned}$$

The corresponding fixed field $K = L^{\mathfrak{S}_n}$ is referred to as the field of *symmetric rational functions* in n variables with coefficients in k. As we read from 4.1/4, the extension L/K is Galois of degree $n!$ and admits $\mathfrak{S}_n$ as Galois group.

To specify the "equation" of the extension L/K, we choose a variable X and consider the polynomial

$$\begin{aligned} f(X) &= \prod_{i=1}^{n}(X - T_i) \\ &= \sum_{j=0}^{n}(-1)^j \cdot s_j(T_1, \ldots, T_n) \cdot X^{n-j} \in k[T_1, \ldots, T_n][X], \end{aligned}$$

where s_j is obtained by expanding the product of the factors $X - T_i$ and by collecting coefficients of $(-1)^j X^{n-j}$; it is called the jth *elementary symmetric polynomial* (or the jth *elementary symmetric function*) in $T_1, \ldots, T_n$. More explicitly, the elementary symmetric polynomials are given by

$$\begin{aligned} s_0 &= 1, \\ s_1 &= T_1 + \ldots + T_n, \\ s_2 &= T_1T_2 + T_1T_3 + \ldots + T_{n-1}T_n, \\ &\ldots \\ s_n &= T_1 \ldots T_n. \end{aligned}$$

Viewing f as a polynomial in $L[X]$, we see that it is invariant under the action of $\mathfrak{S}_n$ and hence admits coefficients already in K. Therefore, we have $k(s_1, \ldots, s_n) \subset K$, and it follows that L is a splitting field of f over $k(s_1, \ldots, s_n)$, resp. K. Moreover, using $\deg f = n$ and $[L : K] = n!$, we see from Proposition 1 that f is *irreducible* in $K[X]$.

Proposition 3. *Every symmetric rational function in $k(T_1, \ldots, T_n)$ can be uniquely written as a rational function over k in the elementary symmetric polynomials $s_1, \ldots, s_n$. In more precise terms*:

(i) $k(s_1, \ldots, s_n) = K$.

(ii) *$s_1, \ldots, s_n$ are algebraically independent over k.*

Proof. To justify (i), observe that

$$[L : K] = \operatorname{ord} \mathfrak{S}_n = n!,$$

and that $k(s_1, \ldots, s_n) \subset K$. Therefore, it is enough to establish the estimate

$$\big[L : k(s_1, \ldots, s_n)\big] \leq n!.$$

However, this is a consequence of Proposition 1, since L is a splitting field of $f = \prod(X - T_i)$ over $k(s_1, \ldots, s_n)$.

To show that the elementary symmetric polynomials $s_1, \ldots, s_n$ are algebraically independent over k, consider the field $k(S_1, \ldots, S_n)$ of all rational

functions in n variables $S_1, \ldots, S_n$, as well as a splitting field $\tilde{L}$ of the polynomial

$$\tilde{f}(X) = \sum_{j=0}^{n} (-1)^j \cdot S_j \cdot X^{n-j} \in k(S_1, \ldots, S_n)[X],$$

where we put $S_0 = 1$ by convention. Let $t_1, \ldots, t_n$ be the zeros of $\tilde{f}$ in $\tilde{L}$, allowing repetitions according to their possible multiplicities. Then we see that

$$\tilde{L} = k(S_1, \ldots, S_n)(t_1, \ldots, t_n) = k(t_1, \ldots, t_n),$$

since the elements $S_1, \ldots, S_n$ may be written as elementary symmetric functions in $t_1, \ldots, t_n$ and as such belong to $k(t_1, \ldots, t_n)$. Now the homomorphism

$$k[T_1, \ldots, T_n] \longrightarrow k[t_1, \ldots, t_n], \qquad \sum a_\nu T^\nu \longmapsto \sum a_\nu t^\nu,$$

maps elementary symmetric functions in $T_1, \ldots, T_n$ to elementary symmetric functions in $t_1, \ldots, t_n$ and thus restricts to a homomorphism

$$k[s_1, \ldots, s_n] \longrightarrow k[S_1, \ldots, S_n], \qquad \sum a_\nu s^\nu \longmapsto \sum a_\nu S^\nu.$$

Since $S_1, \ldots, S_n$ are variables, this map is necessarily injective and therefore an isomorphism. In particular, $s_1, \ldots, s_n$ may be viewed as variables and hence are algebraically independent over k. □

The idea we have just employed, namely to make use of generic polynomials, i.e., of polynomials with variables as coefficients, leads immediately to the generic equation of degree n. Indeed, fixing variables $S_1, \ldots, S_n$, the polynomial

$$p(X) = X^n + S_1 X^{n-1} + \ldots + S_n \in k(S_1, \ldots, S_n)[X]$$

is referred to as the *generic polynomial* of degree n over k. The corresponding equation $p(x) = 0$ is called the *generic equation* of degree n. We want to determine its Galois group by showing that we may identify $p(X)$ up to isomorphism with the polynomial $f(X)$ discussed above.

Proposition 4. *The generic polynomial $p(X) \in k(S_1, \ldots, S_n)[X]$ of degree n is separable and irreducible. It admits $\mathfrak{S}_n$ as its Galois group.*

Proof. We consider the rational function field $L = k(T_1, \ldots, T_n)$ in n variables $T_1, \ldots, T_n$ over k, as well as the fixed field

$$K = L^{\mathfrak{S}_n} = k(s_1, \ldots, s_n)$$

of all symmetric rational functions; cf. Proposition 3. Since the elementary symmetric polynomials $s_1, \ldots, s_n$ are algebraically independent over k, we can view them as variables and therefore introduce a k-isomorphism

$$k(S_1, \ldots, S_n) \xrightarrow{\sim} k(s_1, \ldots, s_n) = K$$

via $S_j \longmapsto (-1)^j s_j$. Interpreting this as an identification, $p(X)$ is transformed into the familiar polynomial

$$f(X) = \sum_{j=0}^{n} (-1)^j \cdot s_j \cdot X^{n-j} = \prod_{j=0}^{n} (X - T_j) \in K[X]$$

that was studied before. Therefore, p, just like f, is separable and irreducible, and admits $\mathfrak{S}_n$ as its Galois group. Furthermore, L is obtained as a splitting field of p over $k(S_1, \ldots, S_n)$. □

Similarly as we did with symmetric rational functions, one can study symmetric polynomials. For this we restrict the automorphisms of $k(T_1, \ldots, T_n)$ given by permutations of the variables to the subring $k[T_1, \ldots, T_n]$. Just as in the case of rational functions, a polynomial $f \in k[T_1, \ldots, T_n]$ is called *symmetric* if f is left fixed by all permutations $\pi \in \mathfrak{S}_n$. Clearly, the elementary symmetric polynomials $s_0, \ldots, s_n$ are examples of symmetric polynomials. As a generalization of Proposition 3, we will prove the *fundamental theorem on symmetric polynomials*, although at this place only for coefficients in a field k. See 4.4/1 for a more general version.

Proposition 5. *For every symmetric polynomial $f \in k[T_1, \ldots, T_n]$, there exists a unique polynomial $g \in k[S_1, \ldots, S_n]$ in n variables $S_1, \ldots, S_n$ such that $f = g(s_1, \ldots, s_n)$.*

Proof. The uniqueness assertion follows directly from the algebraic independence of the polynomials $s_1, \ldots, s_n$ over k, as established in Proposition 3.

To settle the existence part, consider the *lexicographic order* on $\mathbb{N}^n$, where we write $\nu < \nu'$ for two tuples $\nu = (\nu_1, \ldots, \nu_n)$, $\nu' = (\nu'_1, \ldots, \nu'_n) \in \mathbb{N}^n$ if there is an index $i_0 \in \{1, \ldots, n\}$ such that $\nu_{i_0} < \nu'_{i_0}$, as well as $\nu_i = \nu'_i$ for $i < i_0$. Then, for every nontrivial polynomial $f = \sum_{\nu \in \mathbb{N}^n} c_\nu T^\nu \in k[T_1, \ldots, T_n]$, the set $\{\nu \in \mathbb{N}^n \,;\, c_\nu \neq 0\}$ contains a lexicographically biggest element. Such an element is unique; it is called the *lexicographic degree* of f and is denoted by $\operatorname{lexdeg}(f)$. Now let $f = \sum_{\nu \in \mathbb{N}^n} c_\nu T^\nu$ be a symmetric polynomial with lexicographic degree $\operatorname{lexdeg}(f) = \mu = (\mu_1, \ldots, \mu_n)$. Then we have $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_n$, since f is symmetric. Furthermore,

$$f_1 = c_\mu s_1^{\mu_1 - \mu_2} s_2^{\mu_2 - \mu_3} \ldots s_n^{\mu_n} \in k[s_1, \ldots, s_n]$$

is a symmetric polynomial of total degree

$$(\mu_1 - \mu_2) + 2(\mu_2 - \mu_3) + 3(\mu_3 - \mu_4) + \ldots + n\mu_n = \sum_{i=1}^{n} \mu_i = |\mu|,$$

which, just like f, starts with $c_\mu T^\mu$ as lexicographically highest term. This implies

$$\operatorname{lexdeg}(f - f_1) < \operatorname{lexdeg}(f), \qquad \deg(f - f_1) \leq \deg(f).$$

If f is different from f_1, we can repeat this step once more, replacing f by $f - f_1$. Continuing in this way, we end up with a sequence of elements $f_1, f_2, \ldots \in k[s_1, \ldots, s_n]$ such that the lexicographic degree of the sequence

$$f,\ f - f_1,\ f - f_1 - f_2,\ \ldots$$

decreases step by step, while at the same time, the total degree is bounded by $\deg(f)$. Therefore, the sequence will end after finitely many steps with the zero polynomial, thereby yielding a representation of f as a polynomial in the elementary symmetric functions $s_1, \ldots, s_n$. □

The proof of Proposition 5 yields a very effective principle to determine for a given symmetric polynomial f a polynomial g satisfying $f = g(s_1, \ldots, s_n)$. Note that the principle works quite generally over an arbitrary ring R instead of k as coefficient domain. For some examples one may consult Section 6.2, where we have to write certain special symmetric polynomials that occur in solving algebraic equations of degree 3 and 4 as polynomial expressions in the elementary symmetric polynomials.

Finally, observe that the argument concerning the uniqueness part in the proof of Proposition 5 remains valid if we replace the coefficient field k by an integral domain R, for example by $R = \mathbb{Z}$. This is enough (granting the existence assertion) to define for monic polynomials their *discriminant* as a polynomial in the corresponding coefficients; see Section 4.4, in particular 4.4/3.

Exercises

1. *Show for every finite group G that there is a Galois extension L/K satisfying* $\mathrm{Gal}(L/K) \simeq G$.

2. Consider a subfield $L \subset \mathbb{C}$ such that $L/\mathbb{Q}$ is a cyclic Galois extension of degree 4. Show that $L/\mathbb{Q}$ admits a unique nontrivial intermediate field E, and that E is contained in $\mathbb{R}$.

3. Let K be a field of characteristic $\neq 2$ and $f \in K[X]$ a separable irreducible polynomial with zeros $\alpha_1, \ldots, \alpha_n$ in a splitting field L of f over K. Assume that the Galois group of f is cyclic of even order and show:
   (i) The discriminant $\Delta = \prod_{i<j}(\alpha_i - \alpha_j)^2$ does not admit a square root in K.
   (ii) There is a unique intermediate field E of L/K satisfying $[E : K] = 2$, namely $E = K(\sqrt{\Delta})$.

4. Let $\alpha, \beta \in \mathbb{C}$, $\alpha \neq \beta$, $\alpha \neq -\beta$, be zeros of the polynomial $(X^3-2)(X^2+3) \in \mathbb{Q}[X]$ and let $L = \mathbb{Q}(\alpha, \beta)$. Show that $L/\mathbb{Q}$ is a Galois extension, and determine its Galois group, as well as all intermediate fields of L/K.

5. Consider $L = \mathbb{Q}\left(\dfrac{1+i}{\sqrt{2}}\right)$ as a subfield of $\mathbb{C}$ and show that $L/\mathbb{Q}$ is a Galois extension. Determine the corresponding Galois group, as well as all intermediate fields of $L/\mathbb{Q}$.

6. Determine the Galois groups of the following polynomials in $\mathbb{Q}[X]$:
   (i) $X^3 + 6X^2 + 11X + 7$,
   (ii) $X^3 + 3X^2 - 1$,
   (iii) $X^4 + 2X^2 - 2$.
7. Determine a splitting field and the Galois group of the polynomial $X^4 - 5$ over $\mathbb{Q}$, resp. $\mathbb{Q}(i)$, as well as all intermediate fields of the arising extensions.
8. Determine the Galois groups of the following polynomials:
   (i) $X^4 - X^2 - 3 \in \mathbb{F}_5[X]$,
   (ii) $X^4 + 7X^2 - 3 \in \mathbb{F}_{13}[X]$.

4.4 Symmetric Polynomials, Discriminant, Resultant*

In the present section we extend the fundamental theorem on symmetric polynomials 4.3/5 to general coefficient domains, using an alternative method. As an application, we study the discriminant of a polynomial, characterizing it in terms of the resultant. The discriminant of a monic polynomial $f \in K[X]$ with coefficients in a field K vanishes if and only if f admits multiple zeros in an algebraic closure $\overline{K}$ of K; see Remark 3. Similarly, the vanishing of the resultant of two polynomials $f, g \in K[X]$ indicates that f and g have a common zero in $\overline{K}$; see Corollary 8.

In the following we use methods from linear algebra and work with modules over rings, as introduced in Section 2.9. Modules generalize the concept of vector spaces over fields to arbitrary rings as coefficient domains. However, in the present section it is enough to know for a ring extension $R \subset R'$, or more generally, a ring homomorphism $\varphi\colon R \longrightarrow R'$, that an additive subgroup $M \subset R'$ is called an *R-module* if $r \in R$, $x \in M$ always imply $\varphi(r)x \in M$, where for simplicity, we will write rx instead of $\varphi(r)x$. In particular, R' itself can be viewed as an R-module. A system $(x_i)_{i\in I}$ of elements in M is called a *free system of generators* of M if every $x \in M$ admits a representation $x = \sum_{i\in I} r_i x_i$ with coefficients $r_i \in R$ that are unique and, of course, vanish for almost all indices i. Free systems of generators of modules correspond to bases of vector spaces, although their existence is not always guaranteed in the module case.

Let us consider the polynomial ring $R[T_1, \ldots, T_n]$ in n variables T_i over a ring R. As in Section 4.3, we view the permutation group $\mathfrak{S}_n$ as a group of automorphisms of $R[T_1, \ldots, T_n]$, interpreting every $\pi \in \mathfrak{S}_n$ as the corresponding R-automorphism

$$\begin{aligned} R[T_1, \ldots, T_n] &\longrightarrow R[T_1, \ldots, T_n], \\ f(T_1, \ldots, T_n) &\longmapsto f(T_{\pi(1)}, \ldots, T_{\pi(n)}). \end{aligned}$$

A polynomial $f \in R[T_1, \ldots, T_n]$ is called *symmetric* if it is invariant under all elements $\pi \in \mathfrak{S}_n$. As special examples of symmetric polynomials, we know already the elementary symmetric polynomials

$$\begin{aligned} s_0 &= 1, \\ s_1 &= T_1 + \ldots + T_n, \\ s_2 &= T_1T_2 + T_1T_3 + \ldots + T_{n-1}T_n, \\ &\ldots \\ s_n &= T_1 \ldots T_n, \end{aligned}$$

which, using an additional variable X, are defined in terms of the equation

$$\prod_{i=1}^{n}(X - T_i) = \sum_{j=0}^{n}(-1)^j s_j X^{n-j}. \tag{1}$$

Clearly, the symmetric polynomials in $R[T_1, \ldots, T_n]$ form a subring containing R as well as all s_j.

Theorem 1 (Fundamental theorem on symmetric polynomials). *As before, consider the polynomial ring $R[T] = R[T_1, \ldots, T_n]$ in n variables over a ring R and let $s_1, \ldots, s_n$ be the corresponding elementary symmetric polynomials.*

(i) *The ring $R[s_1, \ldots, s_n]$ equals the subring of all symmetric polynomials in $R[T]$, i.e., every symmetric polynomial in $R[T]$ is a polynomial in the elementary symmetric polynomials $s_1, \ldots, s_n$.*

(ii) *The elements $s_1, \ldots, s_n \in R[T]$ are algebraically independent over R (in the sense of* 2.5/6).

(iii) *Let $N \subset \mathbb{N}^n$ be the set of all n-tuples $\nu = (\nu_1, \ldots, \nu_n)$, where $0 \leq \nu_i < i$ for $1 \leq i \leq n$. Then the system $(T^\nu)_{\nu \in N}$ is a free system of generators of $R[T]$ as a module over $R[s_1, \ldots, s_n]$.*

Proof. We conclude by induction on n. The case $n = 1$ is trivial, since in this case, $s_1 = T_1$ and every polynomial in $R[T_1]$ is symmetric. Therefore, assume $n > 1$, and let $s'_0, \ldots, s'_{n-1}$ be the elementary symmetric polynomials in $R[T_1, \ldots, T_{n-1}]$. Then one deduces from the equation

$$\sum_{j=0}^{n}(-1)^j s_j X^{n-j} = (X - T_n) \cdot \sum_{j=0}^{n-1}(-1)^j s'_j X^{n-1-j}$$

the relations

$$s_j = s'_j + s'_{j-1}T_n, \qquad 1 \leq j \leq n-1, \tag{2}$$

as well as $s'_0 = s_0 = 1$ and $s'_{n-1}T_n = s_n$. Using an inductive argument, this shows that $s'_1, \ldots, s'_{n-1}$ can be represented in a similar way as linear combinations of the $s_1, \ldots, s_{n-1}$ with coefficients in $R[T_n]$. Thereby we see that

$$R[s'_1, \ldots, s'_{n-1}, T_n] = R[s_1, \ldots, s_{n-1}, T_n]. \tag{3}$$

Moreover, we make the following claim:

(4) *The systems $s'_1, \dots, s'_{n-1}, T_n$, as well as $s_1, \dots, s_{n-1}, T_n$, are algebraically independent over R.*

To justify this claim we replace R by $R[T_n]$ and thereby may assume, due to the induction hypothesis, that $s'_1, \dots, s'_{n-1}$ are algebraically independent over $R[T_n]$, or what amounts to the same, that $s'_1, \dots, s'_{n-1}, T_n$ are algebraically independent over R. To obtain the corresponding fact for $s_1, \dots, s_{n-1}, T_n$, consider a nontrivial polynomial f in $n-1$ variables and with coefficients in $R[T_n]$ such that $f(s_1, \dots, s_{n-1})$ vanishes in $R[T_1, \dots, T_n]$. Since T_n is not a zero divisor in $R[T_1, \dots, T_n]$, we may assume that not all coefficients of f are divisible by T_n. Now apply the homomorphism $\tau\colon R[T_1, \dots, T_n] \longrightarrow R[T_1, \dots, T_{n-1}]$ substituting T_n by 0, which satisfies $\tau(s_j) = s'_j$ for $j = 1, \dots, n-1$, due to the relations (2). Using the fact that not all coefficients of f are divisible by T_n and hence are mapped to 0 by τ, we obtain from $f(s_1, \dots, s_{n-1}) = 0$ a nontrivial relation of type $g(s'_1, \dots, s'_{n-1}) = 0$ in $R[T_1, \dots, T_{n-1}]$. However, this contradicts the fact that $s'_1, \dots, s'_{n-1}$ are algebraically independent over R by the induction hypothesis. Therefore, our claim (4) is justified.

Let us turn now to the assertions of the fundamental theorem. To derive (i) consider a symmetric polynomial f in $R[T_1, \dots, T_n]$. Since all homogeneous parts of f are symmetric as well, we may assume f to be symmetric of a certain degree $m > 0$. Using that f is invariant under all permutations of the variables $T_1, \dots, T_{n-1}$, we can read from the induction hypothesis that it belongs to $R[s'_1, \dots, s'_{n-1}, T_n]$ and hence, by (3), to $R[s_1, \dots, s_{n-1}, T_n]$. Now write f as

$$f = \sum f_i T_n^i \tag{5}$$

with coefficients $f_i \in R[s_1, \dots, s_{n-1}]$. Then every f_i is a symmetric polynomial in $T_1, \dots, T_n$, which, as we claim, is homogeneous of degree $m - i$. To justify this, write the f_i explicitly as sums of terms of type $c s_1^{\nu_1} \dots s_{n-1}^{\nu_{n-1}}$. Viewing such a term as a polynomial in $T_1, \dots, T_n$, it is homogeneous of degree $\sum_{j=1}^{n-1} j\nu_j$, the so-called *weight* of the term. Furthermore, multiplication by T_n^i yields a homogeneous polynomial of degree $i + \sum_{j=1}^{n-1} j\nu_j$. Let us write f'_i for the sum of all terms $c s_1^{\nu_1} \dots s_{n-1}^{\nu_{n-1}}$ in f_i that are of weight $m-i$. Then we get $f = \sum f'_i T_n^i$, since f is homogeneous of degree m. However, $s_1, \dots, s_{n-1}, T_n$ are algebraically independent over R by (4), so that the representation (5) is unique. Therefore, we must have $f_i = f'_i$ for all i, and f_i, as a polynomial in $T_1, \dots, T_n$, is homogeneous of degree $m - i$.

In particular, the coefficient $f_0 \in R[s_1, \dots, s_{n-1}]$ is symmetric and homogeneous of degree m in $T_1, \dots, T_n$. If (5) reduces to $f = f_0$, we are done. Otherwise, consider the difference $f - f_0$, which is symmetric and homogeneous of degree m in $T_1, \dots, T_n$ as well, while T_n divides $f - f_0$ by construction. Then, by a symmetry argument, $s_n = T_1 \dots T_n$ divides $f - f_0$, and we can write

$$f = f_0 + g s_n, \tag{6}$$

where g is symmetric and homogeneous of some degree $< m$ in $T_1, \dots, T_n$. Finally, induction on m yields $f \in R[s_1, \dots, s_n]$, as asserted in (i).

Next, we need an auxiliary result, whose proof will be given further below:

Lemma 2. *Consider the polynomial ring $A[X]$ in a variable X over a ring A, and let $h = c_0X^n + c_1X^{n-1} + \ldots + c_n$ be a polynomial in $A[X]$ whose leading coefficient c_0 is a unit in A. Then every $f \in A[X]$ admits a representation $f = \sum_{i=0}^{n-1} f_i X^i$ with unique coefficients $f_i \in A[h]$. Furthermore, every f_i is of type $f_i = \sum_{j\geq 0} a_{ij}h^j$ with unique coefficients $a_{ij} \in A$.*

In particular, h is algebraically independent over A, and $X^0, X^1, \ldots, X^{n-1}$ form a free system of generators of $A[X]$ viewed as a module over $A[h]$.

To approach assertion (ii) of Theorem 1 we go back to the polynomial (1) above, which admits T_n as a zero. Substituting X by T_n, we get the equation

$$(-1)^{n+1}s_n = \sum_{j=0}^{n-1}(-1)^j s_j T_n^{n-j} = T_n^n - s_1T_n^{n-1} + \ldots + (-1)^{n-1}s_{n-1}T_n.$$

Now apply Lemma 2 for $A = R[s_1, \ldots, s_{n-1}]$, $X = T_n$, and $h = s_n$. We thereby see that s_n is algebraically independent over $R[s_1, \ldots, s_{n-1}]$ and hence that $s_1, \ldots, s_n$ are algebraically independent over R, since $s_1, \ldots, s_{n-1}$, are algebraically independent over R by (4). Thus, assertion (ii) is clear. Deducing assertion (iii) from the lemma is just as easy. Indeed, the system

$$\mathfrak{F}' = \big\{T_1^{\nu_1} \ldots T_{n-1}^{\nu_{n-1}}\,;\, 0 \leq \nu_i < i \;\text{ for }\; 1 \leq i \leq n-1\big\}$$

forms a free system of generators of $R[T_1, \ldots, T_n]$ over $R[s_1, \ldots, s_{n-1}, T_n]$, due to the induction hypothesis; to check this, use $R[T_n]$ as coefficient ring and apply (3). Furthermore, due to the lemma, $\mathfrak{F}'' = \{T_n^0, \ldots, T_n^{n-1}\}$ is a free system of generators of $R[s_1, \ldots, s_{n-1}, T_n]$ over $R[s_1, \ldots, s_n]$. But then, by a standard argument, $\mathfrak{F} = \{a'a''\,;\, a' \in \mathfrak{F}', a'' \in \mathfrak{F}''\}$ is a free system of generators of $R[T_1, \ldots, T_n]$ over $R[s_1, \ldots, s_n]$. □

It remains to supply the *proof of Lemma* 2. To do this, we have to show that every polynomial $f \in A[X]$ admits a representation

$$f = \sum_{i=0}^{n-1}\Big(\sum_{j\geq 0} a_{ij}h^j\Big)X^i = \sum_{j\geq 0}\Big(\sum_{i=0}^{n-1} a_{ij}X^i\Big)h^j$$

with unique coefficients $a_{ij} \in A$, or in other words, a representation

$$f = \sum_{j\geq 0} r_j h^j \tag{7}$$

with unique polynomials $r_j \in A[X]$ of degree $\deg r_j < n$. To achieve this we use Euclidean division by h, which exists in $A[X]$, since the leading coefficient of h is a unit; cf. 2.1/4. Hence, we can consider the sequence of decompositions

$$f = f_1h + r_0, \qquad f_1 = f_2h + r_1, \qquad f_2 = f_3h + r_2, \qquad \ldots,$$

where $r_0, r_1, \ldots \in A[X]$ are polynomials of degree $< n$. The degree of the f_j strictly decreases, until we reach a point where $\deg f_j < n = \deg h$ and hence $f_j = r_j$. Going backward and putting everything together, we arrive at a decomposition as stated in (7). To verify the uniqueness, consider an equation $0 = \sum_{j\geq 0} r_j h^j$ for polynomials $r_0, r_1, \ldots \in A[X]$ of degree $< n$. Using the uniqueness of Euclidean division, the decomposition

$$0 = r_0 + h \cdot \sum_{j>0} r_j h^{j-1}$$

yields $r_0 = 0$, as well as $\sum_{j>0} r_j h^{j-1} = 0$. By induction we can conclude that $r_j = 0$ for all j. □

The proof of Theorem 1 suggests another practical principle for representing symmetric polynomials in terms of elementary symmetric polynomials, a procedure that looks a bit more complicated than the one given in the proof of 4.3/5. Substituting T_n by 0 in the equation $f = f_0(s_1, \ldots, s_{n-1}) + g s_n$, see (6), we get

$$f(T_1, \ldots, T_{n-1}, 0) = f_0(s'_1, \ldots, s'_{n-1})$$

using (2). This means that the problem of representing f as a polynomial in the elementary symmetric polynomials $s_1, \ldots, s_n$ is reduced to the following subproblems:

(a) Consider the symmetric polynomial $f(T_1, \ldots, T_{n-1}, 0)$ in $n-1$ variables and write it as a polynomial $f_0(s'_1, \ldots, s'_{n-1})$ in the elementary symmetric polynomials $s'_1, \ldots, s'_{n-1}$ in $T_1, \ldots, T_{n-1}$ with coefficients in R.

(b) Replace $s'_1, \ldots, s'_{n-1}$ in f_0 by the corresponding elementary symmetric polynomials $s_1, \ldots, s_{n-1}$ in $T_1, \ldots, T_n$, divide the difference $f - f_0(s_1, \ldots, s_{n-1})$ by s_n, and write $s_n^{-1} \cdot (f - f_0(s_1, \ldots, s_{n-1}))$ as a polynomial in the elementary symmetric polynomials $s_1, \ldots, s_n$.

Step (a) allows one to reduce the number of variables, whereas (b) will reduce the degree of the symmetric polynomial under consideration. Thus, after finitely many steps of type (a) or (b) we end up with the desired representation of f.

To give a first application of Theorem 1, we can prove again the assertion of 4.3/3 and show that every symmetric rational function in n variables $T_1, \ldots, T_n$ with coefficients in a field k can be written as a rational function in $s_1, \ldots, s_n$ with coefficients in k. Indeed, consider a symmetric rational function $q \in k(T_1, \ldots, T_n)$, say $q = f/g$ with polynomials $f, g \in k[T_1, \ldots, T_n]$. Multiplying f and g by the product $\prod_{\pi \in \mathfrak{S}_n - \{\mathrm{id}\}} \pi(g)$, we may assume that g is symmetric. But then $f = q \cdot g$ is symmetric as well. In particular, q is a quotient of symmetric polynomials, and hence by Theorem 1 (i), a rational function in $s_1, \ldots, s_n$. Also note that the free system of generators of Theorem 1 (iii) gives rise to a basis of $k(T_1, \ldots, T_n)$ over $k(s_1, \ldots, s_n)$.

To give another application of the fundamental theorem on symmetric polynomials, we want to discuss the *discriminant* of a monic polynomial. Working over the coefficient domain $R = \mathbb{Z}$, consider

$$\prod_{i<j}(T_i - T_j)^2$$

as a symmetric polynomial in $\mathbb{Z}[T_1, \ldots, T_n]$. The latter is a polynomial Δ in the elementary symmetric polynomials $s_1, \ldots, s_n$, as we can read from Theorem 1. One calls $\Delta = \Delta(s_1, \ldots, s_n)$ the discriminant of the polynomial

$$\prod_{i=1}^{n}(X - T_i) = \sum_{j=0}^{n}(-1)^j s_j X^{n-j},$$

where X is viewed as a variable over the ring $\mathbb{Z}[T_1, \ldots, T_n]$. Now, to define the discriminant of any monic polynomial $f = X^n + c_1 X^{n-1} + \ldots + c_n \in R[X]$ over any coefficient ring R, set

$$\Delta_f = \Delta\big(-c_1, c_2, \ldots, (-1)^n c_n\big).$$

In this way, Δ_f equals the image of Δ under the ring homomorphism

$$\varphi\colon \mathbb{Z}[s_1, \ldots, s_n] \longrightarrow R, \qquad s_j \longmapsto (-1)^j c_j,$$

which extends the canonical homomorphism $\mathbb{Z} \longrightarrow R$. Observe when setting up φ that the elementary symmetric polynomials $s_1, \ldots, s_n$ are algebraically independent over $\mathbb{Z}$ and thus $\mathbb{Z}[s_1, \ldots, s_n]$ can be viewed as the polynomial ring in the n "variables" $s_1, \ldots, s_n$. Therefore, existence and uniqueness of φ follow from 2.5/5. Furthermore, note that the definition of the discriminant Δ_f is meaningful also for $n = 0$, i.e., for the monic polynomial $f = 1$. Indeed, we get $\Delta_f = 1$ in this case, since empty products assume the value 1 by convention. All in all, we see that the discriminant Δ_f of a monic polynomial f is a polynomial expression in the coefficients of f and as such is an element of the coefficient ring R.

Remark 3. *Let R be a ring and $f = X^n + c_1 X^{n-1} + \ldots + c_n \in R[X]$ a monic polynomial with discriminant Δ_f. If $f = \prod_{i=1}^{n}(X - \alpha_i)$ is a factorization over a ring R' extending R, then*

$$\Delta_f = \prod_{i<j}(\alpha_i - \alpha_j)^2.$$

Proof. Assuming there is a factorization of f as stated, consider the ring homomorphism

$$\varphi\colon \mathbb{Z}[T_1, \ldots, T_n] \longrightarrow R', \qquad T_i \longmapsto \alpha_i,$$

which extends the canonical homomorphism $\mathbb{Z} \longrightarrow R$. It maps the polynomial $F = \prod_{i=1}^{n}(X - T_i)$ to the polynomial $F^\varphi = f$, so that $\varphi(s_j) = (-1)^j c_j$ and hence $\varphi(\Delta_F) = \Delta_f$. Therefore, we get

$$\Delta_f = \varphi\Big(\prod_{i<j}(T_i - T_j)^2\Big) = \prod_{i<j}(\alpha_i - \alpha_j)^2,$$

as desired. □

In particular, we thereby see for a monic polynomial $f \in K[X]$ over a field K that its discriminant Δ_f vanishes if and only if f has multiple zeros in an algebraic closure $\overline{K}$ of K. This, in turn, is equivalent to the fact that f and its derivative f' have common zeros in $\overline{K}$, see 2.6/2, a fact that can be checked by looking at the resultant $\operatorname{res}(f, f')$. More specifically, we want to show in the following that the discriminant Δ_f coincides with the resultant $\operatorname{res}(f, f')$, up to sign. As we will see, such a characterization is of particular interest for explicit computations of discriminants.

Now to introduce the resultant $\operatorname{res}(f, g)$ of two polynomials f, g, consider polynomials

$$f = a_0X^m + a_1X^{m-1} + \ldots + a_m, \qquad g = b_0X^n + b_1X^{n-1} + \ldots + b_n$$

in a variable X and with coefficients in a ring R. Since we do not require a_0 or b_0 to be nonzero, we call m, resp. n, the *formal degree* of f, resp. g, as well as the pair (m, n) the formal degree of (f, g). Then the *resultant* $\operatorname{res}(f, g)$ of formal degree (m, n) is defined as the determinant of the following matrix consisting of $m + n$ rows and columns:

$$(*) \qquad \begin{array}{c} \uparrow \\ \\ n \\ \\ \downarrow \\ \uparrow \\ \\ m \\ \\ \downarrow \end{array} \begin{pmatrix} a_0 & a_1 & . & . & . & a_m & & & & \\ & a_0 & a_1 & . & . & . & a_m & & & \\ & & \multicolumn{6}{c}{\dotfill} & & \\ & & & a_0 & a_1 & . & . & . & a_m & \\ & & & & a_0 & a_1 & . & . & . & a_m \\ b_0 & b_1 & . & . & . & b_n & & & & \\ & b_0 & b_1 & . & . & . & b_n & & & \\ & & \multicolumn{6}{c}{\dotfill} & & \\ & & & b_0 & b_1 & . & . & . & b_n & \\ & & & & b_0 & b_1 & . & . & . & b_n \end{pmatrix},$$

where we have to enter zeros at free places. Note that the trivial case $m = n = 0$ is not excluded, since the determinant of the empty matrix is 1 by convention. In cases in which the formal degree of f, resp. g, is not mentioned explicitly, we assume for $f \neq 0 \neq g$ that it is given by the degree of f, resp. g. Applying the common rules for handling determinants,[3] we can see immediately that the following must hold:

Remark 4. *Consider polynomials $f, g \in R[X]$ in a variable X over a ring R, as before. Then*:

(i) $\operatorname{res}(f, g) = (-1)^{m \cdot n} \operatorname{res}(g, f)$.

(ii) $\operatorname{res}(af, bg) = a^n b^m \operatorname{res}(f, g)$ *for constants $a, b \in R$.*

[3] Formal rules for computations in terms of determinants remain valid if instead of coefficients in a field, we admit coefficients in a more general ring R. Indeed, in a first step, we can view the necessary coefficients, say $c_1, \ldots, c_r$, as variables and observe that the corresponding rule holds over the rational function field $\mathbb{Q}(c_1, \ldots, c_r)$ or its subring $\mathbb{Z}[c_1, \ldots, c_r]$. The rule can then be transported to arbitrary rings R by considering suitable ring homomorphisms $\mathbb{Z}[c_1, \ldots, c_r] \longrightarrow R$.

(iii) *For a ring homomorphism* $\varphi\colon R \longrightarrow R'$ *and the polynomials* f^φ, g^φ *transported via* φ *to* $R'[X]$, *we get* $\operatorname{res}(f^\varphi, g^\varphi) = \varphi(\operatorname{res}(f,g))$.

Let S be the transpose of the matrix $(*)$. We want to realize S as an R-linear map and accordingly interpret $\operatorname{res}(f,g)$ as the determinant of this map. To do so, let $R[X]_i$ for $i \in \mathbb{N}$ be the R-module of all polynomials in $R[X]$ of degree $< i$.

Lemma 5. *Let* $f, g \in R[X]$ *be polynomials as before.*

(i) *Fix* $X^{i-1}, \ldots, X^0$ *as a free system of generators of* $R[X]_i$ *over* R *for each* i. *Then the transpose* S *of the matrix* $(*)$ *corresponds to the* R*-linear map*

$$\Phi\colon R[X]_n \times R[X]_m \longrightarrow R[X]_{m+n}, \qquad (u,v) \longmapsto uf + vg.$$

(ii) *If* f *is monic of degree* m, *then along with* $\mathfrak{F} = (X^{m+n-1}, \ldots, X^0)$, *also* $\mathfrak{F}' = (fX^{n-1}, \ldots, fX^0, X^{m-1}, \ldots, X^0)$ *is a free system of generators of* $R[X]_{m+n}$ *over* R. *The corresponding change of basis matrix from* $\mathfrak{F}$ *to* $\mathfrak{F}'$ *admits determinant* 1.

(iii) *If* f *is monic of degree* m, *then* $\operatorname{res}(f,g)$ *equals the determinant of the* R*-endomorphism* $\Phi'\colon R[X]_{m+n} \longrightarrow R[X]_{m+n}$ *that is given by multiplying* $X^{m-1}, \ldots, X^0$ *by* g *and otherwise leaving* $fX^{n-1}, \ldots, fX^0$ *fixed.*

Proof. Assertion (i) is immediately clear, since the coefficient vectors of the Φ-images of

$$(X^{n-1}, 0), \ldots, (X^0, 0), (0, X^{m-1}), \ldots, (0, X^0)$$

coincide with the rows of the matrix $(*)$ and hence with the columns of the matrix S.

Turning to assertion (ii), represent the members of $\mathfrak{F}'$ as linear combinations of the free system of generators $\mathfrak{F}$. The corresponding coefficient matrix is a lower triangular matrix, whose diagonal entries are all equal to 1; use the fact that f is monic. In particular, this matrix has determinant 1 and hence is invertible. Therefore, $\mathfrak{F}'$ is a free system of generators of $R[X]_{m+n}$, and the basis change matrix from $\mathfrak{F}$ to $\mathfrak{F}'$ admits determinant equal to 1.

In order to verify (iii), check that the map Φ' is described by the matrix S if we fix on the source $\mathfrak{F}'$ and on the target $\mathfrak{F}$ as free systems of generators. To compute the determinant of Φ' in terms of the determinant of a corresponding matrix, we have to consider, on both the source and the target, the same free system of generators. For example, we can change from $\mathfrak{F}'$ to $\mathfrak{F}$ on the source. However, using (ii), the corresponding basis change matrix is of determinant 1, so that we get $\det \Phi' = \det S = \operatorname{res}(f,g)$. □

Let us draw some conclusions from the lemma.

Proposition 6. *Let* $f, g \in R[X]$ *be polynomials as before and assume that* $m + n \geq 1$. *Then there are polynomials* $p, q \in R[X]$, $\deg p < n$, $\deg q < m$, *such that* $\operatorname{res}(f,g) = pf + qg$.

Proof. We consider the map Φ of Lemma 5 and claim that $\operatorname{res}(f,g)$, viewed as a constant polynomial in $R[X]$, belongs to the image of Φ. To justify this, we use Cramer's rule

$$S \cdot S^* = (\det S) \cdot E,$$

where S^* is the adjoint matrix of S, and E the unit matrix with $m+n$ rows and columns; see, for example, [4], Satz 4.4/3, in conjunction with the generalization we have given in the proof of 3.3/1. Translated to R-linear maps, the above means: there exists an R-linear map $\Phi^*\colon R[X]_{m+n} \longrightarrow R[X]_m \times R[X]_n$ whose composition with the map $\Phi\colon R[X]_m \times R[X]_n \longrightarrow R[X]_{m+n}$ yields the map $(\det S)\cdot \mathrm{id}$. In particular, the constant polynomial $\Phi \circ \Phi^*(1) = \det S = \operatorname{res}(f,g)$ belongs to the image of Φ, which justifies our claim. □

In particular, we can read from Proposition 6 that $\operatorname{res}(f,g)$ vanishes if f and g have a common zero. However, the resultant can vanish in additional cases, for example, when the leading coefficients a_0, b_0 of f and g are both zero. Next we want to prove a characterization that serves as a key ingredient for the further study of resultants.

Proposition 7. *Let $f \in R[X]$ be a monic polynomial of degree m. View the residue class ring $A = R[X]/(f)$ as an R-module under the canonical map $R \longrightarrow R[X]/(f)$ and write x for the residue class of X. Then the powers $x^{m-1}, \ldots, x^0$ form a free system of generators of A as an R-module. Furthermore, let $g \in R[X]$ be a polynomial of degree $\leq n$ and $g(x)$ its residue class in A. Then the resultant of formal degree (m,n) satisfies*

$$\operatorname{res}(f,g) = \mathrm{N}_{A/R}\big(g(x)\big),$$

where $\mathrm{N}_{A/R}(g(x))$ is the norm *of $g(x)$, i.e., the determinant of the R-linear map $A \longrightarrow A$, $a \longmapsto g(x)\cdot a$.*

In particular, $\operatorname{res}(f,g)$ is independent of the choice of the formal degree n of g.

Proof. Since f is monic, we can use Euclidean division by f in $R[X]$, which is unique; see 2.1/4. Therefore, the projection $R[X] \longrightarrow A$ induces an isomorphism of R-modules $R[X]_m \xrightarrow{\sim} A$, and it follows that the elements $x^{m-1}, \ldots, x^0$, being the images of $X^{m-1}, \ldots, X^0$, form a free system of generators of A over R. This settles the first assertion of the proposition. To derive the second, consider the map Φ' of Lemma 5 (iii). The projection $R[X]_{m+n} \longrightarrow A$ admits $fR[X]_n$ as its kernel. Since Φ' maps this kernel into itself, it follows that Φ' induces an R-linear map $\overline{\Phi}'\colon A \longrightarrow A$, which clearly is multiplication by $g(x)$. Since Φ' restricts on $fR[X]_n$ to the identity map, we get $\det \Phi' = \det \overline{\Phi}'$, i.e., we have $\operatorname{res}(f,g) = \mathrm{N}_{A/R}(g(x))$, due to Lemma 5 (iii). □

Corollary 8. *Let f, g be nontrivial polynomials with coefficients in a field K. Furthermore, let $\deg f = m$ and $\deg g \leq n$. Then the following conditions are equivalent*:

(i) *The resultant* $\operatorname{res}(f,g)$ *of formal degree* (m,n) *is nonzero.*
(ii) *The polynomials* f *and* g *do not admit common zeros in an algebraic closure* $\overline{K}$ *of* K.

Proof. Following Remark 4, we may assume f to be monic. If $\operatorname{res}(f,g) \neq 0$, we see from Proposition 7 that the determinant of the multiplication by $g(x)$ is nonzero on $K[X]/(f)$ and hence invertible. Then the multiplication by $g(x)$ is invertible itself, which implies that $g(x)$ is a unit in $K[X]/(f)$. Therefore, f and g generate the unit ideal in $K[X]$, and it is impossible that these polynomials have a common zero in an extension field of K. Conversely, assume that f and g are without common zeros in $\overline{K}$. Then f and g are prime to each other in $\overline{K}[X]$ and therefore also in $K[X]$. Hence, there is an equation of type $uf+vg=1$ for certain polynomials $u,v \in K[X]$, and we see that $g(x)$ is a unit in $K[X]/(f)$. In particular, the multiplication by $g(x)$ is invertible on $K[X]/(f)$, and the corresponding determinant is nonzero. But this implies $\operatorname{res}(f,g) \neq 0$, due to Proposition 7. □

Corollary 9. *Let*

$$f = \alpha \prod_{i=1}^{m} (X-\alpha_i), \qquad g = \beta \prod_{j=1}^{n} (X-\beta_j),$$

be factorizations of two polynomials $f,g \in R[X]$ *with constants* $\alpha,\beta \in R$, *as well as zeros* $\alpha_1,\dots,\alpha_m$, $\beta_1,\dots,\beta_n$ *belonging to a ring* R' *extending* R. *Then the resultant of formal degree* (m,n) *satisfies*

$$\operatorname{res}(f,g) = \alpha^n \prod_{i=1}^{m} g(\alpha_i) = \alpha^n \beta^m \prod_{\substack{1\le i\le m\\ 1\le j\le n}} (\alpha_i - \beta_j).$$

Proof. Making use of Remark 4, we may assume $R = R'$, as well as $\alpha = 1 = \beta$, and hence that f and g are monic polynomials admitting a factorization into linear factors in $R[X]$. It is easily checked that $\operatorname{res}(X-\alpha_i, g) = g(\alpha_i)$, either by going back to the definition of the resultant or by applying Proposition 7. Moreover, we conclude from the multiplicativity of the norm of elements in $R[X]/(f)$ over R, in other words from the multiplicativity of the determinant, that any two polynomials $g_1, g_2 \in R[X]$ satisfy the relation

$$\operatorname{res}(f, g_1 g_2) = \operatorname{res}(f,g_1) \cdot \operatorname{res}(f,g_2).$$

Then, applying Remark 4 (i), we can derive for polynomials f_1, f_2 in $R[X]$ of degree $\deg f_i \le m_i$ the relation

$$\operatorname{res}(f_1 f_2, g) = \operatorname{res}(f_1,g) \cdot \operatorname{res}(f_2,g),$$

where the resultants are meant with respect to the formal degrees (m_1+m_2,n), resp. (m_1,n) and (m_2,n). Applying the above equation repeatedly to the factorization of f yields the desired assertion. □

Now let us characterize the discriminant in terms of the resultant, as was already announced before.

Corollary 10. *Let $f \in R[X]$ be a monic polynomial of degree $m > 0$ and f' its derivative. Then the discriminant Δ_f is related to the resultant* $\operatorname{res}(f, f')$ *of formal degree $(m, m-1)$ by*

$$\Delta_f = (-1)^{m(m-1)/2} \operatorname{res}(f, f').$$

Alternatively, let $A = R[X]/(f)$ and write $x \in A$ for the residue class of the variable $X \in R[X]$. Then

$$\Delta_f = (-1)^{m(m-1)/2} \mathrm{N}_{A/R}\big(f'(x)\big).$$

Proof. The second equation follows from the first one by means of Proposition 7. To derive the first equation, we may replace R by a suitable extension ring; use the definition of the discriminant in conjunction with Remark 4 (iii). Thereby we may assume that f decomposes over R into a product of linear factors. Indeed, using the idea of Kronecker's construction 3.4/1, we can replace R in a first step by $R' = R[X]/(f)$. Then f admits at least one zero in R', namely the residue class $\overline{x}$ of X. Dividing out the linear factor $X - \overline{x}$ from f, we obtain a monic polynomial of degree $m-1$, which can be treated in the same way again, so that after finitely many steps, we arrive at a ring extending R over which f decomposes completely into a product of linear factors.

Therefore assume $f = \prod_{i=1}^{m}(X - \alpha_i)$, so that

$$\operatorname{res}(f, f') = \prod_{i=1}^{m} f'(\alpha_i)$$

by Corollary 9. Furthermore, the product rule for derivatives yields

$$f' = \sum_{i=1}^{m} (X - \alpha_1) \ldots (X - \alpha_{i-1})(X - \alpha_{i+1}) \ldots (X - \alpha_m)$$

and hence

$$f'(\alpha_i) = (\alpha_i - \alpha_1) \ldots (\alpha_i - \alpha_{i-1})(\alpha_i - \alpha_{i+1}) \ldots (\alpha_i - \alpha_m).$$

However, this means that

$$\operatorname{res}(f, f') = \prod_{i \neq j} (\alpha_i - \alpha_j) = (-1)^{m(m-1)/2} \prod_{i<j} (\alpha_i - \alpha_j)^2 = (-1)^{m(m-1)/2} \Delta_f,$$

as claimed. □

Observe that Corollary 10 provides in particular an explicit formula for determining discriminants. Let us look at a simple example and compute the

discriminant of the polynomial $f = X^3 + aX + b \in R[X]$, which was needed in Section 4.3 for the case of a field $K = R$ in order to study the Galois group of f. Following Corollary 10, we get

$$\Delta_f = -\operatorname{res}(f, f') = -\det \begin{pmatrix} 1 & 0 & a & b & 0 \\ 0 & 1 & 0 & a & b \\ 3 & 0 & a & 0 & 0 \\ 0 & 3 & 0 & a & 0 \\ 0 & 0 & 3 & 0 & a \end{pmatrix}.$$

To simplify the computation of the determinant, one should subtract 3 times row 1 from row 3, as well as 3 times row 2 from row 4. Thereby we get

$$\Delta_f = -\det \begin{pmatrix} -2a & -3b & 0 \\ 0 & -2a & -3b \\ 3 & 0 & a \end{pmatrix} = -4a^3 - 27b^2.$$

Following the same procedure, one shows for $m \geq 2$ that the discriminant of the polynomial $f = X^m + aX + b$ is given by

$$\Delta_f = (-1)^{m(m-1)/2}\big((1-m)^{m-1}a^m + m^m b^{m-1}\big).$$

Also the case $f = X^4 + aX^2 + bX + c$ is quite simple. We obtain

$$\Delta_f = 144ab^2c - 128a^2c^2 - 4a^3b^2 + 16a^4c - 27b^4 + 256c^3,$$

where similarly as in the preceding examples, one reduces the computation of the corresponding seven-row determinant to a four-row determinant, using elementary row transformations.

Finally, let us point out that discriminants are usually considered in a somewhat more general setup. Let $R \subset A$ be a ring extension such that A, as an R-module, admits a finite free system of generators $e_1, \ldots, e_n$. Then, just as for vector spaces, one can define the *trace* of R-linear maps $\varphi\colon A \longrightarrow A$. Indeed, if φ is described by the matrix $(a_{ij}) \in R^{n\times n}$, with respect to a free system of generators of A over R, then one defines the trace of φ by $\operatorname{tr}(\varphi) = \sum_{i=1}^n a_{ii} \in R$. In particular, one can consider for $a \in A$ the R-linear map

$$\varphi_a\colon A \longrightarrow A, \qquad x \longmapsto ax,$$

given by multiplication by a. Its trace is denoted by $\operatorname{tr}_{A/R}(a)$ and, more specifically, is called the *trace* of the element $a \in A$ with respect to the extension A/R; see also Section 4.7.

Now, if $x_1, \ldots, x_n$ is a system of elements in A, one looks at the matrix $(\operatorname{tr}_{A/R}(x_i x_j))_{i,j=1,\ldots,n} \in R^{n\times n}$ and defines the *discriminant* of $x_1, \ldots, x_n$ with respect to the ring extension $R \subset A$ by

$$D_{A/R}(x_1, \ldots, x_n) = \det\big(\operatorname{tr}_{A/R}(x_i x_j)\big)_{i,j=1,\ldots,n}.$$

To relate the discriminant of a polynomial Δ_f, as defined before, to the discriminant of a system of elements $D_{A/R}(x_1, \ldots, x_n)$, we prove the following result:

Proposition 11. *Let $f \in R[X]$ be a monic polynomial of degree $n > 0$. As in Proposition 7, consider $A = R[X]/(f)$ as an R-module with the free system of generators $x^0, \ldots, x^{n-1}$, where $x \in A$ is the residue class of $X \in R[X]$. Then*

$$\Delta_f = D_{A/R}(x^0, \ldots, x^{n-1}).$$

Proof. We start with a reduction step showing that the asserted equation needs to be established only for certain rings R and polynomials f. Consider a ring homomorphism $\tau\colon \tilde{R} \longrightarrow R$ together with a monic polynomial $\tilde{f} \in \tilde{R}[X]$ of degree n that is transported via τ to f. Then τ gives rise to a ring homomorphism

$$\tau'\colon \tilde{A} = \tilde{R}[X]/(\tilde{f}) \longrightarrow R[X]/(f) = A.$$

Writing $\tilde{x}$ for the residue class of X in $\tilde{R}[X]/(\tilde{f})$, we can view $\tilde{x}^0, \ldots, \tilde{x}^{n-1}$ as a free system of generators of $\tilde{A}$ as an $\tilde{R}$-module, just as $x^0, \ldots, x^{n-1}$ form a free system of generators of A as an R-module. Furthermore, we have $\tau'(\tilde{x}^i) = x^i$ for $i = 0, \ldots, n-1$. We want to relate the multiplication by an element $\tilde{a} \in \tilde{A}$, i.e., the $\tilde{R}$-linear map $\varphi_{\tilde{a}}\colon \tilde{A} \longrightarrow \tilde{A}$, to the corresponding R-linear map $\varphi_a\colon A \longrightarrow A$, where $a = \tau'(\tilde{a})$. Since τ' is a ring homomorphism, we see that the matrix M_a describing φ_a (with respect to the free system of generators $x^0, \ldots, x^{n-1}$) is obtained from the corresponding matrix $M_{\tilde{a}}$ of $\varphi_{\tilde{a}}$ by transporting it via τ from $\tilde{R}^{n\times n}$ to $R^{n\times n}$. We therefore get $\operatorname{tr}_{A/R}(a) = \tau(\operatorname{tr}_{\tilde{A}/\tilde{R}}(\tilde{a}))$, and in particular,

$$D_{A/R}(x^0, \ldots, x^{n-1}) = \tau\big(D_{\tilde{A}/\tilde{R}}(\tilde{x}^0, \ldots, \tilde{x}^{n-1})\big).$$

In addition, the definition of discriminants shows that

$$\Delta_f = \tau(\Delta_{\tilde{f}}).$$

As a conclusion, if the assertion of the proposition is known for $\tilde{R}$ and $\tilde{f}$, it is known for R and f as well.

Now consider the diagram of ring and field extensions

$$\begin{array}{ccc} \mathbb{Z}[T_1, \ldots, T_n] & \subset & \mathbb{Q}(T_1, \ldots, T_n) \\ \bigcup & & \bigcup \\ \mathbb{Z}[s_1, \ldots, s_n] & \subset & \mathbb{Q}(s_1, \ldots, s_n), \end{array}$$

where $s_1, \ldots, s_n$ are the elementary symmetric polynomials in $T_1, \ldots, T_n$. Furthermore, consider a monic polynomial $f = X^n + c_1X^{n-1} + \ldots + c_n \in R[X]$, for an arbitrary ring R, as required in the assumption of the proposition. Then we can define a ring homomorphism $\mathbb{Z}[s_1, \ldots, s_n] \longrightarrow R$ by $s_j \longmapsto (-1)^j c_j$. The reduction step discussed before shows that the assertion of the proposition needs to be proved only for $R = \mathbb{Z}[s_1, \ldots, s_n]$ and $f = \sum_{j=0}^{n} (-1)^j s_j X^{n-j}$. In

addition, since $\mathbb{Z}[s_1,\dots,s_n]$ is a subring of $\mathbb{Q}(s_1,\dots,s_n)$, we may even assume $R=\mathbb{Q}(s_1,\dots,s_n)$. Then, as a polynomial in $\mathbb{Q}(s_1,\dots,s_n)[X]$, we know from 4.3 that f is irreducible. Therefore, all in all, it is enough to consider the case that $R=K$ is a field and f is a monic irreducible polynomial in $K[X]$. Also note that $L=A=K[X]/(f)$ is then a field that is finite over K. In the following, we will consider such a situation.

By its definition, the discriminant of $x^0,\dots,x^{n-1}$ is given by

$$D_{L/K}(x^0,\dots,x^{n-1})=\det\big(\mathrm{tr}_{L/K}(x^{i+j})\big)_{i,j=0,\dots,n-1}.$$

In order to compute the trace $\mathrm{tr}_{L/K}(x^{i+j})$, we decompose f over an algebraic closure of L into linear factors, say $f=\prod_{k=1}^{n}(X-\alpha_k)$, and anticipate a result to be proved later in Section 4.7, namely 4.7/4, from which we can read

$$\mathrm{tr}_{L/K}(x^{i+j})=\sum_{k=1}^{n}\alpha_k^{i+j}.$$

Now consider Vandermonde's matrix $V=(\alpha_{i+1}^{j})_{i,j=0,\dots,n-1}$ that is attached to $\alpha_1,\dots,\alpha_n$. Since $\det V=\prod_{i<j}(\alpha_j-\alpha_i)$ and

$$\big(\mathrm{tr}_{L/K}(x^{i+j})\big)_{i,j=0,\dots,n-1}=V^t\cdot V,$$

it follows that

$$D_{L/K}(x^0,\dots,x^{n-1})=(\det V)^2=\prod_{i<j}(\alpha_i-\alpha_j)^2=\Delta_f.$$

□

Exercises

1. *Explain the role played by the fundamental theorem on symmetric polynomials in defining the discriminant Δ_f of a polynomial $f=X^n+a_1X^{n-1}+\ldots+a_n$. Why should one consider symmetric polynomials with coefficients in rings, even if the discriminant is needed only for polynomials over fields?*

2. Look at the polynomial ring $R[T_1,T_2,T_3]$ over a ring R and write the symmetric polynomial $T_1^3+T_2^3+T_3^3$ as a polynomial in the corresponding elementary symmetric polynomials.

3. For a ring R, verify the following formulas for discriminants Δ_f of polynomials $f\in R[X]$:
   (i) $f=X^2+aX+b$,
   $\Delta_f=a^2-4b$.
   (ii) $f=X^m+aX+b$ for $m\geq 2$,
   $\Delta_f=(-1)^{m(m-1)/2}((1-m)^{m-1}a^m+m^mb^{m-1})$.
   (iii) $f=X^3+aX^2+bX+c$,
   $\Delta_f=a^2b^2+18abc-4a^3c-4b^3-27c^2$.

(iv) $f = X^4 + aX^2 + bX + c$,
$\Delta_f = 144ab^2c - 128a^2c^2 - 4a^3b^2 + 16a^4c - 27b^4 + 256c^3$.

4. Let R be a ring. For two polynomials $f, g \in R[X]$ of formal degree m, n, determine the corresponding resultant $\operatorname{res}(f, g)$ in the following cases:
   (i) $g = g_0 \in R$ (constant polynomial), $m = 0$.
   (ii) $g = g_0 \in R$ (constant polynomial), $m = 1$.
   (iii) $f = a_0X + a_1$, $g = b_0X + b_1$, $m = n = 1$.

5. Show for polynomials over a ring R that

$$\operatorname{res}(a_0X^2 + a_1X + a_2, b_0X^2 + b_1X + b_2) = (a_0b_2 - a_2b_0)^2 + (a_1b_2 - a_2b_1)(a_1b_0 - a_0b_1).$$

6. Show for monic polynomials $f, g \in R[X]$ over a ring R that

$$\Delta_{fg} = \Delta_f \cdot \Delta_g \cdot \operatorname{res}(f, g)^2.$$

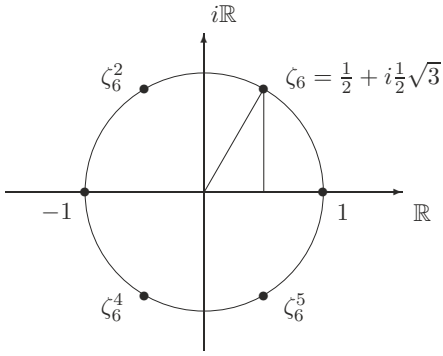
7. Let R be a ring and $f \in R[X]$ a monic polynomial. Show for $c \in R$ that the discriminants of the polynomials f and $g = f(X + c)$ coincide.

4.5 Roots of Unity

Let us fix a field K and an algebraic closure $\overline{K}$ of it. The zeros of the polynomial $X^n - 1$, where $n \in \mathbb{N} - \{0\}$, are called the nth *roots of unity* (in $\overline{K}$). These roots form a subgroup $U_n \subset \overline{K}^*$. If $\operatorname{char} K = 0$, or more generally, if $\operatorname{char} K$ does not divide n, then $X^n - 1$ and its derivative $D(X^n - 1) = nX^{n-1}$ do not admit common zeros. This shows that $X^n - 1$ is separable in this case, so that $\operatorname{ord} U_n = n$. Otherwise, for $p = \operatorname{char} K > 0$, consider a factorization $n = p^r n'$ where $p \nmid n'$. Then the polynomial $X^{n'} - 1$ is separable, as explained before, and its zeros coincide with those of $X^n - 1$, since $X^n - 1 = (X^{n'} - 1)^{p^r}$. Hence, we get $U_n = U_{n'}$ and therefore $\operatorname{ord} U_n = n'$. In particular, for the study of the groups U_n we may restrict ourselves to the case $\operatorname{char} K \nmid n$. Our main objective in the following is to study field extensions that occur by adjoining roots of unity. To begin with, let us reformulate the result 3.6/14:

Proposition 1. *Let $n \in \mathbb{N} - \{0\}$ be an integer such that $\operatorname{char} K \nmid n$. Then the group U_n of nth roots of unity in $\overline{K}$ is cyclic of order n.*

A root of unity $\zeta \in U_n$ is called a *primitive nth root of unity* if it generates the group U_n. For example, the fourth roots of unity in $\mathbb{C}$ are given by $1, i, -1, -i$, while i and $-i$ are the primitive ones among them. All roots of unity in $\mathbb{C}$ are of absolute value 1 and hence are situated on the unit circle $\{z \in \mathbb{C}\,;\, |z| = 1\}$. The nth roots of unity in $\mathbb{C}$ can easily be described in terms of the complex exponential function. They consist of all powers of the primitive nth root of unity $\zeta_n = e^{2\pi i/n}$, as is illustrated for $n = 6$ by the following figure:



In view of the complex case, we can say that the roots of unity split the unit circle line into equal parts. Extension fields of $\mathbb{Q}$ constructed by adjoining a root of unity are referred to as *cyclotomic fields*, from Greek "kúklos" (circle) and "tomé" (a cutting).

Remark 2. *Let $m, n \in \mathbb{N} - \{0\}$ be prime to each other. Then the map*

$$h\colon U_m \times U_n \longrightarrow U_{mn}, \qquad (\zeta, \eta) \longmapsto \zeta\eta,$$

*is an isomorphism of groups. If $\zeta_m \in U_m$ is a primitive m*th *root of unity and $\zeta_n \in U_n$ a primitive n*th *root of unity, then $\zeta_m\zeta_n$ is a primitive mn*th *root of unity.*

Proof. We may assume that $\operatorname{char} K$ does not divide mn. Since U_m and U_n are subgroups of U_{mn}, we see that h is a well-defined homomorphism of commutative groups. Furthermore, U_m and U_n are cyclic by Proposition 1, of orders m and n. Now if $\zeta_m \in U_m$ is a primitive mth root of unity and $\zeta_n \in U_n$ a primitive nth root of unity, we see from 3.6/13 that $(\zeta_m, \zeta_n) \in U_m \times U_n$, as a product of $(\zeta_m, 1)$ and $(1, \zeta_n)$, is of order mn and thus generates the group $U_m \times U_n$. In the same way we conclude from 3.6/13 that $\zeta_m\zeta_n \in U_{mn}$ is of order mn and thus is a primitive mnth root of unity generating U_{mn}. In particular, we thereby see that h is an isomorphism. □

Next we want to look more closely at the primitive ones among the nth roots of unity. To determine their number we need to introduce a function that is referred to as Euler's φ-function.

Definition 3. *For $n \in \mathbb{N} - \{0\}$ let $\varphi(n)$ denote the order of the multiplicative group $(\mathbb{Z}/n\mathbb{Z})^*$, i.e., of the group of units of the residue class ring $\mathbb{Z}/n\mathbb{Z}$. The map $\varphi\colon \mathbb{N} - \{0\} \longrightarrow \mathbb{N}$ is called* Euler's φ-function.

Remark 4. (i) *Let $n \in \mathbb{N} - \{0\}$. Then*

$$(\mathbb{Z}/n\mathbb{Z})^* = \left\{\overline{a} \in \mathbb{Z}/n\mathbb{Z}\,;\, a \in \mathbb{Z} \text{ such that } \gcd(a, n) = 1\right\}$$

and hence

$$\varphi(n) = \#\big\{a \in \mathbb{N}\,;\, 0 \le a < n \text{ such that } \gcd(a,n) = 1\big\}.$$

(ii) *If $m, n \in \mathbb{N} - \{0\}$ are prime to each other, then $\varphi(mn) = \varphi(m)\varphi(n)$; this property is referred to as the* multiplicativity of the φ-function.

(iii) *Consider a prime factorization $n = p_1^{\nu_1} \dots p_r^{\nu_r}$ with exponents $\nu_\rho > 0$, as well as distinct prime numbers p_ρ. Then*

$$\varphi(n) = \prod_{\rho=1}^{r} p_\rho^{\nu_\rho - 1}(p_\rho - 1).$$

Proof. For elements $a \in \mathbb{Z}$ and $n \in \mathbb{N} - \{0\}$, we have $\gcd(a,n) = 1$ if and only if $a\mathbb{Z} + n\mathbb{Z} = \mathbb{Z}$, i.e., if and only if there are elements $c, d \in \mathbb{Z}$ such that $ac + nd = 1$; cf. 2.4/13. This is equivalent to the fact that the residue class of a is a unit in $\mathbb{Z}/n\mathbb{Z}$, which verifies assertion (i).

To establish (ii), use the Chinese remainder theorem in the version 2.4/14. It provides for numbers $m, n \in \mathbb{N} - \{0\}$ that are prime to each other, an isomorphism of rings

$$\mathbb{Z}/mn\mathbb{Z} \xrightarrow{\sim} \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z},$$

and thereby an isomorphism between the corresponding groups of units

$$(\mathbb{Z}/mn\mathbb{Z})^* \xrightarrow{\sim} (\mathbb{Z}/m\mathbb{Z})^* \times (\mathbb{Z}/n\mathbb{Z})^*,$$

which implies the multiplicativity of the φ-function.

To verify (iii) we may apply (ii). Then it is only necessary to determine $\varphi(p^\nu)$ for a prime number p and a positive exponent ν. Since the products $0 \cdot p, 1 \cdot p, \dots, (p^{\nu-1} - 1) \cdot p$ represent precisely those natural numbers d such that $0 \le d < p^\nu$ and d is divisible by p, we conclude from (i) that

$$\varphi(p^\nu) = p^\nu - p^{\nu-1} = p^{\nu-1}(p-1),$$

as desired. □

Proposition 5. *Let $n \in \mathbb{N}$. An element $\overline{a}$ generates the additive cyclic group $\mathbb{Z}/n\mathbb{Z}$ if and only if $\overline{a}$ is a unit in the residue class ring $\mathbb{Z}/n\mathbb{Z}$. In particular, if $n \neq 0$, then $\mathbb{Z}/n\mathbb{Z}$ contains precisely $\varphi(n)$ elements generating $\mathbb{Z}/n\mathbb{Z}$ as a cyclic group.*

Proof. Clearly, $\mathbb{Z}/n\mathbb{Z}$ is generated by $\overline{a}$ if and only if the residue class $\overline{1}$ of $1 \in \mathbb{Z}$ is contained in the cyclic subgroup generated by $\overline{a}$. This is the case if and only if there is some $r \in \mathbb{Z}$ satisfying $\overline{1} = r \cdot \overline{a} = \overline{r} \cdot \overline{a}$, i.e., if and only if $\overline{a}$ is a unit in $\mathbb{Z}/n\mathbb{Z}$. □

Corollary 6. *Let K be a field and $n \in \mathbb{N} - \{0\}$ an integer such that* $\operatorname{char} K \nmid n$. *Then the group U_n of* nth *roots of unity contains precisely $\varphi(n)$ elements that*

are primitive. If $\zeta \in U_n$ is such a primitive nth root of unity, then ζ^r for some integer $r \in \mathbb{Z}$ is a primitive nth root of unity as well if and only if the residue class of r modulo n is a unit in $\mathbb{Z}/n\mathbb{Z}$, i.e., if and only if $\gcd(r, n) = 1$.

Proof. Following Proposition 1, the group U_n is isomorphic to $\mathbb{Z}/n\mathbb{Z}$. Therefore, U_n contains precisely $\varphi(n)$ primitive nth roots of unity, as we can read from Proposition 5. Furthermore, for a primitive nth root of unity $\zeta \in U_n$, the homomorphism $\mathbb{Z}/n\mathbb{Z} \longrightarrow U_n$ mapping the residue class $\overline{1} \in \mathbb{Z}/n\mathbb{Z}$ to ζ is an isomorphism. Hence, a power ζ^r for some $r \in \mathbb{Z}$ is a primitive nth root of unity if and only if its preimage in $\mathbb{Z}/n\mathbb{Z}$, namely the residue class $\overline{r}$, generates the group $\mathbb{Z}/n\mathbb{Z}$. Due to Proposition 5, the latter is the case if and only if $\overline{r}$ is a unit. □

If $\zeta_n \in \overline{K}$ is a primitive nth root of unity, the field $K(\zeta_n)$ contains all nth roots of unity. Hence, it is a splitting field of the polynomial $X^n - 1 \in K[X]$ and is therefore *normal* over K. In addition, since $X^n - 1$ is separable for $\operatorname{char} K \nmid n$, we recognize $K(\zeta_n)/K$ as a finite Galois extension. For $K = \mathbb{Q}$, the field $\mathbb{Q}(\zeta_n)$ is called the nth *cyclotomic field*.

Proposition 7. *Let K be a field and $\zeta_n \in \overline{K}$ a primitive nth root of unity, where $\operatorname{char} K \nmid n$. Then:*

(i) *$K(\zeta_n)/K$ is a finite abelian Galois extension of a degree dividing $\varphi(n)$.*

(ii) *Every $\sigma \in \operatorname{Gal}(K(\zeta_n)/K)$ induces by restriction an automorphism of the group U_n of nth roots of unity, and the map*

$$\psi\colon \operatorname{Gal}\big(K(\zeta_n)/K\big) \longrightarrow \operatorname{Aut}(U_n), \qquad \sigma \longmapsto \sigma|_{U_n},$$

is a monomorphism of groups.

(iii) *The automorphism group $\operatorname{Aut}(U_n)$ is canonically isomorphic to the multiplicative group $(\mathbb{Z}/n\mathbb{Z})^*$, namely via the map*

$$(\mathbb{Z}/n\mathbb{Z})^* \longrightarrow \operatorname{Aut}(U_n), \qquad \overline{r} \longmapsto (\zeta \longmapsto \zeta^r),$$

which is well defined.

Proof. We know already that $K(\zeta_n)/K$ is a finite Galois extension. Using assertions (ii) and (iii), we may view the Galois group $\operatorname{Gal}(K(\zeta_n)/K)$ as a subgroup of $(\mathbb{Z}/n\mathbb{Z})^*$. In particular, $\operatorname{Gal}(K(\zeta_n)/K)$ is abelian of some order dividing $\operatorname{ord}(\mathbb{Z}/n\mathbb{Z})^* = \varphi(n)$. Therefore, assertion (i) is a consequence of (ii) and (iii).

To approach (ii) observe that every Galois automorphism $\sigma \in \operatorname{Gal}(K(\zeta_n)/K)$ maps zeros of $X^n - 1$ to such zeros again. Since in addition, σ is multiplicative and injective, we recognize $\sigma|_{U_n}$ as an automorphism of the finite group U_n. Thus, $\sigma \longmapsto \sigma|_{U_n}$ defines a group homomorphism $\operatorname{Gal}(K(\zeta_n)/K) \longrightarrow \operatorname{Aut}(U_n)$, which is injective, since every automorphism $\sigma \in \operatorname{Gal}(K(\zeta_n)/K)$ leaving $\zeta_n \in U_n$ fixed must coincide with the identity map.

Finally, to establish assertion (iii), we use the fact that U_n is cyclic of order n and hence is isomorphic to $\mathbb{Z}/n\mathbb{Z}$. Then, writing $\overline{r}$ and $\overline{a}$ for the residue classes

in $\mathbb{Z}/n\mathbb{Z}$ of elements $r, a \in \mathbb{Z}$, we have to show that the map

$$\psi' \colon (\mathbb{Z}/n\mathbb{Z})^* \longrightarrow \mathrm{Aut}(\mathbb{Z}/n\mathbb{Z}), \qquad \overline{r} \longmapsto (\overline{a} \longmapsto r \cdot \overline{a}),$$

is a well-defined group isomorphism. First, observe that ψ' is well defined, since we have $r \cdot \overline{a} = \overline{ra} = \overline{r} \cdot \overline{a}$. Furthermore, it is clear that multiplication by a unit $\overline{r} \in (\mathbb{Z}/n\mathbb{Z})^*$ yields an automorphism of the additive group $\mathbb{Z}/n\mathbb{Z}$ and that ψ' is injective. To show that ψ' is surjective as well, consider any automorphism ρ of the additive group $\mathbb{Z}/n\mathbb{Z}$ and let $\overline{r} = \rho(\overline{1})$. Then, for $a \in \mathbb{Z}$, we can write

$$\rho(\overline{a}) = \rho(\overline{1} \cdot a) = \rho(\overline{1}) \cdot a = \overline{r} \cdot \overline{a},$$

and it turns out that ρ coincides with multiplication by $\overline{r}$. Since such a multiplication can be surjective only for units $\overline{r} \in \mathbb{Z}/n\mathbb{Z}$, we get $\overline{r} \in (\mathbb{Z}/n\mathbb{Z})^*$ and hence $\rho = \psi'(\overline{r})$, so that ψ' is seen to be surjective. □

Proposition 8. *Let $\zeta_n \in \overline{\mathbb{Q}}$ be a primitive nth root of unity. The nth cyclotomic field $\mathbb{Q}(\zeta_n)$ is a finite Galois extension of $\mathbb{Q}$ of degree $\big[\mathbb{Q}(\zeta_n) : \mathbb{Q}\big] = \varphi(n)$ with Galois group*

$$\mathrm{Gal}\big(\mathbb{Q}(\zeta_n)/\mathbb{Q}\big) \xrightarrow{\sim} \mathrm{Aut}(U_n) \simeq (\mathbb{Z}/n\mathbb{Z})^*.$$

Proof. Since $\big[\mathbb{Q}(\zeta_n) : \mathbb{Q}\big] = \deg f$, where $f \in \mathbb{Q}[X]$ is the minimal polynomial of ζ_n over $\mathbb{Q}$, it is enough to show that $\deg f = \varphi(n)$. The characterization of the Galois group $\mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q})$ is then a consequence of Proposition 7. Also note that

$$\big[\mathbb{Q}(\zeta_n) : \mathbb{Q}\big] = \deg f \leq \varphi(n),$$

again by Proposition 7.

We claim that every primitive nth root of unity $\zeta \in U_n$ is a zero of f, which then implies $\deg f = \varphi(n)$ by means of Corollary 6. To justify this claim observe that f, being the minimal polynomial of ζ_n, will divide $X^n - 1$. Hence, there is a polynomial $h \in \mathbb{Q}[X]$ satisfying

$$X^n - 1 = f \cdot h.$$

Since f is monic by definition, the same is true for h, and we have $f, h \in \mathbb{Z}[X]$, due to 2.7/6.

Now let p be a prime number such that $p \nmid n$. Then ζ_n^p is a primitive nth root of unity by Corollary 6, and we want to show that it is a zero of f. If such is not the case, i.e., if $f(\zeta_n^p) \neq 0$, we must have $h(\zeta_n^p) = 0$. In other words, ζ_n is then a zero of $h(X^p)$. This, in turn, implies $f | h(X^p)$, say $h(X^p) = f \cdot g$, where just as before, g is a monic polynomial in $\mathbb{Z}[X]$ by 2.7/6. Now, reducing coefficients modulo p, we consider the homomorphism

$$\mathbb{Z}[X] \longrightarrow \mathbb{Z}/p\mathbb{Z}[X] = \mathbb{F}_p[X], \qquad \sum c_i X^i \longmapsto \sum \overline{c}_i X^i,$$

extending the canonical projection $\mathbb{Z} \longrightarrow \mathbb{Z}/p\mathbb{Z}$. Then $\overline{h}^p = \overline{h}(X^p) = \overline{f} \cdot \overline{g}$ shows that $\overline{h}$ and $\overline{f}$ cannot be prime to each other in $\mathbb{F}_p[X]$. Therefore, the

polynomial $X^n - 1 = \overline{f} \cdot \overline{h} \in \mathbb{F}_p[X]$ admits multiple zeros in an algebraic closure of $\mathbb{F}_p$, which, however, is in contradiction to $p \nmid n$. Consequently, the assumption $f(\zeta_n^p) \neq 0$ above is untenable, and ζ_n^p must be a zero of f.

Now we can show for an arbitrary primitive nth root of unity ζ that it is a zero of f. Assume $\zeta = \zeta_n^m$, where $\gcd(m, n) = 1$, due to Corollary 6. Then we can obtain ζ from ζ_n by taking successive powers with prime exponents not dividing n. Therefore, a repeated application of the argument explained above shows that $f(\zeta) = 0$. □

For relatively prime integers $m, n \in \mathbb{N} - \{0\}$ and primitive mth, nth, and mnth roots of unity $\zeta_m, \zeta_n, \zeta_{mn} \in \overline{\mathbb{Q}}$, the decomposition

$$(\mathbb{Z}/mn\mathbb{Z})^* \xrightarrow{\sim} (\mathbb{Z}/m\mathbb{Z})^* \times (\mathbb{Z}/n\mathbb{Z})^*$$

discussed in the proof of Remark 4 (ii), together with the assertion of Proposition 8, yields a decomposition of Galois groups

$$\mathrm{Gal}\big(\mathbb{Q}(\zeta_{mn})/\mathbb{Q}\big) \xrightarrow{\sim} \mathrm{Gal}\big(\mathbb{Q}(\zeta_m)/\mathbb{Q}\big) \times \mathrm{Gal}\big(\mathbb{Q}(\zeta_m)/\mathbb{Q}\big),$$

which we want to study in more detail.

Corollary 9. *Let $\zeta_m, \zeta_n \in \overline{\mathbb{Q}}$ be primitive m*th *and n*th *roots of unity, where* $\gcd(m, n) = 1$. *Then*

$$\mathbb{Q}(\zeta_m) \cap \mathbb{Q}(\zeta_n) = \mathbb{Q},$$

and the map

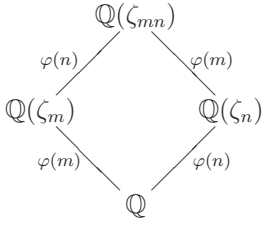
$$\begin{aligned} \mathrm{Gal}\big(\mathbb{Q}(\zeta_m, \zeta_n)/\mathbb{Q}\big) &\longrightarrow \mathrm{Gal}\big(\mathbb{Q}(\zeta_m)/\mathbb{Q}\big) \times \mathrm{Gal}\big(\mathbb{Q}(\zeta_n)/\mathbb{Q}\big), \\ \sigma &\longmapsto (\sigma|_{\mathbb{Q}(\zeta_m)}, \sigma|_{\mathbb{Q}(\zeta_n)}), \end{aligned}$$

is an isomorphism.

Proof. We know from Remark 2 that $\zeta_{mn} = \zeta_m \zeta_n$ is a primitive mnth root of unity. Therefore, the composite field of $\mathbb{Q}(\zeta_m)$ and $\mathbb{Q}(\zeta_n)$ in $\overline{\mathbb{Q}}$ is given by

$$\mathbb{Q}(\zeta_m) \cdot \mathbb{Q}(\zeta_n) = \mathbb{Q}(\zeta_{mn}).$$

There is the following diagram of field extensions,



where the degrees of $\mathbb{Q}(\zeta_{mn})$, $\mathbb{Q}(\zeta_m)$, and $\mathbb{Q}(\zeta_n)$ over $\mathbb{Q}$ are given by $\varphi(mn)$, $\varphi(m)$, and $\varphi(n)$, due to Proposition 8. Furthermore, $\varphi(mn) = \varphi(m) \cdot \varphi(n)$ implies

$$\big[\mathbb{Q}(\zeta_{mn}) : \mathbb{Q}(\zeta_m)\big] = \varphi(n), \qquad \big[\mathbb{Q}(\zeta_{mn}) : \mathbb{Q}(\zeta_n)\big] = \varphi(m).$$

Now let $L = \mathbb{Q}(\zeta_m) \cap \mathbb{Q}(\zeta_n)$. Since we have $\mathbb{Q}(\zeta_{mn}) = \mathbb{Q}(\zeta_m, \zeta_n)$, it becomes clear that ζ_m is of degree $\varphi(m)$ over $\mathbb{Q}(\zeta_n)$ and hence of degree $\geq \varphi(m)$ over L. Then the estimate

$$\varphi(m) = \big[\mathbb{Q}(\zeta_m) : \mathbb{Q}\big] = \big[\mathbb{Q}(\zeta_m) : L\big] \cdot \big[L : \mathbb{Q}\big] \geq \varphi(m) \cdot \big[L : \mathbb{Q}\big]$$

implies $\big[L : \mathbb{Q}\big] = 1$ and thus $L = \mathbb{Q}$. In particular, applying 4.1/12 (ii), we see that the map specified in the assertion provides an isomorphism between $\mathrm{Gal}(\mathbb{Q}(\zeta_{mn})/\mathbb{Q})$ and the Cartesian product of the Galois groups $\mathrm{Gal}(\mathbb{Q}(\zeta_m)/\mathbb{Q})$ and $\mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q})$. □

Next let us factorize the polynomial $X^n - 1$, which admits the nth roots of unity as its zeros, into so-called cyclotomic polynomials.

Definition 10. *Let K be a field. For $n \in \mathbb{N}-\{0\}$ and $\mathrm{char}\, K \nmid n$, let $\zeta_1, \ldots, \zeta_{\varphi(n)}$ be the primitive nth roots of unity in $\overline{K}$. Then*

$$\Phi_n = \prod_{i=1}^{\varphi(n)} (X - \zeta_i)$$

is called the nth cyclotomic polynomial *over K.*

Proposition 11. (i) *Φ_n is a monic separable polynomial in $K[X]$ satisfying $\deg \Phi_n = \varphi(n)$.*

(ii) *If $K = \mathbb{Q}$, we have $\Phi_n \in \mathbb{Z}[X]$, and Φ_n is irreducible in $\mathbb{Z}[X]$, as well as in $\mathbb{Q}[X]$.*

(iii) $X^n - 1 = \displaystyle\prod_{d|n, d>0} \Phi_d.$

(iv) *If $\Phi_n \in \mathbb{Z}[X]$ is the nth cyclotomic polynomial over $\mathbb{Q}$, then the nth cyclotomic polynomial over any other field K where $\mathrm{char}\, K \nmid n$ is obtained from Φ_n by applying the canonical homomorphism $\mathbb{Z} \longrightarrow K$ to the coefficients of Φ_n.*

Proof. Concerning (i), we have only to show that Φ_n, which does not have multiple zeros and hence is separable, admits coefficients in K. To justify this, observe that $L = K(\zeta_1) = K(\zeta_1, \ldots, \zeta_{\varphi(n)})$ is a finite Galois extension of K by Proposition 7, and that we have $\Phi_n \in L[X]$ by definition. Since every Galois automorphism $\sigma \in \mathrm{Gal}(L/K)$ induces a self-map and hence a permutation on the set of primitive nth roots of unity, we conclude that Φ_n is invariant under $\mathrm{Gal}(L/K)$. But then $\Phi_n \in K[X]$, due to 4.1/5(ii).

Now assume $K = \mathbb{Q}$. Since every primitive nth root of unity ζ is of degree $\varphi(n)$ over $\mathbb{Q}$ and since $\Phi_n \in \mathbb{Q}[X]$ is a monic polynomial of degree $\varphi(n)$ satisfying $\Phi_n(\zeta) = 0$, we see that Φ_n must coincide with the minimal polynomial

of ζ over $\mathbb{Q}$; this polynomial is irreducible. Furthermore, Φ_n is a monic polynomial dividing $(X^n - 1)$, and we get $\Phi_n \in \mathbb{Z}[X]$ using 2.7/6. Of course, Φ_n is irreducible in $\mathbb{Z}[X]$ as well, see 2.7/7, so that assertion (ii) is clear.

Next, the formula in (iii) is obtained by combining factors of the decomposition

$$X^n - 1 = \prod_{\zeta \in U_n} (X - \zeta).$$

Indeed, let P_d for $d \in \mathbb{N} - \{0\}$ be the set of primitive dth roots of unity in $\overline{K}$. Then U_n is the disjoint union of all P_d, where $d \,|\, n$, $d > 0$. Therefore, we get

$$X^n - 1 = \prod_{d|n, d>0} \prod_{\zeta \in P_d} (X - \zeta) = \prod_{d|n, d>0} \Phi_d.$$

Finally, to derive assertion (iv), we write Φ_n for the nth cyclotomic polynomial over $\mathbb{Q}$, as well as $\tilde{\Phi}_n$ for the corresponding polynomial over any other field K under consideration. We have to show that the canonical homomorphism $\tau \colon \mathbb{Z}[X] \longrightarrow K[X]$ maps Φ_n to $\tilde{\Phi}_n$, i.e., that $\tau(\Phi_n) = \tilde{\Phi}_n$, which we shall prove by induction on n. The case $n = 1$ is obvious, since

$$\tau(\Phi_1) = X - 1 = \tilde{\Phi}_1.$$

Therefore, assume $n > 1$. Then we have over $\mathbb{Z}$ the equation

$$X^n - 1 = \Phi_n \cdot \prod_{d|n, 0<d<n} \Phi_d,$$

as well as over K the equation

$$X^n - 1 = \tilde{\Phi}_n \cdot \prod_{d|n, 0<d<n} \tilde{\Phi}_d.$$

Using the induction hypothesis in conjunction with the fact that $K[X]$ is an integral domain, we get $\tau(\Phi_n) = \tilde{\Phi}_n$, as desired. □

Departing from $\Phi_1 = X - 1$, the formula of Proposition 11 (iii) can be used to compute cyclotomic polynomials in a recursive way. For $K = \mathbb{Q}$, the formula provides, in fact, the prime factorization of $X^n - 1$ in $\mathbb{Z}[X]$ or in $\mathbb{Q}[X]$, since then the factors Φ_d are irreducible by (ii). Let us give some examples. For a prime number p, we get

$$X^p - 1 = \Phi_1 \cdot \Phi_p$$

and hence

$$\Phi_p = X^{p-1} + X^{p-2} + \ldots + 1.$$

Furthermore, if p and q are distinct prime numbers, then

$$X^{pq} - 1 = \Phi_1 \cdot \Phi_p \cdot \Phi_q \cdot \Phi_{pq}$$

and therefore

$$\Phi_{pq} = \frac{X^{pq}-1}{(X-1)\cdot(X^{p-1}+\ldots+1)\cdot(X^{q-1}+\ldots+1)}.$$

For instance, we get

$$\Phi_6 = \frac{X^6-1}{(X-1)\cdot(X+1)\cdot(X^2+X+1)} = X^2-X+1.$$

Let us list the first 12 cyclotomic polynomials explicitly:

$$\begin{aligned}
\Phi_1 &= X-1\\
\Phi_2 &= X+1\\
\Phi_3 &= X^2+X+1\\
\Phi_4 &= X^2+1\\
\Phi_5 &= X^4+X^3+X^2+X+1\\
\Phi_6 &= X^2-X+1\\
\Phi_7 &= X^6+X^5+X^4+X^3+X^2+X+1\\
\Phi_8 &= X^4+1\\
\Phi_9 &= X^6+X^3+1\\
\Phi_{10} &= X^4-X^3+X^2-X+1\\
\Phi_{11} &= X^{10}+X^9+X^8+X^7+X^6+X^5+X^4+X^3+X^2+X+1\\
\Phi_{12} &= X^4-X^2+1
\end{aligned}$$

Looking at the these examples, one might guess that 1 and -1 are the only nonzero coefficients that can occur in cyclotomic polynomials. However, that is not the case. In fact,

$$\begin{aligned}
\Phi_{105} = {} & X^{48}+X^{47}+X^{46}-X^{43}-X^{42}-2X^{41}-X^{40}-X^{39}\\
& +X^{36}+X^{35}+X^{34}+X^{33}+X^{32}+X^{31}-X^{28}-X^{26}-X^{24}-X^{22}-X^{20}\\
& +X^{17}+X^{16}+X^{15}+X^{14}+X^{13}+X^{12}-X^9-X^8-2X^7-X^6-X^5\\
& +X^2+X+1
\end{aligned}$$

is the first cyclotomic polynomial whose coefficients are not all of absolute value ≤ 1. The next three polynomials of this type are Φ_{165}, Φ_{195}, and Φ_{210}, where, just as for Φ_{105}, all their coefficients are of absolute value ≤ 2. However, it is known from a result of I. Schur that the coefficients of cyclotomic polynomials are unbounded. Indeed, for $n = p_1\cdot\ldots\cdot p_m$ with prime numbers $p_1 < \ldots < p_m$ such that $p_m < p_1 + p_2$, it turns out that the coefficient of X^{p_m} in Φ_n is $1-m$. Furthermore, one can show using arguments from number theory that for m odd, there always exist prime numbers $p_1, \ldots, p_m$ satisfying the preceding conditions.

Finally, let us have a special look at finite fields, considering the field $\mathbb{F}_q$ of q elements for a prime power q. Recall from 3.8/6 that the Galois group of a finite extension $\mathbb{F}/\mathbb{F}_q$ is cyclic of order $[\mathbb{F}:\mathbb{F}_q]$, generated by the relative Frobenius homomorphism $\mathbb{F} \longrightarrow \mathbb{F}$, $a \longmapsto a^q$.

Proposition 12. *Let $\zeta \in \overline{\mathbb{F}}_q$ be a primitive nth root of unity and assume that $\gcd(n, q) = 1$, where q is a prime power.*

(i) *Look at the injection $\psi\colon \mathrm{Gal}(\mathbb{F}_q(\zeta)/\mathbb{F}_q) \hookrightarrow \mathrm{Aut}(U_n)$ of Proposition 7 (ii) and use the identification $\mathrm{Aut}(U_n) \simeq (\mathbb{Z}/n\mathbb{Z})^*$ of Proposition 7 (iii). The relative Frobenius homomorphism of $\mathbb{F}_q(\zeta)/\mathbb{F}_q$ is mapped under ψ to the residue class $\overline{q} \in (\mathbb{Z}/n\mathbb{Z})^*$ attached to q. In particular, ψ gives rise to an isomorphism between $\mathrm{Gal}(\mathbb{F}_q(\zeta)/\mathbb{F}_q)$ and the subgroup $\langle \overline{q} \rangle \subset (\mathbb{Z}/n\mathbb{Z})^*$.*

(ii) *The degree $[\mathbb{F}_q(\zeta) : \mathbb{F}_q]$ coincides with the order of $\overline{q}$ in $(\mathbb{Z}/n\mathbb{Z})^*$.*

(iii) *The nth cyclotomic polynomial Φ_n is irreducible in $\mathbb{F}_q[X]$ if and only if $\overline{q}$ generates the group $(\mathbb{Z}/n\mathbb{Z})^*$.*

Proof. The relative Frobenius homomorphism over $\mathbb{F}_q$ is given on U_n by the map $\zeta \longrightarrow \zeta^q$ and therefore, using the canonical isomorphism $\mathrm{Aut}(U_n) \simeq (\mathbb{Z}/n\mathbb{Z})^*$ of Proposition 7 (iii), corresponds to the residue class $\overline{q} \in (\mathbb{Z}/n\mathbb{Z})^*$. Furthermore, assertion (ii) is a consequence of (i), due to

$$\big[\mathbb{F}_q(\zeta) : \mathbb{F}_q\big] = \mathrm{ord}\,\mathrm{Gal}\big(\mathbb{F}_q(\zeta)/\mathbb{F}_q\big) = \mathrm{ord}\,\langle \overline{q} \rangle = \mathrm{ord}\,\overline{q}.$$

Finally, to verify (iii) observe that Φ_n is irreducible if and only if the equation $[\mathbb{F}_q(\zeta) : \mathbb{F}_q] = \deg \Phi_n = \varphi(n)$ holds. By (ii), that is equivalent to the condition that $\overline{q}$ generates the group $(\mathbb{Z}/n\mathbb{Z})^*$. □

Therefore, we see that a necessary condition for the nth cyclotomic polynomial Φ_n to be irreducible is that the group $(\mathbb{Z}/n\mathbb{Z})^*$ be cyclic. For example, $(\mathbb{Z}/n\mathbb{Z})^*$ is cyclic for a prime number $n = p$, see 3.6/14, or more generally, also for a prime power $n = p^r$ of a prime number $p \neq 2$; cf. Exercise 7 below.

Exercises

1. *Consider a primitive nth root of unity ζ, as well as the nth cyclotomic polynomial $\Phi_n \in K[X]$ over a field K, where $\mathrm{char}\, K \nmid n$. Show that Φ_n decomposes over K into $\varphi(n)/s$ distinct irreducible factors of degree $s = [K(\zeta) : K]$.*

2. *Let $\zeta_m \in \overline{\mathbb{Q}}$ be a primitive mth root of unity. Determine all integers n such that the nth cyclotomic polynomial Φ_n is irreducible over $\mathbb{Q}(\zeta_m)$.*

3. Show: $\varphi(n) = n \cdot \prod_{p|n,\ p \text{ prime}} (1 - p^{-1})$.

4. Determine the Galois group of the polynomial $X^5 - 1 \in \mathbb{F}_7[X]$.

5. Let ζ be a primitive 12th root of unity over $\mathbb{Q}$. Determine all intermediate fields of $\mathbb{Q}(\zeta)/\mathbb{Q}$.

6. Let p be a prime number such that $p - 1 = \prod_{\nu=1}^{n} p_\nu$ factorizes into distinct prime factors p_ν, and consider a primitive pth root of unity $\zeta_p \in \overline{\mathbb{Q}}$. Show that $\mathbb{Q}(\zeta_p)/\mathbb{Q}$ is a cyclic Galois extension and that $\mathbb{Q}(\zeta_p)/\mathbb{Q}$ admits precisely 2^n intermediate fields.

7. Let p be an odd prime number. Show that the group $(\mathbb{Z}/p^r\mathbb{Z})^*$ is cyclic for $r > 0$ and conclude from this that the p^rth cyclotomic field $\mathbb{Q}(\zeta_{p^r})$ is a cyclic Galois extension of $\mathbb{Q}$. *Hint*: Consider the canonical homomorphism $(\mathbb{Z}/p^r\mathbb{Z})^* \longrightarrow (\mathbb{Z}/p\mathbb{Z})^*$ together with its kernel W and show by induction that the residue class of $1+p$ is an element of order p^{r-1} in W, hence in particular, that W is cyclic.
8. Verify the following formulas for cyclotomic polynomials Φ_n:
   (i) $\Phi_{p^r}(X) = \Phi_p(X^{p^{r-1}})$, for p prime, $r > 0$.
   (ii) $\Phi_n(X) = \Phi_{p_1 \dots p_s}(X^{p_1^{r_1-1} \dots p_s^{r_s-1}})$, for a prime factorization $n = p_1^{r_1} \dots p_s^{r_s}$ with distinct prime factors p_ν and exponents $r_\nu > 0$.
   (iii) $\Phi_{2n}(X) = \Phi_n(-X)$, for $n \geq 3$ odd.
   (iv) $\Phi_{pn}(X) = \dfrac{\Phi_n(X^p)}{\Phi_n(X)}$, for a prime number p such that $p \nmid n$.
9. Determine all roots of unity that are contained in the fields $\mathbb{Q}(\sqrt{2})$, $\mathbb{Q}(i)$, $\mathbb{Q}(i\sqrt{2})$, and $\mathbb{Q}(i\sqrt{3})$.

4.6 Linear Independence of Characters

In the next two sections we will discuss some methods from linear algebra that are of special interest for applications in Galois theory, in particular for the study of cyclic extensions in 4.8. The "linear" point of view in Galois theory was suggested by E. Artin, who effectively used it in [1], [2] to develop an alternative approach to the theory. In a first step, we will study characters. Related to Galois theory, characters will occur in the form of homomorphisms $K^* \longrightarrow L^*$ between the multiplicative groups of two fields K and L. The main objective of the present section is to show that different characters are linearly independent.

Definition 1. *Let G be a group and K a field. A K-valued* character *of G is a group homomorphism $\chi\colon G \longrightarrow K^*$.*

For a group G and a field K, there exists always the *trivial* character $G \longrightarrow K^*$, mapping every element $g \in G$ to the unit element $1 \in K^*$. Furthermore, the K-valued characters of G form a group, whose law of composition is induced from the multiplication on K^*. Indeed, the product of two characters $\chi_1, \chi_2\colon G \longrightarrow K^*$ is given by

$$\chi_1 \cdot \chi_2\colon G \longrightarrow K^*, \qquad g \longmapsto \chi_1(g) \cdot \chi_2(g).$$

Also note that the K-valued characters of G can be viewed as special elements of the K-vector space $\mathrm{Map}(G, K)$, consisting of all maps from G to K. In particular, it is meaningful to talk about linear dependence or independence of characters.

Proposition 2 (E. Artin). *Distinct characters $\chi_1, \dots, \chi_n$ on a group G with values in a field K are linearly independent in* $\mathrm{Map}(G, K)$.

Proof. We proceed indirectly and assume that the assertion of the proposition is false. Then there is a minimal number $n \in \mathbb{N}$ such that there exists a linearly dependent system of K-valued characters $\chi_1, \ldots, \chi_n$ on G. Of course, we must have $n \geq 2$, since every character assumes values in K^* and therefore cannot coincide with the zero map. Now let

$$a_1\chi_1 + \ldots + a_n\chi_n = 0$$

be a nontrivial relation in $\mathrm{Map}(G, K)$ with coefficients $a_i \in K$. Then $a_i \neq 0$ for all i, due to the minimality of n, and we get

$$a_1\chi_1(gh) + \ldots + a_n\chi_n(gh) = 0$$

for $g, h \in G$. Choose g with the property that $\chi_1(g) \neq \chi_2(g)$; this is possible, since $\chi_1 \neq \chi_2$. Then, varying h over G, we see that

$$a_1\chi_1(g) \cdot \chi_1 + \ldots + a_n\chi_n(g) \cdot \chi_n = 0$$

is a new nontrivial relation in $\mathrm{Map}(G, K)$. Multiplying the initial one by $\chi_1(g)$ and subtracting the new one from it, we get a third relation:

$$a_2\big(\chi_1(g) - \chi_2(g)\big)\chi_2 + \ldots + a_n\big(\chi_1(g) - \chi_n(g)\big)\chi_n = 0.$$

This is a nontrivial relation of length $n-1$, since $a_2(\chi_1(g)-\chi_2(g)) \neq 0$. However, this contradicts the minimality of n, and it follows that the assertion of the proposition is true. □

The preceding proposition can be applied in various settings. For example, if L/K is an algebraic field extension, we see that the system $\mathrm{Aut}_K(L)$ of all K-automorphisms of L is linearly independent in the L-vector space of all maps $L \longrightarrow L$. To justify this, restrict K-homomorphisms $L \longrightarrow L$ to group homomorphisms $L^* \longrightarrow L^*$.

Corollary 3. *Let L/K be a finite separable field extension and $x_1, \ldots, x_n$ a basis of L as a K-vector space. Furthermore, let $\sigma_1, \ldots, \sigma_n$ denote the K-homomorphisms of L to an algebraic closure $\overline{K}$ of K. Then the vectors*

$$\begin{aligned} \xi_1 &= \big(\sigma_1(x_1), \ldots, \sigma_1(x_n)\big), \\ &\ldots \\ &\ldots \\ &\ldots \\ \xi_n &= \big(\sigma_n(x_1), \ldots, \sigma_n(x_n)\big), \end{aligned}$$

give rise to a system that is linearly independent over $\overline{K}$.

Proof. The linear dependence of the ξ_i would imply the linear dependence of the σ_i. However, as we can read from Proposition 2, the σ_i form a linearly independent system. □

To give another example, consider characters of type

$$\mathbb{Z} \longrightarrow K^*, \qquad \nu \longmapsto a^\nu,$$

for fixed $a \in K^*$. If there are distinct elements $a_1, \ldots, a_n \in K^*$, as well as further elements $c_1, \ldots, c_n \in K$ such that

$$c_1 a_1^\nu + \ldots + c_n a_n^\nu = 0$$

for all $\nu \in \mathbb{Z}$, then Proposition 2 shows that $c_1 = \ldots = c_n = 0$.

Exercises

1. *Let G be a cyclic group and $\mathbb{F}$ a finite field. Determine all $\mathbb{F}$-valued characters of G, and in particular, specify their number.*
2. Let L/K and M/K be field extensions and consider distinct K-homomorphisms $\sigma_1, \ldots, \sigma_r$ from L to M. Prove the existence of elements $x_1, \ldots, x_r \in L$ such that, similarly as in Corollary 3, the vectors $\xi_i = (\sigma_i(x_1), \ldots, \sigma_i(x_r)) \in M^r$, $i = 1, \ldots, r$, are linearly independent over M. *Hint*: Look at the map $L \longrightarrow M^r$, $x \longmapsto (\sigma_1(x), \ldots, \sigma_r(x))$, and show that M^r, as an M-vector space, is generated by the image of this map.
3. Let L/K and M/K be field extensions and $\sigma_1, \ldots, \sigma_r$ distinct K-homomorphisms from L to M. Furthermore, consider a polynomial $f \in M[X_1, \ldots, X_r]$ such that $f(\sigma_1(x), \ldots, \sigma_r(x)) = 0$ for all $x \in L$. Use Exercise 2 and show that f is the zero polynomial if K contains infinitely many elements. *Hint*: Choose elements $x_1, \ldots, x_r \in L$ as in Exercise 2 and verify in a first step that the polynomial $g(Y_1, \ldots, Y_r) = f(\sum_{i=1}^r \sigma_1(x_i)Y_i, \ldots, \sum_{i=1}^r \sigma_r(x_i)Y_i)$ is the zero polynomial.

4.7 Norm and Trace

In linear algebra one defines the determinant and the trace for endomorphisms of finite-dimensional vector spaces over fields. Since we will use these notions in the sequel, we give a brief review of them. Let K be a field, V an n-dimensional K-vector space, and $\varphi\colon V \longrightarrow V$ an endomorphism. The *characteristic polynomial* of φ is given by

$$\chi_\varphi(X) = \det(X \cdot \mathrm{id} - \varphi) = \sum_{i=0}^{n} c_i X^{n-i},$$

where $(-1)^n c_n = \det(\varphi)$ is the *determinant* and $-c_1 = \mathrm{trace}(\varphi)$ the *trace* of φ. If the matrix $A = (a_{ij}) \in K^{n\times n}$ represents φ with respect to a certain basis of V, then

$$\det(\varphi) = \det(A) = \sum_{\pi \in \mathfrak{S}_n} \mathrm{sgn}(\pi) a_{1,\pi(1)} \ldots a_{n,\pi(n)},$$

$$\mathrm{trace}(\varphi) = \mathrm{trace}(A) = \sum_{i=1}^{n} a_{ii}.$$

Furthermore, for two endomorphisms $\varphi, \psi\colon V \longrightarrow V$ and constants $a, b \in K$ we have

$$\begin{aligned} \operatorname{trace}(a\varphi + b\psi) &= a \cdot \operatorname{trace}(\varphi) + b \cdot \operatorname{trace}(\psi), \\ \det(\varphi \circ \psi) &= \det(\varphi) \cdot \det(\psi). \end{aligned}$$

Definition 1. *Let L/K be a finite field extension. For elements $a \in L$, consider the multiplication map*

$$\varphi_a\colon L \longrightarrow L, \qquad x \longmapsto ax,$$

as an endomorphism of L as a K-vector space. Then

$$\operatorname{tr}_{L/K}(a) := \operatorname{trace}(\varphi_a), \qquad \operatorname{N}_{L/K}(a) := \det(\varphi_a)$$

are called the trace *and the* norm *of a with respect to the extension L/K.*

In particular, $\operatorname{tr}_{L/K}\colon L \longrightarrow K$ is a homomorphism of K-vector spaces, or in more precise terms, a linear functional on L viewed as a K-vector space. Likewise, $\operatorname{N}_{L/K}\colon L^* \longrightarrow K^*$ is a group homomorphism, hence a character on L^* with values in K. For example, we have

$$\operatorname{N}_{\mathbb{C}/\mathbb{R}}(z) = |z|^2.$$

Indeed, if $z = x + iy$ is the decomposition of z into its real and imaginary parts, then the multiplication by z on $\mathbb{C}$ is described relative to the $\mathbb{R}$-basis $1, i$ by the matrix

$$\begin{pmatrix} x & -y \\ y & x \end{pmatrix}.$$

In the following, let us discuss some methods for computing the trace and the norm.

Lemma 2. *Let L/K be a finite field extension of degree $n = [L : K]$, and consider an element $a \in L$.*

(i) *If $a \in K$, then*

$$\operatorname{tr}_{L/K}(a) = na, \qquad \operatorname{N}_{L/K}(a) = a^n.$$

(ii) *If $L = K(a)$ and $X^n + c_1X^{n-1} + \ldots + c_n$ is the minimal polynomial of a over K, then*

$$\operatorname{tr}_{L/K}(a) = -c_1, \qquad \operatorname{N}_{L/K}(a) = (-1)^n c_n.$$

Proof. For $a \in K$ the linear map $\varphi_a\colon L \longrightarrow L$ is described by a times the unit matrix of $K^{n\times n}$. This justifies the formulas in (i). Furthermore, if $L = K(a)$, the minimal polynomial of a coincides with the minimal polynomial of the endomorphism φ_a, and hence by reasons of degree, must coincide with the characteristic

polynomial of φ_a. Therefore, the formulas in (ii) follow from the description of $\mathrm{trace}(\varphi_a)$ and $\det(\varphi_a)$ in terms of the coefficients of the characteristic polynomial of φ_a. □

The two cases of Lemma 2 can be combined, thus showing how to compute the norm and the trace of elements when one is dealing with arbitrary field extensions.

Lemma 3. *Consider an element $a \in L$ of a finite field extension L/K, and let $s = [L : K(a)]$. Then*

$$\mathrm{tr}_{L/K}(a) = s \cdot \mathrm{tr}_{K(a)/K}(a), \qquad \mathrm{N}_{L/K}(a) = \big(\mathrm{N}_{K(a)/K}(a)\big)^s.$$

Proof. Choose a K-basis $x_1, \ldots, x_r$ of $K(a)$, as well as a $K(a)$-basis $y_1, \ldots, y_s$ of L. Then the products $x_i y_j$ form a K-basis of L. Let $A \in K^{r\times r}$ be the matrix describing the multiplication by a on $K(a)$ relative to the basis $x_1, \ldots, x_r$. It follows that, relative to the basis consisting of the $x_i y_j$, the multiplication by a on L is given by the matrix

$$C = \begin{pmatrix} A & & 0 \\ & \ddots & \\ 0 & & A \end{pmatrix},$$

which consists of s boxes A and of zeros otherwise. Therefore, we get

$$\begin{aligned} \mathrm{tr}_{L/K}(a) &= \mathrm{trace}(C) = s \cdot \mathrm{trace}(A) = s \cdot \mathrm{tr}_{K(a)/K}(a), \\ \mathrm{N}_{L/K}(a) &= \det(C) = \big(\det(A)\big)^s = \big(\mathrm{N}_{K(a)/K}(a)\big)^s, \end{aligned}$$

as claimed. □

Proposition 4. *Let L/K be a finite field extension of degree $[L : K] = qr$, where $r = [L : K]_s$ is the separable degree of L/K. (Note that q is sometimes called the* inseparable degree *of L/K.) If $\sigma_1, \ldots, \sigma_r$ are the K-homomorphisms of L into an algebraic closure $\overline{K}$ of K, the following formulas hold for elements $a \in L$:*

$$\begin{aligned} \mathrm{tr}_{L/K}(a) &= q \sum_{i=1}^{r} \sigma_i(a), \\ \mathrm{N}_{L/K}(a) &= \Big(\prod_{i=1}^{r} \sigma_i(a)\Big)^q. \end{aligned}$$

Furthermore, if the extension L/K is not separable for $p = \mathrm{char}\, K > 0$, then q is a nontrivial power of p, and we get $\mathrm{tr}_{L/K}(a) = 0$ for all $a \in L$.

Before proving the proposition, let us state the transitivity formulas for the trace and the norm; these will be proved together with the assertion of Proposition 4.

Proposition 5. *Let $K \subset L \subset M$ be a chain of finite field extensions. Then*:

$$\mathrm{tr}_{M/K} = \mathrm{tr}_{L/K} \circ \mathrm{tr}_{M/L}, \qquad \mathrm{N}_{M/K} = \mathrm{N}_{L/K} \circ \mathrm{N}_{M/L}.$$

Proof of Propositions 4 *and* 5. In the situation of Proposition 4 we write for elements $a \in L$

$$\mathrm{tr}'_{L/K}(a) = q \sum_{i=1}^{r} \sigma_i(a),$$
$$\mathrm{N}'_{L/K}(a) = \Big(\prod_{i=1}^{r} \sigma_i(a) \Big)^q,$$

and show that $\mathrm{tr}_{L/K} = \mathrm{tr}'_{L/K}$, as well as $\mathrm{N}_{L/K} = \mathrm{N}'_{L/K}$. To do this, we consider the special cases of Lemma 2 and apply the transitivity formulas to settle the general case.

First, assume $a \in K$. Since $[L : K] = qr$ and $\sigma_i(a) = a$ for all i, we conclude from Lemma 2 that

$$\mathrm{tr}_{L/K}(a) = [L : K] \cdot a = q(ra) = \mathrm{tr}'_{L/K}(a),$$
$$\mathrm{N}_{L/K}(a) = a^{[L:K]} = (a^r)^q = \mathrm{N}'_{L/K}(a).$$

Now consider the second special case of Lemma 2 and assume $L = K(a)$. Let

$$X^n + c_1 X^{n-1} + \ldots + c_n \in K[X]$$

be the minimal polynomial of a over K, where $n = qr$. Using 3.4/8 and 3.6/2, the latter polynomial admits the factorization

$$\prod_{i=1}^{r} (X - \sigma_i(a))^q$$

over $\overline{K}$. Therefore, we conclude from Lemma 2 that

$$\mathrm{tr}_{L/K}(a) = -c_1 = q \sum_{i=1}^{r} \sigma_i(a) = \mathrm{tr}'_{L/K}(a),$$
$$\mathrm{N}_{L/K}(a) = (-1)^n c_n = \Big(\prod_{i=1}^{r} \sigma_i(a) \Big)^q = \mathrm{N}'_{L/K}(a).$$

In particular, we thereby see that tr and tr' as well as N and N' coincide in the special cases of Lemma 2.

Now, if $a \in L$ is arbitrary, consider the chain of finite field extensions $K \subset K(a) \subset L$ and use Lemmas 2 and 3. Then the special cases we have just discussed show that

$$\begin{aligned}
\mathrm{tr}_{L/K}(a) &= [L:K(a)]\cdot \mathrm{tr}_{K(a)/K}(a) = \mathrm{tr}_{K(a)/K}\big([L:K(a)]\cdot a\big)\\
&= \mathrm{tr}_{K(a)/K}\big(\mathrm{tr}_{L/K(a)}(a)\big)\\
&= \mathrm{tr}'_{K(a)/K}\big(\mathrm{tr}'_{L/K(a)}(a)\big),\\
\mathrm{N}_{L/K}(a) &= \big(\mathrm{N}_{K(a)/K}(a)\big)^{[L:K(a)]} = \mathrm{N}_{K(a)/K}\big(a^{[L:K(a)]}\big)\\
&= \mathrm{N}_{K(a)/K}\big(\mathrm{N}_{L/K(a)}(a)\big)\\
&= \mathrm{N}'_{K(a)/K}\big(\mathrm{N}'_{L/K(a)}(a)\big).
\end{aligned}$$

Thus, to prove Proposition 4, it is enough to establish the transitivity formulas of Proposition 5 for tr' and N'. The same formulas are then valid for tr and N as well, due to Proposition 4.

Therefore, consider a chain of finite field extensions $K \subset L \subset M$ as in Proposition 5. Embedding M into an algebraic closure $\overline{K}$ of K, we may assume that the chain is contained in $\overline{K}$. For

$$[L:K] = q_1[L:K]_s, \qquad [M:L] = q_2[M:L]_s,$$

the multiplicativity formulas 3.2/2 and 3.6/7 imply

$$[M:K] = q_1q_2[M:K]_s.$$

Assuming

$$\mathrm{Hom}_K(L,\overline{K}) = \big\{\sigma_1,\dots,\sigma_r\big\}, \qquad \mathrm{Hom}_L(M,\overline{K}) = \big\{\tau_1,\dots,\tau_s\big\},$$

where the elements σ_i, resp. τ_j, are distinct, we can apply 3.4/9 and thereby choose extensions $\sigma'_i\colon \overline{K} \longrightarrow \overline{K}$ of the σ_i. It follows that, just as in the the proof of 3.6/7, we get

$$\mathrm{Hom}_K(M,\overline{K}) = \big\{\sigma'_i\circ\tau_j\,;\, i=1,\dots,r,\ j=1,\dots,s\big\}$$

with distinct elements $\sigma'_i\circ\tau_j$. Then it is easy to derive the desired transitivity formulas for elements $a\in M$, since we have

$$\begin{aligned}
\mathrm{tr}'_{M/K}(a) &= q_1q_2\sum_{i,j}\sigma'_i\circ\tau_j(a)\\
&= q_1\sum_i\sigma'_i\Big(q_2\sum_j\tau_j(a)\Big)\\
&= \mathrm{tr}'_{L/K}\big(\mathrm{tr}'_{M/L}(a)\big),
\end{aligned}$$

as well as a similar chain of equalities for $\mathrm{N}'_{M/K}(a)$. However, note that the last line is meaningful only if we know that $\mathrm{tr}'_{M/L}(a)$ is an element of L, or what is enough, that we have $\mathrm{tr}'_{M/L}(a) = \mathrm{tr}_{M/L}(a)$. Going back to the situation faced in the proof of Proposition 4 above, the latter equality is indeed given, so that we can finish the proof of Proposition 4. After this, the general transitivity formulas in Proposition 5 are derived from the corresponding ones for tr' and N' using the assertion of Proposition 4. □

There is an immediate consequence of Proposition 4:

Corollary 6. *Let L/K be a finite Galois extension. Then $\mathrm{tr}_{L/K}$ and $\mathrm{N}_{L/K}$ are compatible with Galois automorphisms of L/K, i.e., we have*

$$\mathrm{tr}_{L/K}(a) = \mathrm{tr}_{L/K}\big(\sigma(a)\big), \qquad \mathrm{N}_{L/K}(a) = \mathrm{N}_{L/K}\big(\sigma(a)\big)$$

for all $a \in L$, $\sigma \in \mathrm{Gal}(L/K)$.

We want to derive some further consequences from Proposition 4. Given a finite field extension L/K, we can view L as a K-vector space and consider the symmetric bilinear form

$$\mathrm{tr}\colon L \times L \longrightarrow K, \qquad (x, y) \longmapsto \mathrm{tr}_{L/K}(xy).$$

By Proposition 4, this map vanishes identically if L/K fails to be separable.

Proposition 7. *A finite field extension L/K is separable if and only if the K-linear map $\mathrm{tr}_{L/K}\colon L \longrightarrow K$ is nontrivial and hence surjective. If L/K is separable, the symmetric bilinear map*

$$\mathrm{tr}\colon L \times L \longrightarrow K, \qquad (x, y) \longmapsto \mathrm{tr}_{L/K}(xy),$$

is nondegenerate. In other words, tr *induces then an isomorphism*

$$L \longrightarrow \hat{L}, \qquad x \longmapsto \mathrm{tr}(x, \cdot),$$

of L onto its dual space $\hat{L}$.

Proof. We assume that L/K is separable. If $\sigma_1, \ldots, \sigma_r$ are the K-homomorphisms of L into an algebraic closure of K, we get

$$\mathrm{tr}_{L/K} = \sigma_1 + \ldots + \sigma_r$$

by Proposition 4. Furthermore, Proposition 4.6/2 on the linear independence of characters shows that $\mathrm{tr}_{L/K}$ is not identically zero. Now consider an element x of the kernel of $L \longrightarrow \hat{L}$, i.e., an element satisfying $\mathrm{tr}(x, \cdot) = 0$. Then we get $\mathrm{tr}_{L/K}(xL) = 0$ and therefore necessarily $x = 0$, since otherwise, we would have $xL = L$, and $\mathrm{tr}_{L/K}$ would vanish on L. Hence, the map $L \longrightarrow \hat{L}$ is injective, and since $\dim L = \dim \hat{L} < \infty$, also surjective. □

Corollary 8. *Let L/K be a finite separable field extension with a K-basis $x_1, \ldots, x_n$ of L. Then there exists a unique K-basis $y_1, \ldots, y_n$ of L such that $\mathrm{tr}_{L/K}(x_i y_j) = \delta_{ij}$ for $i, j = 1, \ldots, n$.*

Proof. Use the existence and uniqueness of the dual basis of $x_1, \ldots, x_n$. □

Exercises

1. *Let L/K be a field extension of degree $n < \infty$. Describe the properties of the set $\{a \in L\,;\, \mathrm{tr}_{L/K}(a) = 0\}$.*

2. *Let $\mathbb{F}'/\mathbb{F}$ be an extension of finite fields. Describe the kernel and the image of the attached norm map* $\mathrm{N}\colon \mathbb{F}'^{*} \longrightarrow \mathbb{F}^{*}$.

3. Let K be a field and $L = K(a)$ a simple algebraic field extension with minimal polynomial $f \in K[X]$ of a. Show $f(x) = \mathrm{N}_{L/K}(x - a)$ for $x \in K$.

4. For relatively prime positive integers m, n, consider a field extension L/K of degree m. Then every element $a \in K$ admitting an nth root in L already admits an nth root in K.

5. Let L/K be a finite Galois extension with K-basis $x_1, \ldots, x_n$. Show for a subgroup $H \subset \mathrm{Gal}(L/K)$ that its corresponding fixed field L^H is characterized by $L^H = K(\mathrm{tr}_{L/L^H}(x_1), \ldots, \mathrm{tr}_{L/L^H}(x_n))$.

6. Let L/K be a finite field extension in characteristic $p > 0$. Show for elements $a \in L$ that $\mathrm{tr}_{L/K}(a^p) = (\mathrm{tr}_{L/K}(a))^p$.

4.8 Cyclic Extensions

In order to solve algebraic equations by radicals it is necessary to study extensions of a given field K that are obtained by adjoining an nth root of some element $c \in K$. The aim of the present section is to characterize such extensions in terms of Galois theory. We will base our study on the famous Theorem 90 of D. Hilbert [9], which we will prove first. Recall that a Galois extension L/K is called *cyclic* if its Galois group $\mathrm{Gal}(L/K)$ is cyclic.

Theorem 1 (Hilbert 90). *Let L/K be a finite cyclic Galois extension and let $\sigma \in \mathrm{Gal}(L/K)$ be a generating element. Then the following conditions are equivalent for elements $b \in L$:*

(i) $\mathrm{N}_{L/K}(b) = 1$.

(ii) *There exists an element $a \in L^*$ such that $b = a \cdot \sigma(a)^{-1}$.*

Proof. If $b = a \cdot \sigma(a)^{-1}$ for some $a \in L^*$, we conclude from 4.7/6 that

$$\mathrm{N}_{L/K}(b) = \frac{\mathrm{N}_{L/K}(a)}{\mathrm{N}_{L/K}\big(\sigma(a)\big)} = 1.$$

Conversely, consider an element $b \in L$ satisfying $\mathrm{N}_{L/K}(b) = 1$. Let $n = [L : K]$. Using the linear independence of characters 4.6/2, it follows that

$$\sigma^0 + b\sigma^1 + b \cdot \sigma(b) \cdot \sigma^2 + \ldots + b \cdot \sigma(b) \cdot \ldots \cdot \sigma^{n-2}(b) \cdot \sigma^{n-1},$$

as a map $L^* \longrightarrow L$, is not identically zero. Therefore, there is an element $c \in L^*$ such that

$$a := c + b\sigma(c) + b \cdot \sigma(b) \cdot \sigma^2(c) + \ldots + b \cdot \sigma(b) \cdot \ldots \cdot \sigma^{n-2}(b) \cdot \sigma^{n-1}(c) \neq 0.$$

Applying σ and then multiplying by b, we get

$$b \cdot \sigma(a) = b\sigma(c) + b \cdot \sigma(b) \cdot \sigma^2(c) + \ldots + b \cdot \sigma(b) \cdot \ldots \cdot \sigma^{n-1}(b) \cdot \sigma^n(c) = a,$$

since $\sigma^n = \mathrm{id}$ and $b \cdot \sigma(b) \cdot \ldots \cdot \sigma^{n-1}(b) = \mathrm{N}_{L/K}(b) = 1$, due to 4.7/4. □

The preceding theorem can also be interpreted within the more general context of Galois cohomology. We want to briefly explain this; for more details one may consult Serre [14], Chaps. VII, X. In the following, consider a group G together with an abelian group A, as well as an action of G on A, by which is meant a group homomorphism $G \longrightarrow \mathrm{Aut}(A)$. Given a finite (not necessarily cyclic) Galois extension L/K, we are mainly interested in the case that $G = \mathrm{Gal}(L/K)$ and $A = L^*$, with $G \longrightarrow \mathrm{Aut}(L^*)$ being the canonical homomorphism. For $\sigma \in G$ and $a \in A$, write $\sigma(a)$ for the image of a with respect to the automorphism of A that is attached to σ. Then we can consider the following subgroups of $\mathrm{Map}(G, A)$, the abelian group consisting of all maps from G to A:

$$Z^1(G, A) = \Big\{ f \,;\, f(\sigma \circ \sigma') = \sigma\big(f(\sigma')\big) \cdot f(\sigma) \text{ for all } \sigma, \sigma' \in G \Big\},$$

$$B^1(G, A) = \Big\{ f \,;\, \text{there exists } a \in A \text{ such that } f(\sigma) = a \cdot \sigma(a)^{-1} \text{ for all } \sigma \in G \Big\}.$$

The elements of $B^1(G, A)$ are called 1-*coboundaries*; they constitute a subgroup of $Z^1(G, A)$, the group of 1-*cocycles*. The residue class group

$$H^1(G, A) := Z^1(G, A) \big/ B^1(G, A)$$

is called the *first cohomology group* of G with values in A. Using such terminology, the cohomological version of Hilbert's Theorem 90 reads as follows:

Theorem 2. *If L/K is a finite Galois extension with Galois group G, then $H^1(G, L^*) = \{1\}$, i.e., every 1-cocycle is a 1-coboundary.*

Proof. Let $f\colon G \longrightarrow L^*$ be a 1-cocycle. To show that it is a 1-coboundary, look at the Poincaré series

$$b = \sum_{\sigma' \in G} f(\sigma') \cdot \sigma'(c)$$

for elements $c \in L^*$. Using the linear independence of characters 4.6/2, we can choose c in such a way that $b \neq 0$. Then we see for arbitrary $\sigma \in G$ that

$$\begin{aligned} \sigma(b) &= \sum_{\sigma' \in G} \sigma\big(f(\sigma')\big) \cdot (\sigma \circ \sigma')(c) \\ &= \sum_{\sigma' \in G} f(\sigma)^{-1} \cdot f(\sigma \circ \sigma') \cdot (\sigma \circ \sigma')(c) = f(\sigma)^{-1} \cdot b, \end{aligned}$$

i.e., that f is a 1-coboundary. □

To derive Hilbert's Theorem 90 in its original version from Theorem 2, consider a cyclic Galois extension L/K of degree n and fix a generating element σ of

the Galois group $\mathrm{Gal}(L/K)$. Then one shows for $b \in L^*$ satisfying $\mathrm{N}_{L/K}(b) = 1$ that $f\colon G \longrightarrow L^*$, given by

$$\begin{aligned} \sigma^0 &\longmapsto 1, \\ \sigma^1 &\longmapsto b, \\ &\dots \\ \sigma^{n-1} &\longmapsto b \cdot \sigma(b) \cdot \ldots \cdot \sigma^{n-2}(b), \end{aligned}$$

is a 1-cocycle, and hence by Theorem 2, a 1-coboundary.

We want to apply Hilbert's Theorem 90 in order to characterize cyclic extensions in more detail.

Proposition 3. *Let L/K be a field extension and n an integer > 0 such that* $\mathrm{char}\, K \nmid n$. *Moreover, assume that K contains a primitive* nth *root of unity.*

(i) *Every cyclic Galois extension L/K of degree n is of type $L = K(a)$ for an element $a \in L$, whose minimal polynomial over K equals $X^n - c$ for some element $c \in K$.*

(ii) *Conversely, if $L = K(a)$ for some element $a \in L$ that is a zero of a polynomial of type $X^n - c \in K[X]$, then L/K is a cyclic Galois extension. Furthermore, $d = [L : K]$ divides n and satisfies $a^d \in K$, which implies that $X^d - a^d \in K[X]$ is the minimal polynomial of a over K.*

Proof. Let $\zeta \in K$ be a primitive nth root of unity. If L/K is a cyclic Galois extension of degree n, then $\mathrm{N}_{L/K}(\zeta^{-1}) = \zeta^{-n} = 1$ by 4.7/2. Furthermore, using Hilbert's Theorem 90, there exists an element $a \in L^*$ such that $\sigma(a) = \zeta a$, where σ is a generating element of $\mathrm{Gal}(L/K)$. Then we get

$$\sigma^i(a) = \zeta^i a, \qquad i = 0, \dots, n-1.$$

In particular, the elements $\sigma^0(a), \dots, \sigma^{n-1}(a)$ are distinct, and we see that $[K(a) : K] \geq n$, in fact that $L = K(a)$, since $K(a) \subset L$ and $[L : K] = n$.[4] Now observe that

$$\sigma(a^n) = \sigma(a)^n = \zeta^n a^n = a^n,$$

i.e., that $a^n \in K$. Therefore, a is a zero of the polynomial

$$X^n - a^n \in K[X].$$

Since a is of degree n over K, this polynomial is already the minimal polynomial of a over K. This justifies assertion (i).

Now turning to assertion (ii), assume $L = K(a)$, where a is a zero of a polynomial of type $X^n - c \in K[X]$. We may assume $a \neq 0$, since the case $a = 0$ is trivial. Then $\zeta^0 a, \dots, \zeta^{n-1} a$ are n distinct zeros of $X^n - c$, and we see

[4] Note that in this way we have constructed a special generating element of the field extension L/K. However, that this extension is simple follows just as well from the primitive element theorem 3.6/12.

that $L = K(a)$ is a splitting field over K of this polynomial. Since $X^n - c$ is a separable polynomial, due to $\operatorname{char} K \nmid n$, it follows that L/K is even a Galois extension. Now, just as a is a zero of $X^n - c$, the same is true for $\sigma(a)$ for every $\sigma \in \operatorname{Gal}(L/K)$. Therefore, we can associate to σ in each case an nth root of unity $w_\sigma \in U_n$ such that $\sigma(a) = w_\sigma a$. It follows that

$$\operatorname{Gal}(L/K) \longrightarrow U_n, \qquad \sigma \longmapsto w_\sigma,$$

is an injective group homomorphism, and hence due to the theorem of Lagrange 1.2/3 that $d := [L : K] = \operatorname{ord}(\operatorname{Gal}(L/K))$ divides $n = \operatorname{ord} U_n$. Since U_n is cyclic by 4.5/1, every subgroup of U_n admits this property as well. In particular, $\operatorname{Gal}(L/K)$ is cyclic. If $\sigma \in \operatorname{Gal}(L/K)$ generates this cyclic group of order d, then w_σ is a primitive dth root of unity, and we have

$$\sigma(a^d) = \sigma(a)^d = w_\sigma^d a^d = a^d,$$

so that $a^d \in K$. Clearly, a is a zero of $X^d - a^d \in K[X]$, and we see by reasons of degree that this is the minimal polynomial of $a \in L$ over K. □

Next we want to prove an additive version of Hilbert's Theorem 90. Also in this case there is a generalization in terms of Galois cohomology; cf. Exercise 5 below.

Theorem 4 (Hilbert 90, additive version). *Let L/K be a finite cyclic Galois extension and $\sigma \in \operatorname{Gal}(L/K)$ a generating element. The following conditions are equivalent for elements $b \in L$:*

(i) $\operatorname{tr}_{L/K}(b) = 0$.

(ii) *There exists an element $a \in L$ such that $b = a - \sigma(a)$.*

Proof. We proceed similarly as in the proof of Theorem 1. If $b = a - \sigma(a)$ for some element $a \in L$, then

$$\operatorname{tr}_{L/K}(b) = \operatorname{tr}_{L/K}(a) - \operatorname{tr}_{L/K}\big(\sigma(a)\big) = 0$$

by 4.7/6. Conversely, consider an element $b \in L$ such that $\operatorname{tr}_{L/K}(b) = 0$, and let $n = [L : K]$. Since the trace map $\operatorname{tr}_{L/K}$ is not identically zero by 4.7/7, there exists an element $c \in L$ such that $\operatorname{tr}_{L/K}(c) \neq 0$. Define $a \in L$ by

$$\begin{aligned} a \cdot \big(\operatorname{tr}_{L/K}(c)\big) &= b \cdot \sigma(c) + \big(b + \sigma(b)\big) \cdot \sigma^2(c) + \ldots \\ &\quad + \big(b + \sigma(b) + \ldots + \sigma^{n-2}(b)\big) \cdot \sigma^{n-1}(c). \end{aligned}$$

Applying σ yields

$$\begin{aligned} \sigma(a) \cdot \big(\operatorname{tr}_{L/K}(c)\big) &= \sigma(b)\sigma^2(c) + \big(\sigma(b) + \sigma^2(b)\big) \cdot \sigma^3(c) + \ldots \\ &\quad + \big(\sigma(b) + \sigma^2(b) + \ldots + \sigma^{n-1}(b)\big) \cdot \sigma^n(c). \end{aligned}$$

Then, using the relations

$$\begin{aligned}\mathrm{tr}_{L/K}(b) &= b + \sigma(b) + \ldots + \sigma^{n-1}(b) = 0,\\ \mathrm{tr}_{L/K}(c) &= c + \sigma(c) + \ldots + \sigma^{n-1}(c),\end{aligned}$$

see 4.7/4, we get

$$\begin{aligned}\big(a - \sigma(a)\big) \cdot \mathrm{tr}_{L/K}(c) &= b\sigma(c) + b\sigma^2(c) + \ldots + b\sigma^{n-1}(c)\\ &\quad - \big(\sigma(b) + \sigma^2(b) + \ldots + \sigma^{n-1}(b)\big) \cdot \sigma^n(c)\\ &= b \cdot \big(\sigma(c) + \sigma^2(c) + \ldots + \sigma^{n-1}(c) + c\big)\\ &= b \cdot \mathrm{tr}_{L/K}(c)\end{aligned}$$

and hence $b = a - \sigma(a)$. □

We want to apply the additive version of Hilbert's Theorem 90 in order to study cyclic extensions of degree p for $p = \mathrm{char}\, K > 0$, a case that is not covered by Proposition 3.

Theorem 5 (Artin–Schreier). *Let L/K be a field extension in characteristic $p > 0$.*

(i) *Every cyclic Galois extension L/K of degree p is of type $L = K(a)$ for an element $a \in L$ whose minimal polynomial over K equals $X^p - X - c$ for some $c \in K$.*

(ii) *Conversely, if $L = K(a)$ for some element $a \in L$ that is a zero of a polynomial of type $X^p - X - c \in K[X]$, then L/K is a cyclic Galois extension. Furthermore, if the polynomial $X^p - X - c$ does not split completely into linear factors over K, it is irreducible. In this case, L/K is a cyclic Galois extension of degree p.*

Proof. Assume first that L/K is a cyclic Galois extension of degree p. Then $\mathrm{tr}_{L/K}(c) = 0$ for all $c \in K$ by 4.7/2. In particular, using the additive version of Hilbert's Theorem 90, there is an element $a \in L$ satisfying $\sigma(a) - a = 1$, where σ is a generating element of $\mathrm{Gal}(L/K)$. Therefore, we get

$$\sigma^i(a) = a + i, \qquad i = 0, \ldots, p-1.$$

Since the elements $\sigma^0(a), \ldots, \sigma^{p-1}(a)$ are distinct, we conclude that the degree of a over K is at least p, and hence that $L = K(a)$. Furthermore, we obtain

$$\sigma(a^p - a) = \sigma(a)^p - \sigma(a) = (a+1)^p - (a+1) = a^p - a$$

and thereby see that $c := a^p - a \in K$. In particular, a is a zero of the polynomial $X^p - X - c \in K[X]$. By reasons of degree, this is the minimal polynomial of a over K.

Conversely, let us assume $L = K(a)$, where a is a zero of a polynomial of type $f = X^p - X - c \in K[X]$. Since a is a zero of f, the same is true for $a+1$, and we see that

$$a, a+1, \ldots, a+p-1 \in L$$

are the p distinct zeros of f. In particular, if one of the zeros of f is contained in K, then all of them belong to K, and f splits completely into linear factors over K. The same argument shows that L is a splitting field of the separable polynomial f over K and hence that the extension L/K is Galois. For $L = K$, the extension is cyclic for trivial reasons. Therefore, assume that f does not admit a zero in K. We claim that then f is irreducible over K. Indeed, if such is not the case, then there exists a factorization $f = gh$ into nonconstant monic polynomials g and h. Over L we have the factorization

$$f = \prod_{i=0}^{p-1} (X - a - i),$$

and it follows that g is a product of some of these factors. Let $d = \deg g$. The coefficient of X^{d-1} in g is of type $-da + j$ for some element j belonging to the prime subfield $\mathbb{F}_p \subset K$. However, from $-da + j \in K$ and $p \nmid d$ we get $a \in K$, so that f would have a zero in K. Since this was excluded, f must be irreducible over K. Now use 3.4/8 and choose an element $\sigma \in \mathrm{Gal}(L/K)$ such that $\sigma(a) = a+1$. Then σ is of order $\geq p$, and since $\operatorname{ord} \mathrm{Gal}(L/K) = \deg f = p$, we see that L/K is a cyclic Galois extension of degree p. □

Exercises

1. *In the situation of Theorem* 1, *fix* $b \in L^*$ *and consider elements* $a \in L^*$ *satisfying* $b = a \cdot \sigma(a)^{-1}$. *Is there a uniqueness assertion? Furthermore, study the corresponding question in the situation of Theorem* 4.

2. *Illustrate the assertion of Hilbert's Theorem* 90 *for the extension* $\mathbb{C}/\mathbb{R}$.

3. Let L be a splitting field of a polynomial of type $X^n - a$ over a field K such that $\operatorname{char} K \nmid n$. Check whether the extension L/K is always cyclic. Also discuss the case $K = \mathbb{Q}$.

4. Show for a finite Galois extension L/K with Galois group $G = \mathrm{Gal}(L/K)$ that $H^1(G, \mathrm{GL}(n, L)) = \{1\}$. *Hint*: Although this is not really necessary, assume that K admits infinitely many elements. Then proceed as in Theorem 2 and use Exercise 3 from Section 4.6. Furthermore, observe for the definition of $H^1(G, \mathrm{GL}(n, L))$ that a priori, this object has to be viewed as a "cohomology set," since the group $\mathrm{GL}(n, L)$ is not abelian for $n > 1$. Therefore, it is not clear from the outset that the corresponding group of 1-coboundaries constitutes a normal subgroup of the group of 1-cocycles and hence that the set of residue classes forms a group.

5. Consider a finite Galois extension L/K with Galois group $G = \mathrm{Gal}(L/K)$ and equip the additive group L with the canonical action of G. Show that $H^1(G, L) = 0$. *Hint*: Adapt the proof of Theorem 2 to the additive point of view.

6. Use Hilbert's Theorem 90 and show for two rational numbers $a, b \in \mathbb{Q}$ that the relation $a^2 + b^2 = 1$ is equivalent to the existence of integers $m, n \in \mathbb{Z}$ satisfying

$$a = \frac{m^2 - n^2}{m^2 + n^2}, \qquad b = \frac{2mn}{m^2 + n^2}.$$

4.9 Multiplicative Kummer Theory*

Recall that a Galois extension L/K is called *abelian* if the corresponding Galois group $G = \mathrm{Gal}(L/K)$ is abelian. More specifically, it is said to be *abelian of exponent* d for an integer $d > 0$ if in addition, G is of exponent d, i.e., if $\sigma^d = 1$ for every $\sigma \in G$ and d is minimal with this property. In the following we want to generalize cyclic extensions by studying abelian extensions of exponents that divide a given number $n \in \mathbb{N} - \{0\}$. Such extensions are referred to as *Kummer extensions*, named after E. Kummer, who considered these extensions for number-theoretic reasons.[5]

In the present section, we assume $\mathrm{char}\, K \nmid n$ and furthermore that K contains the group U_n of all nth roots of unity. Given $c \in K$, we write $K(c^{1/n})$ for the extension that is obtained by adjoining an nth root of c to K. However, observe that $c^{1/n}$, as an element of an algebraic closure of K, is unique only up to an nth root of unity, while the field $K(c^{1/n})$ itself is well defined. Indeed, it is the splitting field of the polynomial $X^n - c$, since K is supposed to contain all nth roots of unity. Furthermore, it follows from 4.8/3 that $K(c^{1/n})/K$ is a cyclic extension of a degree dividing n. Similarly, fixing a subset $C \subset K$, we can define the Galois extension $K(C^{1/n})$ that is obtained from K by adjoining all nth roots $c^{1/n}$ of elements $c \in C$. The resulting field can be viewed as the composite field (in an algebraic closure of K) of all extensions $K(c^{1/n})$, where c varies over C. In particular, we can consider the restriction homomorphisms $\mathrm{Gal}(K(C^{1/n})/K) \longrightarrow \mathrm{Gal}(K(c^{1/n})/K)$ of 4.1/2, giving rise to a monomorphism

$$\mathrm{Gal}\big(K(C^{1/n})/K\big) \longrightarrow \prod_{c \in C} \mathrm{Gal}\big(K(c^{1/n})/K\big).$$

Thereby we recognize $K(C^{1/n})/K$ as a (not necessarily finite) abelian extension of some exponent dividing n. This fact will be re-proved in Proposition 1 (i) below in a direct way, without referring to the characterization of cyclic extensions as given in 4.8/3.

Let us write G_C for the Galois group of the extension $K(C^{1/n})/K$. Given $\sigma \in G_C$ and an nth root $c^{1/n}$ of some element $c \in C$, it follows that $\sigma(c^{1/n})$ is also an nth root of c. Therefore, there is an nth root of unity $w_\sigma \in U_n$ such that $\sigma(c^{1/n}) = w_\sigma c^{1/n}$. As is easily checked, $w_\sigma = \sigma(c^{1/n}) \cdot c^{-1/n}$ is independent of the choice of the nth root $c^{1/n}$ of c. Therefore, we get a well-defined pairing

$$\langle \cdot, \cdot \rangle \colon G_C \times C \longrightarrow U_n, \qquad (\sigma, c) \longmapsto \frac{\sigma(c^{1/n})}{c^{1/n}}.$$

In the sequel we will assume that C is a subgroup of K^*. Then $\langle \cdot, \cdot \rangle$ is bimultiplicative in the sense that

[5] Strictly speaking, an abelian extension L/K of exponent d for some $d > 0$ is called a *Kummer extension* if $\mathrm{char}\, K \nmid d$ and K contains the group U_d of all dth roots of unity.

$$\langle \sigma\circ\tau, c\rangle = \frac{\sigma\circ\tau(c^{1/n})}{c^{1/n}} = \frac{\sigma\circ\tau(c^{1/n})}{\tau(c^{1/n})}\cdot\frac{\tau(c^{1/n})}{c^{1/n}} = \langle\sigma,c\rangle\cdot\langle\tau,c\rangle,$$
$$\langle \sigma, c\cdot c'\rangle = \frac{\sigma(c^{1/n}c'^{1/n})}{c^{1/n}c'^{1/n}} = \frac{\sigma(c^{1/n})}{c^{1/n}}\cdot\frac{\sigma(c'^{1/n})}{c'^{1/n}} = \langle\sigma,c\rangle\cdot\langle\sigma,c'\rangle,$$

for $\sigma,\tau\in G_C$ and $c,c'\in C$. Furthermore, we get $\langle\sigma,c^n\rangle = 1$ for $\sigma\in G_C$ and $c\in K^*$. Therefore, if we assume that $C\subset K^*$ is a subgroup containing the group K^{*n} of all nth powers of elements in K^*, it follows that $\langle\cdot,\cdot\rangle$ gives rise to a bimultiplicative map

$$G_C\times C/K^{*n}\longrightarrow U_n,\qquad (\sigma,\overline{c})\longmapsto\frac{\sigma(c^{1/n})}{c^{1/n}},$$

which will be denoted by $\langle\cdot,\cdot\rangle$ again.

Proposition 1. *As before, consider a field K and an integer $n>0$ such that $\operatorname{char}K\nmid n$ and $U_n\subset K^*$. Furthermore, let $C\subset K^*$ be a subgroup containing K^{*n}. Then*:

(i) *The extension $K(C^{1/n})/K$ is Galois and abelian of some exponent dividing n. Let G_C be the corresponding Galois group.*

(ii) *The bimultiplicative map*

$$\langle\cdot,\cdot\rangle\colon G_C\times C/K^{*n}\longrightarrow U_n,\qquad (\sigma,\overline{c})\longmapsto\frac{\sigma(c^{1/n})}{c^{1/n}},$$

is nondegenerate in the sense that it gives rise to monomorphisms

$$\varphi_1\colon\quad G_C\longrightarrow \operatorname{Hom}(C/K^{*n},U_n),\qquad \sigma\longmapsto\langle\sigma,\cdot\rangle,$$
$$\varphi_2\colon\quad C/K^{*n}\longrightarrow\operatorname{Hom}(G_C,U_n),\qquad \overline{c}\longmapsto\langle\cdot,\overline{c}\rangle,$$

into the group of all homomorphisms $C/K^{*n}\longrightarrow U_n$, resp. of all homomorphisms $G_C\longrightarrow U_n$. More precisely, φ_1 is an isomorphism, while φ_2 restricts to an isomorphism $C/K^{*n}\xrightarrow{\sim}\operatorname{Hom}_{\text{cont}}(G_C,U_n)$ onto the group of all continuous homomorphisms $G_C\longrightarrow U_n$.[6]

(iii) *The extension $K(C^{1/n})/K$ is finite if and only if the index $(C:K^{*n})$ is finite. If such is the case, both maps φ_1 and φ_2 of* (ii) *are isomorphisms, and one obtains $\big[K(C^{1/n}):K\big]=(C:K^{*n})$.*

Proof. Assertion (i) follows from the injectivity of φ_1 in (ii). To show that φ_1 is injective, consider an element $\sigma\in G_C$ such that $\sigma(c^{1/n})=c^{1/n}$ for all $c\in C$. Then $\sigma(a)=a$ for all $a\in K(C^{1/n})$, and we get $\sigma=\mathrm{id}$. Hence, φ_1 is injective. On the other hand, look at an element $c\in C$ satisfying $\sigma(c^{1/n})=c^{1/n}$ for all

[6] Within this context, consider G_C as a topological group as explained in Section 4.2, and equip U_n with the discrete topology. In this way, a homomorphism $f\colon G_C\longrightarrow U_n$ is continuous if and only if $H=\ker f$ is an open subgroup in G_C, i.e., according to 4.2/3 and 4.2/5, if and only if there exists a finite Galois extension K'/K in $K(C^{1/n})$ such that $H=\operatorname{Gal}(K(C^{1/n})/K')$, or what is enough, such that $H\supset\operatorname{Gal}(K(C^{1/n})/K')$.

$\sigma \in G_C$. Then $c^{1/n} \in K$ and therefore $c \in K^{*n}$, and it follows that also φ_2 is injective. This settles the nondegenerateness of the pairing in (ii).

Next we establish assertion (iii), relying on the injectivity of φ_1 and φ_2 in (ii). If $\big[K(C^{1/n}) : K\big]$ is finite, then the same is true for G_C, and we can conclude that $\operatorname{Hom}(G_C, U_n)$ is finite. Due to the injectivity of φ_2, it follows that C/K^{*n} is finite. Conversely, if C/K^{*n} is finite, then the same is true for $\operatorname{Hom}(C/K^{*n}, U_n)$, and due to the injectivity of φ_1, also for G_C and $\big[K(C^{1/n}) : K\big]$. Furthermore, if the finiteness is given, there exist (noncanonical) isomorphisms

$$C/K^{*n} \xrightarrow{\sim} \operatorname{Hom}(C/K^{*n}, U_n), \qquad G_C \xrightarrow{\sim} \operatorname{Hom}(G_C, U_n),$$

as we will show in Lemma 2 below. Therefore, the estimate

$$\begin{aligned}\big[K(C^{1/n}) : K\big] = \operatorname{ord} G_C &\le \operatorname{ord} \operatorname{Hom}(C/K^{*n}, U_n) = \operatorname{ord} C/K^{*n} \\ &\le \operatorname{ord} \operatorname{Hom}(G_C, U_n) = \operatorname{ord} G_C = \big[K(C^{1/n}) : K\big]\end{aligned}$$

yields the desired equality $\big[K(C^{1/n}) : K\big] = (C : K^{*n})$ and in addition shows that φ_1, φ_2 are isomorphisms. Thus, we are done with the proof of Proposition 1 in the special case in which $\big[K(C^{1/n}) : K\big]$ and $(C : K^{*n})$ are finite.

In the nonfinite case consider the system $(C_i)_{i\in I}$ of all subgroups of C such that $C_i \supset K^{*n}$ and $(C_i : K^{*n}) < \infty$. Then we get $C = \bigcup_{i\in I} C_i$, as well as $K(C^{1/n}) = \bigcup_{i\in I} K(C_i^{1/n})$, viewing these fields as subfields of an algebraic closure of K. For every $i \in I$ there is a commutative diagram

$$\begin{array}{ccc} G_C & \xrightarrow{\varphi_1} & \operatorname{Hom}(C/K^{*n}, U_n) \\ \big\downarrow & & \big\downarrow \\ G_{C_i} & \xrightarrow{\varphi_{1,i}} & \operatorname{Hom}(C_i/K^{*n}, U_n), \end{array}$$

where the vertical map on the left is restriction of Galois automorphisms of $K(C^{1/n})/K$ to automorphisms of $K(C_i^{1/n})/K$ (see 4.1/2), and where the vertical map on the right is restriction of homomorphisms $C/K^{*n} \longrightarrow U_n$ to C_i/K^{*n}. As we have seen, all maps $\varphi_{1,i}$ are bijective. Therefore, given a homomorphism $f\colon C/K^{*n} \longrightarrow U_n$, there exist unique elements $\sigma_i \in G_{C_i}$ such that $\varphi_{1,i}(\sigma_i) = f|_{C_i/K^{*n}}$ for all i, and it is easily checked that the σ_i make up a Galois automorphism $\sigma \in G_C$ satisfying $\varphi_1(\sigma) = f$. Thereby it follows that φ_1 is surjective and hence bijective.

To obtain the stated assertion on φ_2, consider for all $i \in I$ the commutative diagram

$$\begin{array}{ccc} C_i/K^{*n} & \xrightarrow{\varphi_{2,i}} & \operatorname{Hom}(G_{C_i}, U_n) \\ \big\downarrow & & \big\downarrow \\ C/K^{*n} & \xrightarrow{\varphi_2} & \operatorname{Hom}(G_C, U_n), \end{array}$$

where the vertical map on the left is the canonical inclusion, and where the vertical map on the right is induced by the restriction map $G_C \longrightarrow G_{C_i}$, which

was considered before. Since every continuous homomorphism $f\colon G_C \longrightarrow U_n$ is induced from a homomorphism of type $f_i\colon G_{C_i} \longrightarrow U_n$, the assertion on φ_2 in (ii) follows from the bijectivity of $\varphi_{2,i}$. □

It remains to establish the duality isomorphisms that were needed in the proof above. Note that U_n is cyclic of order n due to 4.5/1 and hence that it is isomorphic to $\mathbb{Z}/n\mathbb{Z}$.

Lemma 2. *For an integer $n \in \mathbb{N} - \{0\}$, consider a finite abelian group H of some exponent dividing n. Then there exists a (noncanonical) isomorphism $H \xrightarrow{\sim} \mathrm{Hom}(H, \mathbb{Z}/n\mathbb{Z})$.*

Proof. Since $\mathrm{Hom}(\cdot, \mathbb{Z}/n\mathbb{Z})$ is compatible with finite direct sums, we can apply the fundamental theorem of finitely generated abelian groups 2.9/9 and thereby assume that H is cyclic of some order d, where $d \,|\, n$. Then we have to construct an isomorphism

$$\mathbb{Z}/d\mathbb{Z} \xrightarrow{\sim} \mathrm{Hom}\big(\mathbb{Z}/d\mathbb{Z}, \mathbb{Z}/n\mathbb{Z}\big).$$

To do this, let us reduce to the case $d = n$. We know for every divisor d of n from the solution of Exercise 2 in Section 1.3 that there exists a unique subgroup $H_d \subset \mathbb{Z}/n\mathbb{Z}$ of order d and that it is cyclic. Clearly, we have $H_{d'} \subset H_d$ for $d' \mid d$, and it follows that every homomorphism $\mathbb{Z}/d\mathbb{Z} \longrightarrow \mathbb{Z}/n\mathbb{Z}$ factors through H_d. In particular, we see that the canonical map $\mathrm{Hom}(\mathbb{Z}/d\mathbb{Z}, H_d) \longrightarrow \mathrm{Hom}(\mathbb{Z}/d\mathbb{Z}, \mathbb{Z}/n\mathbb{Z})$ is an isomorphism. Therefore, using the fact that $H_d \simeq \mathbb{Z}/d\mathbb{Z}$, it is enough to specify an isomorphism $\mathbb{Z}/d\mathbb{Z} \longrightarrow \mathrm{Hom}(\mathbb{Z}/d\mathbb{Z}, \mathbb{Z}/d\mathbb{Z})$ in order to settle the assertion of the lemma. But that is easy, since

$$\mathbb{Z} \longrightarrow \mathrm{Hom}\big(\mathbb{Z}/d\mathbb{Z}, \mathbb{Z}/d\mathbb{Z}\big), \qquad 1 \longmapsto \mathrm{id},$$

is an epimorphism with kernel $d\mathbb{Z}$ and therefore induces an isomorphism, as desired. □

Theorem 3. *Let K be a field and $n > 0$ an integer such that $\mathrm{char}\, K \nmid n$ as well as $U_n \subset K^*$. The maps*

$$\left\{ \begin{array}{l} \text{subgroups } C \subset K^* \\ \text{such that } K^{*n} \subset C \end{array} \right\} \underset{\Psi}{\overset{\Phi}{\rightleftarrows}} \left\{ \begin{array}{l} \text{abelian extensions } L/K \\ \text{of exponents dividing } n \end{array} \right\},$$

$$\begin{array}{rcl} C & \longmapsto & K(C^{1/n}), \\ L^n \cap K^* & \longmapsfrom & L, \end{array}$$

are inclusion-preserving, bijective, and mutually inverse to each other.[7] *Furthermore, the Galois group G_C of an extension $K(C^{1/n})/K$ is characterized by the isomorphism*

[7] In order to be able to talk about the *set* of all abelian extensions of K, we consider these extensions as subfields of a chosen algebraic closure $\overline{K}$ of K.

$$\varphi_1 \colon G_C \longrightarrow \mathrm{Hom}\big(C/K^{*n}, U_n\big), \qquad \sigma \longmapsto \langle \sigma, \cdot \rangle,$$

of Proposition 1 (ii). *If* C/K^{*n} *is finite, then* $\mathrm{Hom}(C/K^{*n}, U_n)$ *and hence also* G_C *are (noncanonically) isomorphic to* C/K^{*n}.

Proof. Due to Proposition 1 and Lemma 2, it remains only to show that the maps Φ, Ψ are bijective and mutually inverse to each other. Starting with the relation $\Psi \circ \Phi = \mathrm{id}$, we consider a subgroup $C \subset K^*$ satisfying $C \supset K^{*n}$, where in a first step, we assume $(C : K^{*n}) < \infty$. Then $C' = (K(C^{1/n}))^n \cap K^*$ satisfies $C \subset C'$, and furthermore $K(C^{1/n}) = K(C'^{1/n})$. Applying Proposition 1 (iii) yields $C = C'$, as desired.

If the index $(C : K^{*n})$ is not necessarily finite, we can apply the preceding argument to all subgroups $C_i \subset C$ that are of finite index over K^{*n}. Since C is the union of these subgroups and since we have $K(C^{1/n}) = \bigcup_i K(C_i^{1/n})$, it follows also in the general case that $C = (K(C^{1/n}))^n \cap K^*$ and therefore $\Psi \circ \Phi = \mathrm{id}$.

To justify the remaining relation $\Phi \circ \Psi = \mathrm{id}$, we consider an abelian extension L/K of an exponent dividing n. Writing $C = L^n \cap K^*$, we get $K(C^{1/n}) \subset L$ and must show that these fields coincide. Since we can write L as a union of finite Galois and hence abelian extensions, we may assume that L/K is finite. Now look at the epimorphism

$$q \colon \mathrm{Gal}\big(L/K\big) \longrightarrow G_C, \qquad \sigma \longmapsto \sigma|_{K(C^{1/n})},$$

which exists by 4.1/2. It is enough to show that the associated homomorphism

$$q^* \colon \mathrm{Hom}\big(G_C, U_n\big) \longrightarrow \mathrm{Hom}\big(\mathrm{Gal}(L/K), U_n\big), \qquad f \longmapsto f \circ q,$$

is an isomorphism. Indeed, if such is the case, then the Galois groups of $K(C^{1/n})/K$ and L/K will have the same order due to Lemma 2, and we can conclude that $[L : K] = [K(C^{1/n}) : K]$, as well as $L = K(C^{1/n})$.

Of course, q^* is injective, since q is surjective. To see that q^* is surjective as well, consider a homomorphism $g \colon \mathrm{Gal}(L/K) \longrightarrow U_n$. Since it satisfies

$$g(\sigma \circ \sigma') = g(\sigma) \cdot g(\sigma') = \sigma \circ g(\sigma') \cdot g(\sigma)$$

for $\sigma, \sigma' \in \mathrm{Gal}(L/K)$, it is a 1-cocycle, in the terminology of Section 4.8. Then we can conclude from 4.8/2 that g is already a 1-coboundary, i.e., that there exists an element $a \in L^*$ such that $g(\sigma) = a \cdot \sigma(a)^{-1}$ for all $\sigma \in \mathrm{Gal}(L/K)$. Since $g(\sigma)^n = 1$ and therefore $\sigma(a^n) = \sigma(a)^n = a^n$ for all $\sigma \in \mathrm{Gal}(L/K)$, we conclude that $a^n \in C = L^n \cap K^*$ and hence that $a \in K(C^{1/n})$. Finally, look at the homomorphism

$$f \colon G_C \longrightarrow U_n, \qquad \sigma \longmapsto a \cdot \sigma(a)^{-1},$$

and observe that it satisfies $g = f \circ q = q^*(f)$. In particular, q^* is surjective. □

For an abelian extension L/K of some exponent dividing n, as dealt with in Theorem 3, we can easily specify a K-basis of L as follows. Write $C = L^n \cap K^*$

and consider a system $(c_i)_{i\in I}$ of elements in C giving rise to a system of representatives of C/K^{*n}. Then, choosing arbitrary nth roots of the c_i, the system $(c_i^{1/n})_{i\in I}$ is a K-basis of L/K. Indeed, it is clearly a generating system of L/K. Moreover, if $[L:K]<\infty$, then it consists of precisely $(C:K^{*n})=[L:K]$ elements and therefore is linearly independent as well. In the general case we can exhaust L by finite abelian extensions of K to see that $(c_i^{1/n})_{i\in I}$ is linearly independent and thus is a K-basis of L.

Exercises

Let K be a field, $\overline{K}$ an algebraic closure of K, and $n>0$ an integer such that $\operatorname{char} K\nmid n$ and K contains a primitive nth root of unity.

1. *Deduce the characterization of cyclic extensions of K, as given in* 4.8/3, *from Kummer theory.*
2. *Consider in $\overline{K}$ all abelian extensions L/K of exponents dividing n and show that there is a largest extension L_n/K among these. Give a characterization of the corresponding Galois group* $\operatorname{Gal}(L_n/K)$.
3. Set $K=\mathbb{Q}$ and $n=2$ in Exercise 2. Show that $L_2=\mathbb{Q}(i,\sqrt{2},\sqrt{3},\sqrt{5},\ldots)$ and determine the Galois group of the extension $L_2/\mathbb{Q}$.
4. For $c,c'\in K^*$, consider the splitting fields $L,L'\subset\overline{K}$ of the polynomials X^n-c and X^n-c' over K. Show that $L=L'$ is equivalent to the fact that there exists an integer $r\in\mathbb{N}$ that is relatively prime to n and satisfies $c^r\cdot c'\in K^{*n}$.
5. Show for every finite Galois extension L/K that there is a canonical isomorphism of groups
$$(L^n\cap K^*)/K^{*n}\xrightarrow{\sim}\operatorname{Hom}\big(\operatorname{Gal}(L/K),U_n\big).$$

4.10 General Kummer Theory and Witt Vectors*

In the preceding section we developed Kummer theory for a field K and an exponent n, where $\operatorname{char} K\nmid n$. In a similar way, one can study Kummer theory for a field K of characteristic $p>0$ and p as exponent. The resulting theory is referred to as *Artin–Schreier theory*. More generally, for $p=\operatorname{char} K>0$, there is a Kummer theory for exponents of type p^r, where $r\geq 1$, that goes back to E. Witt. All these Kummer theories are based on a common skeleton and are, so to speak, special cases of a general Kummer theory, which is the subject of the present section. Let K_s be a separable algebraic closure of K. For example, choose an algebraic closure of K and consider its subfield consisting of all elements that are separable over K, where for the moment, we do not require any restrictions on the characteristic of K. Then K_s/K is a Galois extension. The corresponding Galois group $G=\operatorname{Gal}(K_s/K)$ is referred to as the *absolute Galois group* of K. It is viewed as a topological group in the sense of Section 4.2.

Kummer theory. — As main ingredient, we need for a Kummer theory over K a continuous G-module A, where G is the absolute Galois group of K. By

a G-module we mean an abelian group A equipped with the discrete topology, together with a continuous G-action

$$G \times A \longrightarrow A, \qquad (\sigma, a) \longmapsto \sigma(a),$$

respecting the group law on A; for the definition of a group action one may consult 5.1/1 and 5.1/2. In particular, we can view such an action as the homomorphism $G \longrightarrow \mathrm{Aut}(A)$ mapping an element $\sigma \in G$ to the automorphism $\sigma(\cdot)$ of A. Also note that the continuity of the action means for every element $a \in A$ that the subgroup

$$G(A/a) = \big\{\sigma \in G\,;\, \sigma(a) = a\big\}$$

is open in G. According to 4.2/5, this is equivalent to the fact that $G(A/a)$ is closed in G and the fixed field $K_s^{G(A/a)}$ is finite over K.

As is known from the fundamental theorem of Galois theory 4.2/3, the intermediate fields of K_s/K correspond bijectively to the closed subgroups of G via the mapping $L \longmapsto \mathrm{Gal}(K_s/L)$. Therefore, we can associate to an intermediate field L of K_s/K, resp. to a closed subgroup $\mathrm{Gal}(K_s/L) \subset G$, the fixed group

$$A_L = \big\{a \in A\,;\, \sigma(a) = a \text{ for all } \sigma \in \mathrm{Gal}(K_s/L)\big\}.$$

If L is Galois over K, or equivalently, if $\mathrm{Gal}(K_s/L)$ is a normal subgroup in G, it is easily seen that the G-action on A restricts to a G-action on A_L. Thereby we get an action of $G/\mathrm{Gal}(K_s/L)$ on A_L, where due to 4.1/7, this quotient can be identified with $\mathrm{Gal}(L/K)$. Thus, given a Galois extension L/K, the G-action on A gives rise to an action of the corresponding Galois group $\mathrm{Gal}(L/K)$ on A_L, and we can define the cohomology group $H^1(\mathrm{Gal}(L/K), A_L)$ in the same way as we did in Section 4.8. An essential prerequisite of any kind of Kummer theory of a given exponent n is the condition that the cohomological version of Hilbert's Theorem 90 be valid for cyclic extensions, in the following form:

(**Hilbert 90**) *Let L/K be a cyclic Galois extension of a degree dividing n. Then $H^1(\mathrm{Gal}(L/K), A_L) = 0$.*

Of course, this condition is not automatically fulfilled; it serves, so to speak, as an axiom on which Kummer theory is based.

Having associated to an intermediate field L of K_s/K the fixed group A_L, there is another construction that goes in the reverse direction. Indeed, for a subset $\Delta \subset A$, consider the subgroup

$$G\big(A/\Delta\big) = \big\{\sigma \in G\,;\, \sigma(a) = a \text{ for all } a \in \Delta\big\}$$

of G. It is closed in G, since $G(A/\Delta) = \bigcap_{a\in\Delta} G(A/a)$, where all groups $G(A/a)$ are open and hence closed in G, due to the continuity of A as a G-module. Therefore, $G(A/\Delta)$ can be interpreted as the absolute Galois group of a well-defined intermediate field $K(\Delta)$ of K_s/K, namely of

$$K(\Delta) = K_s^{G(A/\Delta)} = \big\{\alpha \in K_s\,;\, \sigma(\alpha) = \alpha \text{ for all } \sigma \in G(A/\Delta)\big\}.$$

Kummer theory of a given exponent n relies, furthermore, on the choice of a surjective G-homomorphism $\wp\colon A \longrightarrow A$ whose kernel, denoted by μ_n in the following, is a cyclic group of order n satisfying $\mu_n \subset A_K$. Clearly, a G-homomorphism $\wp\colon A \longrightarrow A$ is meant as a homomorphism that is compatible with the G-action in the sense that $\sigma(\wp(a)) = \wp(\sigma(a))$ for all $\sigma \in G$ and $a \in A$.

In Section 4.9 we considered the multiplicative group $A = K_s^*$ under the natural action of the Galois group G and with $\wp\colon A \longrightarrow A$, $a \longmapsto a^n$, as a G-homomorphism, assuming $\operatorname{char} K \nmid n$. This implies that $\mu_n = \ker \wp$ coincides with the group U_n of nth roots of unity and therefore is cyclic of order n. Using the notation above, we have $A_L = L^*$ for intermediate fields L of K_s/K, as well as $K(\wp^{-1}(C)) = K(C^{1/n})$ for $C \subset K^*$. Also note that we assumed $U_n \subset K^*$ in Section 4.9 and thereby $\mu_n \subset A_K$. Finally, in 4.9/3 we derived a characterization of abelian extensions of exponents dividing n, in the style of the fundamental theorem of Galois theory, in fact, in terms of subgroups $C \subset A_K$ containing $\wp(A_K)$. The proof required Hilbert's Theorem 90 in the version of 4.8/2.

Now we want to show that the results 4.9/1 and 4.9/3 can be extended to the context of general Kummer theory. To do this, consider a subset $C \subset A_K$, as well as the subgroup $G(A/\wp^{-1}(C)) \subset G$ consisting of all $\sigma \in G$ that are trivial on $\wp^{-1}(C)$. Let $K(\wp^{-1}(C))$ be the corresponding intermediate field of K_s/K. Next, writing the group law on A additively, we see for $\sigma \in G$ and $a \in \wp^{-1}(C)$ that

$$\wp \circ \sigma(a) = \sigma \circ \wp(a) = \wp(a), \qquad \text{i.e.,} \qquad \sigma(a) - a \in \ker \wp = \mu_n.$$

Therefore, every $\sigma \in G$ restricts to a bijection $\wp^{-1}(C) \longrightarrow \wp^{-1}(C)$, and it follows that $G(A/\wp^{-1}(C))$ is a normal subgroup in G. Then $K(\wp^{-1}(C))/K$ is a Galois extension, due to 4.2/3, resp. 4.1/7, and even an abelian extension, as we will see further below. Let G_C be the corresponding Galois group. By 4.1/7, it can be identified with the quotient $G/G(A/\wp^{-1}(C))$.

For $c \in C$ and $a \in \wp^{-1}(c)$, the difference $\sigma(a) - a \in \mu_n$ will in general depend on c, but not on the choice of a special preimage $a \in \wp^{-1}(c)$. Indeed, if we consider another preimage $a' \in \wp^{-1}(c)$, say $a' = a+i$ for some $i \in \ker \wp = \mu_n$, then

$$\sigma(a') - a' = \big(\sigma(a) + \sigma(i)\big) - (a + i) = \sigma(a) - a.$$

Therefore, the map

$$\langle \cdot, \cdot \rangle\colon G_C \times C \longrightarrow \mu_n, \qquad (\sigma, c) \longmapsto \sigma(a) - a, \qquad \text{where } a \in \wp^{-1}(c),$$

is well defined. Restricting to subgroups $C \subset A_K$ such that $\wp(A_K) \subset C$, we obtain similarly as in Section 4.9 a pairing

$$\langle \cdot, \cdot \rangle\colon G_C \times C/\wp(A_K) \longrightarrow \mu_n, \qquad (\sigma, \overline{c}) \longmapsto \sigma(a) - a, \qquad \text{where } a \in \wp^{-1}(c),$$

that is homomorphic in both variables.

Theorem 1. *Let K be a field and G its absolute Galois group. Consider a continuous G-module A together with a surjective G-homomorphism $\wp\colon A \longrightarrow A$, whose kernel μ_n is a finite cyclic subgroup in A_K of order n. Assume for every cyclic Galois extension L/K of degree dividing n that $H^1(\mathrm{Gal}(L/K), A_L) = 0$. Then*:

(i) *Viewing the abelian extensions of K as subfields of K_s, the maps*

$$\left\{\begin{array}{c}\text{subgroups } C \subset A_K \\ \text{such that } \wp(A_K) \subset C\end{array}\right\} \underset{\Psi}{\overset{\Phi}{\rightleftarrows}} \left\{\begin{array}{c}\text{abelian extensions } L/K \\ \text{of exponents dividing } n\end{array}\right\},$$

$$\begin{aligned} C &\longmapsto K\big(\wp^{-1}(C)\big), \\ \wp(A_L) \cap A_K &\longmapsfrom L, \end{aligned}$$

are inclusion-preserving, bijective, and mutually inverse to each other.

(ii) *For subgroups $C \subset A_K$ such that $\wp(A_K) \subset C$, the bihomomorphic map*

$$\langle \cdot, \cdot \rangle\colon G_C \times C/\wp(A_K) \longrightarrow \mu_n, \qquad (\sigma, \overline{c}) \longmapsto \sigma(a) - a, \qquad \text{where } a \in \wp^{-1}(c),$$

is nondegenerate in the sense that it gives rise to monomorphisms

$$\begin{aligned} \varphi_1\colon \quad & G_C \longrightarrow \mathrm{Hom}\big(C/\wp(A_K), \mu_n\big), && \sigma \longmapsto \langle \sigma, \cdot \rangle, \\ \varphi_2\colon \quad & C/\wp(A_K) \longrightarrow \mathrm{Hom}\big(G_C, \mu_n\big), && \overline{c} \longmapsto \langle \cdot, \overline{c} \rangle. \end{aligned}$$

More precisely, φ_1 is an isomorphism, and φ_2 restricts to an isomorphism $C/\wp(A_K) \xrightarrow{\sim} \mathrm{Hom}_{\mathrm{cont}}(G_C, \mu_n)$ onto the group of all continuous homomorphisms $G_C \longrightarrow \mu_n$.

(iii) *The extension $K(\wp^{-1}(C))/K$ is finite if and only if $(C : \wp(A_K))$, the index of $\wp(A_K)$ in C, is finite. If such is the case, both maps φ_1 and φ_2 of* (ii) *are isomorphisms, and one obtains $[K(\wp^{-1}(C)) : K] = (C : \wp(A_K))$.*

Proof. Similarly as in the proof of 4.9/1, we start by showing that φ_1 and φ_2 are injective. To do this, consider an element $\sigma \in G_C$ such that $\langle \sigma, \overline{c} \rangle = 0$ for all $c \in C$. Then we see that $\sigma(a) = a$ for all $a \in \wp^{-1}(C)$, and if we choose a representative $\sigma' \in G$ of σ, that $\sigma'(a) = a$ for all $a \in \wp^{-1}(C)$. However, this implies $\sigma' \in G(A/\wp^{-1}(C))$ and hence that σ is trivial. Therefore, φ_1 is injective. On the other hand, consider an element $c \in C$ such that $\langle \sigma, \overline{c} \rangle = 0$ for all $\sigma \in G_C$, i.e., such that $\sigma(a) - a = 0$ for all $\sigma \in G_C$ and for preimages $a \in \wp^{-1}(c)$. Then every such a is invariant under G_C, resp. G, so that $a \in A_K$ and hence $c = \wp(a) \in \wp(A_K)$. Thus, φ_2 is injective. As a by-product, the injectivity of φ_1 shows that the map Φ in (i) is well defined. Indeed, for a subgroup $C \subset A_K$ satisfying $C \supset \wp(A_K)$, we see that the extension $K(\wp^{-1}(C))/K$ is abelian of some exponent dividing n.

Next, turning to the assertions of (iii), observe that the exponents of G_C and $C/\wp(A_K)$ divide n, due to the injectivity of φ_1 and φ_2. Therefore, we can proceed in literally the same way as in the proof of 4.9/1 (iii), including the application of the auxiliary Lemma 4.9/2. Furthermore, to derive from (iii) the

isomorphism properties of φ_1 and φ_2 stated in (ii), we consider the system $(C_i)_{i\in I}$ of all subgroups in C that are of finite index over $\wp(A_K)$. Then

$$C = \sum_{i\in I} C_i, \qquad G\big(A/\wp^{-1}(C)\big) = \bigcap_{i\in I} G\big(A/\wp^{-1}(C_i)\big),$$

and therefore

$$G\Big(K_s/K\big(\wp^{-1}(C)\big)\Big) = \bigcap_{i\in I} G\Big(K_s/K\big(\wp^{-1}(C_i)\big)\Big),$$

so that we can interpret $K(\wp^{-1}(C))$ as the composite field of the subfields $K(\wp^{-1}(C_i))$. Now observe that the system $(C_i)_{i\in I}$ is directed, i.e., for $i, j \in I$ there is always an index $k \in I$ such that $C_i, C_j \subset C_k$. Hence, we can even write $K(\wp^{-1}(C)) = \bigcup_{i\in I} K(\wp^{-1}(C_i))$. As a result, the argument given in the proof of 4.9/1 (ii) carries over to the present situation and yields the isomorphism properties stated in (ii) for φ_1 and φ_2.

Also concerning assertion (i), we base our argument on the corresponding approach given in Section 4.9, notably in the proof of 4.9/3. Starting with the equation $\Psi \circ \Phi = \mathrm{id}$, consider a subgroup $C \subset A_K$ such that $C \supset \wp(A_K)$, and let $L = K(\wp^{-1}(C))$. We have to show that $C' = \wp(A_L) \cap A_K$ coincides with C. By its definition, $A_L \subset A$ is the fixed group of $G(A/\wp^{-1}(C))$, so that we get $\wp^{-1}(C) \subset A_L$ and therefore $C \subset \wp(A_L) \cap A_K = C'$. Moreover, since $G(A/A_L) = G(A/\wp^{-1}(C))$, we conclude that

$$L = K\big(\wp^{-1}(C)\big) \subset K\big(\wp^{-1}(C')\big) \subset K(A_L) = L,$$

and therefore $L = K(\wp^{-1}(C)) = K(\wp^{-1}(C'))$. Now if C is of finite index over $\wp(A_K)$, we obtain $C = C'$ directly from (iii). Otherwise, consider again the directed system $(C_i)_{i\in I}$ of all subgroups in C that are of finite index over $\wp(A_K)$. As we have seen, the system of all fields $L_i = K(\wp^{-1}(C_i))$ is directed as well, and we get $L = \bigcup_{i\in I} L_i$.

Moreover, we claim that

$$(*) \qquad A_L = \bigcup_{i\in I} A_{L_i}.$$

Of course, we have $A_L \supset \bigcup_{i\in I} A_{L_i}$. To derive the reverse inclusion, consider an element $a \in A_L$, as well as the corresponding subgroup $G(A/a) \subset G$ leaving a fixed. This subgroup is open in G, since the action of G on A is continuous. Using 4.2/5, we see that $G(A/a)$ corresponds to an intermediate field E of K_s/K that is finite over K. We have even $E \subset L$, since $G(A/a) \supset G(A/A_L)$, where the group $G(A/A_L)$ coincides with $G(A/\wp^{-1}(C))$. Since the system $(L_i)_{i\in I}$ is directed, there is an index $j \in I$ such that $E \subset L_j$. In particular, this implies

$$a \in A_E \subset A_{L_j} \subset \bigcup_{i\in I} A_{L_i}$$

and thereby the equation $(*)$ above.

Now observe that $\wp(A_{L_i}) \cap A_K = C_i$ for all i, since the indices $(C_i : \wp(A_K))$ are finite. Using $(*)$, this implies $\wp(A_L) \cap A_K = C$ and hence $\Psi \circ \Phi = \mathrm{id}$.

To justify the relation $\Phi \circ \Psi = \mathrm{id}$, consider an abelian extension L/K of some exponent dividing n. Writing $C = \wp(A_L) \cap A_K$, we get $\wp^{-1}(C) \subset A_L$. In particular, $\wp^{-1}(C)$ is fixed by $\mathrm{Gal}(K_s/L)$, which implies $K(\wp^{-1}(C)) \subset L$, and we have to show that this inclusion is, in fact, an equality. To achieve this, interpret L as the composite field of finite and necessarily abelian Galois extensions L'/K. Each of these extensions L'/K can be written as the composite field of finitely many cyclic extensions. To justify this, it is enough to specify subgroups H_j of the Galois group $H = \mathrm{Gal}(L'/K)$ such that in each case, H/H_j is cyclic and such that $\bigcap_j H_j = \{1\}$. Applying the fundamental theorem of finitely generated abelian groups 2.9/9 shows that this is indeed possible. Therefore, L is the composite field of a family $(L_i)_{i\in I}$ of finite *cyclic* extensions, and it is clearly enough to show that $L_i \subset K(\wp^{-1}(C_i))$ for $C_i = \wp(A_{L_i}) \cap A_K$. In other words, we may assume L/K to be a finite cyclic extension of some exponent dividing n.

Therefore, let L/K be such an extension and consider for $C = \wp(A_L) \cap A_K$ the epimorphism

$$q\colon \mathrm{Gal}\big(L/K\big) \longrightarrow G_C, \qquad \sigma \longmapsto \sigma|_{K(\wp^{-1}(C))},$$

as well as the corresponding homomorphism

$$q^*\colon \mathrm{Hom}\big(G_C, \mu_n\big) \longrightarrow \mathrm{Hom}\big(\mathrm{Gal}(L/K), \mu_n\big), \qquad f \longmapsto f \circ q.$$

It is enough to show that q^* is an isomorphism, since 4.9/2 then implies $\mathrm{ord}\,\mathrm{Gal}(L/K) = \mathrm{ord}\,G_C$ and hence $L = K(\wp^{-1}(C))$.

Of course q^* is injective, since q is surjective. To see that q^* is surjective, fix a homomorphism $g\colon \mathrm{Gal}(L/K) \longrightarrow \mu_n$. Then the relation

$$g(\sigma \circ \sigma') = g(\sigma) + g(\sigma') = \sigma \circ g(\sigma') + g(\sigma), \qquad \sigma, \sigma' \in \mathrm{Gal}\big(L/K\big),$$

shows that g is a 1-cocycle with respect to the action of $\mathrm{Gal}(L/K)$ on A_L. Hence, it is a 1-coboundary as well, due to our assumption on $H^1(\mathrm{Gal}(L/K), A_L)$. Thus, there is an element $a \in A_L$ such that $g(\sigma) = a - \sigma(a)$ for all $\sigma \in \mathrm{Gal}(L/K)$. Now, using $\ker \wp = \mu_n$, we get $\sigma \circ \wp(a) = \wp \circ \sigma(a) = \wp(a)$ for $\sigma \in \mathrm{Gal}(L/K)$, which implies $\wp(a) \in \wp(A_L) \cap A_K = C$. But then we can look at the homomorphism

$$f\colon G_C \longrightarrow \mu_n, \qquad \sigma \longmapsto a - \sigma(a),$$

which satisfies $g = f \circ q = q^*(f)$, and it follows that q^* is surjective. □

As a first example in which Theorem 1 is applicable, we studied in Section 4.9 Kummer theory for an exponent n not divisible by the characteristic of the field K under consideration. From now on we assume $p = \mathrm{char}\,K > 0$ in order to develop Kummer theory for exponents of type $n = p^r$. The case $n = p$ (Artin–Schreier theory) is quite simple. Here one considers the additive group $A = K_s$ with the canonical action of G as a G-module, together with

$$\wp\colon A \longrightarrow A, \qquad a \longmapsto a^p - a,$$

as a G-homomorphism. Then $\mu_p = \ker \wp$ equals the prime field in $A_K = K$ and thus is a cyclic subgroup in A_K of order p, as required. To make Theorem 1 applicable, it remains only to establish Hilbert's Theorem 90. We will do this on a more general scale in Proposition 11 below.

Witt vectors. — Kummer theory in characteristic $p > 0$ for general exponents $n = p^r$, $r \geq 1$, is quite involved and relies on the formalism of *Witt vectors*, introduced by E. Witt, a theory that we want to present now. Given a prime number p, the Witt vectors with coefficients in a ring R form a ring $W(R)$, the *Witt ring* on R. Characterizing $W(R)$ as a *set*, it is given by $W(R) = R^{\mathbb{N}}$, the countably infinite Cartesian product of R with itself. However, the sum and the product of two elements $x, y \in W(R)$ are defined in a nonstandard way by expressions of type

$$x + y = \big(S_n(x,y)\big)_{n\in\mathbb{N}}, \qquad x \cdot y = \big(P_n(x,y)\big)_{n\in\mathbb{N}},$$

where $S_n(x,y), P_n(x,y)$ for $n \in \mathbb{N}$ are polynomials in $x_0, \ldots, x_n$ and $y_0, \ldots, y_n$ with coefficients in $\mathbb{Z}$, hence polynomials in the first $n+1$ components of x, resp. y.[8] If $p = p \cdot 1$ is invertible in R, we will see that $W(R)$, as a ring, is isomorphic to $R^{\mathbb{N}}$ with componentwise addition and multiplication.

To set up the polynomials $S_n, P_n \in \mathbb{Z}[X_0, \ldots, X_n, Y_0, \ldots, Y_n]$ for $n \in \mathbb{N}$, consider the *Witt polynomials*

$$W_n = \sum_{i=0}^{n} p^i X_i^{p^{n-i}} = X_0^{p^n} + pX_1^{p^{n-1}} + \ldots + p^n X_n \in \mathbb{Z}[X_0, \ldots, X_n].$$

They satisfy the recurrence formulas

$$(*) \qquad W_n = W_{n-1}(X_0^p, \ldots, X_{n-1}^p) + p^n X_n, \qquad n > 0,$$

and it is seen by induction that in each case, X_n can be written as a polynomial in $W_0, \ldots, W_n$ with coefficients in $\mathbb{Z}[\frac{1}{p}]$, say

$$X_0 = W_0, \qquad X_1 = p^{-1}W_1 - p^{-1}W_0^p, \qquad \ldots \ .$$

Lemma 2. *The substitution endomorphism*

$$\omega_n\colon \mathbb{Z}[\tfrac{1}{p}][X_0, \ldots, X_n] \longrightarrow \mathbb{Z}[\tfrac{1}{p}][X_0, \ldots, X_n],$$
$$f(X_0, \ldots, X_n) \longmapsto f(W_0, \ldots, W_n),$$

mapping the variables $X_0, \ldots, X_n$ to the polynomials $W_0, \ldots, W_n$, is bijective. In particular, the maps ω_n, $n \in \mathbb{N}$, give rise to an automorphism

[8] The multiplication of elements in R by integers in $\mathbb{Z}$ is defined in the usual way, for example, with the aid of the canonical homomorphism $\mathbb{Z} \longrightarrow R$.

$$\omega\colon \mathbb{Z}[\tfrac{1}{p}][X_0, X_1, \dots] \longrightarrow \mathbb{Z}[\tfrac{1}{p}][X_0, X_1, \dots],$$
$$f(X_0, X_1, \dots) \longmapsto f(W_0, W_1, \dots).$$

Proof. Indeed, ω_n is surjective, since each of the variables $X_0, \dots, X_n$ can be written as a polynomial in $W_0, \dots, W_n$. But then ω_n is injective as well for general reasons; for example, extend coefficients from $\mathbb{Z}[\frac{1}{p}]$ to $\mathbb{Q}$ and apply 7.1/9.

Let us add here an alternative argument showing that ω_n is injective. We proceed by induction on n. The case $n = 0$ is trivial, since $W_0 = X_0$. Therefore, assume $n > 0$ and let

$$f = \sum_{i=0}^{r} f_i \cdot X_n^i, \qquad f_i \in \mathbb{Z}[\tfrac{1}{p}][X_0, \dots, X_{n-1}],$$

be a nontrivial polynomial in $X_0, \dots, X_n$ with coefficients in $\mathbb{Z}[\frac{1}{p}]$ such that $f_r \neq 0$. Then we obtain

$$\omega_n(f) = \sum_{i=0}^{r} f_i(W_0, \dots, W_{n-1}) \cdot W_n^i,$$

where each $f_i(W_0, \dots, W_{n-1})$ is a polynomial in $X_0, \dots, X_{n-1}$ and where furthermore, $f_r(W_0, \dots, W_{n-1})$ is nonzero by the induction hypothesis. Now write $\omega_n(f)$ as a polynomial in X_n with coefficients in $\mathbb{Z}[\frac{1}{p}][X_0, \dots, X_{n-1}]$. Since $p^n X_n$ is the only term in W_n containing the variable X_n, the leading term of $\omega_n(f)$ turns out to be $p^{nr} f_r(W_0, \dots, W_{n-1}) \cdot X_n^r$, and we get $\omega_n(f) \neq 0$. Therefore, ω_n is injective. □

In most cases, we will view the polynomials W_n as elements of the polynomial ring $\mathbb{Z}[X_0, X_1, \dots]$, although they actually are polynomials in finitely many variables. Proceeding like this, the values $W_n(x)$ will be meaningful for points $x \in R^{\mathbb{N}}$, for any ring R.

Lemma 3. *Assume that p is invertible in R. Then the map*

$$w\colon W(R) = R^{\mathbb{N}} \longrightarrow R^{\mathbb{N}}, \qquad x \longmapsto \big(W_n(x)\big)_{n \in \mathbb{N}},$$

is bijective.

Proof. The homomorphisms ω_n and ω of Lemma 2 are substitution homomorphisms, and the same is true for their inverses ω_n^{-1} and ω^{-1}, due to the universal property of polynomial rings dealt with in 2.5/5, resp. 2.5/1. Therefore, there exist polynomials $\widetilde{W}_n \in \mathbb{Z}[\frac{1}{p}][X_0, \dots, X_n]$, $n \in \mathbb{N}$, such that

$$W_n(\widetilde{W}_0, \dots, \widetilde{W}_n) = X_n, \qquad \widetilde{W}_n(W_0, \dots, W_n) = X_n$$

for all n. Since p is invertible in R, the canonical homomorphism $\mathbb{Z} \longrightarrow R$ extends (uniquely) to a homomorphism $\mathbb{Z}[\frac{1}{p}] \longrightarrow R$. Furthermore, the preceding relations remain valid if we replace $\mathbb{Z}[\frac{1}{p}]$ by R as coefficient ring. As a consequence, the map w admits an inverse and therefore is bijective.

Alternatively, we can view the map

$$R^{\mathbb{N}} \longrightarrow \mathrm{Hom}\Big(\mathbb{Z}[\tfrac{1}{p}][X_0, X_1, \dots], R\Big), \qquad x \longmapsto \big(f \longmapsto f(x)\big),$$

as an identification and then interpret $w\colon R^{\mathbb{N}} \longrightarrow R^{\mathbb{N}}$ as the map

$$\mathrm{Hom}\Big(\mathbb{Z}[\tfrac{1}{p}][X_0, X_1, \dots], R\Big) \longrightarrow \mathrm{Hom}\Big(\mathbb{Z}[\tfrac{1}{p}][X_0, X_1, \dots], R\Big),$$
$$\varphi \longmapsto \varphi \circ \omega,$$

which is induced by the isomorphism ω of Lemma 2; here $\mathrm{Hom}(C, R)$, for rings C and R, means the set of all ring homomorphisms $C \longrightarrow R$. □

Now consider a ring R such that p is invertible in R. Then, departing from $R^{\mathbb{N}}$ as a ring with componentwise addition "$+_c$" and componentwise multiplication "$\cdot_c$", we can introduce laws of composition "$+$" and "$\cdot$" on $W(R)$ by means of the formulas

$$x + y = w^{-1}\big(w(x) +_c w(y)\big), \qquad x \cdot y = w^{-1}\big(w(x) \cdot_c w(y)\big).$$

It is immediately clear that $W(R)$ is a ring under these laws. Indeed, addition and multiplication on $W(R)$ are defined in such a way that the map $w\colon W(R) \longrightarrow R^{\mathbb{N}}$ becomes an isomorphism of rings.

It can easily be checked that the nth components of a sum $x + y$ or of a product $x \cdot y$ of elements $x, y \in W(R)$ depend in a polynomial way on the ith components of x and y, where $i \leq n$. More precisely, w is given in terms of polynomial expressions with coefficients in $\mathbb{Z}$, and similarly, w^{-1} is given by expressions of the same type with coefficients in $\mathbb{Z}[\frac{1}{p}]$, so that all in all, coefficients in $\mathbb{Z}[\frac{1}{p}]$ are needed. However, we will see at once that coefficients from $\mathbb{Z}$ are sufficient to characterize the laws of composition "$+$" and "$\cdot$" on $W(R)$. This will enable us to define the Witt ring $W(R)$ also for those rings R in which p is not invertible. To explain this we need an auxiliary assertion on Witt polynomials.

Lemma 4. *Let R be a ring such that $p = p \cdot 1$ is not a zero divisor in R. The following conditions are equivalent for elements $a_0, \dots, a_n, b_0, \dots, b_n \in R$ and $r \in \mathbb{N} - \{0\}$:*

(i) $a_i \equiv b_i \mod (p^r)$ *for* $i = 0, \dots, n$.

(ii) $W_i(a_0, \dots, a_i) \equiv W_i(b_0, \dots, b_i) \mod (p^{r+i})$ *for* $i = 0, \dots, n$.

Proof. We proceed by induction on n, the case $n = 0$ being trivial. Therefore, assume $n > 0$. Conditions (i) and (ii) are equivalent for $n-1$ in place of n by the induction hypothesis. Therefore, if one of the conditions (i) and (ii) holds, we

may in either case assume that both conditions are satisfied for $i = 0, \dots, n-1$. Taking the pth power of the congruences in (i) yields

$$a_i^p \equiv b_i^p \mod (p^{r+1}), \qquad i = 0, \dots, n-1,$$

since $r \geq 1$ and p divides the binomial coefficients $\binom{p}{1}, \dots, \binom{p}{p-1}$. In particular, using the induction hypothesis we get

$$W_{n-1}(a_0^p, \dots, a_{n-1}^p) \equiv W_{n-1}(b_0^p, \dots, b_{n-1}^p) \mod (p^{r+n}),$$

and by the recurrence formulas $(*)$,

$$W_n(a_0, \dots, a_n) - W_n(b_0, \dots, b_n) \equiv p^n a_n - p^n b_n \mod (p^{r+n}).$$

Therefore, the congruence

$$W_n(a_0, \dots, a_n) \equiv W_n(b_0, \dots, b_n) \mod (p^{r+n})$$

is equivalent to $p^n a_n \equiv p^n b_n \mod (p^{r+n})$, hence to $a_n \equiv b_n \mod (p^r)$, since p is not a zero divisor in R. □

Lemma 5. *Let $\Phi \in \mathbb{Z}[\zeta, \xi]$ be a polynomial in two variables ζ and ξ. Then there exist unique polynomials $\varphi_n \in \mathbb{Z}[X_0, \dots, X_n, Y_0, \dots, Y_n]$, $n \in \mathbb{N}$, such that*

$$W_n(\varphi_0, \dots, \varphi_n) = \Phi\big(W_n(X_0, \dots, X_n), W_n(Y_0, \dots, Y_n)\big)$$

for all n.

Proof. Set $\mathfrak{X} = (X_0, X_1, \dots)$, as well as $\mathfrak{Y} = (Y_0, Y_1, \dots)$, and consider the commutative diagram

$$\begin{array}{ccc} \mathbb{Z}[\frac{1}{p}][\mathfrak{X}] & \xrightarrow{\omega} & \mathbb{Z}[\frac{1}{p}][\mathfrak{X}] \\ {\scriptstyle\tau}\downarrow & & \downarrow{\scriptstyle\tau'} \\ \mathbb{Z}[\frac{1}{p}][\mathfrak{X}, \mathfrak{Y}] & \xrightarrow{\omega\otimes\omega} & \mathbb{Z}[\frac{1}{p}][\mathfrak{X}, \mathfrak{Y}] \end{array}$$

that is determined by

$$\begin{aligned} \omega: &\quad X_n \longmapsto W_n, \\ \omega\otimes\omega: &\quad X_n \longmapsto W_n(X_0, \dots, X_n),\ Y_n \longmapsto W_n(Y_0, \dots, Y_n), \\ \tau: &\quad X_n \longmapsto \Phi(X_n, Y_n), \\ \tau': &\quad = (\omega\otimes\omega) \circ \tau \circ \omega^{-1}. \end{aligned}$$

For this, observe that ω is an isomorphism by Lemma 2 and that the same is true for $\omega \otimes \omega$. In particular, τ' is well defined as the unique homomorphism making the diagram commutative, i.e., such that the equation $\tau' \circ \omega = (\omega \otimes \omega) \circ \tau$ holds. Now set $\varphi_n = \tau'(X_n)$ for $n \in \mathbb{N}$. Thereby we get unique polynomials $\varphi_n \in \mathbb{Z}[\frac{1}{p}][X_0, \dots, X_n, Y_0, \dots, Y_n]$ satisfying

$$W_n(\varphi_0,\dots,\varphi_n) = \Phi\big(W_n(X_0,\dots,X_n), W_n(Y_0,\dots,Y_n)\big), \qquad n \in \mathbb{N}.$$

In particular, in order to verify the assertion of the lemma, it remains to show that all polynomials φ_n have coefficients in $\mathbb{Z}$.

To justify this claim we use induction on n. If $n = 0$, we have $W_0 = X_0$ and therefore $\varphi_0 = \Phi$, so that φ_0 admits coefficients in $\mathbb{Z}$. Now let $n > 0$. We may assume by the induction hypothesis that $\varphi_0,\dots,\varphi_{n-1}$ have coefficients in $\mathbb{Z}$. Furthermore, consider the element

$$W_n(\varphi_0,\dots,\varphi_n) = \tau' \circ \omega(X_n) = (\omega \otimes \omega) \circ \tau(X_n),$$

which, by the definition of ω and τ, is a polynomial in $\mathfrak{X}$ and $\mathfrak{Y}$ with coefficients in $\mathbb{Z}$. Using the induction hypothesis, the same is true for $W_{n-1}(\varphi_0^p,\dots,\varphi_{n-1}^p)$, and the recurrence formula $(*)$ yields

$$W_n(\varphi_0,\dots,\varphi_n) = W_{n-1}(\varphi_0^p,\dots,\varphi_{n-1}^p) + p^n\varphi_n.$$

Thus, to show that φ_n admits coefficients in $\mathbb{Z}$, it is enough to show that

$$W_n(\varphi_0,\dots,\varphi_n) \equiv W_{n-1}(\varphi_0^p,\dots,\varphi_{n-1}^p) \mod (p^n).$$

Every polynomial $\varphi \in \mathbb{Z}[\mathfrak{X},\mathfrak{Y}]$ satisfies $\varphi^p \equiv \varphi(\mathfrak{X}^p,\mathfrak{Y}^p) \mod (p)$, as is easily seen by applying the reduction homomorphism $\mathbb{Z}[\mathfrak{X},\mathfrak{Y}] \longrightarrow \mathbb{F}_p[\mathfrak{X},\mathfrak{Y}]$ and using 3.1/3; note that $\mathfrak{X}^p$, resp. $\mathfrak{Y}^p$, is meant as the system of all pth powers of the components of $\mathfrak{X}$, resp. $\mathfrak{Y}$. In particular, we have

$$\varphi_i^p \equiv \varphi_i(\mathfrak{X}^p,\mathfrak{Y}^p) \mod (p), \qquad i = 0,\dots,n-1,$$

which implies by means of Lemma 4 that

$$W_{n-1}(\varphi_0^p,\dots,\varphi_{n-1}^p) \equiv W_{n-1}\big(\varphi_0(\mathfrak{X}^p,\mathfrak{Y}^p),\dots,\varphi_{n-1}(\mathfrak{X}^p,\mathfrak{Y}^p)\big) \mod (p^n).$$

Now, combining the commutativity of the above diagram with the recurrence formula $(*)$, we get the following congruence modulo (p^n):

$$\begin{aligned} W_n(\varphi_0,\dots,\varphi_n) &= \Phi\big(W_n(\mathfrak{X}), W_n(\mathfrak{Y})\big) \\ &\equiv \Phi\big(W_{n-1}(\mathfrak{X}^p), W_{n-1}(\mathfrak{Y}^p)\big) \\ &= W_{n-1}\big(\varphi_0(\mathfrak{X}^p,\mathfrak{Y}^p),\dots,\varphi_{n-1}(\mathfrak{X}^p,\mathfrak{Y}^p)\big) \\ &\equiv W_{n-1}(\varphi_0^p,\dots,\varphi_{n-1}^p). \end{aligned}$$

This shows, as explained before, that φ_n admits coefficients in $\mathbb{Z}$. □

Applying Lemma 5 to the polynomials

$$\Phi(\zeta,\xi) = \zeta + \xi, \quad \text{resp.} \quad \Phi(\zeta,\xi) = \zeta \cdot \xi,$$

we obtain corresponding polynomials

$$S_n, P_n \in \mathbb{Z}[X_0, \ldots, X_n, Y_0, \ldots, Y_n], \qquad n \in \mathbb{N},$$

in place of the φ_n, where, for example,

$$\begin{aligned} S_0 &= X_0 + Y_0, & S_1 &= X_1 + Y_1 + \tfrac{1}{p}\big(X_0^p + Y_0^p - (X_0 + Y_0)^p\big), \\ P_0 &= X_0 \cdot Y_0, & P_1 &= X_1 Y_0^p + X_0^p Y_1 + p X_1 Y_1. \end{aligned}$$

We will utilize the polynomials S_n, P_n in order to define the addition and multiplication in Witt rings $W(R)$ over arbitrary rings R. As already mentioned, $W(R)$ as a set equals the Cartesian product $R^{\mathbb{N}}$. Therefore, we can introduce laws of composition on $W(R)$ by setting

$$x + y = \big(S_n(x, y)\big)_{n \in \mathbb{N}}, \qquad x \cdot y = \big(P_n(x, y)\big)_{n \in \mathbb{N}}, \qquad x, y \in W(R).$$

Looking at Lemma 5, we see that the map

$$w \colon W(R) \longrightarrow R^{\mathbb{N}}, \qquad x \longmapsto \big(W_n(x)\big)_{n \in \mathbb{N}},$$

satisfies the compatibility relations

$$w(x + y) = w(x) +_c w(y), \qquad w(x \cdot y) = w(x) \cdot_c w(y), \qquad x, y \in W(R),$$

and we can state that w is a ring homomorphism, provided we know that $W(R)$ is a ring with respect to the laws of composition given by "$+$" and "$\cdot$". On the other hand, if p is invertible in R, we know from Lemma 3 that w is bijective. In such a case we get

$$x + y = w^{-1}\big(w(x) +_c w(y)\big), \qquad x \cdot y = w^{-1}\big(w(x) \cdot_c w(y)\big), \qquad x, y \in W(R),$$

and it follows, as already explained, that $W(R)$ is indeed a ring with respect to the laws of composition "$+$" and "$\cdot$". In particular, $w \colon W(R) \longrightarrow R^{\mathbb{N}}$ is a ring isomorphism in such a case.

Proposition 6. *Consider an arbitrary ring R. Then the laws of composition "$+$" and "$\cdot$" derived from the polynomials S_n, P_n as given above make $W(R)$ a ring with the following properties*:

(i) $(0, 0, \ldots) \in W(R)$ *is the zero element, and* $(1, 0, 0, \ldots) \in W(R)$ *the unit element.*

(ii) $w \colon W(R) \longrightarrow R^{\mathbb{N}}$, $x \longmapsto (W_n(x))_{i \in \mathbb{N}}$, *is a ring homomorphism, even a ring isomorphism if p is invertible in R.*

(iii) *For every ring homomorphism $f \colon R \longrightarrow R'$, the induced map*

$$W(f) \colon W(R) \longrightarrow W(R'), \qquad (a_n)_{n \in \mathbb{N}} \longmapsto \big(f(a_n)\big)_{n \in \mathbb{N}},$$

is a ring homomorphism as well.

Proof. First, assume that p is invertible in R. Then, as we have seen, $W(R)$ is a ring and $w \colon W(R) \longrightarrow R^{\mathbb{N}}$ is an isomorphism of rings. Since apparently

$$w(0,0,\ldots) = (0,0,\ldots), \qquad w(1,0,0,\ldots) = (1,1,1,\ldots),$$

it follows that $(0,0,\ldots)$ is the zero element and $(1,0,0,\ldots)$ the unit element in $W(R)$.

Next we look at the particular case that $R = \mathbb{Z}[\frac{1}{p}][\mathfrak{X},\mathfrak{Y},\mathfrak{Z}]$ for countably infinite systems of variables

$$\mathfrak{X} = (X_0, X_1, \ldots), \qquad \mathfrak{Y} = (Y_0, Y_1, \ldots), \qquad \mathfrak{Z} = (Z_0, Z_1, \ldots).$$

Then $\mathfrak{X},\mathfrak{Y},\mathfrak{Z}$ may be viewed as elements of $W(R)$, and the associativity conditions

$$(\mathfrak{X}+\mathfrak{Y})+\mathfrak{Z} = \mathfrak{X}+(\mathfrak{Y}+\mathfrak{Z}), \qquad (\mathfrak{X}\cdot\mathfrak{Y})\cdot\mathfrak{Z} = \mathfrak{X}\cdot(\mathfrak{Y}\cdot\mathfrak{Z})$$

represent certain polynomial identities among the S_n, resp. P_n, where only coefficients from $\mathbb{Z}$ are involved. As a consequence, these identities must hold already in the polynomial ring $\mathbb{Z}[\mathfrak{X},\mathfrak{Y},\mathfrak{Z}]$. Concerning further ring axioms, we can proceed in a similar way. For example, we can apply Lemma 5 for $\Phi(\zeta,\xi) = -\zeta$ in order to see that the addition in $W(R)$ admits a process of forming inverses, defined in terms of coefficients in $\mathbb{Z}$. As a consequence, it follows that the laws of composition "$+$" and "$\cdot$" satisfy, so to speak in a generic way, the axioms of a ring structure on $W(R)$, where they are represented by certain formal polynomial identities with coefficients in $\mathbb{Z}$. If we substitute the variables by values of an arbitrary ring R, the laws of composition "$+$" and "$\cdot$" retain the properties of a ring structure, and we can conclude that $W(R)$ is ring, regardless of the type of underlying ring R.

It still remains to verify assertion (iii). Using the universal property of polynomial rings as proven in 2.5/1, we can identify $W(R)$ for arbitrary rings R with the set $\mathrm{Hom}(\mathbb{Z}[\mathfrak{X}],R)$ of all ring homomorphisms $\mathbb{Z}[\mathfrak{X}] \longrightarrow R$ and likewise $W(R)\times W(R)$ with $\mathrm{Hom}(\mathbb{Z}[\mathfrak{X},\mathfrak{Y}],R)$. Then addition and multiplication on $W(R)$ are to be interpreted as the maps

$$\mathrm{Hom}\big(\mathbb{Z}[\mathfrak{X},\mathfrak{Y}],R\big) \longrightarrow \mathrm{Hom}\big(\mathbb{Z}[\mathfrak{X}],R\big), \qquad \varphi \longmapsto \varphi\circ g,$$

that are induced by

$$g\colon \mathbb{Z}[\mathfrak{X}] \longrightarrow \mathbb{Z}[\mathfrak{X},\mathfrak{Y}], \qquad X_n \longmapsto S_n, \text{ resp. } X_n \longmapsto P_n.$$

Furthermore, given a ring homomorphism $f\colon R \longrightarrow R'$, we obtain for the addition, as well as for the multiplication, in each case a canonical commutative diagram

$$\begin{array}{ccc} \mathrm{Hom}\big(\mathbb{Z}[\mathfrak{X},\mathfrak{Y}],R\big) & \longrightarrow & \mathrm{Hom}\big(\mathbb{Z}[\mathfrak{X}],R\big) \\ \downarrow & & \downarrow \\ \mathrm{Hom}\big(\mathbb{Z}[\mathfrak{X},\mathfrak{Y}],R'\big) & \longrightarrow & \mathrm{Hom}\big(\mathbb{Z}[\mathfrak{X}],R'\big), \end{array}$$

where the vertical maps are given by composition with f, and in addition can be interpreted as the map

$$W(f)\colon W(R) \longrightarrow W(R'), \qquad (a_n)_{n\in\mathbb{N}} \longmapsto \big(f(a_n)\big)_{n\in\mathbb{N}},$$

resp. as the Cartesian product of this map with itself. The commutativity of the diagram then corresponds to the homomorphism property of $W(f)$. $\square$

The ring $W(R)$ is called the *Witt ring* attached to the ring R, and its elements are referred to as *Witt vectors* with coefficients in R. For any element $a \in W(R)$, its image $w(a) \in R^{\mathbb{N}}$ is called the associated vector of *ghost components* of a. The reason is that addition and multiplication of these components in the usual way (at least in the case that p is invertible in R) determine the ring structure of $W(R)$, although these components themselves are not visible in $W(R)$.

We want to add some simple rules for doing computations in $W(R)$. Considering again the homomorphism $w\colon W(R) \longrightarrow R^{\mathbb{N}}$, we have

$$w\big((\alpha \cdot \beta, 0, 0, \ldots)\big) = w\big((\alpha, 0, 0, \ldots)\big) \cdot w\big((\beta, 0, 0, \ldots)\big)$$

for elements $\alpha, \beta \in R$, since $W_n(\gamma, 0, 0, \ldots) = \gamma^{p^n}$ for $\gamma \in R$. From this we deduce the rule

$$(\alpha, 0, 0, \ldots) \cdot (\beta, 0, 0, \ldots) = (\alpha \cdot \beta, 0, 0, \ldots)$$

for the multiplication in $W(R)$, in a first step for the case that p is invertible in R, and then, using an argument as in the proof of Proposition 6, also for arbitrary rings R. In the same way one verifies the decomposition rule

$$(a_0, a_1, \ldots) = (a_0, \ldots, a_n, 0, 0, \ldots) + (0, \ldots, 0, a_{n+1}, a_{n+2}, \ldots)$$

for the addition in $W(R)$. In the sequel the multiplication in $W(R)$ by p will be of special interest, particularly for the case in which we have $p \cdot 1 = 0$ in R. For example, for $R = \mathbb{F}_p$ and $a \in W(\mathbb{F}_p)$ we get $w(p \cdot a) = p \cdot w(a) = 0$, although the p-multiplication on $W(\mathbb{F}_p)$ is nontrivial, as we will see below. In particular, this shows that the homomorphism $w\colon W(\mathbb{F}_p) \longrightarrow \mathbb{F}_p^{\mathbb{N}}$ cannot be injective. In other words, elements in $W(\mathbb{F}_p)$ are not uniquely characterized by their ghost components.

In order to study the p-multiplication on $W(R)$, we introduce the *Frobenius operator*

$$F\colon W(R) \longrightarrow W(R), \qquad (a_0, a_1, \ldots) \longmapsto (a_0^p, a_1^p, \ldots),$$

as well the *Verschiebung operator* [9]

$$V\colon W(R) \longrightarrow W(R), \qquad (a_0, a_1, \ldots) \longmapsto (0, a_0, a_1, \ldots).$$

Both operators commute with each other, i.e., $V \circ F = F \circ V$. Also note that the Frobenius operator $F\colon W(R) \longrightarrow W(R)$ is a ring homomorphism over rings R satisfying $p \cdot 1 = 0$. Indeed, $R \longrightarrow R$, $a \longmapsto a^p$, is then a ring homomorphism, and as we have explained, it induces a ring homomorphism $W(R) \longrightarrow W(R)$,

[9] It is common to use the German word *Verschiebung* for this operator; it means shifting.

which coincides with F. The Verschiebung operator V does not admit such a property, but is always additive. To justify this, we may assume, similarly as in the proof of Proposition 6, that p is invertible in R. Then $w\colon W(R) \longrightarrow R^{\mathbb{N}}$ is an isomorphism, and we have $W_{n+1}(V(a)) = pW_n(a)$, resp.

$$w\big(V(a)\big) = \big(0, pW_0(a), pW_1(a), \ldots\big).$$

This means that V is transported via w to the map

$$R^{\mathbb{N}} \longrightarrow R^{\mathbb{N}}, \qquad (x_0, x_1, \ldots) \longmapsto (0, px_0, px_1, \ldots),$$

which clearly is componentwise additive.

Now the p-multiplication on $W(R)$, which in most cases is simply denoted by p, can be characterized in terms of the Frobenius and the Verschiebung operators as follows:

Lemma 7. *For $a \in W(R)$, let $(p \cdot a)$ be the p-fold sum of a in $W(R)$ and $(p \cdot a)_n$ its associated component of index n. Likewise, let $((V \circ F)(a))_n$ be the component of $(V \circ F)(a)$ of index n. Then*

$$\big((V \circ F)(a)\big)_n \equiv \big(p \cdot a\big)_n \mod (p), \qquad n \in \mathbb{N}.$$

In particular, if $p \cdot 1 = 0$ in R, we get the relation

$$V \circ F = F \circ V = p.$$

Proof. Using a similar argument to that applied in the proof of Proposition 6, we may assume that p is not a zero divisor in R. Then, by Lemma 4, the stated congruences are equivalent to

$$W_n\big((V \circ F)(a)\big) \equiv W_n(p \cdot a) \mod p^{n+1}, \qquad n \in \mathbb{N}.$$

Furthermore, the recurrence formulas $(*)$ imply

$$W_n\big((V \circ F)(a)\big) = W_n\Big(F\big(V(a)\big)\Big) \equiv W_{n+1}\big(V(a)\big) \mod p^{n+1},$$

and we get

$$W_{n+1}\big(V(a)\big) = p \cdot W_n(a) = W_n(p \cdot a).$$

Indeed, the left equality of the preceding equation was used above to show that V is additive, while the right one follows from the fact that $w\colon W(R) \longrightarrow R^{\mathbb{N}}$ is a homomorphism. Putting everything together, the desired congruences follow. □

In view of Kummer theory in characteristic $p > 0$ and for an exponent p^r, we need rings of Witt vectors of *finite length* $r \geq 1$. So far we have considered the set $R^{\mathbb{N}}$ and have looked, so to speak, at Witt vectors of infinite length.

However, we can just as well restrict ourselves to vectors $(a_0, \ldots, a_{r-1}) \in R^r$ of finite length r. Since the polynomials S_n, P_n contain only variables X_i, Y_i for indices $i \leq n$, they give rise to laws of composition on each R^r, and one shows as in the case of Witt vectors of infinite length that they define a ring structure on R^r. The resulting ring is denoted by $W_r(R)$ and is called the ring of *Witt vectors of length* r over R. It is easily checked that the assertions of Proposition 6 carry over mutatis mutandis, where $W_1(R)$ is canonically isomorphic to R. If V is the Verschiebung operator on $W(R)$, as considered above, then the projection

$$W(R) \longrightarrow W_r(R), \qquad (a_0, a_1, \ldots) \longmapsto (a_0, \ldots, a_{r-1}),$$

is a surjective ring homomorphism with kernel

$$V^r W(R) = \big\{(a_0, a_1, \ldots) \in W(R)\,;\, a_0 = \ldots = a_{r-1} = 0\big\}$$

and hence induces an isomorphism $W(R)/V^r W(R) \xrightarrow{\sim} W_r(R)$. In particular, $V^r W(R)$ is an ideal in $W(R)$. Furthermore, $V^r \colon W(R) \longrightarrow W(R)$ is an injective homomorphism of additive groups mapping $V^k W(R)$ onto $V^{r+k} W(R)$ for every $k \in \mathbb{N}$. Therefore, V^r gives rise to an r-fold Verschiebung operator $V_k^r \colon W_k(R) \longrightarrow W_{r+k}(R)$, which is an injective homomorphism of additive groups as well. Clearly, we get

$$\operatorname{im} V_k^r = \big\{(a_0, \ldots, a_{r+k-1}) \in W_{r+k}(R)\,;\, a_0 = \ldots = a_{r-1} = 0\big\},$$

and this image coincides with the kernel of the projection

$$W_{r+k}(R) \longrightarrow W_r(R), \qquad (a_0, \ldots, a_{r+k-1}) \longmapsto (a_0, \ldots, a_{r-1}).$$

It follows that V_k^r induces an isomorphism $W_{r+k}(R)/V_k^r W_k(R) \xrightarrow{\sim} W_r(R)$ and hence gives rise to an exact sequence of abelian groups

$$0 \longrightarrow W_k(R) \xrightarrow{V_k^r} W_{r+k}(R) \longrightarrow W_r(R) \longrightarrow 0.$$

Alternatively, we can consider on $W_r(R)$ the map

$$W_r(R) \xrightarrow{V_r^1} W_{r+1}(R) \longrightarrow W_r(R), \qquad (a_0, \ldots, a_{r-1}) \longmapsto (0, a_0, \ldots, a_{r-2}),$$

as a Verschiebung operator. This operator, in the following denoted by V again, is additive, and its kth power V^k for $0 \leq k \leq r$ admits

$$V^{r-k} W_r(R) = \big\{(a_0, \ldots, a_{r-1}) \in W_r(R)\,;\, a_0 = \ldots = a_{r-k-1} = 0\big\}$$

as its kernel.

Kummer theory of exponent p^r. — Let us consider an exponent p^r for $r \geq 1$, as well as a field K of characteristic $p > 0$, together with a separable algebraic closure K_s of K and its absolute Galois group $G = \operatorname{Gal}(K_s/K)$. Then every Galois automorphism $\sigma \colon K_s \longrightarrow K_s$ induces an automorphism of rings

$$W_r(K_s) \longrightarrow W_r(K_s), \qquad (a_0, \ldots, a_{r-1}) \longmapsto \big(\sigma(a_0), \ldots, \sigma(a_{r-1})\big).$$

Thereby we obtain a homomorphism $G \longrightarrow \mathrm{Aut}(W_r(K_s))$, which we will view as an action of G on $W_r(K_s)$. This action is continuous, since the action of G is continuous on the particular components of $W_r(K_s)$. Thus, writing A for the additive group of $W_r(K_s)$, we see that A is equipped with a continuous G-action, where, in the sense of general Kummer theory, we have $A_L = W_r(L)$ for intermediate fields L of K_s/K.

Furthermore,

$$\wp\colon A \longrightarrow A, \qquad a \longmapsto F(a) - a,$$

is an endomorphism of A that is compatible with the G-action on A.

Theorem 8. *Assume* $\mathrm{char}\, K = p > 0$. *Then* $A = W_r(K_s)$, *viewed as a* G*-module, together with the* G*-homomorphism*

$$\wp\colon A \longrightarrow A, \qquad a \longmapsto F(a) - a,$$

satisfies the conditions of Theorem 1 *for Kummer theory of exponent* p^r *over* K.

We divide the proof into several steps.

Lemma 9. $\wp\colon W_r(K_s) \longrightarrow W_r(K_s)$ *is surjective.*

Proof. For $r = 1$, we get $A = W_1(K_s) = K_s$, and it has to be shown that the map

$$\wp\colon K_s \longrightarrow K_s, \qquad \alpha \longmapsto \alpha^p - \alpha,$$

is surjective. However, this is clear, since polynomials of type $X^p - X - c$, where $c \in K_s$, are separable. For general r, it is easily checked that $\wp$ is compatible with the Verschiebung operator, as well as with the projection $W_r(K_s) \longrightarrow W_1(K_s)$. Therefore, if $r > 1$, we get a commutative diagram

$$\begin{array}{ccccccccc}
0 & \longrightarrow & W_{r-1}(K_s) & \xrightarrow{V^1_{r-1}} & W_r(K_s) & \longrightarrow & W_1(K_s) & \longrightarrow & 0 \\
 & & \wp\downarrow & & \wp\downarrow & & \wp\downarrow & & \\
0 & \longrightarrow & W_{r-1}(K_s) & \xrightarrow{V^1_{r-1}} & W_r(K_s) & \longrightarrow & W_1(K_s) & \longrightarrow & 0\,,
\end{array}$$

and we can conclude from the surjectivity of $\wp$ on $W_1(K_s)$ and on $W_{r-1}(K_s)$ that it is surjective on $W_r(K_s)$ as well. □

As a next ingredient for Kummer theory, we determine the kernel of $\wp$. To do this we view $\mathbb{F}_p$ as the prime subfield of our field K.

Lemma 10. *The kernel of* $\wp\colon W_r(K_s) \longrightarrow W_r(K_s)$ *satisfies* $\ker \wp = W_r(\mathbb{F}_p)$. *This group is cyclic of order* p^r *and generated by the unit element* $e \in W_r(\mathbb{F}_p)$.

Proof. The solutions of the equation $x^p = x$ in K_s consist precisely of the elements of the prime subfield $\mathbb{F}_p \subset K_s$. Therefore, we get $\ker \wp = W_r(\mathbb{F}_p)$, due to the definition of $\wp$. Thus, $\ker \wp$ is a group of order p^r, and we claim that the unit element $e = (1, 0, \ldots, 0) \in W_r(\mathbb{F}_p)$ is of this order. Indeed, the order of e divides p^r and thus is a p-power. Using the formula $V \circ F = p$ from Lemma 7, repeated multiplication by p moves the component 1 in e at each step by one position to the right, so that indeed, p^r turns out to be the order of e. □

Finally, in order to make Theorem 1 applicable and thereby characterize all abelian extensions of an exponent dividing p^r, it remains to establish Hilbert's Theorem 90.

Proposition 11. *Let L/K be a finite Galois extension in characteristic $p > 0$ with Galois group G. On the ring of Witt vectors $W_r(L)$ over L of given length r, consider the componentwise action of G. Then*

$$H^1\big(G, W_r(L)\big) = 0,$$

i.e., every 1-cocycle is already a 1-coboundary.

Proof. We proceed similarly as in the proof of 4.8/2, but in addition, must make use of the trace map

$$\mathrm{tr}_{L/K} \colon W_r(L) \longrightarrow W_r(K), \qquad a \longmapsto \sum_{\sigma \in G} \sigma(a).$$

Since every $\sigma \in G$ defines a $W_r(K)$-automorphism of $W_r(L)$, we see immediately that the trace map is $W_r(K)$-linear. In addition, $\mathrm{tr}_{L/K}$ is compatible with the projection $W_r(L) \longrightarrow W_1(L) = L$, where the trace map on $W_1(L)$ coincides by 4.7/4 with the usual trace map $\mathrm{tr}_{L/K} \colon L \longrightarrow K$. Proceeding by induction on r, we want to show that $\mathrm{tr}_{L/K} \colon W_r(L) \longrightarrow W_r(K)$ is surjective.

If $r = 1$, we have to deal with the usual trace map, as defined for finite field extensions. The assertion then follows from 4.7/7. Otherwise, we can use the fact that the trace map on $W_r(L)$ is compatible with the Verschiebung operator, and hence for $r > 1$, leads to a commutative diagram of the following type:

$$\begin{array}{ccccccccc}
0 & \longrightarrow & W_{r-1}(L) & \xrightarrow{V^1_{r-1}} & W_r(L) & \longrightarrow & W_1(L) & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle \mathrm{tr}_{L/K}} & & \downarrow{\scriptstyle \mathrm{tr}_{L/K}} & & \downarrow{\scriptstyle \mathrm{tr}_{L/K}} & & \\
0 & \longrightarrow & W_{r-1}(K) & \xrightarrow{V^1_{r-1}} & W_r(K) & \longrightarrow & W_1(K) & \longrightarrow & 0
\end{array}$$

As we know, the trace map is surjective on $W_1(L)$, and by the induction hypothesis, also on $W_{r-1}(L)$. Therefore, it will be surjective on $W_r(L)$ as well. In particular, there exists an element $a \in W_r(L)$ such that $\mathrm{tr}_{L/K}(a) = 1$.

Now let $f \colon G \longrightarrow W_r(L)$ be a 1-cocycle. Considering the Poincaré series

$$b = \sum_{\sigma' \in G} f(\sigma') \cdot \sigma'(a),$$

we obtain for arbitrary $\sigma \in G$ the equation

$$\begin{aligned}\sigma(b) &= \sum_{\sigma' \in G} \sigma\big(f(\sigma')\big) \cdot (\sigma \circ \sigma')(a) \\ &= \sum_{\sigma' \in G} \big(f(\sigma \circ \sigma') - f(\sigma)\big) \cdot (\sigma \circ \sigma')(a) \\ &= \sum_{\sigma' \in G} f(\sigma \circ \sigma') \cdot (\sigma \circ \sigma')(a) - \sum_{\sigma' \in G} f(\sigma) \cdot (\sigma \circ \sigma')(a) \\ &= b - f(\sigma) \cdot \mathrm{tr}_{L/K}(a) = b - f(\sigma),\end{aligned}$$

i.e., f is a 1-coboundary. □

This concludes the proof of Theorem 8.

Exercises

1. *Within the context of general Kummer theory for some exponent n, characterize all cyclic Galois extensions of a degree dividing n.*
2. *Let K be a perfect field of characteristic $p > 0$. Prove the following properties of the Witt ring $W(K)$:*
   (i) *The map*
   $$K^* \longrightarrow W(K)^*, \qquad \alpha \longmapsto (\alpha, 0, 0, \ldots),$$
   is a monomorphism of multiplicative groups. Is a similar assertion valid for the additive group K as well?
   (ii) *The canonical map $W(K) \longrightarrow \varprojlim W(K)/p^n W(K)$ is an isomorphism of rings. In particular, $W(\mathbb{F}_p)$ coincides with the ring $\mathbb{Z}_p$ of integral p-adic numbers*; *see Section* 4.2.
   (iii) *$W(K)$ is a principal ideal domain with maximal ideal $p \cdot W(K) = V^1 W(K)$. Every other nontrivial ideal in $W(K)$ is a power of this maximal ideal and hence is of type $p^n \cdot W(K) = V^n W(K)$.*
3. Let p be a prime number and $q = p^r$ a nontrivial power of p. Show:
   (i) Every $a \in W(\mathbb{F}_q)$ admits a representation
   $$a = \sum_{i \in \mathbb{N}} c_i p^i$$
   with unique coefficients $c_i \in \mathbb{F}_q$ that are to be interpreted as Witt vectors $(c_i, 0, 0, \ldots) \in W(\mathbb{F}_q)$.
   (ii) $W(\mathbb{F}_q) = \mathbb{Z}_p[\zeta]$ for a primitive $(q-1)$th root of unity ζ.

   Furthermore, determine the degree of the field of fractions $Q(W(\mathbb{F}_q))$ over $Q(\mathbb{Z}_p)$.
4. Let G be the absolute Galois group of a field K. Using the notions of general Kummer theory, consider for a G-module A the maps
   $$\Phi\colon \Delta \longmapsto G(A/\Delta), \qquad \Psi\colon H \longmapsto A^H,$$
   for subgroups $\Delta \subset A$ and $H \subset G$. Show that
   $$\Phi \circ \Psi \circ \Phi(\Delta) = \Phi(\Delta), \qquad \Psi \circ \Phi \circ \Psi(H) = \Psi(H).$$

4.11 Galois Descent*

Let K'/K be a field extension. Given a K-vector space V, say with basis $(v_i)_{i\in I}$, we can extend coefficients and construct from V a K'-vector space $V' = V \otimes_K K'$, for example by viewing $(v_i)_{i\in I}$ as a K'-basis and admitting coefficients from K'. Then V is called a *K-form* of V'. In a similar way, it is possible to derive from a K-homomorphism $\varphi\colon V \longrightarrow W$ a K'-homomorphism $\varphi'\colon V' \longrightarrow W'$ by means of extending coefficients. The subject of descent theory for K'/K is the reverse problem. Its aim is to describe K-vector spaces and their homomorphisms in terms of the corresponding extended objects over K' together with so-called *descent data* on them. It is quite easy to specify K-forms V, W of given K'-vector spaces V', W'. However, for a K'-homomorphism $\varphi'\colon V' \longrightarrow W'$ and fixed K-forms V, W, it is not true in general that φ' is obtained from a K-homomorphism $\varphi\colon V \longrightarrow W$ by extending coefficients. For this to work well it is necessary that φ' respect the descent data given on V' and W'.

In the present section we will study descent theory only for the case that K is the fixed field of a finite group of automorphisms of K'/K, i.e., for finite Galois extensions K'/K; cf. 4.1/4. Then the necessary descent data can be described in terms of group actions. However, let us point out that in algebraic geometry, descent theory is developed in much more generality. For example, consult the foundational work of Grothendieck [6], or see [3], Chap. 4.

Before we actually start studying descent, let us put the process of coefficient extension onto a solid basis, by introducing tensor products. We restrict ourselves to the special case of vector spaces. More general tensor products will be dealt with in Section 7.2.

Definition 1. *Let K'/K be a field extension and V a K-vector space. A* tensor product *of K' with V over K is a K'-vector space V', together with a K-linear map $\tau\colon V \longrightarrow V'$, admitting the following universal property*:

Given a K-linear map $\varphi\colon V \longrightarrow W'$ into a K'-vector space W', there exists a unique K'-homomorphism $\varphi'\colon V' \longrightarrow W'$ such that $\varphi = \varphi' \circ \tau$, in other words, such that φ' is a K'-linear "continuation" of φ.

Due to the defining universal property, tensor products are unique, up to canonical isomorphism. In the situation of the preceding definition, one writes $K' \otimes_K V$ or $V \otimes_K K'$ for V', depending on whether one likes to view V' as a left or a right vector space under the scalar multiplication by elements from K'. Furthermore, for $(a, v) \in K' \times V$, the product $a \cdot \tau(v)$ is usually denoted by $a \otimes v$; this element is called a *tensor*. The elements of $K' \otimes_K V$ are finite sums of such tensors, as we will see further below. Of course, the corresponding fact remains true if V' is viewed as a right vector space.

Remark 2. *The tensor product $V' = K' \otimes_K V$, as specified in Definition* 1, *always exists.*

Proof. Fix a K-basis $(v_i)_{i\in I}$ of V and consider the K'-vector space $V' = K'^{(I)}$ with its canonical basis $(e_i)_{i\in I}$. Mapping the basis vector $v_i \in V$ to the basis vector $e_i \in K'^{(I)}$ for each $i \in I$, and using K-linear extension, we obtain an injective K-linear map $\tau\colon V \longrightarrow V'$. Now let $\varphi\colon V \longrightarrow W'$ be a K-linear map to an arbitrary K'-vector space W'. If there exists a K'-linear map $\varphi'\colon V' \longrightarrow W'$ satisfying $\varphi = \varphi' \circ \tau$, then we get necessarily $\varphi'(e_i) = \varphi'(\tau(v_i)) = \varphi(v_i)$. In particular, φ' is uniquely determined by φ on the K'-basis (e_i) of V', and thereby as a K'-linear map, on all of V'. On the other hand, we can define a K'-linear map $\varphi'\colon V' \longrightarrow W'$ satisfying $\varphi = \varphi' \circ \tau$ by mapping $e_i \longmapsto \varphi(v_i)$ and using K'-linear extension. It follows that V' together with the map τ admits the properties of a tensor product of K' with V over K. □

The proof shows that indeed, $V' = K' \otimes_K V$ arises from V by "extending coefficients on V." Using the injective K-linear map $\tau\colon V \longrightarrow V' = K' \otimes_K V$ as an identification, we may view V as a K-linear subspace of $K' \otimes_K V$. In the proof we have fixed a K-basis $(v_i)_{i\in I}$ of V and defined the tensor product $K' \otimes_K V$ in such a way that it admits the same system $(v_i)_{i\in I}$ as a K'-basis. Also note that the resulting tensor product $K' \otimes_K V$ is independent of the chosen K-basis $(v_i)_{i\in I}$ of V, due to the universal property of the tensor product. Furthermore, one can easily show that every K-homomorphism $\varphi\colon V \longrightarrow W$ between K-vector spaces V and W extends to a K'-homomorphism $(K' \otimes \varphi)\colon K' \otimes_K V \longrightarrow K' \otimes_K W$ between the corresponding K'-vector spaces. However, this fact can just as well be seen using the universal property of tensor products, since

$$V \longrightarrow K' \otimes_K W, \qquad v \longmapsto 1 \otimes \varphi(v),$$

is a K-linear map and hence gives rise to a well-defined K'-homomorphism $(K' \otimes \varphi)\colon K' \otimes_K V \longrightarrow K' \otimes_K W$.[10]

We are now able to make the terms "K-form" and "defined over K," as used above, more precise. To this end, consider any field extension K'/K. A K-linear subspace V of a K'-vector space V' is called a *K-form* of V' if the K'-linear map $K' \otimes_K V \longrightarrow V'$ induced by $V \hookrightarrow V'$ is an isomorphism. Fixing a K-form V of V', the preceding isomorphism is usually viewed as an identification. Furthermore, a K'-linear subspace $U' \subset V'$ is said to be *defined over* K if U' is the K'-extension of a K-linear subspace $U \subset V$, or in other words, if there exists a K-linear subspace $U \hookrightarrow V$ such that the induced K'-linear map $K' \otimes_K U \longrightarrow K' \otimes_K V = V'$ (it is always injective!) identifies $K' \otimes_K U$ with U'. In particular, U is then a K-form of U'. Finally, a K'-homomorphism $\varphi'\colon V' \longrightarrow W'$ between K'-vector spaces with K-forms V and W is said to be *defined over* K if φ' is the K'-extension of a K-homomorphism $\varphi\colon V \longrightarrow W$, i.e., if there exists a K-homomorphism $\varphi\colon V \longrightarrow W$ such that φ', using the identifications $V' = K' \otimes_K V$ and $W' = K' \otimes_K W$, coincides with $K' \otimes \varphi$.

[10] In dealing with general tensor products, it is common practice to use the notation $\mathrm{id}_{K'} \otimes \varphi$ instead of $K' \otimes \varphi$. Actually, we are dealing here with the tensor product of two K-linear maps, namely the identity map on K' and the map φ; see also Section 7.2.

Next we want to approach the case of finite Galois extensions K'/K, assuming however, for the moment only that K is the fixed field of a (not necessarily finite) subgroup $G \subset \mathrm{Aut}(K')$; see 4.1/4. Consider a K'-vector space V' together with a K-form V, where we identify V' with $K' \otimes_K V$. Then we can define for every $\sigma \in G$ a K-linear map $f_\sigma \colon K' \otimes_K V \longrightarrow K' \otimes_K V$, namely the one characterized by $a \otimes v \longmapsto \sigma(a) \otimes v$. Indeed, fixing a K-basis $(v_i)_{i \in I}$ of V and viewing it as a K'-basis of V' as well, f_σ can be defined by

$$f_\sigma \colon V' \longrightarrow V', \qquad \sum a_i v_i \longmapsto \sum \sigma(a_i) v_i.$$

The map f_σ is called σ-*linear*, since it satisfies the relations

$$f_\sigma(v' + w') = f_\sigma(v') + f_\sigma(w'), \qquad f_\sigma(a'v') = \sigma(a') f_\sigma(v'),$$

for $v', w' \in V'$ and $a' \in K'$. Furthermore, we have $f_\sigma \circ f_\tau = f_{\sigma\tau}$ for $\sigma, \tau \in G$, as well as $f_\varepsilon = \mathrm{id}_{V'}$ for the unit element $\varepsilon \in G$. This means that the maps f_σ give rise to an action

$$G \times V' \longrightarrow V', \qquad (\sigma, v) \longmapsto f_\sigma(v),$$

of G on V' in the sense of 5.1/1. This action, characterized by $f = (f_\sigma)_{\sigma \in G}$, is referred to as the *canonical G-action* attached to the K-form V of V'.

Proposition 3. *Let K'/K be a field extension such that K is the fixed field of a subgroup $G \subset \mathrm{Aut}(K')$. Furthermore, consider a K'-vector space V', together with a K-form V and its corresponding canonical G-action f.*

(i) *An element $v \in V'$ belongs to V if and only if $f_\sigma(v) = v$ for all $\sigma \in G$.*

(ii) *A K'-linear subspace $U' \subset V'$ is defined over K if and only if we have $f_\sigma(U') \subset U'$ for all $\sigma \in G$.*

(iii) *A K'-homomorphism $\varphi' \colon V' \longrightarrow W'$ between K'-vector spaces with K-forms V, W and corresponding G-actions f, g, is defined over K if and only if φ' is compatible with all $\sigma \in G$, i.e., if and only if $\varphi'(f_\sigma(v)) = g_\sigma(\varphi'(v))$ for all $\sigma \in G$ and all $v \in V'$.*

Proof. Assertion (i) is easy to obtain. Fix a K-basis $(v_i)_{i \in I}$ of V and write $v = \sum_i a_i v_i$ with coefficients $a_i \in K'$. Since $f_\sigma(v) = \sum_i \sigma(a_i) v_i$, we see that v is invariant under all f_σ if and only if the coefficients a_i are invariant under all $\sigma \in G$, i.e., if and only if all a_i belong to K and hence v is an element of V. Just as easily we can derive assertion (iii). Certainly, the compatibility condition given in (iii) is necessary. On the other hand, the condition implies $\varphi'(V) \subset W$ if we use (i).

Turning to assertion (ii), the condition $f_\sigma(U') \subset U'$ for all $\sigma \in G$ is clearly necessary. To see that it is also sufficient, consider a K-basis $(v_i)_{i \in I}$ of V, as well as the residue classes $\overline{v}_i \in W' = V'/U'$ of the elements v_i. There is a subsystem $(\overline{v}_i)_{i \in I'}$ of the system of all $\overline{v}_i$ forming a K'-basis of W'. Therefore, we can view $W = \sum_{i \in I'} K \overline{v}_i$ as a K-form of W' and consider on W' the G-action g that is attached to W. We claim that the canonical projection $\varphi' \colon V' \longrightarrow W'$ is defined over K. To justify this, observe that every $v \in V'$ can be written as

$$v = u + \sum_{i \in I'} a_i v_i,$$

where $u \in U'$ and $a_i \in K'$ for all $i \in I'$. Now use the fact that by our assumption, $f_\sigma(U') \subset U' = \ker \varphi'$ for all $\sigma \in G$, and that $f_\sigma(v_i) = v_i$ for all $i \in I'$. This shows that $\varphi'(f_\sigma(v)) = g_\sigma(\varphi'(v))$ for all $\sigma \in G$ and hence that φ' is defined over K, due to (iii). Then it is not difficult to see that together with φ', also $U' = \ker \varphi'$ is defined over K. □

Now we want to show that K-forms of vector spaces can be characterized in terms of group actions.

Proposition 4. *Let K'/K be a field extension such that K is the fixed field of a subgroup $G \subset \mathrm{Aut}(K')$. Furthermore, consider a K'-vector space V'. For each $\sigma \in G$, let $f_\sigma \colon V' \longrightarrow V'$ be a σ-linear map satisfying $f_\sigma \circ f_\tau = f_{\sigma\tau}$ for $\sigma, \tau \in G$, as well as $f_\varepsilon = \mathrm{id}_{V'}$ for the unit element $\varepsilon \in G$. Thus, the maps f_σ set up an action $f = (f_\sigma)$ of G on V'. Let $V \subset V'$ be the corresponding fixed set.*

(i) *V is a K-linear subspace of V', and $\lambda \colon V \hookrightarrow V'$ induces a K'-linear map $\lambda' \colon K' \otimes_K V \longrightarrow V'$, which is injective.*

(ii) *If G is finite and hence K'/K a finite Galois extension, then λ' is surjective and therefore bijective. In particular, V is a K-form of V'.*

Proof. Of course, V is a K-form of $K' \otimes_K V$ for trivial reasons. Let h be the canonical action of G on $K' \otimes_K V$, where $h_\sigma \colon K' \otimes_K V \longrightarrow K' \otimes_K V$ is characterized by $a \otimes v \longmapsto \sigma(a) \otimes v$. Then λ' is compatible with the actions h and f, since we have

$$\lambda'\big(h_\sigma(a \otimes v)\big) = \lambda'\big(\sigma(a) \otimes v\big) = \sigma(a)v = f_\sigma(av) = f_\sigma\big(\lambda'(a \otimes v)\big).$$

This implies $h_\sigma(\ker \lambda') \subset \ker \lambda'$, and we conclude from Proposition 3 (ii) that $\ker \lambda'$ is defined over K. Hence, there is a K-linear subspace $U \subset V$ whose K'-extension in $K' \otimes_K V$ coincides with $\ker \lambda'$. However, for $u \in U$ we have $u = \lambda(u) = \lambda'(u) = 0$ and hence $u = 0$, so that λ' is injective. This settles assertion (i).

To verify (ii), assume that G is finite. It is enough to show that every linear functional $\varphi' \colon V' \longrightarrow K'$ vanishing on V is identically zero on V'. Therefore, consider such a linear functional φ', where $\varphi'(V) = 0$, and let $v \in V'$. Then, for variable $a \in K'$, the elements $v_a = \sum_{\sigma \in G} f_\sigma(av)$ are invariant under the action of G on V' and thus belong to V. Since $\varphi'(V) = 0$, we get $\sum_{\sigma \in G} \sigma(a)\varphi'(f_\sigma(v)) = 0$ for all $a \in K'$. Now view the preceding sum as a linear combination of the characters $\sigma \in G$ and apply the linear independence result 4.6/2. It shows that all coefficients $\varphi'(f_\sigma(v)) \in K'$ vanish. In particular, for $\sigma = \varepsilon$, the unit element of G, we obtain $\varphi'(v) = 0$. Therefore, every linear functional on V' that is trivial on V is identically zero on V'. □

As a summary, we can learn from Propositions 3 and 4 for a finite Galois extension K'/K with Galois group G that the theory of K-vector spaces is

equivalent to the theory of K'-vector spaces with G-actions, as studied in this section. Within this setup, K-homomorphisms of K-vector spaces correspond to those K'-homomorphisms between attached K'-objects that are compatible with the G-actions under consideration. In particular, the G-actions play the role of the descent data, which were mentioned earlier.

Finally, let us point out that the linear independence of characters 4.6/2 was used in the proof of Proposition 4 (ii) in a similar way to what we did in the proof of the cohomological version of Hilbert's Theorem 90; see 4.8/2. Moreover, 4.8/2 implies for a finite Galois extension K'/K the assertion of Proposition 4 in the case $\dim_{K'} V' = 1$, resp. $V' = K'$. Just check for fixed $v \in K'^*$ that the map

$$G \longrightarrow K'^*, \qquad \sigma \longmapsto \frac{f_\sigma(v)}{v},$$

is a 1-cocycle, and thus a 1-coboundary by 4.8/2. Therefore, there exists an element $a \in K'^*$ such that $f_\sigma(v) \cdot v^{-1} = a \cdot \sigma(a)^{-1}$, and we get

$$f_\sigma(av) = \sigma(a) \cdot f_\sigma(v) = av,$$

i.e., $av \in V'$ is fixed by all f_σ. Then it is easily seen that $V = K \cdot av$ equals the fixed set of the action of G on V' and hence that it is a K-form of V'.

Exercises

1. *Let K'/K be a field extension and A a K-algebra, i.e., a ring together with a ring homomorphism $K \longrightarrow A$. Show that $A \otimes_K K'$ is naturally a K'-algebra.*

2. *Give an alternative proof of Proposition 4 as follows. Use an inductive argument to verify assertion* (i) *in a direct way. To establish* (ii), *choose a K-basis $\alpha_1, \ldots, \alpha_n$ of K' and show that every $v \in V'$ admits a representation of type*

$$v = \sum_{i=1}^{n} c_i \Big(\sum_{\sigma \in G} f_\sigma(\alpha_i v) \Big)$$

with coefficients $c_i \in K'$.

3. Let K'/K be a field extension and assume that K is the fixed field of a subgroup $G \subset \mathrm{Aut}(K')$. Furthermore, consider a K'-vector space V'. For every $\sigma \in G$ introduce a K'-vector space V'_σ as follows. Take V' as its additive group, and define the scalar multiplication by $a \cdot v := \sigma(a)v$ for $a \in K'$ and $v \in V'$, where the product on the right is meant in the sense of the K'-vector space V'. Then view the diagonal embedding $\lambda\colon V' \longrightarrow \prod_{\sigma \in G} V'_\sigma$ as a K-linear map, and look at the induced K'-linear map $\Lambda\colon V' \otimes_K K' \longrightarrow \prod_{\sigma \in G} V'_\sigma$. Show that Λ is injective, and even bijective if $[K' : K] < \infty$. *Hint*: Introduce a suitable action of G on $\prod_{\sigma \in G} V'_\sigma$ such that V' is the corresponding fixed set.

4. Let K'/K be a field extension and V a K-vector space. Consider the K-linear maps

$$\begin{aligned} V &\longrightarrow V \otimes_K K', & v &\longmapsto v \otimes 1, \\ V \otimes_K K' &\longrightarrow V \otimes_K K' \otimes_K K', & v \otimes a &\longmapsto v \otimes a \otimes 1, \\ V \otimes_K K' &\longrightarrow V \otimes_K K' \otimes_K K', & v \otimes a &\longmapsto v \otimes 1 \otimes a, \end{aligned}$$

and show that the diagram

$$V \to V \otimes_K K' \rightrightarrows V \otimes_K K' \otimes_K K'$$

is exact in the sense that the map on the left is injective and its image equals the kernel of the difference of the two maps on the right.

5. Let K'/K be a finite Galois extension with Galois group G. In the setting of Exercise 4 write $V' = V \otimes_K K'$ and identify $V' \otimes_K K'$ with $\prod_{\sigma \in G} V'_\sigma$ in the sense of Exercise 3. Then describe the two maps $V' \rightrightarrows \prod_{\sigma \in G} V'_\sigma$ of Exercise 4 in as simple a way as possible.

5. More Group Theory

Background and Overview

Returning for a moment to the problem of solving algebraic equations, let us look at a monic polynomial $f \in K[X]$ with coefficients in a field K. Furthermore, let L be a splitting field of f, where we assume that L/K is separable. Solving the algebraic equation $f(x) = 0$ by radicals amounts to constructing a chain of fields of type

$$(*) \qquad K = K_0 \subsetneq K_1 \subsetneq \ldots \subsetneq K_r$$

such that $L \subset K_r$, and in each case, K_{i+1} is obtained from K_i by adjoining a root of some element in K_i. Indeed, it is precisely if such a chain exists that the solutions of the equation $f(x) = 0$ that generate the extension L/K can be characterized in terms of rational operations on elements of K, combined with the process of extracting roots. To simplify, let us assume in the following that the extension K_r/K is Galois. Then the fundamental theorem of Galois theory 4.1/6 is applicable, and every chain of fields of type $(*)$ is equivalent to a chain of subgroups

$$(**) \qquad \mathrm{Gal}(K_r/K) = G_0 \supsetneq G_1 \supsetneq \ldots \supsetneq G_r = \{1\}.$$

Furthermore, in 4.5 and 4.8 we characterized extensions that arise through the adjunction of nth roots. If we restrict ourselves to fields of characteristic 0 and assume that K contains sufficiently many roots of unity, we can conclude from 4.8/3 and 4.1/6 that a chain of fields as in $(*)$ is given by successively adjoining nth roots of elements for variable n if and only if the corresponding chain $(**)$ admits the property that in each case, G_{i+1} is a normal subgroup of G_i and the residue class group G_i/G_{i+1} is cyclic. More precisely, we will see in 6.1 that the equation $f(x) = 0$ is solvable by radicals if and only if there exists a chain $(**)$ admitting these properties for the Galois group $\mathrm{Gal}(L/K)$.

The above considerations in terms of Galois theory show that the problem of solving algebraic equations by radicals can be reduced to a group-theoretic problem. For example, from the fundamental theorem for finitely generated abelian groups 2.9/9 we conclude that algebraic equations with abelian Galois group are always solvable. However, to arrive at more specific results on the solvability of algebraic equations it is necessary to further complete the theory

of finite (not necessarily commutative) groups. In particular, we want to characterize all groups G admitting a chain of subgroups $(**)$ such that as before, G_{i+1} is a normal subgroup in G_i and the residue class group G_i/G_{i+1} is cyclic. Such a group G is called *solvable*, where instead of "cyclic" we can just as well require the quotients G_i/G_{i+1} to be abelian; see 5.4/3 and 5.4/7.

In order to approach the subject of solvable groups, we start in 5.1 with some basic material on *group actions*. Interpreting the Galois group of an algebraic equation $f(x) = 0$ as a group of permutations on the corresponding set of solutions, cf. 4.3/1, we get a prototype of such an action. Using the concept of group actions, we prove in 5.2 the so-called Sylow theorems on finite groups, named after the mathematician L. Sylow. They provide information on the existence of subgroups whose order is a prime power. In special situations, Sylow theorems can be used to check whether a given group is solvable. Furthermore, we have assembled in 5.3 some basic facts on permutation groups, while finally, in 5.4 we study solvable groups. In particular, we prove that the symmetric group $\mathfrak{S}_n$ is not solvable for $n \geq 5$, which will imply in 6.1 that the generic equation of degree n is not solvable by radicals for $n \geq 5$.

5.1 Group Actions

In the chapter on Galois theory we have already worked with group actions. However, this concept was not introduced explicitly, since we had to consider only the canonical action of a Galois group on its field or on the zero set of a polynomial. In the following we want to free ourselves from such a restricted setting and establish some combinatorial properties for general group actions.

Definition 1. *Let G be a (multiplicatively written) group and X a set. An* action *or an* operation *of G on X is a map*

$$G \times X \longrightarrow X, \qquad (g, x) \longmapsto g \cdot x,$$

such that:

(i) $1 \cdot x = x$ *for the unit element* $1 \in G$ *and for elements* $x \in X$.

(ii) $(gh) \cdot x = g \cdot (h \cdot x)$ *for* $g, h \in G$, $x \in X$.

To begin with, let us list some examples of group actions.

(1) For a group G and a set X, there is always the trivial action of G on X. It is given by the map

$$G \times X \longrightarrow X, \qquad (g, x) \longmapsto x.$$

(2) Let X be a set and write $S(X)$ for the group of bijective self-maps $X \longrightarrow X$. Then every subgroup $G \subset S(X)$ acts on X via the map

$$G \times X \longrightarrow X, \qquad (\sigma, x) \longmapsto \sigma(x).$$

In particular, we can consider for a Galois extension L/K the action of the Galois group $\mathrm{Gal}(L/K) = \mathrm{Aut}_K(L)$ on L. This action was studied thoroughly in Chapter 4, on Galois theory.

(3) For every group G the group multiplication

$$G \times G \longrightarrow G, \qquad (g,h) \longmapsto gh,$$

may be viewed as an action of G on itself. In fact, G acts on itself via left translation, where, as mentioned earlier, the *left translation* by $g \in G$ is given by the map

$$\tau_g \colon G \longrightarrow G, \qquad h \longmapsto gh.$$

Similarly, we can use the right translation to define an action of G on itself, namely via

$$G \times G \longrightarrow G, \qquad (g,h) \longmapsto hg^{-1}.$$

Looking at g instead of g^{-1}, the map

$$\tau'_g \colon G \longrightarrow G, \qquad h \longmapsto hg,$$

is called the *right translation* by g on G.

(4) Another action of G on itself is given by the *conjugation action*

$$G \times G \longrightarrow G, \qquad (g,h) \longmapsto ghg^{-1}.$$

The map

$$\mathrm{int}_g = \tau_g \circ \tau'_{g^{-1}} \colon G \longrightarrow G, \qquad h \longmapsto ghg^{-1},$$

is a group automorphism of G, the *conjugation* with g. Automorphisms of type int_g are called *inner automorphisms* of G ("int" stands for "interior"), and it is easily checked that the canonical map $G \longrightarrow \mathrm{Aut}(G),\ g \longmapsto \mathrm{int}_g$, is a group homomorphism. Two elements $h, h' \in G$ are said to be *conjugate* if there exists an element $g \in G$ such that $h' = \mathrm{int}_g(h)$. In the same way, two subgroups $H, H' \subset G$ are called *conjugate* if there is an element $g \in G$ such that $H' = \mathrm{int}_g(H)$. The relation of being conjugate is an equivalence relation for elements or subgroups in G. Of course, the conjugation action is trivial if G is commutative.

Similarly as in (3) we can define for a group action

$$G \times X \longrightarrow X$$

its (left) translation by an element $g \in G$ by

$$\tau_g \colon X \longrightarrow X, \qquad x \longmapsto g \cdot x.$$

The family of translations $(\tau_g)_{g \in G}$ fully characterizes a given action of G on X. In addition, $G \longrightarrow S(X),\ g \longmapsto \tau_g$, is a group homomorphism, as is easily checked.

On the other hand, proceeding as in example (2), every group homomorphism $\varphi\colon G \longrightarrow S(X)$ gives rise to an action of G on X, namely to

$$G \times X \longrightarrow X, \qquad (g,x) \longmapsto \varphi(g)(x).$$

Both mappings are inverse to each other and we see that the following is true:

Remark 2. *Let G be a group and X a set. Then, using the above mappings, the group actions $G \times X \longrightarrow X$ correspond bijectively to the group homomorphisms $G \longrightarrow S(X)$.*

If we consider for a group G the action $G \times G \longrightarrow G$ via left translation, then the corresponding group homomorphism $G \longrightarrow S(G)$ is injective, since $\tau_g = \tau_{g'}$ is equivalent to $g = g'$. In particular, we can view G as a subgroup of $S(G)$.

Definition 3. *Let $G \times X \longrightarrow X$ be an action of a group G on a set X. The following notions are commonly used for points $x \in X$:*

(i) *$Gx := \{gx\,;\, g \in G\}$ is called the* orbit *of x in G.*

(ii) *$G_x := \{g \in G\,;\, gx = x\}$ is called the* stabilizer *or* isotropy subgroup *of x in G.*

That G_x is a subgroup of G is easily checked. Indeed, note that G_x contains the unit element of G and that for $g, h \in G$ satisfying $gx = x = hx$ we have

$$(gh^{-1})x = (gh^{-1})(hx) = g\big(h^{-1}(hx)\big) = g\big((h^{-1}h)x\big) = gx = x.$$

Remark 4. *Let $G \times X \longrightarrow X$ be an action of a group G on a set X. If x, y are two points of one and the same G-orbit in X, then the stabilizer subgroups $G_x, G_y \subset G$ are conjugate.*

Proof. It is enough to consider the case $y \in Gx$. Therefore, let $h \in G$ be an element such that $y = hx$. Then, for $g \in G_x$, we get

$$(hgh^{-1})y = (hgh^{-1})hx = h(gx) = hx = y$$

and hence $hgh^{-1} \in G_y$, so that $hG_xh^{-1} \subset G_y$. Likewise, we can derive from $x = h^{-1}y$ the inclusion $h^{-1}G_yh \subset G_x$, showing that in fact, $G_y = hG_xh^{-1}$. □

Furthermore, we want to show that two orbits $Gx, Gy \subset X$ coincide as soon as $Gx \cap Gy \neq \emptyset$. Indeed, if there is an element $z \in Gx \cap Gy$, say $z = gx = hy$ for some elements $g, h \in G$, we get $x = g^{-1}z = g^{-1}hy$ and therefore $Gx \subset Gy$. Similarly we obtain $Gx \supset Gy$ and thus $Gx = Gy$. Thereby we obtain the following result:

Remark 5. *Let $G \times X \longrightarrow X$ be an action of a group G on a set X. Then X is the disjoint union of its G-orbits.*

Given a group action $G \times X \longrightarrow X$ and a G-orbit $B \subset X$, every element $x \in B$ will be referred to as a *representative* of this orbit. Likewise, a *system of representatives* of a family $(B_i)_{i\in I}$ of disjoint G-orbits is a family $(x_i)_{i\in I}$ of elements of X such that $x_i \in B_i$ for all $i \in I$. The action $G \times X \longrightarrow X$ is called *transitive* if there is only a single G-orbit.

We want to characterize the orbits of a group action in more specific terms. As usual, $\operatorname{ord} M$ denotes the number of elements of a set M, and $(G : H)$ denotes the index of a subgroup H in a group G.

Remark 6 (Orbit-stabilizer lemma). *Let $G \times X \longrightarrow X$ be a group action. For a point $x \in X$, the map $G \longrightarrow X$, $g \longmapsto gx$, induces a bijection $G/G_x \xrightarrow{\sim} Gx$ from the* set *of left cosets of G modulo the stabilizer subgroup G_x onto the orbit of x under G. In particular, we get*

$$\operatorname{ord} Gx = \operatorname{ord} G/G_x = (G : G_x).$$

Proof. Look at the surjective map

$$\varphi\colon G \longrightarrow Gx, \qquad g \longmapsto gx,$$

and observe for $g, h \in G$ the following equivalences:

$$\begin{aligned}\varphi(g) = \varphi(h) &\iff gx = hx \iff h^{-1}gx = x \iff h^{-1}g \in G_x \\ &\iff gG_x = hG_x.\end{aligned}$$

This shows that φ induces, similarly as in the case of the fundamental theorem on homomorphisms 1.2/7, a bijection $G/G_x \xrightarrow{\sim} Gx$. □

As a direct consequence, we can conclude the following from Remark 5 and Remark 6:

Proposition 7 (Orbit equation). *Let $G \times X \longrightarrow X$ be an action of a group G on a finite set X, and let $x_1, \ldots, x_n$ be a system of representatives of the orbits of X. Then*

$$\operatorname{ord} X = \sum_{i=1}^{n} \operatorname{ord}(Gx_i) = \sum_{i=1}^{n} (G : G_{x_i}).$$

We will apply the orbit equation especially for $X = G$ and the conjugation action $G \times G \longrightarrow G$, in order to derive the so-called class equation; see Proposition 9 below. In the following let G be a group and $S \subset G$ a subset. The *centralizer* of S in G is given by

$$Z_S = \big\{x \in G\,;\, xs = sx \text{ for all } s \in S\big\}.$$

Furthermore, the *center* of G is defined as the centralizer of G, i.e., as

$$Z = Z_G = \big\{x \in G\,;\, xs = sx \text{ for all } s \in G\big\}.$$

Finally, the *normalizer* of S in G is given by

$$N_S = \big\{x \in G\,;\, xS = Sx\big\}.$$

Remark 8. (i) *Z is a normal subgroup in G.*
(ii) *Z_S and N_S are subgroups in G.*
(iii) *If S is a subgroup in G, then N_S is the largest of all subgroups $H \subset G$ such that S is a normal subgroup of H.*

All these assertions are easy to check. As an example, let us consider the case Z_S in (ii). If S consists of a single element s, then $Z_S = N_S$ equals the stabilizer group of s with respect to the conjugation action on G. Thereby we see for general S that $Z_S = \bigcap_{s\in S} Z_{\{s\}}$ is a subgroup of G. Also note that always $Z_S \subset N_S$.

Proposition 9 (Class equation). *Let G be a finite group with center Z. Furthermore, consider the conjugation action on G and let $x_1, \dots, x_n$ be a system of representatives of the orbits contained in $G-Z$. Then*

$$\operatorname{ord} G = \operatorname{ord} Z + \sum_{i=1}^{n} \big(G : Z_{\{x_i\}}\big).$$

Proof. The orbit of an element $x \in Z$ consists only of the element x itself. On the other hand, we can identify the orbit of an element $x \in G-Z$ with $G/Z_{\{x\}}$; cf. Remark 6. Therefore, the assertion follows from the orbit equation in Proposition 7. □

Finally, let us add two results on the center Z of a group G. Since Z equals the kernel of the homomorphism

$$G \longrightarrow \operatorname{Aut}(G), \qquad g \longmapsto \operatorname{int}_g,$$

we can conclude the following from the fundamental theorem on homomorphisms in the version of 1.2/7:

Remark 10. *The group of inner automorphisms of G is isomorphic to G/Z.*

Remark 11. *If G/Z is cyclic, then G is abelian.*

Proof. Fix an element $a \in G$ such that G/Z is generated by the residue class $\overline{a}$ of a. Furthermore, consider elements $g, h \in G$ with residue classes $\overline{g} = \overline{a}^m$, $\overline{h} = \overline{a}^n$. Then there are elements $b, c \in Z$ such that $g = a^m b$, $h = a^n c$. Since

$$gh = a^m b a^n c = a^{m+n} bc, \qquad hg = a^n c a^m b = a^{m+n} cb = a^{m+n} bc,$$

it follows that $gh = hg$. □

Exercises

1. *Let G be a finite group and $H \subset G$ a subgroup. Consider the action of H on G via left translation (resp. right translation) and interpret the corresponding orbit equation in terms of elementary group theory.*

2. *Let L/K be a finite Galois extension with Galois group G. Consider the natural action of G on L and interpret the stabilizer group G_a for $a \in L$, as well as the orbit Ga, in terms of Galois theory. Furthermore, determine the orders of G_a and Ga.*

3. Let G be a group and X the set of all subgroups of G. Show:
   (i) $G \times X \longrightarrow X$, $(g, H) \longmapsto gHg^{-1}$, defines an action of G on X.
   (ii) The orbit of an element $H \in X$ consists of H itself if and only if H is a normal subgroup of G.
   (iii) If the order of G is a power of a prime number p, then the number of subgroups in G differs from the number of normal subgroups in G by a multiple of p.

4. Let G be a finite group, H a subgroup, and N_H its normalizer. Furthermore, write $M := \bigcup_{g \in G} gHg^{-1}$ and show:
   (i) $\operatorname{ord} M \leq (G : N_H) \cdot \operatorname{ord} H$.
   (ii) $H \neq G$ implies $M \neq G$.

5. Let G be a group, H a subgroup, as well as N_H and Z_H the corresponding normalizer and centralizer of H in G. Show that Z_H is a normal subgroup of N_H and that the group N_H/Z_H is isomorphic to a subgroup of the automorphism group $\operatorname{Aut}(H)$.

6. *Burnside's lemma*: Let $G \times X \longrightarrow X$ be an action of a finite group G on a set X. Write X/G for the set of orbits, as well as $X^g = \{x \in X \,;\, gx = x\}$ for the set of elements in X that are left fixed by an element $g \in G$. Show that

$$\operatorname{ord}(X/G) = \frac{1}{\operatorname{ord} G} \cdot \sum_{g \in G} \operatorname{ord} X^g.$$

5.2 Sylow Groups

The fundamental theorem of finitely generated abelian groups 2.9/9 gives precise information on the structure of such groups, in particular, on the structure of finite abelian groups. In the following we will study finite groups without the commutativity condition. Our main objective is to derive the theorems named after L. Sylow on the existence of certain subgroups, called p-Sylow subgroups (or Sylow p-subgroups). We start by introducing the notion of Sylow groups, and in particular of p-groups.

Definition 1. *Let G be a finite group and p a prime number.*
(i) *G is called a* p-group *if the order of G is a power of p.*

(ii) *A subgroup $H \subset G$ is called a* p-Sylow subgroup *if H is a p-group such that p does not divide the index $(G : H)$, in other words, if there exist integers $k, m \in \mathbb{N}$ satisfying* $\operatorname{ord} H = p^k$, *as well as* $\operatorname{ord} G = p^k m$, *where* $p \nmid m$ (*use* 1.2/3).

It follows from the theorem of Lagrange 1.2/3 that the order of every element of a p-group is a power of p. Similarly, the same result shows that a p-Sylow subgroup cannot be strictly contained in a p-subgroup of G and therefore is a maximal p-subgroup of G. The converse of this fact will follow later from the Sylow theorems; cf. Corollary 11. The trivial subgroup $\{1\} \subset G$ is an example of a p-group, and for $p \nmid \operatorname{ord} G$, even an example of a p-Sylow subgroup in G. Furthermore, we can read from the fundamental theorem 2.9/9 that a finite *abelian* group G contains a unique p-Sylow subgroup $S_p \neq \{1\}$ for each prime p dividing $\operatorname{ord} G$ and furthermore, that G is the direct sum of all these Sylow subgroups. Although this is not needed later on, let us illustrate in Remark 2 below how to prove the existence of Sylow subgroups in this case by means of elementary arguments. On the other hand, Remark 2 can be viewed as a simple consequence of the Sylow theorems, once they have been established; cf. Exercise 1.

Remark 2. *Let G be a finite* abelian *group. For every prime number p, there is exactly one p-Sylow subgroup in G, namely*

$$S_p = \big\{a \in G\,;\, a^{p^t} = 1 \text{ for some } t \in \mathbb{N}\big\}.$$

Proof. First we have to show that S_p is a subgroup of G. To do this consider elements $a, b \in S_p$, say where $a^{p^{t'}} = 1$ and $b^{p^{t''}} = 1$, and write $t = \max(t', t'')$. Using the commutativity of G, we get

$$(ab^{-1})^{p^t} = a^{p^t} \cdot b^{-p^t} = 1$$

and thus $ab^{-1} \in S_p$. Since we have $1 \in S_p$ anyway, S_p is indeed a subgroup of G. By its definition, S_p contains all elements in G whose order is a power of p. In particular, it will contain all p-subgroups of G. Thus, if we can show that S_p is a p-Sylow subgroup of G, it will be unique.

To justify that S_p is a p-Sylow group, we proceed by induction on $n = \operatorname{ord} G$. For $n = 1$ nothing has to be shown. Therefore, assume $n > 1$ and choose an element $x \neq 1$ in G. Replacing x by a suitable power of itself, we may assume that $q = \operatorname{ord} x$ is prime. Then look at the cyclic subgroup $\langle x \rangle \subset G$ generated by x, and consider the projection $\pi\colon G \longrightarrow G' = G/\langle x \rangle$, where $\operatorname{ord} G' = \frac{1}{q} \operatorname{ord} G$ by the theorem of Lagrange 1.2/3.

If $S'_p \subset G'$ denotes the subgroup consisting of all elements in G' whose order is a power of p, we know from the induction hypothesis that S'_p is a p-Sylow subgroup in G'. Furthermore, $\pi(S_p) \subset S'_p$, and we claim that we have even $\pi(S_p) = S'_p$. To justify this, consider an element $\overline{a} \in S'_p$ with π-preimage $a \in G$. If p^t is the order of $\overline{a}$, we get $a^{p^t} \in \langle x \rangle$ and hence $a^{p^t q} = 1$. For $p = q$ we can

conclude that $a \in S_p$. On the other hand, if $p \neq q$, we see that p and q are relatively prime. Hence, there is an equation $rp^t + sq = 1$ for some integers r, s. This implies

$$\pi(a^{sq}) = \overline{a}^{sq} = \overline{a}^{rp^t}\overline{a}^{sq} = \overline{a}^{rp^t+sq} = \overline{a}.$$

In addition, we have $a^{sq} \in S_p$, since $a^{p^t q} = 1$. Thus, in either case, π induces a surjective map $\pi_p \colon S_p \longrightarrow S'_p$ satisfying $\ker \pi_p = \langle x \rangle \cap S_p$.

Now let $n = \operatorname{ord} G = p^k m$, where $p \nmid m$. If $p = q$, the order of G' is given by $\operatorname{ord} G' = \frac{1}{p} \operatorname{ord} G = p^{k-1}m$, and we read $\operatorname{ord} S'_p = p^{k-1}$ from the induction hypothesis. Furthermore, we have $\langle x \rangle \subset S_p$ and therefore $\ker \pi_p = \langle x \rangle$, so that π_p induces an isomorphism $S_p/\langle x \rangle \xrightarrow{\sim} S'_p$. This shows that $\operatorname{ord} S_p = p \cdot \operatorname{ord} S'_p = p^k$ by 1.2/3, and it follows that S_p is a p-Sylow subgroup in G. On the other hand, if $p \neq q$, we have $\operatorname{ord} G' = p^k \cdot \frac{m}{q}$ and hence $\operatorname{ord} S'_p = p^k$. Since $\langle x \rangle$ cannot contain an element whose order is a nontrivial p-power, we get $\ker \pi_p = \langle x \rangle \cap S_p = \{1\}$, and it follows that π_p restricts to an isomorphism $S_p \xrightarrow{\sim} S'_p$. In particular, $\operatorname{ord} S_p = \operatorname{ord} S'_p = p^k$, so that also in this case, S_p is a p-Sylow subgroup in G. □

In the noncommutative case, the theory of p-groups, and in particular of p-Sylow groups, is more complicated. We start by considering p-groups.

Proposition 3. *Let G be a p-group of order p^k, for a prime number p and an exponent $k \geq 1$. Then p divides the order of the center Z of G, so that $Z \neq \{1\}$.*

Proof. Look at the class equation 5.1/9 for the conjugation action of G on itself

$$\operatorname{ord} G = \operatorname{ord} Z + \sum_{i=1}^{n} \big(G : Z_{\{x_i\}}\big),$$

where $x_1, \ldots, x_n$ is a system of representatives of the G-orbits in $G - Z$. By 1.2/3, the index $(G : Z_{\{x_i\}})$ is a p-power for each i, since $\operatorname{ord} G$ is a p-power. Furthermore, $(G : Z_{\{x_i\}})$ is even a nontrivial p-power, since $Z_{\{x_i\}}$ is a proper subgroup of G, due to $x_i \notin Z$. Consequently, we get $p \,|\, \operatorname{ord} Z$. □

Corollary 4. *Let G be a p-group of order p^k, for a prime number p. Then there is a descending chain of subgroups*

$$G = G_k \supset G_{k-1} \supset \ldots \supset G_0 = \{1\}$$

such that $\operatorname{ord} G_\ell = p^\ell$ and $G_{\ell-1}$ is a normal subgroup in G_ℓ for $\ell = 1, \ldots, k$.[1]

In particular, for every divisor p^ℓ of p^k there is a p-subgroup $H \subset G$ such that $\operatorname{ord} H = p^\ell$. If $k \geq 1$, it follows that G admits an element of order p.

Proof. We conclude by induction on k, the case $k = 0$ being trivial. Therefore, assume $k > 0$. Applying Proposition 3, the center $Z \subset G$ is nontrivial and there

[1] The quotients $G_\ell/G_{\ell-1}$ are of order p and hence cyclic as well as abelian. Thereby it is seen that every finite p-group G is *solvable* in the sense of Definition 5.4/3.

is an element $a \neq 1$ in Z. If p^r is its order, we see that $a^{p^{r-1}}$ is of order p. Hence, we may assume $\operatorname{ord} a = p$. Since a belongs to the center of G, the subgroup $\langle a \rangle \subset G$ generated by a is normal in G. Then $\overline{G} = G/\langle a \rangle$ is of order p^{k-1}, using 1.2/3, and we can apply the induction hypothesis to this group. Hence, there exists a chain of subgroups

$$\overline{G} = \overline{G}_k \supset \overline{G}_{k-1} \supset \ldots \supset \overline{G}_1 = \{1\}, \qquad \operatorname{ord} \overline{G}_\ell = p^{\ell-1},$$

such that $\overline{G}_{\ell-1}$ is a normal subgroup in $\overline{G}_\ell$ for $\ell = 2, \ldots, k$. Now consider the projection $\pi \colon G \longrightarrow G/\langle a \rangle$ and set $G_\ell = \pi^{-1}(\overline{G}_\ell)$ for $\ell = 1, \ldots, k$. Then clearly,

$$G = G_k \supset G_{k-1} \supset \ldots \supset G_1 \supset \{1\}$$

is a chain of subgroups in G as desired. □

Proposition 5. *For a prime number p, let G be a group of order p^2. Then G is abelian. More precisely, we have*

$$G \simeq \mathbb{Z}/p^2\mathbb{Z} \quad \text{or} \quad G \simeq \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}.$$

Proof. To start with we show that G is abelian. From Proposition 3 we conclude that $p \mid \operatorname{ord} Z$ for the center Z of G and hence that Z is of order p or p^2. If $\operatorname{ord} Z = p^2$, then $G = Z$ and G is abelian. On the other hand, if $\operatorname{ord} Z = p$, it follows that G cannot be abelian. However, G/Z is of order p then, in fact cyclic of order p, and we could conclude from 5.1/11 that G is abelian, in contradiction to the fact that $\operatorname{ord} Z = p$.

Now use Corollary 4 and choose an element $a \in G$ such that $\operatorname{ord} a = p$. Furthermore, let $b \in G$ belong to the complement of the cyclic subgroup $\langle a \rangle \subset G$ generated by a. Then b is of order p or p^2, where in the latter case G is generated by b, implying $G = \langle b \rangle \simeq \mathbb{Z}/p^2\mathbb{Z}$. Therefore, assume $\operatorname{ord} b = p$. We claim that the map

$$\varphi \colon \langle a \rangle \times \langle b \rangle \longrightarrow G, \qquad (a^i, b^j) \longmapsto a^i b^j,$$

is a group isomorphism. First, φ is a group homomorphism, since we know already that G is abelian. Furthermore, we see that $\langle a \rangle \cap \langle b \rangle$ is a proper subgroup of $\langle b \rangle$, since $b \notin \langle a \rangle$. Hence, we must have $\langle a \rangle \cap \langle b \rangle = \{1\}$, and φ is injective. But then φ is surjective as well, since

$$\operatorname{ord}\big(\langle a \rangle \times \langle b \rangle\big) = p^2 = \operatorname{ord} G.$$

Using $\langle a \rangle \simeq \mathbb{Z}/p\mathbb{Z} \simeq \langle b \rangle$, we get $G \simeq \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$, as desired. Alternatively, we could have based our argument on the fundamental theorem of finitely generated abelian groups 2.9/9. □

After these preliminaries, let us derive the Sylow theorems that were mentioned before; they correspond to the different items of the following result:

Theorem 6 (Sylow theorems). *Let G be a finite group and p a prime number.*

(i) *The group G contains at least one p-Sylow subgroup. More precisely, for every p-subgroup $H \subset G$ there is a p-Sylow subgroup $S \subset G$ such that $H \subset S$.*

(ii) *If $S \subset G$ is a p-Sylow subgroup, then every subgroup in G that is conjugate to S is a p-Sylow subgroup of G as well. Conversely, any two p-Sylow subgroups in G are conjugate to each other.*

(iii) *The number s of p-Sylow subgroups in G satisfies*

$$s \,|\, \operatorname{ord} G, \qquad s \equiv 1 \mod (p).$$

We divide the proof of the theorem into several parts and start with a fundamental lemma. Its proof will be given following an idea of H. Wielandt, similarly as in [10], Kap. I, Satz 7.2.

Lemma 7. *Let G be a finite group of order $n = p^k m$, where p is prime, but not necessarily relatively prime to m. Then the number s of p-subgroups $H \subset G$ having order $\operatorname{ord} H = p^k$ satisfies the relation*

$$s \equiv \binom{n-1}{p^k-1} = \frac{1}{m}\binom{n}{p^k} \mod (p).$$

Proof. We write X for the set of all subsets in G that consist of precisely p^k elements. Then

$$\operatorname{ord} X = \binom{n}{p^k},$$

and G acts on X by "left translation" via

$$G \times X \longrightarrow X, \qquad (g, U) \longmapsto gU = \{gu\,;\, u \in U\}.$$

Different from our previous notation, we write $G(U)$ for the G-orbit of an element $U \in X$; as usual, G_U stands for the stabilizer subgroup of U in G. Viewing U as a subset of G, the left translation of G on itself gives rise to an action of G_U on U. Therefore, U consists of certain right cosets of G_U in G. These are disjoint and consist of $\operatorname{ord} G_U$ elements each. Therefore, $\operatorname{ord} G_U$ divides $\operatorname{ord} U = p^k$ and hence is of type $p^{k'}$ for some $k' \le k$. In particular, U itself is a right coset of G_U if and only if $\operatorname{ord} G_U = p^k$.

Now let $(U_i)_{i\in I}$ be a system of elements in X representing all G-orbits of X. Then the orbit equation 5.1/7 yields

$$\binom{n}{p^k} = \operatorname{ord} X = \sum_{i\in I} \operatorname{ord} G(U_i) = \sum_{i\in I} (G : G_{U_i}).$$

We want to exploit this equation by taking equivalence classes modulo (pm). As we have seen, G_{U_i} is a p-group of order p^{k_i} for some $k_i \le k$. Using the

Theorem of Lagrange 1.2/3, this implies $(G : G_{U_i}) = p^{k-k_i}m$. Then, writing $I' = \{i \in I\,;\, k_i = k\}$, we get

$$(\operatorname{ord} I') \cdot m = \sum_{i \in I'} (G : G_{U_i}) \equiv \binom{n}{p^k} \mod (pm),$$

and it is enough for the proof of the lemma to show that $\operatorname{ord} I'$ coincides with the number s of all p-subgroups $H \subset G$ of order p^k.

To justify this assertion, recall that an index $i \in I$ belongs to I' if and only if $\operatorname{ord} G(U_i) = (G : G_{U_i}) = m$, hence if and only if the orbit $G(U_i)$ consists of precisely m elements. Now consider for a p-subgroup $H \subset G$ of order p^k the G-orbit $G(H) \subset X$; it consists of the left cosets of H in G, hence, by the theorem of Lagrange 1.2/3, of precisely m elements. Two different such subgroups $H, H' \subset G$ induce different G-orbits, since $gH = H'$ for some element $g \in G$ implies $g \in H$ and therefore $H = H'$, due to $1 \in H'$. On the other hand, it is easy to see that every G-orbit $G(U_i)$, $i \in I'$, is of type $G(H)$ for a p-subgroup $H \subset G$ of order p^k. Indeed, for $i \in I'$ we have $\operatorname{ord} G_{U_i} = p^k$, and as seen above, U_i is a right coset of G_{U_i} in G, say $U_i = G_{U_i} \cdot u_i$ for some $u_i \in U_i$. Then the G-orbit of U_i in X satisfies

$$G(U_i) = G(u_i^{-1} \cdot U_i) = G(u_i^{-1} \cdot G_{U_i} \cdot u_i),$$

where now $H = u_i^{-1} \cdot G_{U_i} \cdot u_i$ is a p-subgroup in G of order p^k. Therefore, the elements $i \in I'$ correspond bijectively to the p-subgroups $H \subset G$ of order p^k, and the assertion of the lemma is clear. □

For a cyclic group of order n and a divisor d of n, there is always a unique subgroup of order d; cf. 1.3, Exercise 2 and its solution in the appendix. In this way, we can read from Lemma 7 the nontrivial relation

$$\binom{n-1}{p^k-1} = \frac{1}{m}\binom{n}{p^k} \equiv 1 \mod (p),$$

which leads to the following partial generalization of Corollary 4:

Proposition 8. *Let G be a finite group and p^k a prime power dividing* $\operatorname{ord} G$. *Then the number s of p-subgroups $H \subset G$ of order p^k satisfies $s \equiv 1 \mod (p)$ and hence is nonzero.*

In particular, choosing p^k as a maximal p-power dividing $\operatorname{ord} G$, we see that G contains at least one p-Sylow subgroup, even more specifically, that the number of these subgroups is congruent to 1 modulo p.

Lemma 9. *Let G be a finite group, $H \subset G$ a p-subgroup, and $S \subset G$ a p-Sylow subgroup. Then there is an element $g \in G$ such that $H \subset gSg^{-1}$.*

Proof. On G/S, the set of left cosets of S in G, we consider the H-action

$$H \times G/S \longrightarrow G/S, \qquad (h, gS) \longmapsto (hg)S,$$

and apply the theorem of Lagrange 1.2/3, in conjunction with the orbit-stabilizer lemma 5.1/6 as well as the orbit equation 5.1/7. The order of every H-orbit in G/S divides the order of H and hence is a p-power, since H is a p-group. However, p does not divide $\operatorname{ord} G/S$. Therefore, there must exist an H-orbit whose order is a trivial p-power p^0 and hence is 1. Then this H-orbit is a left coset gS of S, and we have $hgS = gS$ for all $h \in H$. Since $1 \in S$, this implies $hg \in gS$, or $h \in gSg^{-1}$, and therefore $H \subset gSg^{-1}$. □

Since the map $G \longrightarrow G$, $x \longmapsto gxg^{-1}$, is an automorphism for every $g \in G$, we see in the situation of Lemma 9 for every p-Sylow subgroup S in G that gSg^{-1} is a p-Sylow subgroup in G as well. If $H \subset G$ is another p-Sylow subgroup in G, an inclusion $H \subset gSg^{-1}$ as in Lemma 9 implies $H = gSg^{-1}$, due to the fact that $\operatorname{ord} H = \operatorname{ord} S = \operatorname{ord} gSg^{-1}$. As a consequence, Proposition 8 and Lemma 9 together imply the assertions of Theorem 6, except for the fact that $s \,|\, \operatorname{ord} G$ in (iii). However, this remaining part will be a consequence of the following result, using the theorem of Lagrange 1.2/3:

Lemma 10. *Let G be a finite group and S a p-Sylow subgroup in G. Writing N_S for the normalizer of S in G, the index $(G : N_S)$ equals the number of p-Sylow subgroups in G.*

Proof. Let X be the set of p-Sylow subgroups in G. Since all p-Sylow subgroups are conjugate in G, the conjugation action

$$G \times X \longrightarrow X, \qquad (g, S') \longmapsto gS'g^{-1},$$

is transitive. In particular, the orbit-stabilizer lemma 5.1/6 yields

$$\operatorname{ord} X = (G : G_S),$$

where G_S, the stabilizer group with respect to the conjugation action, coincides with the normalizer N_S. □

Thus, summing up, the proof of Theorem 6 and hence of the Sylow theorems is now complete. We want to draw some consequences from these results.

Corollary 11. *Let G be a finite group and p a prime number. Then:*

(i) *If $p \,|\, \operatorname{ord} G$, then G admits an element of order p.*

(ii) *G is a p-group if and only if for every $a \in G$ there exists an exponent $t \in \mathbb{N}$ such that $a^{p^t} = 1$.*

(iii) *A subgroup $H \subset G$ is a p-Sylow subgroup if and only if it is a maximal p-group in G.*

Proof. Assertion (i) follows from Proposition 8, or alternatively, from Theorem 6 (i), in conjunction with Corollary 4.

To verify (ii), assume that every element $a \in G$ admits a p-power as its order. If $\operatorname{ord} G$ is not a p-power, choose a prime number q different from p that divides $\operatorname{ord} G$. Then, as we have seen, G will contain an element of order q, in contradiction to our assumption. Therefore, $\operatorname{ord} G$ is a p-power, and hence G a p-group. Conversely, if G is a p-group, the order of any element $a \in G$ is a p-power, since $\operatorname{ord} a$ divides $\operatorname{ord} G$ by the theorem of Lagrange 1.2/3.

Finally, we conclude from 1.2/3 again that every p-Sylow subgroup of G is a maximal p-subgroup. The converse of this follows from Theorem 6 (i), so that assertion (iii) is clear, too. □

Proposition 12. *Let p, q be prime numbers such that $p < q$ and $p \nmid (q-1)$. Then every group G of order pq is cyclic.*

Proof. Let s be the number of p-Sylow subgroups in G. Then, by Theorem 6 (iii), we have $s \mid \operatorname{ord} G$, i.e., $s \mid pq$, as well as $s \equiv 1(p)$. This implies $p \nmid s$ and hence $s \mid q$. Since $q = s \equiv 1(p)$ is excluded by the condition $p \nmid (q-1)$, we must have $s = 1$. Hence, there is precisely one p-Sylow subgroup S_p in G. It is invariant under conjugation with elements of G and therefore normal in G. Likewise, if s' is the number of q-Sylow groups in G, we conclude that $s' \mid p$. Again, the case $s' = p$ is excluded, since $p = s' \equiv 1(q)$ is not compatible with $p < q$. Therefore, we must have $s' = 1$, and there is a unique q-Sylow subgroup S_q in G, which, as before, is normal in G. Since S_p and S_q do not admit proper subgroups except for the trivial group $\{1\}$, we see that $S_p \cap S_q = \{1\}$.

Now we claim that the map

$$\varphi\colon S_p \times S_q \longrightarrow G, \qquad (a, b) \longmapsto ab,$$

is an isomorphism of groups. Knowing this, we can conclude, for example by the Chinese remainder theorem in the version of 2.4/14, that G, being the Cartesian product of two cyclic groups of relatively prime orders, is itself cyclic. To show that φ is indeed a group homomorphism, choose elements $a \in S_p$, $b \in S_q$ and observe that

$$aba^{-1}b^{-1} = (aba^{-1})b^{-1} \in S_q,$$

as well as

$$aba^{-1}b^{-1} = a(ba^{-1}b^{-1}) \in S_p.$$

Therefore,

$$aba^{-1}b^{-1} \in S_p \cap S_q = \{1\}$$

and hence $ab = ba$, showing that the elements of S_p commute with those of S_q. In particular, for $a, a' \in S_p$ and $b, b' \in S_q$ we can write

$$\begin{aligned}\varphi\big((a,b)\cdot(a',b')\big) &= \varphi(aa', bb') = aa'bb' \\ &= aba'b' = \varphi(a,b)\cdot\varphi(a',b'),\end{aligned}$$

which implies that φ is a group homomorphism. Since $S_p \cap S_q = \{1\}$, it follows that φ is injective, and even bijective, since the orders of $S_p \times S_q$ and G coincide. □

Exercises

1. *Review the Sylow theorems and give an outline of the information they provide for finite abelian groups.*
2. *Let $\varphi\colon G \longrightarrow G'$ be a homomorphism between finite groups. Try to relate the Sylow subgroups of G to those of G'.*
3. Let G be a finite group and $H \subset G$ a p-subgroup, for some prime number p. If H is normal in G, show that H is contained in every p-Sylow subgroup of G.
4. Let $\mathrm{GL}(n, K)$ be the group of all invertible $(n \times n)$ matrices over a finite field K of characteristic $p > 0$. Show that the upper triangular matrices with diagonal elements equal to 1 give rise to a p-Sylow subgroup of $\mathrm{GL}(n, K)$.
5. Show that every group of order 30 or 56 admits a nontrivial Sylow subgroup that is normal.
6. Show that every group of order 45 is abelian.
7. Show that every group G of order 36 admits a nontrivial normal subgroup. *Hint*: Consider the action of G on the set of 3-Sylow subgroups of G.
8. Show that every group G of order $\operatorname{ord} G < 60$ is cyclic or admits a nontrivial normal subgroup.

5.3 Permutation Groups

In the following we want to have a closer look at the group $\mathfrak{S}_n$ of bijective self-maps of $\{1, \ldots, n\}$. As we know already, $\mathfrak{S}_n$ is called the *symmetric group* or the *permutation group* of $\{1, \ldots, n\}$. This group acts naturally on $\{1, \ldots, n\}$ and satisfies $\operatorname{ord} \mathfrak{S}_n = n!$. Elements $\pi \in \mathfrak{S}_n$ are frequently written in the form

$$\begin{pmatrix} 1 & \ldots & n \\ \pi(1) & \ldots & \pi(n) \end{pmatrix},$$

in particular when the images $\pi(1), \ldots, \pi(n)$ are given by explicit expressions. A permutation $\pi \in \mathfrak{S}_n$ is called a *cycle* if there are distinct elements $x_1, \ldots, x_r$ in $\{1, \ldots, n\}$, $r \geq 2$, such that

$$\begin{aligned} \pi(x_i) &= x_{i+1} \quad \text{for} \quad 1 \leq i < r, \\ \pi(x_r) &= x_1, \\ \pi(x) &= x \quad \text{for} \quad x \in \{1, \ldots, n\} - \{x_1, \ldots, x_r\}. \end{aligned}$$

More precisely, in such a situation π is called an *r-cycle*, and one uses the notation $\pi = (x_1, \ldots, x_r)$. Two cycles $(x_1, \ldots, x_r)$ and $(y_1, \ldots, y_s)$ are called *disjoint* if

$$\{x_1, \ldots, x_r\} \cap \{y_1, \ldots, y_s\} = \emptyset.$$

A 2-cycle is called a *transposition*.

Proposition 1. *Let $n \geq 2$.*

(i) *If $\pi_1, \pi_2 \in \mathfrak{S}_n$ are disjoint cycles, then $\pi_1 \circ \pi_2 = \pi_2 \circ \pi_1$.*

(ii) *Every permutation $\pi \in \mathfrak{S}_n$ can be written as a product of disjoint cycles. These are uniquely determined by π, up to ordering.*

(iii) *Every permutation $\pi \in \mathfrak{S}_n$ is a product of transpositions.*

Proof. Assertion (i) is trivial. In the situation of (ii) we write $H = \langle \pi \rangle$ for the cyclic subgroup generated by π in $\mathfrak{S}_n$. Then consider the natural action of H on $\{1, \ldots, n\}$ and look at the corresponding partition of $\{1, \ldots, n\}$ into disjoint H-orbits. Let $B_1, \ldots, B_\ell$ be the orbits that contain at least two elements, i.e., that satisfy $r_\lambda = \operatorname{ord} B_\lambda \geq 2$. Choosing a point $x_\lambda \in B_\lambda$ for each $\lambda = 1, \ldots, \ell$, we get

$$B_\lambda = \big\{x_\lambda, \pi(x_\lambda), \ldots, \pi^{r_\lambda - 1}(x_\lambda)\big\}$$

and

$$\pi = \prod_{\lambda=1}^{\ell} \big(x_\lambda, \pi(x_\lambda), \ldots, \pi^{r_\lambda - 1}(x_\lambda)\big),$$

hence a factorization of π into disjoint cycles, where the ordering of factors does not matter, due to (i). On the other hand, every factorization of π into a product of disjoint cycles corresponds, in the manner as explained before, to the decomposition of $\{1, \ldots, n\}$ into its H-orbits. This settles the uniqueness assertion.

Finally, assertion (iii) follows from (ii) using the factorization

$$(x_1, \ldots, x_r) = (x_1, x_2) \circ (x_2, x_3) \circ \ldots \circ (x_{r-1}, x_r).$$

□

Given a permutation $\pi \in \mathfrak{S}_n$, one defines its *sign* or *signature* by

$$\operatorname{sgn} \pi = \prod_{i<j} \frac{\pi(i) - \pi(j)}{i - j} \, .$$

It is clear that the sign can assume only the values 1 and -1. If $\operatorname{sgn} \pi$ is positive, π is called an *even* permutation, and an *odd* permutation otherwise. The definition of the sign of a permutation counts, so to speak, in a multiplicative way modulo 2, the number of two-element subsets $\{i, j\} \subset \{1, \ldots, n\}$ for which the map π reverses the size relationship between i and j. In particular, the above product may alternatively be executed over any index set I of pairs (i, j) of integers $1 \leq i, j \leq n$ such that the mapping $(i, j) \longmapsto \{i, j\}$ sets up a bijection between I and the set of two-element subsets of $\{1, \ldots, n\}$.

Remark 2. *The map* $\operatorname{sgn} \colon \mathfrak{S}_n \longrightarrow \{1, -1\}$ *is a group homomorphism.*

Proof. Let $\pi, \pi' \in \mathfrak{S}_n$. Then

$$\begin{aligned}\operatorname{sgn}\pi\circ\pi' &= \prod_{i<j}\frac{\pi\circ\pi'(i)-\pi\circ\pi'(j)}{i-j}\\ &= \prod_{i<j}\frac{\pi\circ\pi'(i)-\pi\circ\pi'(j)}{\pi'(i)-\pi'(j)}\cdot\frac{\pi'(i)-\pi'(j)}{i-j}\\ &= \operatorname{sgn}\pi\cdot\operatorname{sgn}\pi'.\end{aligned}$$

□

Clearly, the sign of a transposition in $\mathfrak{S}_n$ equals -1. Thus, if we factorize a permutation $\pi\in\mathfrak{S}_n$ according to Proposition 1 (iii) into a product of transpositions, say $\pi=\tau_1\circ\ldots\circ\tau_\ell$, we get $\operatorname{sgn}\pi=(-1)^\ell$. It follows that the residue class of ℓ modulo 2 is uniquely determined by π. In particular, π is an even or an odd permutation depending on whether π is a product of an even or an odd number of transpositions. Furthermore, we can conclude from Remark 2 that

$$\mathfrak{A}_n=\ker\operatorname{sgn}=\big\{\pi\in\mathfrak{S}_n\,;\ \operatorname{sgn}\pi=1\big\},$$

i.e., the set of all even permutations, defines a normal subgroup of index 2 in $\mathfrak{S}_n$, provided $n>1$. The group $\mathfrak{A}_n$ is called the *alternating group* on $\{1,\ldots,n\}$.

Proposition 3. *For $n\geq 3$, the group $\mathfrak{A}_n$ consists of all permutations $\pi\in\mathfrak{S}_n$ that can be written as a product of 3-cycles.*

Proof. Consider elements $x_1,x_2,x_3,x_4\in\{1,\ldots,n\}$. If x_1,x_2,x_3 are distinct, then the following formula holds:

$$(x_1,x_2)\circ(x_2,x_3)=(x_1,x_2,x_3).$$

Furthermore, if x_1,x_2,x_3,x_4 are distinct, then

$$(x_1,x_2)\circ(x_3,x_4)=(x_1,x_3,x_2)\circ(x_1,x_3,x_4).$$

The first equation shows that every 3-cycle belongs to $\mathfrak{A}_n$, a fact that extends to products of 3-cycles as well. Both equations together imply that the product of an even number of transpositions, in particular every element of $\mathfrak{A}_n$, is a product of 3-cycles. □

To end this section, let us look at some special permutation groups, as well as at some subgroups of these.

(1) The group $\mathfrak{S}_2$ is of order 2, so that $\mathfrak{S}_2\simeq\mathbb{Z}/2\mathbb{Z}$.

(2) The group $\mathfrak{S}_3$ is of order 6 and can be viewed as the *dihedral group* D_3, i.e., as the group of rotations and reflections of an equilateral triangle (three rotations, three reflections). Since $\mathfrak{S}_3$ contains elements of order 1, 2, and 3, but not of order 6, we see that it cannot be isomorphic to the cyclic group $\mathbb{Z}/6\mathbb{Z}$. On the other hand, every abelian group of order 6 is isomorphic to the

product $\mathbb{Z}/2\mathbb{Z}\times\mathbb{Z}/3\mathbb{Z}$ and therefore is cyclic, due to the Chinese remainder theorem 2.4/14. As a consequence, $\mathfrak{S}_3$ is nonabelian of order 6, which alternatively, can just as well be verified by direct inspection.

(3) More generally, the *dihedral group* D_n can be defined for indices $n \geq 3$ as the group of symmetries (rotations and reflections) of a regular (convex) n-gon. Numbering the corners of the n-gon consecutively by $1, \ldots, n$, we can view D_n as the subgroup of $\mathfrak{S}_n$ that is generated by the permutations

$$\sigma = (1, \ldots, n), \qquad \tau = \begin{pmatrix} 1 & 2 & 3 & \ldots & n \\ 1 & n & n-1 & \ldots & 2 \end{pmatrix}.$$

The permutation σ corresponds to the rotation of the regular n-gon by the angle $2\pi/n$, and τ to the reflection across the axis passing through the point 1. It is easy to see that D_n is of order $2n$, and that σ generates a cyclic subgroup of index 2 in D_n. In a similar way one defines symmetry groups for regular convex polyhedra, i.e., for tetrahedra, cubes, octahedra, dodecahedra, and icosahedra.

(4) For later use we want to introduce the *Klein four-group* $\mathfrak{V}_4$ as a subgroup of $\mathfrak{S}_4$:

$$\mathfrak{V}_4 = \big\{\mathrm{id},\ (1,2)(3,4),\ (1,3)(2,4),\ (1,4)(2,3)\big\}.$$

Then

$$\mathfrak{V}_4 \subset \mathfrak{A}_4 \subset \mathfrak{S}_4,$$

and it is easily checked, see Exercise 6, that $\mathfrak{V}_4$ is a normal subgroup of $\mathfrak{S}_4$. Also note that $\mathfrak{V}_4$ is isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$.

Exercises

1. *Transpositions in $\mathfrak{S}_n$ are permutations that exchange two elements of the set $\{1, \ldots, n\}$, leaving everything else fixed. Give a direct argument for $n \geq 2$, showing that every $\pi \in \mathfrak{S}_n$ is a product of transpositions.*
2. *Given a prime number p, specify a p-Sylow subgroup in $\mathfrak{S}_p$.*
3. For an r-cycle $\pi \in \mathfrak{S}_n$, show that $\operatorname{sgn}\pi = (-1)^{r-1}$.
4. Consider a permutation $\pi \in \mathfrak{S}_n$ and write $\langle\pi\rangle \subset \mathfrak{S}_n$ for the corresponding cyclic subgroup generated by π. Let m be the number of $\langle\pi\rangle$-orbits with respect to the natural action of $\langle\pi\rangle$ on $\{1, \ldots n\}$. Show that $\operatorname{sgn}\pi = (-1)^{n-m}$.
5. Write the following permutations as products of cycles and compute the sign:

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix} \in \mathfrak{S}_4, \qquad \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 3 & 1 & 4 & 5 & 2 & 6 & 8 & 7 \end{pmatrix} \in \mathfrak{S}_8.$$

6. Consider an r-cycle $\pi = (x_1, \ldots, x_r) \in \mathfrak{S}_n$ and show for arbitrary $\sigma \in \mathfrak{S}_n$ that

$$\sigma \circ \pi \circ \sigma^{-1} = \big(\sigma(x_1), \ldots, \sigma(x_r)\big).$$

   As an application prove that the Klein four-group is a normal subgroup of $\mathfrak{S}_4$.
7. Show for $n \geq 2$ that the cycles $(1,2)$ and $(1,2,\ldots,n)$ generate the group $\mathfrak{S}_n$.
8. Show for $n \geq 3$ that $\mathfrak{A}_n$ is generated by the cycles $(1,2,3)$, $(1,2,4), \ldots, (1,2,n)$. Conclude from this that a normal subgroup $N \subset \mathfrak{A}_n$ coincides with $\mathfrak{A}_n$ as soon as it contains a 3-cycle.

5.4 Solvable Groups

In order to characterize solvable groups we need the notion of commutator. For two elements a, b of a group G, one calls $[a,b] = aba^{-1}b^{-1}$ the *commutator* of a and b. In a similar way, we define the commutator $[H, H']$ of two subgroups $H, H' \subset G$ as the subgroup that is *generated* in G by all commutators $[a,b]$ for elements $a \in H$, $b \in H'$. In particular, taking $H = H' = G$, we obtain the *commutator group* $[G,G]$ of G. Let us point out that a group G is abelian if and only if $[G,G] = \{1\}$.

Remark 1. (i) $[G,G]$ *consists of all (finite) products of commutators in* G.

(ii) $[G,G]$ *is a normal subgroup in* G, *the smallest of all normal subgroups* $N \subset G$ *such that* G/N *is abelian.*

Proof. For $a, b \in G$ we have

$$[a,b]^{-1} = (aba^{-1}b^{-1})^{-1} = bab^{-1}a^{-1} = [b,a].$$

Therefore, the finite products of commutators form a subgroup in G, the commutator group $[G,G]$. Furthermore, for $a, b, g \in G$ we can write

$$\begin{aligned} g[a,b]g^{-1} &= gaba^{-1}b^{-1}g^{-1} = (gag^{-1})(gbg^{-1})(gag^{-1})^{-1}(gbg^{-1})^{-1} \\ &= [gag^{-1}, gbg^{-1}]. \end{aligned}$$

Hence, $[G,G]$ is a normal subgroup in G. Now write $\overline{x}$ for residue classes in $G/[G,G]$ of elements $x \in G$. Then the equation

$$\overline{a} \cdot \overline{b} \cdot \overline{a}^{\,-1} \cdot \overline{b}^{\,-1} = \overline{aba^{-1}b^{-1}} = 1$$

shows that $G/[G,G]$ is abelian. On the other hand, if $N \subset G$ is a normal subgroup such that G/N is abelian, then N must contain all commutators $[a,b]$ of elements $a, b \in G$. But this implies $[G,G] \subset N$, and it follows that $[G,G]$ is the smallest of all normal subgroups $N \subset G$ such that G/N is abelian. □

For later use, let us determine some special commutator groups.

Remark 2. *The permutation groups* $\mathfrak{S}_n$, $\mathfrak{A}_n$ *satisfy*

$$[\mathfrak{S}_n, \mathfrak{S}_n] = \mathfrak{A}_n \text{ for } n \geq 2,$$

$$[\mathfrak{A}_n, \mathfrak{A}_n] = \begin{cases} \{1\} & \text{for } n = 2, 3, \\ \mathfrak{V}_4 & \text{for } n = 4, \\ \mathfrak{A}_n & \text{for } n \geq 5. \end{cases} \qquad \text{(Klein four-group)}$$

Proof. We start with the computation of $[\mathfrak{S}_n, \mathfrak{S}_n]$. Since $\mathfrak{S}_n/\mathfrak{A}_n \simeq \mathbb{Z}/2\mathbb{Z}$ is abelian, we get $[\mathfrak{S}_n, \mathfrak{S}_n] \subset \mathfrak{A}_n$ by Remark 1 and in particular $[\mathfrak{S}_2, \mathfrak{S}_2] = \mathfrak{A}_2$,

since $\mathfrak{A}_2 = \{1\}$. To justify the inclusion $[\mathfrak{S}_n, \mathfrak{S}_n] \supset \mathfrak{A}_n$ for $n \geq 3$, we use the fact that every element of $\mathfrak{A}_n$ is a product of 3-cycles; cf. 5.3/3. However, every 3-cycle $(x_1, x_2, x_3) \in \mathfrak{S}_n$ is a commutator, since we can write

$$(x_1, x_2, x_3) = (x_1, x_3)(x_2, x_3)(x_1, x_3)^{-1}(x_2, x_3)^{-1}.$$

Thus, we get $[\mathfrak{S}_n, \mathfrak{S}_n] \supset \mathfrak{A}_n$ and hence $[\mathfrak{S}_n, \mathfrak{S}_n] = \mathfrak{A}_n$.

Next observe that the groups $\mathfrak{A}_2$ and $\mathfrak{A}_3$ are of order 1, resp. 3, and therefore are abelian. As a consequence, the commutator group $[\mathfrak{A}_n, \mathfrak{A}_n]$ is trivial for $n = 2, 3$. Furthermore, for $n \geq 5$, consider a 3-cycle (x_1, x_2, x_3) in $\mathfrak{S}_n$. Choosing $x_4, x_5 \in \{1, \ldots, n\}$ such that $x_1, \ldots, x_5$ are distinct, we get

$$(x_1, x_2, x_3) = (x_1, x_2, x_4)(x_1, x_3, x_5)(x_1, x_2, x_4)^{-1}(x_1, x_3, x_5)^{-1}.$$

Since $\mathfrak{A}_n$ consists of all finite products of 3-cycles by 5.3/3, it follows that every element of $\mathfrak{A}_n$ is a product of commutators in $\mathfrak{A}_n$. This implies $\mathfrak{A}_n \subset [\mathfrak{A}_n, \mathfrak{A}_n]$ and hence $[\mathfrak{A}_n, \mathfrak{A}_n] = \mathfrak{A}_n$.

It remains to justify the relation $[\mathfrak{A}_4, \mathfrak{A}_4] = \mathfrak{V}_4$. According to Remark 1 (ii), we know that $[\mathfrak{A}_4, \mathfrak{A}_4]$ equals the smallest normal subgroup $N \subset \mathfrak{A}_4$ such that the factor group $\mathfrak{A}_4/N$ is abelian. Since $\mathfrak{A}_4/\mathfrak{V}_4$ is of order 3 and hence abelian, we conclude that $[\mathfrak{A}_4, \mathfrak{A}_4] \subset \mathfrak{V}_4$. On the other hand, distinct elements $x_1, \ldots, x_4 \in \{1, \ldots, 4\}$ lead to the equation

$$(x_1, x_2)(x_3, x_4) = (x_1, x_2, x_3)(x_1, x_2, x_4)(x_1, x_2, x_3)^{-1}(x_1, x_2, x_4)^{-1},$$

showing that $\mathfrak{V}_4 = \{\mathrm{id},\ (1,2)(3,4),\ (1,3)(2,4),\ (1,4)(2,3)\}$ is contained in $[\mathfrak{A}_4, \mathfrak{A}_4]$. □

In the following we want to use commutators in order to characterize so-called solvable groups. To do this we introduce for a group G and an integer $i \in \mathbb{N}$ its ith *iterated commutator* D^iG inductively by

$$D^0G = G, \qquad D^{i+1}G = [D^iG, D^iG].$$

In this way, we obtain a chain

$$G = D^0G \supset D^1G \supset \ldots \supset D^iG \supset \ldots$$

of subgroups in G, where in each case, $D^{i+1}G$ is normal in D^iG. In addition, $D^iG/D^{i+1}G$ is abelian. More specifically, chains of the latter type are used to define solvable groups:

Definition 3. *Let G be a group. A descending chain of subgroups*

$$G = G_0 \supset G_1 \supset \ldots \supset G_n = \{1\}$$

is called a normal series *of G if in each case, G_{i+1} is a normal subgroup of G_i. The residue class groups G_i/G_{i+1}, $i = 0, \ldots, n-1$, are called the* factors *of the normal series.*

The group G is called solvable *if G admits a normal series with abelian factors.*

Proposition 4. *A group G is solvable if and only if there is an integer $n \in \mathbb{N}$ such that $D^n G = \{1\}$.*

Proof. First, assume that G is solvable and let

$$G = G_0 \supset G_1 \supset \ldots \supset G_n = \{1\}$$

be a normal series with abelian factors. We show by induction that $D^i G \subset G_i$ for $i = 0, \ldots, n$. For $i = 0$ the inclusion holds for trivial reasons. Now assume that $D^i G \subset G_i$ for some $i < n$. Since G_i / G_{i+1} is abelian, we get $[G_i, G_i] \subset G_{i+1}$ by Remark 1 (ii) and hence

$$D^{i+1} G = [D^i G, D^i G] \subset [G_i, G_i] \subset G_{i+1},$$

as desired. In particular, it follows that

$$D^n G \subset G_n = \{1\}.$$

Conversely, if $D^n G = \{1\}$, then

$$G = D^0 G \supset D^1 G \supset \ldots \supset D^n G = \{1\}$$

is a normal series with abelian factors. □

Let us consider some examples. Note that every commutative group is solvable for trivial reasons.

Remark 5. *The symmetric group $\mathfrak{S}_n$ is solvable for $n \leq 4$, but not solvable if $n \geq 5$.*

Proof. There are the following normal series of $\mathfrak{S}_n$ for $n \leq 4$:

$$\begin{gathered} \mathfrak{S}_2 \supset \{1\}, \\ \mathfrak{S}_3 \supset \mathfrak{A}_3 \supset \{1\}, \\ \mathfrak{S}_4 \supset \mathfrak{A}_4 \supset \mathfrak{V}_4 \supset \{1\}. \end{gathered}$$

That the factors of these normal series are abelian is easy to see. The groups $\mathfrak{S}_2$, $\mathfrak{S}_3 / \mathfrak{A}_3$, and $\mathfrak{S}_4 / \mathfrak{A}_4$ are cyclic of order 2, whereas the groups $\mathfrak{A}_3$ and $\mathfrak{A}_4 / \mathfrak{V}_4$ are cyclic of order 3, so that the commutativity is clear in these cases. Furthermore, the Klein four-group $\mathfrak{V}_4$ is commutative as well. Therefore, $\mathfrak{S}_n$ is solvable for $n \leq 4$. For $n \geq 5$ we have $[\mathfrak{S}_n, \mathfrak{S}_n] = \mathfrak{A}_n$ as well as $[\mathfrak{A}_n, \mathfrak{A}_n] = \mathfrak{A}_n$, see Remark 2, so that $\mathfrak{S}_n$ cannot be solvable in this case. □

Remark 6. *For a prime number p, every (finite) p-group, hence every group of order p^n for some $n \in \mathbb{N}$, is solvable.*

The assertion was already proved in 5.2/4. Next we want to derive a special characterization of finite solvable groups that will be of interest later when we study the solvability of algebraic equations.

Proposition 7. *Let G be a finite solvable group. Then every strictly decreasing normal series in G with abelian factors can be refined to a normal series whose factors are cyclic of prime order.*

Proof. Let $G = G_0 \supset \ldots \supset G_n$ be a strictly decreasing normal series with abelian factors. If one of the factors, say G_i/G_{i+1}, is not cyclic of prime order, choose a nontrivial element $\overline{a} \in G_i/G_{i+1}$, where replacing $\overline{a}$ by a suitable power of itself, we may assume that $\operatorname{ord} \overline{a}$ is prime. Since G_i/G_{i+1} is not cyclic, we see that $\langle \overline{a} \rangle$, the cyclic group generated by $\overline{a}$, is strictly contained in G_i/G_{i+1}. Therefore, its preimage with respect to the projection $G_i \longrightarrow G_i/G_{i+1}$ yields a subgroup $H \subset G_i$ satisfying

$$G_i \supsetneq H \supsetneq G_{i+1}.$$

Since $\langle \overline{a} \rangle$ is a normal subgroup of the (abelian) group G_i/G_{i+1}, it follows that its preimage H is a normal subgroup in G_i. Since G_{i+1} is normal in G_i and thus also in H, we may refine the normal series $G_0 \supset \ldots \supset G_n$ by inserting H between G_i and G_{i+1}, thereby obtaining a new normal series of G. It has abelian factors as well, due to the injection $H/G_{i+1} \hookrightarrow G_i/G_{i+1}$ and the epimorphism $G_i/G_{i+1} \longrightarrow G_i/H$, where G_i/G_{i+1} is abelian. Repeating the refining process if necessary and using the finiteness of G, we arrive after finitely many steps at a normal series whose factors are cyclic of prime order. □

Proposition 8. *Let G be a group and $H \subset G$ a subgroup. If G is solvable, the same is true for H as well. If H is normal in G, then G is solvable if and only if H and G/H are solvable.*

Proof. First, assume that G is solvable. Then, since $D^i H \subset D^i G$, we see that H is solvable as well. Furthermore, if H is normal in G, we can consider the canonical epimorphism $\pi\colon G \longrightarrow G/H$. Since $D^i(\pi(G)) = \pi(D^i(G))$, as is easily verified, we see that $G/H = \pi(G)$ is solvable if G is solvable.

Conversely, assume that H and G/H are solvable, say $D^n H = \{1\}$ and $D^n(G/H) = \{1\}$. Then we get

$$\pi(D^n G) = D^n(G/H) = \{1\},$$

i.e., $D^n G \subset H$, and furthermore, $D^{2n} G \subset D^n H = \{1\}$. Therefore G is solvable. □

Corollary 9. *The Cartesian product $\prod_{i=1}^n G_i$ of finitely many groups $G_1, \ldots, G_n$ is solvable if and only if all groups G_i are solvable.*

Proof. Conclude by induction, and for $n = 2$, apply Proposition 8 to the projection $G_1 \times G_2 \longrightarrow G_2$, which admits G_1 as its kernel. □

Exercises

1. *We have seen in Remark* 1 *for a group* G *that its commutator group* $[G, G]$ *equals the smallest of all normal subgroups* $N \subset G$ *such that the factor group* G/N *is abelian. Prove a corresponding assertion for commutators of type* $[G, H]$, *where* H *is a normal subgroup, or alternatively, any subgroup in* G.

2. Let p, q be distinct prime numbers. Show that every group of order pq is solvable.

3. For a finite group G show:
   (i) If H, H' are normal subgroups in G that are solvable, then $H \cdot H'$ is solvable as well.
   (ii) There exists a unique largest subgroup in G that is solvable. This subgroup is invariant under all automorphisms of G.

4. Show that every group of order < 60 is solvable.

5. Show that the alternating group $\mathfrak{A}_5$ does not admit a nontrivial normal subgroup.

6. Let T be the subgroup of upper triangular matrices in $\mathrm{GL}(n, K)$, the group of invertible $(n \times n)$ matrices over a field K. Show that T is solvable.

7. For a group G consider the subgroups $C^i(G)$ that are inductively given by $C^1(G) = G$ and $C^{i+1}(G) = [G, C^i(G)]$. The group G is called *nilpotent* if there is some $n \in \mathbb{N}$ such that $C^n(G) = \{1\}$. Show that every nilpotent group is solvable.

8. In the situation of Exercise 6, consider for $K \neq \mathbb{F}_2$ the group of all upper triangular matrices $T \subset \mathrm{GL}(n, K)$, as well as its subgroup $T_1 \subset T$ consisting of all triangular matrices with diagonal elements equal to 1. Show that T_1 is nilpotent (cf. Exercise 7), but that T is not. Therefore, T is an example of a solvable group that is not nilpotent.

6. Applications of Galois Theory

Background and Overview

In the previous chapters we have developed group theory, field theory, and in particular, Galois theory to a sufficiently high level, allowing us to consider applications to some prominent classical problems. We start in 6.1 with the problem of solving algebraic equations by radicals, which is the problem that motivated E. Galois to work out his "Galois" theory. In particular, we show for a monic separable polynomial f with coefficients in a field K that the algebraic equation $f(x) = 0$ is solvable by radicals if and only if the corresponding Galois group is solvable in the group-theoretic sense.

The basic idea of proof is easy to explain. On the side of fields, the problem is reduced to a setting in which K contains sufficiently many roots of unity. Then one repeatedly extends K by adjoining suitable *radicals*, i.e., zeros of polynomials of type $X^n - c \in K[X]$ for $\operatorname{char} K \nmid n$, and if $p = \operatorname{char} K > 0$, also of type $X^p - X - c \in K[X]$. These extensions represent essentially all cyclic Galois extensions of K; see 4.8/3 and 4.8/5. In a similar way, we use on the level of Galois groups the fact that the finite cyclic groups are, so to speak, the "building blocks" of the finite solvable groups; cf. 5.4/7. Furthermore, let us point out that it is common practice to view the zeros of polynomials of type $X^p - X - c \in K[X]$ for $p = \operatorname{char} K > 0$ as "radicals" as well. Only in this way does the characterization of solvable (separable) algebraic equations in terms of solvable Galois groups remain valid for fields of characteristic > 0. Of course, polynomials of type $X^p - c$ for $p = \operatorname{char} K > 0$ are not separable, which means that their zeros cannot be studied using methods of Galois theory. Finally, in 6.1/10 we give a necessary condition for an irreducible algebraic equation of prime degree to be solvable. The latter criterion, which goes back to E. Galois as well, can be used to easily set up examples of algebraic equations that are not solvable. To further illustrate the solvability problem we work out in Section 6.2 the explicit formulas for the solutions of algebraic equations of degrees 3 and 4.

As a second application, we present in 6.3 a Galois-theoretic proof of the fundamental theorem of algebra. From an algebraic point of view, working on this theorem can be quite tricky, as is clearly visible from first proofs. The difficulties are caused by the fact that the field $\mathbb{C}$ of complex numbers relies on the field of real numbers $\mathbb{R}$, by adjoining a square root of -1. Constructing the field $\mathbb{R}$, however, requires methods from analysis. Therefore, given a polynomial

$f \in \mathbb{C}[X]$, chances are very low that one can realize its zeros in an algebraic way as elements of $\mathbb{C}$. To avoid such difficulties, we proceed indirectly. If $\mathbb{C}$ were not algebraically closed, one could use Kronecker's construction to arrive at a nontrivial extension $L/\mathbb{C}$, which can be assumed to be Galois. Then we use a first (analytic) fact, namely that real polynomials of odd degree admit at least one real zero, and show in terms of Galois theory that we may assume $L/\mathbb{C}$ to be of degree 2. However, such an extension cannot exist. This becomes clear if we use as a second (analytic) fact that positive real numbers admit a square root in $\mathbb{R}$, and hence all complex numbers admit a square root in $\mathbb{C}$. In particular, our proof depends on the mentioned "analytic" properties of the real numbers.

In 6.4 we discuss another application, compass and straightedge constructions in the complex plane. A thorough analysis of the construction steps that are possible in such a setting shows that starting with the points $0, 1 \in \mathbb{C}$, one can construct only points $z \in \mathbb{C}$ that are contained in Galois extensions $L/\mathbb{Q}$ whose degree $[L : \mathbb{Q}]$ is a power of 2. In particular, z is then algebraic over $\mathbb{Q}$, of a degree that is a power of 2. In this way, the constructibility of the cube root $\sqrt[3]{2}$ is excluded, and it follows, for example, that the ancient problem of doubling the cube is not accessible in terms of compass and straightedge constructions. Another topic we elaborate on is the study of C. F. Gauss on the constructibility of regular convex polygons.

6.1 Solvability of Algebraic Equations

Even if the solution formulas for algebraic equations in degrees 1 and 2 look quite simple, the more complicated formulas in degrees 3 and 4, which we will derive in Section 6.2, make it undoubtedly clear that the problem of solving more general algebraic equations is not so easy. In fact, we will see starting from degree 5 on that as a matter of principle, universal solution formulas for algebraic equations cannot exist. In order to look at the corresponding background more closely, let us make the notion of solvability of algebraic equations precise.

Definition 1. *A finite field extension L/K is said to be* solvable by radicals *if L admits an extension field E together with a chain of field extensions*

$$K = E_0 \subset E_1 \subset \ldots \subset E_m = E$$

such that in each case, E_{i+1} is obtained from E_i by adjoining an element of the following type:

(1) *a root of unity, or*
(2) *a zero of a polynomial of type $X^n - a \in E_i[X]$, where $\operatorname{char} K \nmid n$, or*
(3) *a zero of a polynomial of type $X^p - X - a \in E_i[X]$ for $p = \operatorname{char} K > 0$.*

Then L/K is necessarily separable.

The main goal of the present section is to characterize solvability by radicals in terms of the solvability of Galois groups in the group-theoretic sense.

Definition 2. *A finite field extension L/K is called* solvable *if there exists an extension field $E \supset L$ such that E/K is a finite Galois extension with solvable Galois group* $\mathrm{Gal}(E/K)$ *(in the sense of* 5.4/3*).*

Using this definition, observe that a Galois extension L/K is solvable if and only if the Galois group $\mathrm{Gal}(L/K)$ is solvable. Indeed, if we can enlarge L/K to a finite Galois extension E/K with solvable Galois group, then $\mathrm{Gal}(L/K)$ is a quotient of $\mathrm{Gal}(E/K)$ by 4.1/2 and thus solvable by 5.4/8.

The two notions of solvability extend naturally to the context of algebraic equations. If f is a nonconstant (separable) polynomial with coefficients in a field K, we can choose a splitting field L of f over K. Then we say that the algebraic equation $f(x) = 0$ is *solvable over* K, resp. *solvable by radicals*, if the extension L/K admits the corresponding property.

We want to prove some more or less elementary properties of the two solvability notions.

Lemma 3. *Let L/K be a finite field extension and F an arbitrary extension field of K. Embed L via a K-homomorphism into an algebraic closure $\overline{F}$ of F, see* 3.4/9*, and look at the composite field FL in $\overline{F}$. Then, if L/K is solvable (resp. Galois with solvable Galois group, resp. solvable by radicals, resp. exhaustible by a chain of field extensions as in Definition* 1*), then the same is true for the extension FL/F as well.*

Lemma 4. *Given a chain of finite field extensions $K \subset L \subset M$, the extension M/K is solvable (resp. solvable by radicals) if and only if M/L and L/K are solvable (resp. solvable by radicals).*

Proof of Lemma 3. First assume that L/K is solvable. Enlarging L, we may restrict ourselves to the case that L/K is Galois with solvable Galois group $\mathrm{Gal}(L/K)$. Then $FL = F(L)$ is a finite Galois extension of F. Since every $\sigma \in \mathrm{Gal}(FL/F)$ leaves the field K fixed, it follows that $\sigma(L)$ is algebraic over K. In particular, we conclude from 3.5/4 that there is a restriction homomorphism

$$\mathrm{Gal}\big(FL/F\big) \longrightarrow \mathrm{Gal}\big(L/K\big).$$

This homomorphism is injective, since $FL = F(L)$. Therefore, the solvability of $\mathrm{Gal}(FL/F)$ and hence of FL/F follows from 5.4/8. On the other hand, if L/K is solvable by radicals, resp. exhaustible by a chain of field extensions as in Definition 1, then the same is true by trivial reasons for the extension FL/F. □

Proof of Lemma 4. Again we start by considering the property "solvable." Assume first that M/K is solvable. Enlarging M, we may assume M/K to be

Galois with solvable Galois group. Then, by definition, L/K is solvable, too. Furthermore, since $\mathrm{Gal}(M/L)$ can be viewed as a subgroup of $\mathrm{Gal}(M/K)$, we conclude from 5.4/8 that M/L is solvable.

Now assume that the extensions M/L and L/K of the chain $K \subset L \subset M$ are solvable. In a first step we show that both extensions can be assumed to be Galois with solvable Galois group. To achieve this, choose a finite extension L' of L such that L'/K is Galois with solvable Galois group. Then we can use Lemma 3 and thereby replace L by L' as well as M by the composite field $L'M$ (in an algebraic closure of M). Furthermore, there exists a finite extension M' of $L'M$ such that M'/L' is Galois with solvable Galois group. Now, replacing $L'M$ by M', we may assume in the following that both M/L and L/K are Galois with solvable Galois group.

Since M is separable, but not necessarily Galois over K, we pass to a normal closure M' of M/K, see 3.5/7, where now M'/K is a finite Galois extension. To construct M' we consider all K-homomorphisms $\sigma\colon M \longrightarrow \overline{M}$ into an algebraic closure $\overline{M}$ of M and define M' as the composite field of all fields $\sigma(M)$. Since L/K is Galois, we get $\sigma(L) = L$ for all σ, and it follows that every extension $\sigma(M)/L$ is a Galois extension that is isomorphic to M/L. Now we claim that the Galois group $\mathrm{Gal}(M'/K)$ and hence the extension M/K are solvable. To justify this, look at the surjective restriction homomorphism

$$\mathrm{Gal}\big(M'/K\big) \longrightarrow \mathrm{Gal}\big(L/K\big)$$

admitting $\mathrm{Gal}(M'/L)$ as its kernel; cf. 4.1/2 (ii). Since $\mathrm{Gal}(L/K)$ is solvable, it is enough by 5.4/8 to show that $\mathrm{Gal}(M'/L)$ is solvable. However, using 4.1/12 (ii), the latter group can be viewed as a subgroup of the Cartesian product

$$\prod_{\sigma \in \mathrm{Hom}_K(M,\overline{M})} \mathrm{Gal}\big(\sigma(M)/L\big).$$

All groups $\mathrm{Gal}(\sigma(M)/L) = \mathrm{Gal}(\sigma(M)/\sigma(L))$ are canonically isomorphic to $\mathrm{Gal}(M/L)$ and hence are solvable. By 5.4/9, the Cartesian product of these groups is solvable as well, and it follows from 5.4/8 that $\mathrm{Gal}(M'/L)$ is solvable. This finishes the proof of Lemma 4 for the property "solvable."

It remains to look at the property "solvable by radicals." If M/K is solvable by radicals, the same clearly holds for the extensions M/L and L/K as well. Conversely, if M/L and L/K are solvable by radicals, choose an extension L'/L such that L'/K can be exhausted by a chain of field extensions as specified in Definition 1. Then consider the composite field $L'M$ in an algebraic closure of M and use the fact that $L'M/L'$ is solvable by radicals due to Lemma 3. It follows that $L'M/K$ is solvable by radicals and that the same is true for M/K. □

Theorem 5. *A finite field extension L/K is solvable if and only if it is solvable by radicals.*

Proof. Assume first that L/K is solvable. Enlarging L, we may assume that L/K is Galois with solvable Galois group. Let m be the product of all prime

numbers $q \neq \operatorname{char} K$ dividing the degree $[L : K]$. Furthermore, let F be an extension field of K, obtained by adjoining a primitive mth root of unity. Then, by definition, the extension F/K is solvable by radicals. Now look at the chain of fields

$$K \subset F \subset FL,$$

where the composite field FL is constructed in an algebraic closure of K. It is enough to show (see Lemma 4) that FL/F is solvable by radicals. To do this, we know from Lemma 3 that FL/F is solvable, even a Galois extension with solvable Galois group, since the corresponding property was assumed for the extension L/K. Now choose a normal series

$$\operatorname{Gal}\big(FL/F\big) = G_0 \supset G_1 \supset \ldots \supset G_n = \{1\}$$

with factors that are cyclic of prime order; cf. 5.4/7. By the fundamental theorem of Galois theory 4.1/6, this corresponds to a chain of field extensions

$$F = F_0 \subset F_1 \subset \ldots \subset F_n = FL$$

such that in each case, F_{i+1}/F_i is a cyclic Galois extension of prime order, say of some order p_i. Observing that $[FL : F]$ divides $[L : K]$, for example, using 4.1/12 (i), we see for $p_i \neq \operatorname{char} K$ that the prime p_i will divide m. In particular, F and hence F_i contain a primitive p_ith root of unity. Therefore, we can apply 4.8/3 to conclude that F_{i+1} is obtained from F_i by adjoining a zero of a polynomial of type $X^{p_i} - a \in F_i[X]$. On the other hand, if $p_i = \operatorname{char} K$, we see from 4.8/5 that F_{i+1} is constructed from F_i by adjoining a zero of a polynomial of type $X^{p_i} - X - a \in F_i[X]$. In either case, it follows that FL/F and hence L/K are solvable by radicals.

To verify the reverse implication, assume that L/K is solvable by radicals. Then there exists a chain of field extensions $K = K_0 \subset K_1 \subset \ldots \subset K_n$ such that $L \subset K_n$, and in each case, the extension K_{i+1}/K_i is of type (1), (2), or (3) in the sense of Definition 1. Enlarging L, we may assume $L = K_n$. To prove that L/K is solvable it is enough by Lemma 4 to show that all extensions K_{i+1}/K_i are solvable. In other words, we may assume that the extension L/K is of type (1), (2), or (3) in Definition 1. Now observe that the extensions of type (1) and (3) are Galois, in fact abelian by 4.5/7 for type (1) and cyclic by 4.8/5 for type (3). In particular, in both cases L/K is solvable. In the remaining case, in which L/K is of type (2), the field L is obtained from K by adjoining a zero of a polynomial $X^n - c \in K[X]$, for some exponent n not divisible by $\operatorname{char} K$. To handle this case, consider an extension F/K that is generated by a primitive nth root of unity. Then we can consider the chain $K \subset F \subset FL$, where the composite field FL is constructed in an algebraic closure of L. We know from 4.5/7 that F/K is an abelian Galois extension, whereas FL/F is a cyclic Galois extension by 4.8/3. In particular, both extensions are solvable, and we see from Lemma 4 that FL/K and hence L/K are solvable as well. Thus, we are done. □

Corollary 6. *Let L/K be a separable field extension of degree ≤ 4. Then L/K is solvable, and in particular, solvable by radicals.*

Proof. By the primitive element theorem 3.6/12, the extension L/K is simple, say $L = K(a)$. Let $f \in K[X]$ be the minimal polynomial of a over K and let L' be a splitting field of f over K. Then $\deg f = [L : K] \leq 4$, and the Galois group $\mathrm{Gal}(L'/K)$ can be viewed as a subgroup of $\mathfrak{S}_4$, due to 4.3/1. Since $\mathfrak{S}_4$, and therefore all its subgroups are solvable (see 5.4/5 and 5.4/8), we conclude that L'/K and L/K are solvable. □

Corollary 7. *There exist finite separable field extensions that are not solvable by radicals. For example, the generic equation of degree n is not solvable by radicals for $n \geq 5$.*

Proof. It is enough to know that the generic equation of degree n admits the full permutation group $\mathfrak{S}_n$ as its Galois group for $n \geq 2$; cf. Section 4.3, Example (4). Since $\mathfrak{S}_n$ is not solvable for $n \geq 5$ by 5.4/5, we see from Theorem 5 that the corresponding extension L/K cannot be solvable by radicals in this case. □

Let us review once more Example (4) of Section 4.3. Starting out from a field k, we have considered the rational function field $L = k(T_1, \ldots, T_n)$ in a number of variables $T_1, \ldots, T_n$. The permutation group $\mathfrak{S}_n$ acts on L by permuting the T_i, and L was recognized as a Galois extension of the corresponding fixed field K, with Galois group $\mathrm{Gal}(L/K) = \mathfrak{S}_n$. The fixed field K itself could be identified with $K = k(s_1, \ldots, s_n)$, where $s_1, \ldots, s_n$ are the elementary symmetric polynomials in $T_1, \ldots, T_n$. In fact, we have seen that L is a splitting field of the polynomial $f = X^n - s_1 X^{n-1} + \ldots + (-1)^n s_n \in K[X]$. Since the elements $s_1, \ldots, s_n \in K$ are algebraically independent over k by the fundamental theorem on symmetric polynomials 4.3/5, or 4.4/1, the coefficients $-s_1, \ldots, (-1)^n s_n$ can just as well be viewed as variables over k. In this way, we can state for $n \geq 5$ and variables $c_1, \ldots, c_n$ over k that $x^n + c_1 x^{n-1} + \ldots + c_n = 0$, the *generic equation of degree n* over the rational function field $K = k(c_1, \ldots, c_n)$, is not solvable by radicals.

More specifically, we can ask whether there exist algebraic equations that are not solvable by radicals, even over the field $\mathbb{Q}$ of rational numbers. This question will be studied in the following, restricting ourselves to equations of prime degree, however. We start with two auxiliary results on permutations that subsequently will be applied to Galois groups.

Lemma 8. *For a prime number p, let $G \subset \mathfrak{S}_p$ be a subgroup acting transitively on $\{1, \ldots, p\}$. Then G contains a subgroup H of order p. If G is solvable, H is uniquely determined and in particular normal in G.*

Proof. Since G acts transitively on $\{1, \ldots, p\}$, there exists only a single G-orbit. It consists of p elements, and we see, for example using 5.1/6, that p divides

$\operatorname{ord} G$. Since p^2 does not divide the order of $\mathfrak{S}_p$, which is $p!$, we can exclude that p^2 divides $\operatorname{ord} G$. Therefore, G contains a subgroup H of order p, namely a p-Sylow subgroup; see 5.2/6.

Assuming now that G is solvable, it is seen from 5.4/7 that G admits a normal series $G = G_0 \supsetneq \ldots \supsetneq G_n = \{1\}$ whose factors are cyclic of prime order. We want to show by induction that each G_i acts transitively on $\{1, \ldots, p\}$ for $i < n$. This being clear for $i = 0$, let $i > 0$. Since G_i is normal in G_{i-1}, we can look at the relation $g(G_i x) = G_i(gx)$ for elements $g \in G_{i-1}$ and $x \in \{1, \ldots, p\}$. This shows that G_{i-1} acts on the G_i-orbits in $\{1, \ldots, p\}$. Hence, since G_{i-1} acts transitively on $\{1, \ldots, p\}$ by the induction hypothesis, all G_i-orbits in $\{1, \ldots, p\}$ are of the same order. Thus, if $B_1, \ldots, B_r$ are the orbits of the action of G_i on $\{1, \ldots, p\}$, we get $p = \sum_{\rho=1}^{r} \operatorname{ord} B_\rho = r \cdot \operatorname{ord} B_1$, which implies $r = 1$ or $\operatorname{ord} B_1 = 1$. However, since $G_i \neq \{1\}$ for $i < n$ and therefore $\operatorname{ord} B_\rho > 1$, we must have $r = 1$. As a consequence, there is only a single orbit of the action of G_i on $\{1, \ldots, p\}$, and hence G_i acts transitively. In particular, G_i contains a subgroup of order p for $i < n$, as shown before. For $i = n - 1$ this means that G_{n-1} is of order p, since $G_{n-1} \simeq G_{n-1}/G_n$ is of prime order.

A repeated application of the theorem of Lagrange 1.2/3 yields the relation $\operatorname{ord} G = \prod_{i=0}^{n-1} \operatorname{ord} G_i/G_{i+1}$. Since p divides $\operatorname{ord} G$, but p^2 does not, we must have $p \neq \operatorname{ord} G_i/G_{i+1}$ for $i = 0, \ldots, n-2$. Now, departing from $H \subset G_0$, we get $H \subset G_i$ for $i = 0, \ldots, n-1$ by an inductive argument. Indeed, if $H \subset G_i$ for some $i \leq n-2$, then the canonical map

$$H \hookrightarrow G_i \longrightarrow G_i/G_{i+1}$$

is trivial, since $p \nmid \operatorname{ord} G_i/G_{i+1}$, thus implying $H \subset G_{i+1}$. In particular, we get $H \subset G_{n-1}$ and hence $H = G_{n-1}$, since G_{n-1} is of order p. This shows that H is unique. But then H is invariant under conjugation by elements of G and therefore a normal subgroup in G. □

Lemma 9. *In the setting of Lemma* 8, *let G be a solvable group and consider an element $\sigma \in G$. If σ, as a bijective self-map on $\{1, \ldots, p\}$, admits two different fixed points, then $\sigma = \mathrm{id}$.*

Proof. Following Lemma 8, there is a normal subgroup H of order p in G, and H is necessarily cyclic of order p, say generated by some element $\pi \in G \subset \mathfrak{S}_p$. Factoring π into a product of disjoint cycles, see 5.3/1 (ii), and using $\operatorname{ord} \pi = p$, it follows that π is already a p-cycle itself. We write $\pi = (0, \ldots, p-1)$ where for convenience, as we will see below, we view $\mathfrak{S}_p$ as the group of permutations of the elements $0, \ldots, p-1$. Now consider a permutation $\sigma \in G$ admitting two different fixed points. By a renumbering process we may assume that one of them is the element 0. Therefore, let $0, i$ where $0 < i < p$, be two fixed points of σ. Since H is normal in G, the element

$$\sigma \circ \pi \circ \sigma^{-1} = \big(\sigma(0), \ldots, \sigma(p-1)\big)$$

belongs to H as well and hence equals a power π^r for some $0 \leq r < p$, say

$$\big(\sigma(0),\dots,\sigma(p-1)\big) = \big(0,\overline{r\cdot 1},\dots,\overline{r\cdot(p-1)}\,\big),$$

where $\overline{r\cdot j}$ indicates the remainder in $\{0,\dots,p-1\}$ on applying Euclidean division by p to $r\cdot j$. Now $\sigma(0)=0$ and $\sigma(i)=i$ show that $\overline{r\cdot i}=i$, and hence $\overline{r}\cdot\overline{i}=\overline{r\cdot i}=\overline{i}$ in $\mathbb{Z}/p\mathbb{Z}$. However, this implies $\overline{r}=1$ and therefore $r=1$, since $\overline{i}$ is a unit in $\mathbb{Z}/p\mathbb{Z}$, due to $0<i<p$. Consequently, we have $\sigma=\mathrm{id}$. □

Now let us apply the assertion of Lemma 9 to Galois groups and thereby derive some consequences for Galois extensions.

Proposition 10. *Let K be a field and $f\in K[X]$ an irreducible separable polynomial of prime degree p with splitting field L over K. Assume that the corresponding Galois group* $\mathrm{Gal}(L/K)$ *is solvable. Then $L=K(\alpha,\beta)$ for any two different zeros α,β of f.*

Proof. Every element $\sigma\in G=\mathrm{Gal}(L/K)$ induces a permutation of the zeros $\alpha_1,\dots,\alpha_p$ of f, and we may view G as a subgroup of the permutation group $\mathfrak{S}_p$; see 4.3/1. Given two zeros $\alpha,\beta\in L$ of f, the irreducibility of f implies that there is an element $\sigma\in G$ such that $\sigma(\alpha)=\beta$. Therefore, the action of G on $\{\alpha_1,\dots,\alpha_p\}$ is transitive. In addition, G is solvable by assumption and hence satisfies the assumptions of Lemma 9. Thus, if $\alpha\neq\beta$ and $\sigma\in G$ is an automorphism of L leaving $K(\alpha,\beta)$ fixed, then σ, as a permutation of $\alpha_1,\dots,\alpha_p$, admits two different fixed points, namely α and β, and therefore must equal the identity. In particular, we get $\mathrm{Gal}(L/K(\alpha,\beta))=\{1\}$ and hence $L=K(\alpha,\beta)$ by the fundamental theorem of Galois theory 4.1/6. □

The assertion of Proposition 10 allows the construction of a multitude of unsolvable finite field extensions of $\mathbb{Q}$. Indeed, if $f\in\mathbb{Q}[X]$ is an irreducible polynomial of prime degree $p\geq 5$ admitting at least two real zeros and one nonreal zero in $\mathbb{C}$, then the equation $f(x)=0$ cannot be solvable. Otherwise, we could conclude from Proposition 10 that the splitting field of f in $\mathbb{C}$ would be real, in contradiction to the fact that f admits nonreal zeros. As an example, consider for prime numbers $p\geq 5$ the polynomial $f=X^p-4X+2\in\mathbb{Q}[X]$, which is irreducible by Eisenstein's criterion 2.8/1. By curve sketching one realizes that f admits precisely three real zeros. Therefore, the corresponding Galois group cannot be solvable. Alternatively, for $p=5$, we can show that the Galois group G of $f=X^5-4X+2$ is isomorphic to $\mathfrak{S}_5$. Indeed, if we view G as a subgroup of $\mathfrak{S}_5$, see 4.3/1, then G contains an element of order 5, for example by Lemma 8, and hence a 5-cycle. Furthermore, complex conjugation permutes the two nonreal zeros of f, leaving the remaining three real zeros fixed. Therefore, G contains a transposition as well. But this implies $G=\mathfrak{S}_5$; cf. Exercise 7 in Section 5.3. Using such an argument, one can show more generally for every prime number p that there is an irreducible polynomial $f\in\mathbb{Q}[X]$ of degree p whose corresponding Galois group is isomorphic to $\mathfrak{S}_p$; cf. Exercise 5.

Exercises

1. *Let K be a field and $f \in K[X]$ a nonconstant separable polynomial. Let K_0 be the smallest subfield of K containing all coefficients of f. Discuss the relationship between the solvability of the equation $f(x) = 0$ over K and over K_0.*

2. *Let K be a field and $f \in K[X]$ a separable nonconstant polynomial. Using older terminology, an algebraic equation $f(x) = 0$ is called* metacyclic *if it can be reduced to a chain of cyclic equations. This means the following: If L is a splitting field of f over K, then there is a chain of fields $K = K_0 \subset K_1 \subset \ldots \subset K_n$, where $L \subset K_n$ and in each case, K_{i+1}/K_i is a Galois extension given by a cyclic equation, hence with cyclic Galois group. Show that the equation $f(x) = 0$ is metacyclic if and only if it is solvable (resp. solvable by radicals).*

3. Determine the Galois group of the polynomial

$$X^7 - 8X^5 - 4X^4 + 2X^3 - 4X^2 + 2 \in \mathbb{Q}[X]$$

and check whether it is solvable.

4. Verify whether the equation

$$X^7 + 4X^5 - \tfrac{10}{11}X^3 - 4X + \tfrac{2}{11} = 0$$

with coefficients in $\mathbb{Q}$ is solvable by radicals.

5. Show for every prime number $p \geq 5$ that there exists an irreducible polynomial $f_p \in \mathbb{Q}[X]$ of degree p whose corresponding Galois group (over $\mathbb{Q}$) is isomorphic to $\mathfrak{S}_p$. *Hint*: Consider a separable polynomial $h_p \in \mathbb{Q}[X]$ of degree p admitting exactly two nonreal zeros. Then approximate h_p by a suitable irreducible polynomial f_p. In doing so, use the principle of continuity of roots, i.e., that the zeros of h_p change in a continuous way when one makes continuous changes to the coefficients of h_p.

6. For a prime number p and the field $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ consider the group $S(\mathbb{F}_p)$ of bijective self-maps $\mathbb{F}_p \longrightarrow \mathbb{F}_p$. An element $\sigma \in S(\mathbb{F}_p)$ is called *linear* if there are elements $a, b \in \mathbb{F}_p$ such that $\sigma(x) = ax + b$ for all $x \in \mathbb{F}_p$, where necessarily $a \neq 0$. A subgroup $G \subset S(\mathbb{F}_p)$ is called *linear* if all elements $\sigma \in G$ are linear. Finally, a subgroup $G \subset \mathfrak{S}_p$ is called *linear* if there exists a bijection $\{1, \ldots, p\} \longrightarrow \mathbb{F}_p$ transferring G to a linear subgroup of $S(\mathbb{F}_p)$. Show:

   (i) If $\sigma \in S(\mathbb{F}_p)$ is linear and admits at least two different fixed points, then $\sigma = \mathrm{id}$.

   (ii) Every subgroup $G \subset \mathfrak{S}_p$ that is solvable and acts transitively on $\{1, \ldots, p\}$ is linear.

   (iii) Every linear subgroup $G \subset \mathfrak{S}_p$ is solvable.

   (iv) The Galois group of an irreducible polynomial of degree p is linear if it is solvable.

6.2 Algebraic Equations of Degree 3 and 4*

Let K be a field, $f \in K[X]$ a separable monic polynomial, and L a splitting field of f over K. As we have seen, the algebraic equation $f(x) = 0$ is solvable by radicals if and only if the corresponding Galois group $\mathrm{Gal}(L/K)$ is solvable in the group-theoretic sense. This is equivalent to the existence of a normal series

$$\mathrm{Gal}\big(L/K\big) = G_0 \supset G_1 \supset \ldots \supset G_r = \{1\}$$

with (finite) cyclic factors; cf. 5.4/7. On the other hand, starting with such a normal series, the fundamental theorem of Galois theory 4.1/6 shows that the series corresponds to a chain of field extensions

$$K = E_0 \subset E_1 \subset \ldots \subset E_r = L,$$

where E_i/E_{i-1} for $i = 1, \ldots, r$ is a cyclic Galois extension with Galois group G_{i-1}/G_i. In such a setting, the key ingredient for solving the equation $f(x) = 0$ by radicals is given by the characterization 4.8/3 (i) of cyclic extensions: Under the assumption that E_{i-1} contains a root of unity of order $n_i = [E_i : E_{i-1}]$, where $\mathrm{char}\, K$ does not divide the degree n_i, it follows that E_i is obtained from E_{i-1} by adjoining an n_ith root of some element $c_i \in E_i$. However, let us point out that the existence of c_i was obtained only in a nonconstructive way, using Hilbert's Theorem 90.

To arrive at a solution formula for the equation $f(x) = 0$ when dealing with a specific polynomial f, we can proceed as just explained, trying to describe the relevant field extensions in explicit terms. Since we are interested only in polynomials f of degrees $2, 3$, or 4, we can view the Galois group $\mathrm{Gal}(L/K)$ as a subgroup of $\mathfrak{S}_2$, $\mathfrak{S}_3$, resp. $\mathfrak{S}_4$. For these permutation groups the normal series

$$\begin{gathered}\mathfrak{S}_2 \supset \mathfrak{A}_2 = \{1\},\\ \mathfrak{S}_3 \supset \mathfrak{A}_3 \supset \{1\},\\ \mathfrak{S}_4 \supset \mathfrak{A}_4 \supset \mathfrak{V}_4 \supset \mathfrak{Z} \supset \{1\}\end{gathered}$$

are at hand, where the notation is as in Section 5.3. Recall that $\mathfrak{A}_n$ is the alternating group of index n and $\mathfrak{V}_4$ the Klein four-group. Furthermore, $\mathfrak{Z}$ is a cyclic subgroup of order 2 in $\mathfrak{V}_4$.

Now write $x_1, \ldots, x_n \in L$ for the zeros of f and view the Galois group $\mathrm{Gal}(L/K)$ as a subgroup of $\mathfrak{S}_n$. Assume for a moment $\mathrm{Gal}(L/K) = \mathfrak{S}_n$ and consider $\mathfrak{A}_n$ as a subgroup of $\mathrm{Gal}(L/K)$. Then the corresponding intermediate field E_1 of L/K can be described quite easily. One looks at the element $\Delta = \delta^2$ for

$$\delta = \prod_{i<j} (x_i - x_j),$$

where Δ is the *discriminant* of the polynomial f; cf. Section 4.4. Note that $\Delta \neq 0$, since f was assumed to be separable. Furthermore, Δ is invariant under all permutations $\pi \in \mathfrak{S}_n$ and therefore belongs to K. In Section 4.4, see in

particular 4.4/10, we have shown in more detail how to compute Δ from the coefficients of f. In addition, if $\operatorname{char} K \neq 2$, the square root δ of Δ is invariant under a permutation $\pi \in \mathfrak{S}_n$ if and only if π is even and thus belongs to $\mathfrak{A}_n$. This shows that $K(\sqrt{\Delta}) \subset L^{\mathfrak{A}_n} = E_1$ and, due to $\sqrt{\Delta} \notin K$, even $K(\sqrt{\Delta}) = E_1$. Consequently, if $\operatorname{char} K \neq 2$, the step $\mathfrak{S}_n \supset \mathfrak{A}_n$ corresponds on the level of attached fields to the adjunction of a square root of the discriminant Δ.

Proceeding as indicated in the general discussion above, we will derive now the special solution formulas for algebraic equations $f(x) = 0$ of degree ≤ 4. Since it is not necessary, we will make no assumption on f to be irreducible or separable. However, assumptions on the characteristic of K will always ensure that the splitting field L of f is separable and therefore Galois over K. We start with a *quadratic polynomial* $f \in K[X]$, say

$$f = X^2 + aX + b,$$

where we assume $\operatorname{char} K \neq 2$. We could arrive at the solutions quite easily by applying the method of completing the square, as usual. However, for illustration, let us show how to argue in terms of the discriminant. Therefore, let L be a splitting field of f over K, and let x_1, x_2 be the zeros of f in L. The discriminant of f is determined to be $\Delta = a^2 - 4b$. Then $\delta = x_1 - x_2$ is a square root of Δ, and since $x_1 + x_2 = -a$, we get

$$x_1 = \tfrac{1}{2}(-a + \delta), \qquad x_2 = \tfrac{1}{2}(-a - \delta),$$

resp.

$$x_{1/2} = -\frac{a}{2} \pm \sqrt{\frac{a^2}{4} - b}.$$

These are the well-known formulas for the solutions of quadratic equations.

Next we consider a *cubic polynomial* $f \in K[X]$, say

$$f = X^3 + aX^2 + bX + c,$$

where we assume $\operatorname{char} K \neq 2, 3$. Replacing X by $X - \frac{1}{3}a$, we can reduce to the case that f is of the somewhat simpler form

$$f = X^3 + pX + q.$$

Again, let L be a splitting field of f and write x_1, x_2, x_3 for the zeros of f in L. The discriminant of f is $\Delta = -4p^3 - 27q^2$; see the computation following 4.4/10. To solve the equation $f(x) = 0$, it is helpful to look first at the case $\operatorname{Gal}(L/K) = \mathfrak{S}_n$, which will be referred to as the *generic case*. However, our considerations will be valid for arbitrary Galois groups $\operatorname{Gal}(L/K)$, as we will see at the end.

Considering the normal series $\mathfrak{S}_3 \supset \mathfrak{A}_3 \supset \{1\}$, we start by adjoining to K a square root

$$\delta = (x_1 - x_2)(x_1 - x_3)(x_2 - x_3) = \sqrt{\Delta}$$

of the discriminant Δ. Then, in the generic case that we are following for the moment, we recognize $L/K(\delta)$ as a cyclic Galois extension of degree 3. Motivated by the assumptions in 4.8/3 (i), we adjoin to $K(\delta)$, resp. K, a primitive third root of unity ζ and assume from now on for simplicity that $\zeta \in K$. Then $L/K(\delta)$ is obtained by adjoining a third root of some element of $K(\delta)$. Retracing the constructions in 4.8/3 and 4.8/1 shows that we can choose this root as a *Lagrange resolvent*

$$(\zeta, x) = x + \zeta\sigma(x) + \zeta^2\sigma^2(x)$$

for a suitable element $x \in L$, where σ is a generating element of the cyclic group $\mathrm{Gal}(L/K(\delta))$.

Since such an element $x \in L$ cannot be specified in an obvious way, we use the zeros x_1, x_2, x_3 of f to define "resolvents" by

$$\begin{aligned}(1, x) &= x_1 + x_2 + x_3 = 0,\\ (\zeta, x) &= x_1 + \zeta x_2 + \zeta^2 x_3,\\ (\zeta^2, x) &= x_1 + \zeta^2 x_2 + \zeta x_3,\end{aligned}$$

where for reasons of motivation, we could imagine x, $\sigma(x)$, as well as $\sigma^2(x)$ at the place of x_1, x_2, and x_3. Recalling that the primitive third roots

$$\zeta = -\tfrac{1}{2} + \tfrac{1}{2}\sqrt{-3}, \qquad \zeta^2 = -\tfrac{1}{2} - \tfrac{1}{2}\sqrt{-3}, \tag{0}$$

are the zeros of the cyclotomic polynomial $\Phi_3 = X^2 + X + 1$, we can conclude that the zeros x_1, x_2, x_3 of f are determined by

$$\begin{aligned}x_1 &= \tfrac{1}{3}\big((\zeta, x) + (\zeta^2, x)\big),\\ x_2 &= \tfrac{1}{3}\big(\zeta^2(\zeta, x) + \zeta(\zeta^2, x)\big),\\ x_3 &= \tfrac{1}{3}\big(\zeta(\zeta, x) + \zeta^2(\zeta^2, x)\big).\end{aligned} \tag{1}$$

Thus, if we can get hold of the occurring resolvents, we will be done. Due to the fact that in the generic case, the extension $L/K(\delta)$ admits $\mathfrak{A}_3$ as its Galois group, the third powers of the resolvents (ζ, x) and (ζ^2, x) are invariant under $\mathrm{Gal}(L/K(\delta))$ and therefore are contained in $K(\delta)$. We want to justify this also by direct computation. Indeed, writing

$$\begin{aligned}\delta &= (x_1 - x_2)(x_1 - x_3)(x_2 - x_3)\\ &= x_1^2x_2 + x_2^2x_3 + x_3^2x_1 - x_1^2x_3 - x_2^2x_1 - x_3^2x_2\end{aligned}$$

and using (0), we get, independently of the generic case,

$$\begin{aligned}(\zeta, x)^3 &= x_1^3 + x_2^3 + x_3^3 + 3\zeta(x_1^2x_2 + x_2^2x_3 + x_3^2x_1)\\ &\quad + 3\zeta^2(x_1^2x_3 + x_2^2x_1 + x_3^2x_2) + 6x_1x_2x_3\\ &= \textstyle\sum_i x_i^3 - \frac{3}{2}\sum_{i\neq j} x_i^2x_j + 6x_1x_2x_3 + \frac{3}{2}\sqrt{-3}\cdot\delta.\end{aligned}$$

Here a special choice of the square root $\sqrt{-3}$ does not matter, since replacing $\sqrt{-3}$ by $-\sqrt{-3}$ has the effect that the quantities ζ and ζ^2, resp. (ζ, x) and (ζ^2, x), are interchanged. In particular, we can determine $(\zeta^2, x)^3$ by replacing $\sqrt{-3}$ by $-\sqrt{-3}$ in the above formula.

Now let us view $(\zeta, x)^3$ as a symmetric function in x_1, x_2, x_3 and write it as a polynomial in the elementary symmetric polynomials

$$\begin{aligned}\sigma_1 &= s_1(x_1, x_2, x_3) = x_1 + x_2 + x_3 = 0,\\ \sigma_2 &= s_2(x_1, x_2, x_3) = x_1x_2 + x_1x_3 + x_2x_3 = p,\\ \sigma_3 &= s_3(x_1, x_2, x_3) = x_1x_2x_3 = -q,\end{aligned}$$

in order to get an expression in terms of the coefficients p, q of the equation under consideration. To do this, we follow the method of Proposition 4.3/5.

$$\begin{array}{rcl}
(\zeta, x)^3 & = & \sum_i x_i^3 \;-\frac{3}{2}\sum_{i\neq j} x_i^2x_j \;+6x_1x_2x_3 \;+\frac{3}{2}\sqrt{-3}\cdot\delta \\
\sigma_1^3 & = & \sum_i x_i^3 \;+3\sum_{i\neq j} x_i^2x_j \;+6x_1x_2x_3 \\
\hline
 & & -\frac{9}{2}\sum_{i\neq j} x_i^2x_j \qquad\qquad +\frac{3}{2}\sqrt{-3}\cdot\delta \\
-\frac{9}{2}\sigma_1\sigma_2 & = & -\frac{9}{2}\sum_{i\neq j} x_i^2x_j \;-\frac{27}{2}x_1x_2x_3 \\
\hline
 & & \frac{27}{2}x_1x_2x_3 \;+\frac{3}{2}\sqrt{-3}\cdot\delta \\
\frac{27}{2}\sigma_3 & = & \frac{27}{2}x_1x_2x_3 \\
\hline
 & & \frac{3}{2}\sqrt{-3}\cdot\delta
\end{array}$$

Therefore, we obtain $(\zeta, x)^3 = \sigma_1^3 - \frac{9}{2}\sigma_1\sigma_2 + \frac{27}{2}\sigma_3 + \frac{3}{2}\sqrt{-3}\cdot\delta$, and hence due to $\sigma_1 = 0$ and $\sigma_3 = -q$,

$$(\zeta, x)^3 = -\tfrac{27}{2}q + \tfrac{3}{2}\sqrt{-3}\cdot\delta = -\tfrac{27}{2}q + 27\sqrt{(\tfrac{p}{3})^3 + (\tfrac{q}{2})^2}, \tag{2}$$

as well as

$$(\zeta^2, x)^3 = -\tfrac{27}{2}q - \tfrac{3}{2}\sqrt{-3}\cdot\delta = -\tfrac{27}{2}q - 27\sqrt{(\tfrac{p}{3})^3 + (\tfrac{q}{2})^2}. \tag{3}$$

The resolvents (ζ, x) and (ζ^2, x) are uniquely determined by these equations, up to a third root of unity. Furthermore, the equations (1) for x_1, x_2, x_3 show that we may replace (ζ, x) by $\zeta(\zeta, x)$ if at the same time we replace (ζ^2, x) by $\zeta^2(\zeta^2, x)$. This suggests that the third roots of the right-hand parts of the equations (2) and (3) cannot be chosen independently of each other in solving the equation $x^3+px+q = 0$ by means of the relations (1). Indeed, the calculation

$$\begin{aligned}(\zeta, x)(\zeta^2, x) &= (x_1 + \zeta x_2 + \zeta^2 x_3)(x_1 + \zeta^2 x_2 + \zeta x_3)\\ &= x_1^2 + x_2^2 + x_3^2 + (\zeta + \zeta^2)(x_1x_2 + x_1x_3 + x_2x_3)\\ &= \sigma_1^2 - 3\sigma_2 = -3\sigma_2 = -3p\end{aligned}$$

confirms this matter. Therefore, we can state the following:

Proposition 1 (Cardano's formulas). *Let K be a field satisfying* $\operatorname{char} K \neq 2, 3$. *For coefficients $p, q \in K$, the solutions of the algebraic equation $x^3 + px + q = 0$ are given by*

$$x_1 = u + v, \qquad x_2 = \zeta^2 u + \zeta v, \qquad x_3 = \zeta u + \zeta^2 v,$$

where $\zeta \in \overline{K}$ is an arbitrary primitive third root of unity, and where

$$u = \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{p}{3}\right)^3 + \left(\frac{q}{2}\right)^2}}, \qquad v = \sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{p}{3}\right)^3 + \left(\frac{q}{2}\right)^2}},$$

with the side condition that the third roots are chosen to satisfy $uv = -\frac{1}{3}p$.

Proof. If we replace in the above expressions u, v by $\zeta u, \zeta^2 v$, resp. $\zeta^2 u, \zeta v$, this only produces a permutation of x_1, x_2, x_3. Therefore, we may assume without loss of generality that $u = \frac{1}{3}(\zeta, x)$, as well as $v = \frac{1}{3}(\zeta^2, x)$. Then, due to the equations (1), the quantities x_1, x_2, x_3 mentioned in the assertion will coincide with the solutions of the equation $x^3 + px + q = 0$. □

Finally, let us consider a *quartic polynomial* $f \in K[X]$, say

$$f = X^4 + pX^2 + qX + r,$$

where we assume $\operatorname{char} K \neq 2, 3$ again. Every monic polynomial of degree 4 can be transformed to this type by applying a substitution $X \longmapsto X - \frac{1}{4}c$, for a suitable constant $c \in K$. Furthermore, let x_1, x_2, x_3, x_4 be the zeros of f in a splitting field L of f over K. Since the term of degree 3 in f is trivial, we get $x_1 + x_2 + x_3 + x_4 = 0$. Similarly as we did for polynomials of degree 3, we start our consideration by looking at the *generic case* in which $\operatorname{Gal}(L/K) = \mathfrak{S}_4$. In addition, we assume in this case that x_1, x_2, x_3 are algebraically independent over the prime field of K. For example, such a setting is obtained in transforming the generic equation of degree 4 to the special type we are considering here. However, it should be observed that the computations below are valid independent of such special assumptions.

In the generic case we look at the normal series

$$\mathfrak{S}_4 \supset \mathfrak{A}_4 \supset \mathfrak{V}_4 \supset \mathfrak{Z} \supset \{1\}$$

already mentioned before, as well as at the corresponding chain of field extensions

$$K \subset L^{\mathfrak{A}_4} \subset L^{\mathfrak{V}_4} \subset L^{\mathfrak{Z}} \subset L.$$

As usual, we get $L^{\mathfrak{A}_4} = K(\delta)$ for a square root δ of the discriminant Δ of f, where one can show that

$$\Delta = 144pq^2r - 128p^2r^2 - 4p^3q^2 + 16p^4r - 27q^4 + 256r^3$$

using 4.4/10. However, we will not need this in the sequel. The extension $L^{\mathfrak{V}_4}/L^{\mathfrak{A}_4}$ is of degree 3 and hence is generated by an arbitrary element of $L^{\mathfrak{V}_4}$ that is not contained in $L^{\mathfrak{A}_4}$, for example by

$$z_1 = (x_1 + x_2)(x_3 + x_4) \in L.$$

Indeed, z_1 is not left fixed by the permutation $(1,2,3) \in \mathfrak{A}_4$. On the other hand, z_1 is invariant under all elements in $\mathfrak{V}_4$, and in addition, under the permutations $(1,2)$, $(3,4)$, $(1,3,2,4)$, and $(1,4,2,3)$. Together, these are the members of the stabilizer subgroup of z_1 in $\mathfrak{S}_4$. Using the orbit-stabilizer lemma 5.1/6, we can conclude that the orbit of z_1 under the action of $\mathfrak{S}_4$ consists of precisely three elements, namely

$$\begin{aligned} z_1 &= (x_1 + x_2)(x_3 + x_4), \\ z_2 &= (x_1 + x_3)(x_2 + x_4), \\ z_3 &= (x_1 + x_4)(x_2 + x_3). \end{aligned}$$

But then z_1, z_2, z_3 are the solutions of an equation of degree 3 with coefficients in K, namely of

$$z^3 - b_1 z^2 + b_2 z - b_3 = 0,$$

where b_1, b_2, b_3 are the elementary symmetric polynomials in z_1, z_2, z_3; in more detail,[1]

$$\begin{aligned} b_1 = z_1 + z_2 + z_3 &= 2 \textstyle\sum_{i<j} x_i x_j, \\ b_2 = z_1 z_2 + z_1 z_3 + z_2 z_3 &= \textstyle\sum_{i<j} x_i^2 x_j^2 + 3 \sum_{j<k}^{i} x_i^2 x_j x_k + 6 x_1 x_2 x_3 x_4, \\ b_3 = z_1 z_2 z_3 &= \textstyle\sum_{i,j,k} x_i^3 x_j^2 x_k + 2 \sum_{j<k<l}^{i} x_i^3 x_j x_k x_l \\ &\quad + 2 \textstyle\sum_{i<j<k} x_i^2 x_j^2 x_k^2 + 4 \sum_{k<l}^{i<j} x_i^2 x_j^2 x_k x_l. \end{aligned}$$

As we see, b_1, b_2, b_3 are symmetric in x_1, x_2, x_3, x_4. Therefore, we can write the b_i as polynomials in the elementary symmetric polynomials

$$\begin{aligned} \sigma_1 = s_1(x_1,x_2,x_3,x_4) &= \textstyle\sum_i x_i &&= 0, \\ \sigma_2 = s_2(x_1,x_2,x_3,x_4) &= \textstyle\sum_{i<j} x_i x_j &&= p, \\ \sigma_3 = s_3(x_1,x_2,x_3,x_4) &= \textstyle\sum_{i<j<k} x_i x_j x_k &&= -q, \\ \sigma_4 = s_4(x_1,x_2,x_3,x_4) &= x_1 x_2 x_3 x_4 &&= r, \end{aligned}$$

and hence express the b_i in terms of the coefficients p, q, r of the algebraic equation we started with. First, we read $b_1 = 2\sigma_2 = 2p$. Next, to describe b_2, we follow the method from the proof of Proposition 4.3/5:

[1] In subsequent summations, the indices vary over $\{1,2,3,4\}$. Different indices of a sum are allowed to assume only *different* values.

$$\begin{array}{rl}
b_2 = & \sum_{i<j} x_i^2x_j^2 \ +3\sum_{\substack{i\\ j<k}} x_i^2x_jx_k \ +6x_1x_2x_3x_4 \\
\sigma_2^2 = & \sum_{i<j} x_i^2x_j^2 \ +2\sum_{\substack{i\\ j<k}} x_i^2x_jx_k \ +6x_1x_2x_3x_4 \\
\hline
 & \sum_{\substack{i\\ j<k}} x_i^2x_jx_k \\
\sigma_1\sigma_3 = & \sum_{\substack{i\\ j<k}} x_i^2x_jx_k \ +4x_1x_2x_3x_4 \\
\hline
 & -4x_1x_2x_3x_4 \\
-4\sigma_4 = & -4x_1x_2x_3x_4 \\
\hline
 & 0
\end{array}$$

This yields $b_2 = \sigma_2^2 + \sigma_1\sigma_3 - 4\sigma_4 = p^2 - 4r$, since $\sigma_1 = 0$. Finally, we express b_3 in terms of $\sigma_1, \sigma_2, \sigma_3, \sigma_4$:

$$\begin{array}{rl}
b_3 = & \sum_{i,j,k} x_i^3x_j^2x_k \ +2\sum_{\substack{i\\ j<k<l}} x_i^3x_jx_kx_l \ +2\sum_{i<j<k} x_i^2x_j^2x_k^2 \ +4\sum_{\substack{i<j\\ k<l}} x_i^2x_j^2x_kx_l \\
\sigma_1\sigma_2\sigma_3 = & \sum_{i,j,k} x_i^3x_j^2x_k \ +3\sum_{\substack{i\\ j<k<l}} x_i^3x_jx_kx_l \ +3\sum_{i<j<k} x_i^2x_j^2x_k^2 \ +8\sum_{\substack{i<j\\ k<l}} x_i^2x_j^2x_kx_l \\
\hline
 & -\sum_{\substack{i\\ j<k<l}} x_i^3x_jx_kx_l \ -\sum_{i<j<k} x_i^2x_j^2x_k^2 \ -4\sum_{\substack{i<j\\ k<l}} x_i^2x_j^2x_kx_l \\
-\sigma_1^2\sigma_4 = & -\sum_{\substack{i\\ j<k<l}} x_i^3x_jx_kx_l \ -2\sum_{\substack{i<j\\ k<l}} x_i^2x_j^2x_kx_l \\
\hline
 & -\sum_{i<j<k} x_i^2x_j^2x_k^2 \ -2\sum_{\substack{i<j\\ k<l}} x_i^2x_j^2x_kx_l \\
-\sigma_3^2 = & -\sum_{i<j<k} x_i^2x_j^2x_k^2 \ -2\sum_{\substack{i<j\\ k<l}} x_i^2x_j^2x_kx_l \\
\hline
 & 0
\end{array}$$

This shows that $b_3 = \sigma_1\sigma_2\sigma_3 - \sigma_1^2\sigma_4 - \sigma_3^2 = -q^2$, again since $\sigma_1 = 0$. Thereby, independently of the generic case, we recognize z_1, z_2, z_3 as the solutions of the equation

$$z^3 - 2pz^2 + (p^2 - 4r)z + q^2 = 0.$$

This equation (or its inherent polynomial) is sometimes referred to as the *resolvent cubic* of the equation $x^4 + px^2 + qx + r = 0$. Note that the solutions z_1, z_2, z_3 of the resolvent cubic can be determined by means of Cardano's formulas.

Looking at the generic case again, we see that $\mathfrak{V}_4$ is characterized as the subgroup of $\mathfrak{S}_4$ fixing z_1, z_2, z_3. This implies $\mathrm{Gal}(L/K(z_1, z_2, z_3)) = \mathfrak{V}_4$, as well as $K(z_1, z_2, z_3) = L^{\mathfrak{V}_4}$. Thus, to pass from $K(z_1, z_2, z_3)$ to L we would have to adjoin square roots according to the chain $\mathfrak{V}_4 \supset \mathfrak{Z} \supset \{1\}$. Now observe that $x_1 + x_2$ is invariant under the permutations (1) and $(1,2)(3,4)$ of $\mathfrak{V}_4$, but not with respect to the remaining elements of $\mathfrak{V}_4$. Therefore, $x_1 + x_2$ is of degree 2 over $K(z_1, z_2, z_3)$. Indeed, independently of this, the equations

$$(x_1 + x_2)(x_3 + x_4) = z_1, \qquad x_1 + x_2 + x_3 + x_4 = 0$$

yield

$$x_1 + x_2 = \sqrt{-z_1}, \qquad x_3 + x_4 = -\sqrt{-z_1},$$

for a suitable square root of $-z_1$. Likewise, we have

$$\begin{aligned} x_1 + x_3 &= \sqrt{-z_2}, & x_2 + x_4 &= -\sqrt{-z_2}, \\ x_1 + x_4 &= \sqrt{-z_3}, & x_2 + x_3 &= -\sqrt{-z_3}, \end{aligned}$$

and as a consequence,

$$\begin{aligned} x_1 &= \tfrac{1}{2}(\ \ \sqrt{-z_1} + \sqrt{-z_2} + \sqrt{-z_3}), \\ x_2 &= \tfrac{1}{2}(\ \ \sqrt{-z_1} - \sqrt{-z_2} - \sqrt{-z_3}), \\ x_3 &= \tfrac{1}{2}(-\sqrt{-z_1} + \sqrt{-z_2} - \sqrt{-z_3}), \\ x_4 &= \tfrac{1}{2}(-\sqrt{-z_1} - \sqrt{-z_2} + \sqrt{-z_3}). \end{aligned}$$

Similarly as we have done for cubic equations, let us examine how to choose the square roots of $-z_1, -z_2, -z_3$. Observe for this that

$$\begin{aligned} (x_1 + x_2)(x_1 + x_3)(x_1 + x_4) &= x_1^2(x_1 + x_2 + x_3 + x_4) + \textstyle\sum_{i<j<k} x_i x_j x_k \\ &= \textstyle\sum_{i<j<k} x_i x_j x_k \\ &= -q. \end{aligned}$$

It follows that the square roots $\sqrt{-z_1}, \sqrt{-z_2}, \sqrt{-z_3}$ must satisfy the side condition

$$\sqrt{-z_1} \cdot \sqrt{-z_2} \cdot \sqrt{-z_3} = -q.$$

Therefore, we can state the following result:

Proposition 2. *Let K be a field of characteristic* $\operatorname{char} K \neq 2, 3$. *For coefficients $p, q, r \in K$, the solutions of the algebraic equation $x^4 + px^2 + qx + r = 0$ are given by*

$$\begin{aligned} x_1 &= \tfrac{1}{2}(\ \ \sqrt{-z_1} + \sqrt{-z_2} + \sqrt{-z_3}), \\ x_2 &= \tfrac{1}{2}(\ \ \sqrt{-z_1} - \sqrt{-z_2} - \sqrt{-z_3}), \\ x_3 &= \tfrac{1}{2}(-\sqrt{-z_1} + \sqrt{-z_2} - \sqrt{-z_3}), \\ x_4 &= \tfrac{1}{2}(-\sqrt{-z_1} - \sqrt{-z_2} + \sqrt{-z_3}), \end{aligned}$$

where z_1, z_2, z_3 are the solutions of the resolvent cubic

$$z^3 - 2pz^2 + (p^2 - 4r)z + q^2 = 0,$$

and where the square roots are required to satisfy the side condition

$$\sqrt{-z_1} \cdot \sqrt{-z_2} \cdot \sqrt{-z_3} = -q.$$

Finally, let us add that due to the relations

$$z_1 - z_2 = -(x_1 - x_4)(x_2 - x_3),$$
$$z_1 - z_3 = -(x_1 - x_3)(x_2 - x_4),$$
$$z_2 - z_3 = -(x_1 - x_2)(x_3 - x_4),$$

the discriminant of $X^4 + pX^2 + qX + r$ coincides with the discriminant of the resolvent cubic $X^3 - 2pX^2 + (p^2 - 4r)X + q^2$.

Exercises

1. *Let K be a subfield of $\mathbb{R}$, the field of real numbers. Furthermore, let $f, g \in K[X]$ be monic irreducible polynomials of degree 4, resp. 3, such that $g(z) = 0$ is the resolvent cubic of the algebraic equation $f(x) = 0$ (assuming that the term of degree 3 in f is trivial). Determine the Galois group of the equation $f(x) = 0$ under the assumption that f does not admit zeros in $\mathbb{R}$.*
2. Let K be a field and L a splitting field of a quartic polynomial $f \in K[X]$, where we want to assume that the term of degree 3 in f is trivial. Furthermore, let L' for $K \subset L' \subset L$ be a splitting field of the resolvent cubic of f. View the Galois group $G = \mathrm{Gal}(L/K)$ as a subgroup of $\mathfrak{S}_4$ and show that $G \cap \mathfrak{V}_4$ is a normal subgroup in G such that $\mathrm{Gal}(L'/K) = G/(G \cap \mathfrak{V}_4)$.

6.3 Fundamental Theorem of Algebra

In the past, various attempts at studying the algebraic structure of classical fields like $\mathbb{R}$ and $\mathbb{C}$ have given stimulus to further develop the general theory of fields and their extensions. The *fundamental theorem of algebra* is a prominent example of this. Its proof goes back to Euler and Lagrange and will be given below in terms of methods from algebra. But there exist other proofs, for example using techniques from complex analysis.

Theorem 1. *The field $\mathbb{C}$ of complex numbers is algebraically closed.*

Proof. Recall that $\mathbb{C}$ is constructed from the real numbers $\mathbb{R}$ by adjoining a square root of -1. To show that $\mathbb{C}$ is algebraically closed we must rely on certain properties of $\mathbb{R}$. In fact, we will use the following properties:

Every polynomial $f \in \mathbb{R}[X]$ of odd degree admits a zero in $\mathbb{R}$.

Every element $a \in \mathbb{R}$, $a \geq 0$, admits a square root in $\mathbb{R}$.

The second property implies, as we want to show, that every quadratic polynomial in $\mathbb{C}[X]$ admits a zero in $\mathbb{C}$. To prove this it is enough to verify that every element $z \in \mathbb{C}$ admits a square root in $\mathbb{C}$. Therefore, consider an element $z = x + iy \in \mathbb{C}$, where $x, y \in \mathbb{R}$. To write z as a square in $\mathbb{C}$, i.e.,

$$z = x + iy = (a + ib)^2 = a^2 - b^2 + 2iab$$

for elements $a, b \in \mathbb{R}$, we have to solve the equations

$$x = a^2 - b^2, \qquad y = 2ab,$$

in the unknown quantities a and b. A straightforward computation shows that up to the sign of a and b, these equations are equivalent to

$$a^2 = \frac{1}{2}x \pm \frac{1}{2}\sqrt{x^2+y^2}, \qquad b^2 = -\frac{1}{2}x \pm \frac{1}{2}\sqrt{x^2+y^2}.$$

The occurring square roots are meant to be nonnegative, and the $\pm$ sign is, in fact, a $+$ sign, since negative values for a^2 and b^2 are excluded. Therefore, if we use the fact that nonnegative real numbers admit square roots in $\mathbb{R}$, the existence of the desired solutions a and b is assured.

Now consider a chain of field extensions $\mathbb{R} \subset \mathbb{C} \subset L$, where $L/\mathbb{C}$ is finite. To prove that $\mathbb{C}$ is algebraically closed, we have to show that $L = \mathbb{C}$. Enlarging L if necessary, we may assume without loss of generality that $L/\mathbb{R}$ is a Galois extension. Let $G = \mathrm{Gal}(L/\mathbb{R})$ be the corresponding Galois group and assume

$$\big[L : \mathbb{R}\big] = \operatorname{ord} G = 2^k m, \quad \text{where} \quad 2 \nmid m.$$

Then we see that $k \geq 1$. Furthermore, it follows from 5.2/6 that G contains a 2-Sylow group H, hence a subgroup of order 2^k. Therefore, the fixed field L^H under the action of H on L satisfies

$$\big[L : L^H\big] = 2^k, \qquad \big[L^H : \mathbb{R}\big] = m,$$

due to the fundamental theorem of Galois theory 4.1/6. Since every real polynomial of odd degree admits a zero in $\mathbb{R}$, as was mentioned before, we can conclude that $m = 1$, for example by applying the primitive element theorem 3.6/12. But then L is of degree 2^k over $\mathbb{R}$ and hence of degree 2^{k-1} over $\mathbb{C}$. Furthermore, $L/\mathbb{C}$ is a Galois extension. Now, if $L \neq \mathbb{C}$ and therefore $k \geq 2$, we can use 5.2/4 to see that there exists a subgroup $H' \subset G' = \mathrm{Gal}(L/\mathbb{C})$ of order 2^{k-2}. The corresponding fixed field $L^{H'}$ satisfies $\big[L : L^{H'}\big] = 2^{k-2}$ and hence $\big[L^{H'} : \mathbb{C}\big] = 2$. However, this is impossible, since every quadratic complex polynomial admits a zero in $\mathbb{C}$, as we have shown. Consequently, we must have $L = \mathbb{C}$ and $\mathbb{C}$ is algebraically closed. □

The proof of the fundamental theorem of algebra given above is based on the theory of Sylow groups. In Exercise 2 below we suggest a different proof avoiding this theory. Furthermore, let us add that from a purely algebraic point of view, the field $\mathbb{R}$ of real numbers is not unique as a subfield of $\mathbb{C}$, since there exist automorphisms of $\mathbb{C}$ that do not map $\mathbb{R}$ to itself; see, for example, Exercise 2 of Section 7.1. However, there are deeper reasons for the fact that $\mathbb{C}$, as an algebraic closure of $\mathbb{R}$, is of degree 2 over $\mathbb{R}$, as shown by the following result going back to E. Artin.

Proposition 2. *Let K be a field, $\overline{K}$ an algebraic closure of K, and $i \in \overline{K}$ an element satisfying $i^2 = -1$. Then $\overline{K} = K(i)$ if $[\overline{K} : K] < \infty$. If, in addition, $\overline{K}/K$ is a nontrivial extension, we have* $\operatorname{char} K = 0$.

Proof. Assume that the degree $[\overline{K} : K]$ is finite. Since $\overline{K}$ is algebraically closed, the extension $\overline{K}/K$ is normal, and we claim that $\overline{K}/K$ is even Galois. Indeed, assuming $\operatorname{char} K = p > 0$, there is an intermediate field L of $\overline{K}/K$, due to 3.7/4, such that $\overline{K}/L$ is purely inseparable and L/K is separable. Then consider the Frobenius homomorphism $\sigma\colon \overline{K} \longrightarrow \overline{K}$, $x \longmapsto x^p$, and observe that σ is an automorphism, since $\overline{K}$ is algebraically closed. Since $\sigma(L) \subset L$ and the degree

$$[\overline{K} : L] = [\sigma(\overline{K}) : \sigma(L)] = [\overline{K} : \sigma(L)]$$

is finite, we get $\sigma(L) = L$. However, this shows that L does not admit a non-trivial purely inseparable extension. Therefore, we have $L = \overline{K}$, and $\overline{K}/K$ is Galois.

It follows that $\overline{K}/K(i)$ is a finite Galois extension as well, and we have to show that this extension is trivial. Suppose that such is not the case. Then there is a subgroup in $\operatorname{Gal}(\overline{K}/K(i))$ whose order is prime; for example, use 5.2/8 or 5.2/11. If $L \subset \overline{K}$ is the corresponding fixed field, then the degree $[\overline{K} : L]$ is prime as well, say $[\overline{K} : L] = \ell$, and $\overline{K}/L$ is a cyclic Galois extension of degree ℓ. Assuming $p = \operatorname{char} K > 0$, let us first consider the case $\ell = p$. Then $\overline{K} = L(a)$ by 4.8/5 (i) for some element $a \in \overline{K}$ whose minimal polynomial over L is of type $X^p - X - c$. To obtain a contradiction, consider the map $\tau\colon \overline{K} \longrightarrow \overline{K}$, $x \longmapsto x^p - x$. It is surjective, since $\overline{K}$ is algebraically closed. Furthermore, using the relation $\operatorname{tr}_{\overline{K}/L}(x^p) = (\operatorname{tr}_{\overline{K}/L}(x))^p$ derived from 4.7/4, we conclude that

$$\operatorname{tr}_{\overline{K}/L} \circ \tau = \tau|_L \circ \operatorname{tr}_{\overline{K}/L}.$$

Since both τ and $\operatorname{tr}_{\overline{K}/L}$ are surjective, see 4.7/7, it follows that $\tau|_L$ is surjective as well. However, then $X^p - X - c$ admits a zero in L, in contradiction to the fact that this polynomial equals the minimal polynomial of a over L.

Next assume that the prime degree $\ell = [\overline{K} : L]$ is different from the characteristic of K, and choose a primitive ℓth root of unity $\zeta_\ell \in \overline{K}$. Since it is of degree $< \ell$ over L by 4.5/7, the multiplicativity formula 3.2/2 shows that we must have $\zeta_\ell \in L$. In particular, the result 4.8/3 (i) becomes applicable, and we conclude that $\overline{K} = L(a)$ for some element $a \in \overline{K}$ whose minimal polynomial over L is of type $X^\ell - c$. Now let $\alpha \in \overline{K}$ be an ℓth root of a, i.e., an element satisfying $\alpha^\ell = a$. Then the multiplicativity of the norm of $\overline{K}$ over L in conjunction with 4.7/2 (ii) yields

$$\operatorname{N}_{\overline{K}/L}(\alpha)^\ell = \operatorname{N}_{\overline{K}/L}(\alpha^\ell) = \operatorname{N}_{\overline{K}/L}(a) = (-1)^{\ell+1} c.$$

Therefore, if ℓ is odd, $\operatorname{N}_{\overline{K}/L}(\alpha) \in L$ is an ℓth root of c, contradicting the fact that the polynomial $X^\ell - c \in L[X]$ is irreducible. Finally, if $\ell = 2$, we see that $\operatorname{N}_{\overline{K}/L}(\alpha) \in L$ is a square root of $-c$. But then, since $i \in L$, also c admits a square root in L. Again, this contradicts the irreducibility of the polynomial $X^2 - c \in L[X]$, thus establishing the desired equality $\overline{K} = K(i)$.

Now assume $K \subsetneq K(i) = \overline{K}$, so that in particular, -1 is not a square in K. To prove $\operatorname{char} K = 0$, we show that the sum of two squares in K is again a square in K. Indeed, consider two elements $a, b \in K$. Then $a + ib$ admits a

square root in $K(i)$, say $x+iy$, and we have $x^2-y^2+2ixy = a+ib$. This implies $a = x^2 - y^2$, $b = 2xy$, and therefore

$$a^2 + b^2 = (x^2 - y^2)^2 + 4x^2y^2 = (x^2 + y^2)^2.$$

Furthermore, we can use induction to show that the sum of finitely many squares in K is again a square in K. But then, given any field of positive characteristic, the element -1 can be written as a repeated sum of the unit element $1 = 1^2$ and hence is a square. However, since -1 is not a square in K by our assumption, the case of positive characteristic is excluded, and we have $\operatorname{char} K = 0$. □

Exercises

1. *Specify the arguments needed to establish the basic properties of real numbers that were used in the proof of Theorem* 1, *namely that every real polynomial of odd degree admits a zero in* $\mathbb{R}$ *and that every element* $a \in \mathbb{R}$, $a \geq 0$, *admits a square root in* $\mathbb{R}$.

2. Let $f \in \mathbb{R}[X]$ be a nonconstant polynomial of degree $n = 2^k m$, where $2 \nmid m$. Show by induction on k that f admits a zero in $\mathbb{C}$, and deduce the fundamental theorem of algebra from this fact. *Hint*: Assume f to be monic and factorize it over an algebraic closure $\overline{\mathbb{R}}$ of $\mathbb{R}$ into linear factors, say $f = \prod_{\nu=1}^{n}(X - \alpha_\nu)$. Then set $\alpha_{\mu\nu} = \alpha_\mu + \alpha_\nu + b\alpha_\mu\alpha_\nu$ for arbitrary $b \in \mathbb{R}$ and apply the induction hypothesis to the polynomial $g = \prod_{\mu<\nu}(X - \alpha_{\mu\nu})$. The properties of $\mathbb{R}$, as specified in Exercise 1, may be used.

3. For a field K, consider a polynomial $X^n - c \in K[X]$ of degree $n \geq 2$, where $c \neq 0$. Generalizing the methods used in the proof of Proposition 2, show that $X^n - c$ is irreducible if and only if c is not a pth power in K for any prime p dividing n and if, in addition, for $4 \mid n$, the element c is not of type $c = -4a^4$ for some $a \in K$. *Hint*: Start with the case that n is a prime power.

6.4 Compass and Straightedge Construction

In the present section we apply Galois theory to the study of geometric construction problems that are given in the complex plane $\mathbb{C}$. We fix a subset $M \subset \mathbb{C}$ (later it will be $M = \{0, 1\}$) and say that a point $z \in \mathbb{C}$ can be obtained via *compass and straightedge constructions* from M if by means of finitely many *basic construction steps*, M can be enlarged to a subset $M' \subset \mathbb{C}$ such that $z \in M'$. The following basic construction steps are allowed:

(1) Consider two nonparallel straight lines g_1 and g_2 in $\mathbb{C}$, determined by points $z_1, z_2 \in M$, resp. $z_3, z_4 \in M$, and add to M the intersection point of g_1 and g_2.

(2) Consider a circle K in $\mathbb{C}$ with center a point $z_1 \in M$ and with radius given by the distance $|z_3 - z_2|$ between two points $z_2, z_3 \in M$, as well as a

straight line g determined by two points $z_4,\ z_5 \in M$, and add all intersection points of K and g to M.

(3) Consider two distinct circles K_1 and K_2 in $\mathbb{C}$ with centers $z_1, z_2 \in M$, and with radii $|z_4 - z_3|$, resp. $|z_6 - z_5|$, given by the distances between points $z_3, z_4 \in M$, resp. $z_5, z_6 \in M$. Add to M the intersection points of K_1 and K_2.

We write $\mathfrak{K}(M)$ for the set of all points in $\mathbb{C}$ that can be obtained by compass and straightedge constructions from M, where we will always assume $0, 1 \in M$. Furthermore, let $\overline{M}$ be the image of M with respect to the complex conjugation $\mathbb{C} \longrightarrow \mathbb{C},\ z \longmapsto \overline{z}$.[2] Then clearly we get $\mathfrak{K}(M) = \mathfrak{K}(M \cup \overline{M})$, since we can use compass and straightedge constructions to mirror any point $z \in M - \mathbb{R}$ on the real axis $\mathbb{R}$, thereby obtaining the complex conjugate $\overline{z}$ of z. Indeed, if z is purely imaginary, i.e., with real part $\operatorname{Re} z = 0$, we obtain $\overline{z}$ as an intersection point of the line through 0 and z with the circle having center 0 and radius $|z| = |z - 0|$. Otherwise, look at the circle with center z and with radius $|z|$. It has two intersection points z_1, z_2 with the line through 0 and 1, and we obtain $\overline{z}$ as an intersection point of the circles with centers z_1 and z_2, in both cases with radius $|z|$.

To characterize the set $\mathfrak{K}(M)$ in terms of algebraic field extensions, we start out from the smallest subfield in $\mathbb{C}$ containing M and $\overline{M}$, which equals the field $\mathbb{Q}(M \cup \overline{M})$, obtained from $\mathbb{Q}$ by adjoining all elements of $M \cup \overline{M}$.

Proposition 1. *Let $M \subset \mathbb{C}$ satisfy $0, 1 \in M$. The following conditions are equivalent for any point $z \in \mathbb{C}$:*

(i) $z \in \mathfrak{K}(M)$.

(ii) *There exists a chain of field extensions*

$$\mathbb{Q}(M \cup \overline{M}) = L_0 \subset L_1 \subset \ldots \subset L_n \subset \mathbb{C}$$

such that $z \in L_n$ and $[L_i : L_{i-1}] = 2$ for $i = 1, \ldots, n$.

(iii) *z is contained in a Galois extension L of $\mathbb{Q}(M \cup \overline{M})$ whose degree $[L : \mathbb{Q}(M \cup \overline{M})]$ is a power of 2.*

As a direct consequence, this implies the following:

Corollary 2. *Let $M \subset \mathbb{C}$ satisfy $0, 1 \in M$. Then $\mathfrak{K}(M)$ is an algebraic extension field of $\mathbb{Q}(M \cup \overline{M})$. The degree over $\mathbb{Q}(M \cup \overline{M})$ of any element $z \in \mathfrak{K}(M)$ is a power of 2.*

Proof of Proposition 1. We start with the implication (i) $\Longrightarrow$ (ii) and show first of all that M can be assumed to be a field that satisfies $M = \overline{M}$, as well as $i \in M$ for the complex number $i \in \mathbb{C}$, square root of -1. Indeed, the complex conjugation on $\mathbb{C}$ restricts to an automorphism on the field $L_0 = \mathbb{Q}(M \cup \overline{M})$,

[2] In the present section, we use the notation $\overline{M}$ for the image of M with respect to the complex conjugation map, even if M is a field; an algebraic closure of such a field $M \subset \mathbb{C}$, usually denoted by $\overline{M}$, is not needed.

so that $L_0 = \overline{L_0}$. Then $L_1 = L_0(i)$ is an extension field of L_0 of degree ≤ 2 satisfying $L_1 = \overline{L_1}$ and $i \in L_1$. Furthermore, we have $\mathfrak{K}(M) \subset \mathfrak{K}(L_1)$ as well as $\mathbb{Q}(L_1 \cup \overline{L_1}) = L_1$. Hence, it is enough to verify the implication from (i) to (ii) for L_1 instead of M. In other words, we may assume M to be a field satisfying $M = \overline{M}$ and $i \in M$.

From this assumption we conclude for any point $z \in M$ that its real part $\operatorname{Re} z = \frac{1}{2}(z + \overline{z})$, its imaginary part $\operatorname{Im} z = \frac{1}{2i}(z - \overline{z})$, as well as the square $|z|^2 = z\overline{z}$ of its absolute value all belong to M. Now consider a point $z \in \mathfrak{K}(M)$. It is enough to look at the case in which z is obtained from M by a single basic construction step and to show that z is contained in M or in an extension field $L = M(\sqrt{\Delta})$ obtained from M by adjoining a square root of a nonnegative real number $\Delta \in M \cap \mathbb{R}$. Then, due to the fact that $L = \overline{L}$ and $i \in L$, the general case follows using an inductive argument.

We start by assuming that z is obtained from M by means of a basic construction step of type (1). Then z equals the intersection point of two lines

$$g_1 = \big\{z_1 + t(z_2 - z_1)\,;\, t \in \mathbb{R}\big\},$$
$$g_2 = \big\{z_3 + t'(z_4 - z_3)\,;\, t' \in \mathbb{R}\big\},$$

where $z_1, z_2, z_3, z_4 \in M$, and we have to solve the equation

$$z_1 + t(z_2 - z_1) = z_3 + t'(z_4 - z_3)$$

for the parameters $t, t' \in \mathbb{R}$. Splitting the equation into its real and imaginary parts yields a system of two linear equations with coefficients in $\mathbb{R} \cap M$ for the unknown quantities t, t', which admits a unique solution $(t_0, t_0') \in \mathbb{R}^2$. Since this solution can be calculated from the coefficients of the equations by means of rational operations, for example using Cramer's rule, we get $t_0, t_0' \in \mathbb{R} \cap M$ and therefore

$$z = z_1 + t_0(z_2 - z_1) = z_3 + t_0'(z_4 - z_3) \in M,$$

so that in this case it is not necessary to extend M.

Next we assume that z is obtained from M by applying a basic construction step of type (2). Hence, z is an intersection point of a circle

$$K = \big\{\zeta \in \mathbb{C}\,;\, |\zeta - z_1|^2 = |z_3 - z_2|^2\big\}$$

and a straight line

$$g = \big\{z_4 + t(z_5 - z_4)\,;\, t \in \mathbb{R}\big\},$$

where $z_1, \ldots, z_5 \in M$. To determine the intersection points of K and g, the equation

$$|z_4 + t(z_5 - z_4) - z_1|^2 = |z_3 - z_2|^2$$

has to be solved for t. This is a quadratic equation for t with coefficients that are obtained by means of rational operations from the real and imaginary parts of $z_1, \ldots, z_5$ and hence belong to $\mathbb{R} \cap M$. As a consequence, the equation can be written as $t^2 + at + b = 0$ for suitable coefficients $a, b \in M \cap \mathbb{R}$, and we see

that the intersection point $z \in K \cap g$ under consideration corresponds to a real solution t_0 of $t^2 + at + b = 0$. Then the discriminant $\Delta = a^2 - 4b$ of the equation satisfies $\Delta \geq 0$, and we get $t_0 \in (M \cap \mathbb{R})(\sqrt{\Delta})$, as well as $z = z_4 + t_0(z_5 - z_4) \in L$, where $L = M(\sqrt{\Delta})$.

It remains to consider a basic construction step of type (3). Therefore, let z be an intersection point of two distinct circles

$$K_1 = \{\zeta \in \mathbb{C}\,;\, |\zeta - z_1|^2 = r_1^2\},$$
$$K_2 = \{\zeta \in \mathbb{C}\,;\, |\zeta - z_2|^2 = r_2^2\},$$

where $r_1 = |z_4 - z_3|$, $r_2 = |z_6 - z_5|$, $z_1, \ldots, z_6 \in M$. Then z satisfies the equations

$$z\overline{z} - z\overline{z}_1 - \overline{z}z_1 + z_1\overline{z}_1 = r_1^2,$$
$$z\overline{z} - z\overline{z}_2 - \overline{z}z_2 + z_2\overline{z}_2 = r_2^2,$$

and taking their difference, an equation of type

$$az + \overline{az} + b = 0, \quad \text{i.e.,} \quad 2\mathrm{Re}(az) + b = 0,$$

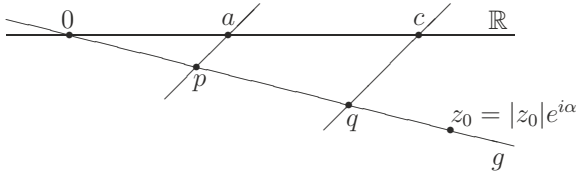
where $a = \overline{z}_2 - \overline{z}_1 \in M$ and $b \in M \cap \mathbb{R}$. Since K_1 and K_2 have different centers, we must have $a \neq 0$. Hence, viewing z as a parameter, the latter equation describes a straight line g, which in particular, contains the points $\frac{-b}{2a}, \frac{-b+i}{2a} \in M$. Furthermore, the intersection points of g with K_1 and K_2 coincide with those of K_1 and K_2. Therefore, we can proceed in the same way as we did for basic construction steps of type (2). This ends the proof of the implication (i) $\Longrightarrow$ (ii).

To prove the reverse implication (ii) $\Longrightarrow$ (i), it is enough to show that $\mathfrak{K}(M)$ (assuming $0, 1 \in M$) is a subfield of $\mathbb{C}$ containing $M \cup \overline{M}$ and hence $\mathbb{Q}(M \cup \overline{M})$, and that for every element $z \in \mathfrak{K}(M)$, its square roots $\pm\sqrt{z}$ belong to $\mathfrak{K}(M)$. To verify these facts, we refer to the relation $\mathfrak{K}(M) = \mathfrak{K}(M \cup \overline{M})$ and establish the following assertions, where some of them are needed only for proof-theoretic reasons:

(a) $z_1, z_2 \in \mathfrak{K}(M) \Longrightarrow z_1 + z_2 \in \mathfrak{K}(M)$,
(b) $z \in \mathfrak{K}(M) \Longrightarrow -z \in \mathfrak{K}(M)$,
(c) $z \in \mathfrak{K}(M) \Longrightarrow |z| \in \mathfrak{K}(M)$,
(d) $e^{\pi i/3} = \frac{1}{2} + \frac{1}{2} \cdot i\sqrt{3} \in \mathfrak{K}(M)$,
(e) $z_1, z_2 \in \mathfrak{K}(M) \Longrightarrow |z_1||z_2| \in \mathfrak{K}(M)$,
(f) $z \in \mathfrak{K}(M), z \neq 0 \Longrightarrow |z|^{-1} \in \mathfrak{K}(M)$,
(g) $z_1, z_2 \in \mathfrak{K}(M) \Longrightarrow z_1 z_2 \in \mathfrak{K}(M)$,
(h) $z \in \mathfrak{K}(M), z \neq 0 \Longrightarrow z^{-1} \in \mathfrak{K}(M)$,
(j) $z \in \mathfrak{K}(M) \Longrightarrow \pm\sqrt{z} \in \mathfrak{K}(M)$.

Each of the preceding assertions can be verified by a simple geometric construction. Concerning (a), interpret the addition of complex numbers as vector addition. The "vector" $z_1 + z_2$ corresponds to the diagonal of the parallelogram

spanned by the "vectors" z_1, z_2. For (b) apply to z a point reflection at the origin. Indeed, $-z$ is situated on the line through 0 and z (assuming $z \neq 0$), as well as on the circle with center 0 and radius $|z| = |z - 0|$. Similarly, in (c) we can interpret $|z|$ as an intersection point of the real axis and the circle with center 0 and radius $|z|$. Next, assertion (d) is needed for the verification of (e) and (f), where we have to know that $\mathfrak{K}(M)$ contains a nonreal point apart form the points $0, 1$. To establish (d) construct an equilateral triangle over the line segment from 0 to 1. Its top vertex, intersection of the circles about 0 and 1 with radius 1, is then the primitive sixth root of unity $e^{\pi i/3} = \frac{1}{2} + \frac{1}{2} \cdot i\sqrt{3}$. Turning to the implications (e) and (f), assume $z_1 \neq 0 \neq z_2$ and consider the following figure:



To draw it, choose a point $z_0 = |z_0|e^{i\alpha}$ in $M - \mathbb{R}$ such that $\mathrm{Re}\, z_0 > 0$, for example $z_0 = e^{\pi i/3}$, and consider the line g through 0 and z_0. Then, on g we can look at the points $p = e^{i\alpha}$ and $q = |z_2|e^{i\alpha}$, as well as on the real axis at the point $a = |z_1|$. All these points belong to $\mathfrak{K}(M)$, as is easily verified. Furthermore, on the real axis consider the point c that is obtained as the intersection point of $\mathbb{R}$ and the parallel of $g_{a,p}$ through q, where $g_{a,p}$ is the line determined by a and p. Also c belongs to $\mathfrak{K}(M)$, as is easily verified using elementary constructions. For example, drop a perpendicular from q to the line $g_{a,p}$, and on this perpendicular, draw another perpendicular at q. Then we can read from the intercept theorem that

$$|q| \cdot |p|^{-1} = |c| \cdot |a|^{-1},$$

whence, due to $|q| = |z_2|$, $|p| = 1$, and $|a| = |z_1|$, that

$$|c| = |a| \cdot |q| = |z_1| \cdot |z_2|$$

and therefore $|z_1| \cdot |z_2| \in \mathfrak{K}(M)$. Furthermore, constructing the parallel of $g_{a,p}$ through $1 \in \mathbb{R}$, its intersection with g yields a complex number of absolute value $|z_1|^{-1}$, where $|z_1|^{-1} \in \mathfrak{K}(M)$ by (c). To settle the implications (g) and (h), observe that in calculating the product of two complex numbers, their absolute values are multiplied and their arguments added. Therefore, it remains to show that addition and negation of angles is possible in terms of basic construction steps; however, this does not pose any problem. Since the same is true for the bisection of angles, it is enough to show for the implication (j) that given a point $z \in \mathfrak{K}(M) - \{0\}$, the square root $\sqrt{|z|}$ can be obtained through basic construction steps. To achieve this, consider on the real axis the line segment from $-|z|$ to 1 and construct Thales's semicircle over it. Then, by the altitude theorem for right triangles, the intersection with the perpendicular erected at 0 yields a complex number of absolute value $\sqrt{|z|}$. Thus, we have shown that

$\mathfrak{K}(M)$ is a subfield of $\mathbb{C}$ that is closed under the extraction of square roots. This ends the proof of the equivalence of conditions (i) and (ii).

It remains to show that conditions (ii) and (iii) are equivalent. Assuming condition (ii), we consider in $\mathbb{C}$ the normal closure L of L_n over $K = \mathbb{Q}(M \cup \overline{M})$; cf. 3.5/7. If $\sigma_1, \ldots, \sigma_r$ are the distinct K-homomorphisms of L_n to $\mathbb{C}$, we can interpret L as the field that is generated over K by all images $\sigma_i(L_n)$, where $i = 1, \ldots, r$. Since L_n is obtained from K by successive adjunction of square roots, the same is true for every $\sigma_i(L_n)$ and hence also for L. Therefore, L/K is a Galois extension whose degree is a power of 2. Since we have $z \in L_n \subset L$, condition (iii) follows.

Conversely, assume that condition (iii) is given. Then the Galois group $\mathrm{Gal}(L/K)$ is a 2-group and therefore solvable by 5.4/6. It follows that $\mathrm{Gal}(L/K)$ admits a normal series with cyclic factors of order 2; see 5.4/7. But then, by the fundamental theorem of Galois theory 4.1/6, this corresponds to a chain of field extensions as required in condition (ii). □

The assertion of Proposition 1 can be used on various occasions to show that certain points of the complex plane, or distances between such points, cannot be obtained from a given subset $M \subset \mathbb{C}$ by means of compass and straightedge constructions. A famous example of this kind is the problem of *squaring the circle*. Its challenge is to use compass and straightedge constructions to turn a circle, given by its center and its radius, into a square whose area is equal to that of the circle. For instance, look at a circle of radius 1 and with center 0. Its area equals the number π, so that a square of the same area will have edge length $\sqrt{\pi}$. Therefore, the problem of squaring the circle consists in figuring out whether $\sqrt{\pi}$ belongs to $\mathfrak{K}(\{0,1\})$. However, $\mathfrak{K}(\{0,1\})$ is an algebraic extension of $\mathbb{Q}$, as we know from Corollary 2. On the other hand, F. Lindemann proved in 1882, see [13], that π, and hence $\sqrt{\pi}$, is transcendental over $\mathbb{Q}$. Consequently, $\sqrt{\pi}$ cannot be obtained from $\{0,1\}$ by means of compass and straightedge constructions, and it follows that squaring the circle in such a way is impossible. In the past, compass and straightedge constructions have been used on several occasions to arrive at excellent approximate solutions for π and $\sqrt{\pi}$. Sometimes such solutions were interpreted as exact solutions, which, however, is not permissible.

Another classical problem that does not admit a solution is the problem of *doubling the cube*: Is it possible to double the volume of a cube by means of compass and straightedge constructions? For example, starting with a cube of edge length 1, doubling its volume leads to a cube of edge length $\sqrt[3]{2}$. However, as we know from Corollary 2, the cube root $\sqrt[3]{2}$ does not belong to $\mathfrak{K}(\{0,1\})$, since its degree over $\mathbb{Q}$ is not a power of 2. In a similar way, one can treat the problem of *angle trisection*; cf. Exercise 2 below.

Finally, let us study applications of compass and straightedge constructions to *regular polygons*.[3] Important contributions to the subject are due to C. F. Gauss. The problem consists in finding out whether for a given integer $n \geq 3$ the regular n-gon is constructible, in the sense that the primitive nth

[3] Regular polygons will always be assumed to be *convex*.

root of unity $e^{2\pi i/n}$ belongs to $\mathfrak{K}(\{0,1\})$. Note that here and in the following, the term "constructible" alludes always to compass and straightedge constructions. In the proof of Proposition 1 we have already shown that $e^{\pi i/3} \in \mathfrak{K}(\{0,1\})$. Therefore, the regular 6-gon is constructible. More generally, we prove the following:

Proposition 3. *For an integer $n \geq 3$, the regular n-gon is constructible if and only if $\varphi(n)$ is a power of 2, where φ is Euler's φ-function* (*cf.* 4.5/3).

Proof. Let ζ_n be a primitive nth root of unity over $\mathbb{Q}$. Then we know from 4.5/8 that $\mathbb{Q}(\zeta_n)/\mathbb{Q}$ is an abelian Galois extension of degree $\varphi(n)$. Assuming that the regular n-gon is constructible, i.e., that $\zeta_n \in \mathfrak{K}(\{0,1\})$, and using Corollary 2, the degree of ζ_n over $\mathbb{Q}$, which equals $\varphi(n)$, is a power of 2. Conversely, if it is known that $\varphi(n) = [\mathbb{Q}(\zeta_n) : \mathbb{Q}]$ is a power of 2, we can conclude that $\zeta_n \in \mathfrak{K}(\{0,1\})$ from the implication (iii) $\Longrightarrow$ (i) in Proposition 1. □

Using 4.5/4 (iii), it is easy to determine the values of the φ-function:

n	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	...
$\varphi(n)$	2	2	4	2	*6*	4	*6*	4	*10*	4	*12*	*6*	8	8	16	...

A number in italics for $\varphi(n)$ in the second row indicates that the regular n-gon is *not* constructible. In particular, the regular 7-gon is the first polygon in this list that is not constructible; the proof of this fact goes back to Gauss. In addition, Gauss was the first to give an explicit description of the (rather complicated) construction of the regular 17-gon; note that $\varphi(17) = 16$ is a power of 2.

Furthermore, let us mention the relationship between the constructibility of the regular n-gon and the decomposition of n into so-called Fermat primes.

Definition 4. *For $\ell \in \mathbb{N}$, the integer $F_\ell = 2^{2^\ell} + 1$ is called the ℓ*th Fermat number. *A* Fermat prime *is a Fermat number that is prime, hence a prime number of type $2^{2^\ell} + 1$.*

The Fermat numbers $F_0 = 3$, $F_1 = 5$, $F_2 = 17$, $F_3 = 257$, $F_4 = 65537$ are prime and hence Fermat primes. So far, these are the only Fermat numbers that are known to be prime.

Proposition 5. *The following conditions are equivalent for $n \geq 2$:*

(i) *$\varphi(n)$ is a power of 2.*

(ii) *There exist distinct Fermat primes $p_1, \ldots, p_r$ and an integer $m \in \mathbb{N}$ such that $n = 2^m p_1 \ldots p_r$.*

Proof. Given a prime p, the expression $\varphi(p^m) = (p-1)p^{m-1}$ is a power of 2 if and only if $p = 2$ or if $p^{m-1} = 1$, i.e., $m = 1$, and $p-1$ is a power of 2. Therefore, using the multiplicativity of the φ-function, the assertion of the proposition can be derived from the following lemma:

Lemma 6. *A prime number $p \geq 3$ is a Fermat prime if and only if $p-1$ is a power of 2.*

Proof. From the definition of Fermat numbers p we see that $p-1$ is a power of 2. Conversely, assume for a prime p that $p-1$ is a power of 2, say $p = (2^{2^\ell})^r + 1$ for an odd exponent r. Then, if $r > 1$, we can factorize p according to the formula

$$1 + a^r = 1 - (-a)^r = \big(1 - (-a)\big)\big((-a)^{r-1} + (-a)^{r-2} + \ldots + 1\big)$$

into a product of nontrivial factors

$$(2^{2^\ell})^r + 1 = (2^{2^\ell} + 1)\big((2^{2^\ell})^{r-1} - \ldots + 1\big).$$

However, since p is prime, we must have $r = 1$. □

To sum up, we can state that the regular n-gon is constructible if and only if n is a product of type $n = 2^m p_1 \ldots p_r$ for distinct Fermat primes $p_1, \ldots, p_r$ and a natural number m.

Exercises

1. *Consider a subset $M \subset \mathbb{C}$ such that $0, 1 \in M$. Discuss the question whether an element $z \in \mathbb{C}$ is contained in $\mathfrak{K}(M)$ if its degree over $\mathbb{Q}(M \cup \overline{M})$ is a power of 2. In particular, assume $M = \{0, 1\}$ and look at the case that z is of degree 4 over $\mathbb{Q}$.*
2. *Look at the problem of angle trisection and check whether it admits a solution in terms of compass and straightedge constructions.*
3. Consider the extension $\mathfrak{K}(M)/\mathbb{Q}$ for $M = \{0, 1\}$ and show:
   (i) $\mathfrak{K}(M)/\mathbb{Q}$ is an infinite Galois extension.
   (ii) $\mathfrak{K}(M)$ can be interpreted as the union of an ascending chain of Galois extensions of $\mathbb{Q}$ whose degree in each case is a power of 2.
   (iii) Give a characterization of the group $\mathrm{Gal}(\mathfrak{K}(M)/\mathbb{Q})$ using the notion of projective limit; see Section 4.2.
4. Exhibit an explicit compass and straightedge construction for the regular 5-gon.

7. Transcendental Field Extensions

Background and Overview

Looking at the domain of rational numbers $\mathbb{Q}$, it was realized quite early that certain familiar "numbers," such as $\sqrt{2}$, are not rational and are hence *irrational*, as one began to say. There were several attempts to classify irrational numbers. For example, Galois theory provided a first means to access at least the algebraic ones among the irrational numbers, hence those satisfying a nontrivial algebraic equation with coefficients in $\mathbb{Q}$. Shortly thereafter it was discovered that algebraic irrational numbers make up only a "very small" part of the realm of all irrational numbers, i.e., that most of them will not satisfy a nontrivial algebraic equation with coefficients in $\mathbb{Q}$. Such numbers were called *transcendental*.

For a transcendental number over $\mathbb{Q}$, such as π, the simple field extension $\mathbb{Q}(\pi)/\mathbb{Q}$ is easy to describe. Indeed, the monomorphism $\mathbb{Q}[X] \hookrightarrow \mathbb{Q}(\pi)$, $X \longmapsto \pi$, gives rise to an isomorphism $\mathbb{Q}(X) \xrightarrow{\sim} \mathbb{Q}(\pi)$, where $\mathbb{Q}(X)$ is the function field in a variable X over $\mathbb{Q}$, thus the field of fractions of $\mathbb{Q}[X]$. But how can we describe the structure of an extension $L/\mathbb{Q}$ from an algebraic point of view for more complicated subfields $L \subset \mathbb{C}$, or even for $L = \mathbb{C}$? An amazingly simple answer to this question was given by E. Steinitz in his groundbreaking work [15]. Indeed, for an arbitrary field extension L/K, there exists a system $\mathfrak{x} = (x_i)_{i\in I}$ of elements in L such that $\mathfrak{x}$ admits the properties of a system of *variables* over K and such that L is algebraic over the "function field" $K(\mathfrak{x})$. The system $\mathfrak{x}$ is referred to as a *transcendence basis* of L/K. However, it has to be observed that the intermediate field $K(\mathfrak{x})$ of L/K depends on the choice of $\mathfrak{x}$. Steinitz showed that transcendence bases somehow behave like bases of vector spaces and in particular, that every two transcendence bases of a field extension L/K are of same cardinality. We will explain all this in more detail in Section 7.1.

The study of field extensions L/K without an algebraicity assumption is of interest not only for the extension $\mathbb{C}/\mathbb{Q}$, but also for questions in algebraic geometry. Fixing a field K with an algebraic closure $\overline{K}$, the elements of the polynomial ring $K[X_1, \ldots, X_n]$ can be viewed as (polynomial) functions on $\overline{K}^n$; cf. 3.9. Likewise, the elements of the "function" field $K(X_1, \ldots, X_n)$ give rise to rational "functions" on $\overline{K}^n$. Indeed, given $h \in K(X_1, \ldots, X_n)$, say $h = f/g$ with $f, g \in K[X_1, \ldots, X_n]$, $g \neq 0$, and looking at points $z \in \overline{K}^n$, where $g(z) \neq 0$, the

fraction $h(z) = f(z)/g(z)$ is well defined in $\overline{K}$. More generally, we can interpret every finitely generated extension field $L = K(x_1, \ldots, x_n)$ of K in a similar way as a field of rational functions. Just consider the subring $A = K[x_1, \ldots, x_n] \subset L$ and use a representation $A \simeq K[X_1, \ldots, X_n]/\mathfrak{p}$ of A as the residue class ring of a polynomial ring modulo a prime ideal $\mathfrak{p}$. As shown in 3.9, the elements of A can be viewed as polynomial functions on the zero set $V(\mathfrak{p}) \subset \overline{K}^n$ of the ideal $\mathfrak{p}$, and likewise, the elements of $L = Q(A)$ as rational functions on $V(\mathfrak{p})$. On the other hand, we will generalize in Section 7.3 the notions *separable* as well as purely inseparable, or *primary* as we prefer to say, from algebraic to more general field extensions. In particular, we will make plausible that these properties correspond to specific geometric properties of the attached algebraic sets $V(\mathfrak{p})$; see the end of Section 7.3 and its Exercise 4.

To handle *separable* and *primary* field extensions in 7.3, we must be familiar with tensor products. Such products have already been studied in 4.11 in a quite restricted setting. However, now it is necessary to develop the corresponding basics in more generality; we do this in 7.2. Finally, in Section 7.4, we give a characterization of separable field extensions in terms of methods from differential calculus.

7.1 Transcendence Bases

In 2.5/6 we introduced the notion of algebraic independence, also referred to as transcendence, for finite systems of elements of a ring R', relative to a ring extension $R \subset R'$. Restricting ourselves to fields, let us recall the definition once again.

Definition 1. *Let L/K be a field extension. A system $(x_1, \ldots, x_n)$ of elements in L is called* algebraically independent *or* transcendental *over K if every equation $f(x_1, \ldots, x_n) = 0$ for a polynomial $f \in K[X_1, \ldots, X_n]$ implies $f = 0$, i.e., if the substitution homomorphism*

$$K[X_1, \ldots, X_n] \longrightarrow L, \qquad \sum c_{\nu_1 \ldots \nu_n} X_1^{\nu_1} \ldots X_n^{\nu_n} \longmapsto \sum c_{\nu_1 \ldots \nu_n} x_1^{\nu_1} \ldots x_n^{\nu_n},$$

is injective.

A system $\mathfrak{X} = (x_i)_{i \in I}$ of (arbitrarily many) elements of L is called algebraically independent *or* transcendental *over K if every finite subsystem of $\mathfrak{X}$ is algebraically independent over K in the above sense.*

Thus, if $\mathfrak{X} = (x_i)_{i \in I}$ is a system of elements in L that is algebraically independent over K, we can view the x_i as *variables* over K. In particular, the field $K(\mathfrak{X})$ generated by $\mathfrak{X}$ over K equals the rational function field in the variables x_i, $i \in I$, which by its definition is the field of fractions of the polynomial ring $K[\mathfrak{X}]$. The extension L/K is called *purely transcendental* if $L = K(\mathfrak{X})$ for a system $\mathfrak{X}$ that is algebraically independent over K.

Definition 2. *Let L/K be a field extension and $\mathfrak{X}$ an algebraically independent system of elements in L. Then $\mathfrak{X}$ is called a* transcendence basis *of L/K if L is algebraic over $K(\mathfrak{X})$.*

Proposition 3. *Let L/K be a field extension. A system $\mathfrak{X}$ of elements in L is a transcendence basis of L/K if and only if $\mathfrak{X}$ is a maximal system in L that is algebraically independent over K. In particular, every field extension L/K admits a transcendence basis.*

Proof. First assume that $\mathfrak{X}$ is a maximal algebraically independent system of L/K. Then, by the maximality of $\mathfrak{X}$, every element of L is algebraic over $K(\mathfrak{X})$, so that $\mathfrak{X}$ is a transcendence basis of L/K. Indeed, for $x \in L$, the system obtained from $\mathfrak{X}$ by adding x is no longer algebraically independent over K. Thus, there exists a finite subsystem $(x_1, \ldots, x_n)$ of $\mathfrak{X}$, as well as a nontrivial polynomial $f \in K[X_1, \ldots, X_{n+1}]$ satisfying $f(x_1, \ldots, x_n, x) = 0$. Since the elements $x_1, \ldots, x_n$ are algebraically independent over K, we see that f, as a polynomial in X_{n+1}, is of degree > 0. However, this means that x is algebraic over $K(x_1, \ldots, x_n)$ and hence over $K(\mathfrak{X})$. In particular, L is algebraic over $K(\mathfrak{X})$, and $\mathfrak{X}$ is a transcendence basis of L/K. On the other hand, every extension L/K admits a maximal algebraically independent system by Zorn's lemma 3.4/5. Hence, it admits a transcendence basis.

Conversely, assume that $\mathfrak{X}$ is algebraically independent over K and that $L/K(\mathfrak{X})$ is algebraic. Then $\mathfrak{X}$ is necessarily a maximal algebraically independent system of L/K. □

Next we want to show that an algebraically independent system of L/K can always be enlarged to a transcendence basis by adding suitable elements from a system generating L/K. Such an "exchange" argument will be used to show that any two transcendence bases of L/K are of same cardinality.

Lemma 4. *Consider a field extension L/K and a system $\mathfrak{Y}$ of elements in L such that L is algebraic over $K(\mathfrak{Y})$. Furthermore, let $\mathfrak{X}' \subset L$ be a system that is algebraically independent over K. Then $\mathfrak{X}'$ can be enlarged by adding elements of $\mathfrak{Y}$ to yield a transcendence basis $\mathfrak{X}$ of L/K.*

In particular, there exists a subsystem in $\mathfrak{Y}$ forming a transcendence basis of L/K.

Proof. Using Zorn's lemma 3.4/5, we choose a maximal subsystem $\mathfrak{X}'' \subset \mathfrak{Y}$ such that the composite system $\mathfrak{X} = \mathfrak{X}' \cup \mathfrak{X}''$ is algebraically independent over K. Similarly as in the proof of Proposition 3, it follows that every element $y \in \mathfrak{Y}$ is algebraic over $K(\mathfrak{X})$. Then $K(\mathfrak{X}, \mathfrak{Y})$ is algebraic over $K(\mathfrak{X})$, and the same is true for $L/K(\mathfrak{X})$. In particular, $\mathfrak{X}$ is a transcendence basis of L/K. □

Theorem 5. *Every two transcendence bases of a field extension L/K have the same cardinality.*

Before turning to the proof of the theorem, let us briefly explain how to compare cardinalities of sets; we will prove two auxiliary results on this matter below. However, since it is not really necessary for our purposes, we will not touch the formal definition of cardinalities using ordinals; instead, we refer to books on set theory for this. Up to now we have simplified things by looking at the order $\operatorname{ord} M$ of a set M from a naive point of view. For example, so far, $\operatorname{ord} M = \infty$ has always indicated the fact that M consists of infinitely many elements, in the sense that it is not finite. However, when dealing with cardinalities one differentiates between different degrees of infinity. Two sets M and N are said to be of *same cardinality*, denoted by $\operatorname{card} M = \operatorname{card} N$, if there exists a bijection $M \longrightarrow N$. Alternatively, we will use the notation $\operatorname{card} M \leq \operatorname{card} N$ if there exists an injection $M \hookrightarrow N$, or equivalently for $M \neq \emptyset$, a surjection $N \longrightarrow M$. However, that a chain $\operatorname{card} M \leq \operatorname{card} N \leq \operatorname{card} M$ already implies $\operatorname{card} M = \operatorname{card} N$ is far from being obvious in the case of nonfinite cardinalities; it is the assertion of the Schröder–Bernstein theorem that we will prove below. As usual, $\operatorname{card} M = n$ (resp. $\operatorname{card} M \leq n$) for a natural number n means that M consists of precisely (resp. at most) n elements.

Lemma 6 (Schröder–Bernstein theorem). *Assume for two sets M and N that there are injections $\sigma\colon M \hookrightarrow N$ and $\tau\colon N \hookrightarrow M$. Then there is a bijection $\rho\colon M \longrightarrow N$.*

Proof. We denote by $M' \subset M$ the set of all elements $x \in M$ satisfying for every $n \in \mathbb{N}$ the implication

$$x \in (\tau \circ \sigma)^n(M) \quad \Longrightarrow \quad x \in (\tau \circ \sigma)^n \circ \tau(N).$$

In other words, an element $x \in M$ belongs to M' if and only if on taking repeated preimages of type x, $\tau^{-1}(x)$, $\sigma^{-1}\tau^{-1}(x)$, $\tau^{-1}\sigma^{-1}\tau^{-1}(x)$, ..., this yields either a chain of infinite length, or one that is finite and ends at an element in N. Hence, we can define a map $\rho\colon M \longrightarrow N$ by

$$\rho(x) = \begin{cases} \tau^{-1}(x) & \text{for } x \in M', \\ \sigma(x) & \text{for } x \notin M'. \end{cases}$$

Then ρ is injective, since the restrictions $\rho|_{M'}$ and $\rho|_{M-M'}$ are injective, and since $\rho(x) = \rho(y)$ for $x \in M'$ and $y \in M - M'$ implies $\sigma(y) = \tau^{-1}(x)$, hence $\tau \circ \sigma(y) = x \in M'$. However, this yields $y \in M'$, contradicting the choice of y. Furthermore, we claim that ρ is surjective as well. Indeed, fix $z \in N$ and look at $x = \tau(z)$. If $x \in M'$, we have $\rho(x) = \tau^{-1}(x) = z$. On the other hand, if $x \notin M'$, we conclude that z admits a preimage $y = \sigma^{-1}(z) = \sigma^{-1}(\tau^{-1}(x))$, since otherwise, x would belong to M'. Since $x \notin M'$, we get $y \notin M'$ as well, and hence $\rho(y) = \sigma(y) = z$. □

Lemma 7. *Every infinite set M is a disjoint union of sets that are countably infinite.*

Proof. Consider the set X of all pairs (A, Z), where A is an infinite subset of M, and where Z is a disjoint decomposition of A into countably infinite subsets. In other words, Z is a system of countably infinite disjoint subsets of A whose union equals A. Since M is infinite, we get $X \neq \emptyset$. Furthermore, we write $(A, Z) \leq (A', Z')$ for two elements of X if A is contained in A' and if Z is a subsystem of Z'. In this way, we have defined a partial order on X, and it is immediately clear that every totally ordered subset of X admits an upper bound in X. Therefore, X admits a maximal element $(\overline{A}, \overline{Z})$ by Zorn's lemma 3.4/5. Now observe that the difference $M - \overline{A}$ is finite, due to the maximality of $(\overline{A}, \overline{Z})$. But then, enlarging an arbitrary element of the decomposition $\overline{Z}$ by the set $M - \overline{A}$, we obtain a decomposition of M into disjoint countably infinite subsets, as desired. □

Now we are able to carry out the *proof of Theorem* 5. Let $\mathfrak{X}$ and $\mathfrak{Y}$ be two transcendence bases of L/K, where for the purposes of our proof, both systems are viewed as subsets of L. First, considering the case that $\mathfrak{X}$ is finite, say $\mathfrak{X} = \{x_1, \ldots, x_n\}$, let us show that $\operatorname{card} \mathfrak{Y} \leq \operatorname{card} \mathfrak{X}$ by induction on $n = \operatorname{card} \mathfrak{X}$. Then we get $\operatorname{card} \mathfrak{Y} = \operatorname{card} \mathfrak{X}$ by reasons of symmetry. The base case $n = 0$ is trivial, since L/K is algebraic then. Therefore assume $n > 0$. In this case, L/K is no longer algebraic, and consequently, $\mathfrak{Y}$ is nonempty. Therefore, there exists an element $y \in \mathfrak{Y}$. Applying Lemma 4, we can enlarge the system $\{y\}$ by adding elements of $\mathfrak{X}$ in order to construct a transcendence basis $\mathfrak{Z}$ of L/K. Then, necessarily, $\operatorname{card} \mathfrak{Z} \leq n$, since $\mathfrak{X}$, as a maximal algebraically independent system of L/K, cannot be contained in $\mathfrak{Z}$ together with y. Next observe that $\mathfrak{Y}$ and $\mathfrak{Z}$ both contain y. Therefore, $\mathfrak{Y} - \{y\}$ and $\mathfrak{Z} - \{y\}$ are two transcendence bases of L over $K(y)$. Since $\operatorname{card}(\mathfrak{Z} - \{y\}) < \operatorname{card} \mathfrak{Z} \leq n$, we conclude from the induction hypothesis that $\operatorname{card}(\mathfrak{Y} - \{y\}) \leq \operatorname{card}(\mathfrak{Z} - \{y\})$ and hence that $\operatorname{card} \mathfrak{Y} \leq \operatorname{card} \mathfrak{Z} \leq n = \operatorname{card} \mathfrak{X}$.

The argument we have just given shows that two transcendence bases $\mathfrak{X}$ and $\mathfrak{Y}$ of L/K are either finite, and then of equal cardinality, or both infinite. Thus, it remains to look at the case that both $\mathfrak{X}$ and $\mathfrak{Y}$ are infinite. To do this, look at any element $x \in \mathfrak{X}$. Since it is algebraic over $K(\mathfrak{Y})$, there exists a finite subset $\mathfrak{Y}_x \subset \mathfrak{Y}$ such that x is algebraic already over $K(\mathfrak{Y}_x)$. Since L cannot be algebraic over $K(\mathfrak{Y}')$ for any proper subsystem $\mathfrak{Y}' \subsetneq \mathfrak{Y}$, we get $\bigcup_{x \in \mathfrak{X}} \mathfrak{Y}_x = \mathfrak{Y}$. Therefore, we can use the inclusions $\mathfrak{Y}_x \hookrightarrow \mathfrak{Y}$ to define a surjective map $\coprod_{x \in \mathfrak{X}} \mathfrak{Y}_x \longrightarrow \mathfrak{Y}$ from the disjoint union of all $\mathfrak{Y}_x$ to $\mathfrak{Y}$. Then we conclude that $\operatorname{card} \mathfrak{Y} \leq \operatorname{card}(\coprod_{x \in \mathfrak{X}} \mathfrak{Y}_x)$ and even $\operatorname{card} \mathfrak{Y} \leq \operatorname{card} \mathfrak{X}$ if we are able to show that $\operatorname{card}(\coprod_{x \in \mathfrak{X}} \mathfrak{Y}_x) = \operatorname{card} \mathfrak{X}$. If $\mathfrak{X}$ is countably infinite, this equality can easily be verified by means of a simple counting argument. The same argument can be used in the general case, where we can apply Lemma 7 and write $\mathfrak{X}$ as a disjoint union of countably infinite subsets. After all this, we get $\operatorname{card} \mathfrak{Y} \leq \operatorname{card} \mathfrak{X}$, and by reasons of symmetry, also $\operatorname{card} \mathfrak{X} \leq \operatorname{card} \mathfrak{Y}$. This implies $\operatorname{card} \mathfrak{X} = \operatorname{card} \mathfrak{Y}$ using Lemma 6. □

The result we have just proved allows us to define for an arbitrary field extension L/K its *transcendence degree* $\operatorname{transdeg}_K L$ as the cardinality of a transcendence basis of L/K. Algebraic extensions are always of transcendence degree 0, while for a polynomial ring $K[X_1, \ldots, X_n]$ over a field K, its field of fractions $K(X_1, \ldots, X_n)$ is a purely transcendental extension of transcendence degree n over K. More generally, for an arbitrary system of variables $\mathfrak{X}$, the field of fractions $K(\mathfrak{X})$ of the polynomial ring $K[\mathfrak{X}]$ is a purely transcendental extension of transcendence degree $\operatorname{card}\mathfrak{X}$ over K. Since K-isomorphisms map transcendence bases into bases of the same type, we have the following corollary:

Corollary 8. *Two purely transcendental field extensions L/K and L'/K admit a K-isomorphism $L \xrightarrow{\sim} L'$ if and only if L and L' are of the same transcendence degree over K.*

In particular, this implies that there cannot exist a K-isomorphism between polynomial rings $K[X_1, \ldots, X_m]$ and $K[Y_1, \ldots, Y_n]$ having different numbers of variables m and n. Otherwise, the corresponding fields of fractions would be isomorphic over K, although they would have different transcendence degrees, which is impossible. We want to exploit the same argument in more detail.

Corollary 9. *Let $\varphi\colon K[X_1, \ldots, X_m] \longrightarrow K[Y_1, \ldots, Y_n]$ be a K-homomorphism between polynomial rings such that every variable $Y \in \{Y_1, \ldots, Y_n\}$ satisfies an equation of type*

$$Y^r + c_1 Y^{r-1} + \ldots + c_r = 0, \qquad c_1, \ldots, c_r \in \operatorname{im}\varphi,$$

i.e., such that φ is integral in the terminology of Section 3.3; *cf.* 3.3/4. *Then, necessarily $m \geq n$. Furthermore, φ is injective if and only if $m = n$.*

Proof. Let R be the image of φ. Being a subring of $K[Y_1, \ldots, Y_n]$, it is an integral domain, and we can view $K(Y_1, \ldots, Y_n)$ as an extension field of the field of fractions $Q(R)$. Since $K(Y_1, \ldots, Y_n) = Q(R)(Y_1, \ldots, Y_n)$ and since the elements $Y_1, \ldots, Y_n$ are algebraic over $Q(R)$ by our assumption, it follows that the extension $K(Y_1, \ldots, Y_n)/Q(R)$ is algebraic. Therefore, $Q(R)$ and $K(Y_1, \ldots, Y_n)$ are of the same transcendence degree over K, in fact of transcendence degree n. On the other hand, since the variables $X_1, \ldots, X_m$ give rise to elements $x_1, \ldots, x_m \in Q(R)$ generating the extension $Q(R)/K$, we conclude that $m \geq n$ by Lemma 4.

Furthermore, φ is injective if and only if the elements $x_1, \ldots, x_m \in Q(R)$ are algebraically independent over K, i.e., if and only if $\operatorname{transdeg}_K Q(R) = m$. Since we know already that $\operatorname{transdeg}_K Q(R) = n$, the injectivity of φ is equivalent to $m = n$, as claimed. □

Considering a chain of field extensions $K \subset L \subset M$, let us add that the transcendence degree behaves additively:

$$\operatorname{transdeg}_K M = \operatorname{transdeg}_K L + \operatorname{transdeg}_L M.$$

Indeed, fixing transcendence bases $\mathfrak{X}$ of L/K and $\mathfrak{Y}$ of M/L, it is easily verified that $\mathfrak{X} \cup \mathfrak{Y}$ is a transcendence basis of M/K. Furthermore, observe that the sum of the cardinalities of $\mathfrak{X}$ and $\mathfrak{Y}$ is by definition the cardinality of the (disjoint) union $\mathfrak{X} \cup \mathfrak{Y}$.

Finally, let us show for a purely transcendental extension L/K that the algebraic closure of K in L coincides with K, in other words, that K is algebraically closed in L.

Remark 10. *Let L/K be a purely transcendental field extension. Then every element $x \in L - K$ is transcendental over K.*

Proof. Consider an element $x \in L$ that is algebraic over K, and choose a transcendence basis $\mathfrak{X}$ of L/K such that $L = K(\mathfrak{X})$. Then there is a finite subsystem $(x_1, \ldots, x_r)$ of $\mathfrak{X}$ such that $x \in K(x_1, \ldots, x_r)$. Therefore, to show that $x \in K$, it is enough to consider the case that $\mathfrak{X} = (x_1, \ldots, x_r)$ and hence is finite. Now let

$$f = X^n + c_1 X^{n-1} + \ldots + c_n \in K[X]$$

be the minimal polynomial of $x \in L = K(\mathfrak{X})$ over K, where we may assume $x \neq 0$ and hence $c_n \neq 0$. Interpreting $K[\mathfrak{X}]$ as a polynomial ring in the variables $x_1, \ldots, x_r$, we conclude from 2.7/3 that $K[\mathfrak{X}]$ is a unique factorization domain. Therefore, we can write x as a reduced fraction, say $x = g/h$, where the polynomials $g, h \in K[\mathfrak{X}]$ are relatively prime. Then the equation $f(x) = 0$ yields

$$g^n + c_1 g^{n-1} h + \ldots + c_n h^n = 0,$$

and we see that every prime element $q \in K[\mathfrak{X}]$ dividing h will divide g as well. However, this implies that h is a unit in $K[\mathfrak{X}]$ and hence that $h \in K^*$. In the same way, we conclude that $g \in K^*$, thus showing that $x \in K$. □

Exercises

1. *Compare the notion of a transcendence basis of a field extension L/K with the notion of a basis of a vector space.*
2. *Show that there exist automorphisms of $\mathbb{C}$ that do not map $\mathbb{R}$ to itself, and furthermore, that $\mathbb{C}$ contains proper subfields that are isomorphic to $\mathbb{C}$.*
3. Show that the transcendence degree of $\mathbb{R}/\mathbb{Q}$ equals the cardinality of $\mathbb{R}$.
4. Show that every field of characteristic 0 is a union of subfields that are isomorphic to subfields of $\mathbb{C}$.
5. Let L/K be a field extension and $\mathfrak{X}$ an algebraically independent system of L/K. Show for every intermediate field K' of L/K that is algebraic over K that $\mathfrak{X}$ is algebraically independent over K'.
6. Let L/K be a finitely generated field extension. Show for every intermediate field L' of L/K that the extension L'/K is finitely generated.

7.2 Tensor Products*

In Section 4.11 we introduced a simplified version of tensor products in order to put the process of coefficient extension for vector spaces onto a solid basis. Now we want to study tensor products from a more general viewpoint, having in mind applications to separable and primary field extensions in Section 7.3. We start with the tensor product of modules over a ring; see Section 2.9 for the definition of modules.

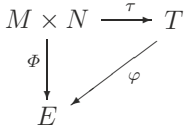
In the following we fix two modules M, N over a ring R. If E is yet another R-module, a map $\Phi\colon M \times N \longrightarrow E$ is called *R-bilinear*, as usual, if for all $x \in M$, $y \in N$ the maps

$$\begin{aligned} \Phi(x,\cdot)\colon N &\longrightarrow E, \qquad z \longmapsto \Phi(x,z), \\ \Phi(\cdot,y)\colon M &\longrightarrow E, \qquad z \longmapsto \Phi(z,y), \end{aligned}$$

are *R-linear*, i.e., are homomorphisms of R-modules. The main point of a tensor product of M and N over R is the fact that it is an R-module T such that the R-bilinear maps of $M \times N$ to an arbitrary R-module E correspond bijectively to the R-linear maps $T \longrightarrow E$. More precisely:

Definition 1. *A* tensor product *of two R-modules M and N over a ring R is given by an R-module T together with an R-bilinear map $\tau\colon M \times N \longrightarrow T$, admitting the following universal property*:

For every R-bilinear map $\Phi\colon M \times N \longrightarrow E$ to an R-module E, there is a unique R-linear map $\varphi\colon T \longrightarrow E$ such that $\Phi = \varphi \circ \tau$, in other words, such that the diagram



is commutative.

By their defining universal property, tensor products are unique, up to canonical isomorphism. They always exist, as we will immediately see below. Furthermore, observe that one usually writes $M \otimes_R N$ instead of T in the situation of Definition 1. Also, given elements $x \in M$ and $y \in N$, it is common practice to write the image of (x, y) under the R-bilinear map $\tau\colon M \times N \longrightarrow T$ in product notation as $x \otimes y$; elements of type $x \otimes y$ are called *tensors* in $M \otimes_R N$. Using such a notation, the R-bilinear map τ is characterized by

$$M \times N \longrightarrow M \otimes_R N, \qquad (x,y) \longmapsto x \otimes y.$$

In particular, tensors are R-bilinear in their factors, which means that

$$\begin{aligned} &(ax + a'x') \otimes (by + b'y') \\ &= ab(x \otimes y) + ab'(x \otimes y') + a'b(x' \otimes y) + a'b'(x' \otimes y') \end{aligned}$$

for $a, a', b, b' \in R,\ x, x' \in M,\ y, y' \in N$. In many cases the defining R-bilinear map $\tau\colon M \times N \longrightarrow M \otimes_R N$ is not mentioned explicitly. Then one talks about $M \otimes_R N$ as being the tensor product of M and N over R, assuming that the tensors $x \otimes y$ in $M \otimes_R N$ are known. Indeed, knowing all tensors in $M \otimes_R N$, the map τ can be reconstructed.

Proposition 2. *The tensor product $T = M \otimes_R N$ exists for arbitrary R-modules M and N.*

Proof. The basic construction idea is quite simple. We start with $R^{(M\times N)}$, the free R-module generated by all pairs $(x, y) \in M \times N$, and divide out the smallest submodule Q such that the residue classes of elements of type (x, y) acquire the properties of tensors.[1] This means that we consider the submodule $Q \subset R^{(M\times N)}$ that is generated by all elements

$$\begin{gathered}(x + x', y) - (x, y) - (x', y),\\ (x, y + y') - (x, y) - (x, y'),\\ (ax, y) - a(x, y),\\ (x, ay) - a(x, y),\end{gathered}$$

where $a \in R$, $x, x' \in M$, $y, y' \in N$. Then, writing $T = R^{(M\times N)}/Q$, the canonical map $\tau\colon M \times N \longrightarrow T$ assigning to a pair (x, y) the residue class of (x, y) in T is R-bilinear. We want to show that τ satisfies the universal property of a tensor product as stated in Definition 1. Therefore, consider an R-bilinear map $\Phi\colon M \times N \longrightarrow E$ to some R-module E. It gives rise canonically to an R-linear map $\hat{\varphi}\colon R^{(M\times N)} \longrightarrow E$, by setting $\hat{\varphi}(x, y) = \Phi(x, y)$ for basis elements of type $(x, y) \in R^{(M\times N)}$ and using R-linear extension. Then one concludes from the R-bilinearity of Φ that $\ker\hat{\varphi}$ contains all generating elements of Q as listed above, hence that $\hat{\varphi}$ induces an R-linear map $\varphi\colon R^{(M\times N)}/Q \longrightarrow E$ satisfying $\Phi = \varphi\circ\tau$. The map φ is uniquely determined by the relation $\Phi = \varphi\circ\tau$, since the residue classes $\overline{(x, y)}$ of the basis elements $(x, y) \in R^{(M\times N)}$ generate $R^{(M\times N)}/Q$ as an R-module and since

$$\varphi\big(\overline{(x, y)}\big) = \varphi\big(\tau(x, y)\big) = \Phi(x, y).$$

In particular, φ is unique on a generating system of $T = R^{(M\times N)}/Q$ and hence unique on all of T. □

Working with tensor products, their explicit construction as given in the proof of Proposition 2 is of only minor interest. In almost all cases it is more efficient and more appropriate to derive the desired properties from the defining universal property of tensor products. For example, we can read from the construction of $M \otimes_R N$ that each element $z \in M \otimes_R N$ can be written as a

[1] Here (x, y) indicates the element $(r_{m,n})_{m\in M, n\in N}$ in $R^{(M\times N)}$ that in terms of Kronecker's symbol is given by $r_{m,n} = \delta_{m,x}\delta_{n,y}$.

finite sum of tensors, say $z = \sum_{i=1}^{n} x_i \otimes y_i$. However, this fact is easily derived from the universal property of $M \otimes_R N$ as well, since the submodule of $M \otimes_R N$ that is generated by all tensors in $M \otimes_R N$ satisfies the universal property of a tensor product of M and N over R, just as does $M \otimes_R N$ itself. Concerning the notation of tensors, a bit of care is recommended. Indeed, for any tensor $x \otimes y$ its ambient tensor product $M \otimes_R N$ where this tensor is built has to be specified, unless this is clear from the context. The reason is that for a submodule $M' \subset M$ the tensor product $M' \otimes_R N$ is not necessarily a submodule of $M \otimes_R N$. In general, there can exist nonzero tensors $x \otimes y$ in $M' \otimes_R N$ that vanish as tensors in $M \otimes_R N$. For example, consider the tensor $2 \otimes 1$ in $(2\mathbb{Z}) \otimes_{\mathbb{Z}} (\mathbb{Z}/2\mathbb{Z})$, as well as in $\mathbb{Z} \otimes_{\mathbb{Z}} (\mathbb{Z}/2\mathbb{Z})$; we will have a closer look at this example later on.

In many cases it is convenient to describe an R-linear map $M \otimes_R N \longrightarrow E$ from a tensor product to an R-module E by specifying the images in E of all tensors $x \otimes y \in M \otimes_R N$, since as we know, these generate $M \otimes_R N$ over R. However, when proceeding like this, the images of the tensors of $M \otimes_R N$ cannot be prescribed arbitrarily, since they must respect the rules of R-bilinearity. Given a family $(z_{x,y})_{x \in M, y \in N}$ of elements in E, there exists an R-linear map $M \otimes_R N \longrightarrow E$ with $x \otimes y \longmapsto z_{x,y}$ precisely in those cases in which $(x, y) \longmapsto z_{x,y}$ defines an R-bilinear map $M \times N \longrightarrow E$.

Remark 3. *For R-modules M, N, P, there exist canonical isomorphisms*

$$\begin{aligned} R \otimes_R M &\xrightarrow{\sim} M, & a \otimes x &\longmapsto ax, \\ M \otimes_R N &\xrightarrow{\sim} N \otimes_R M, & x \otimes y &\longmapsto y \otimes x, \\ (M \otimes_R N) \otimes_R P &\xrightarrow{\sim} M \otimes_R (N \otimes_R P), & (x \otimes y) \otimes z &\longmapsto x \otimes (y \otimes z), \end{aligned}$$

which are uniquely characterized by the stated mapping rules on tensors.

Proof. In all three cases, one proceeds in a similar way. First one shows that the mapping rule given on tensors leads to a well-defined R-linear map, and then one constructs a natural inverse of this map. As an example, let us look at the first isomorphism. Since the map $R \times M \longrightarrow M$, $(a, x) \longmapsto ax$, is R-bilinear, it gives rise to a well-defined R-linear map $\varphi\colon R \otimes_R M \longrightarrow M$, $a \otimes x \longmapsto ax$. To construct an inverse of it, look at the R-linear map $\psi\colon M \longrightarrow R \otimes_R M$, $x \longmapsto 1 \otimes x$. It satisfies $\varphi \circ \psi(x) = x$ for all $x \in M$, and in addition, $\psi \circ \varphi(a \otimes x) = \psi(ax) = 1 \otimes ax = a \otimes x$ for all tensors $a \otimes x$ in $R \otimes_R M$. This implies $\varphi \circ \psi = \mathrm{id}$ and $\psi \circ \varphi = \mathrm{id}$. Therefore, φ is an isomorphism, and $\varphi^{-1} = \psi$. □

In a similar way one proves the following:

Remark 4. *Let $(M_i)_{i \in I}$ be a family of R-modules, and N another R-module. Then there is a canonical isomorphism*

$$\Big(\bigoplus_{i \in I} M_i\Big) \otimes_R N \xrightarrow{\sim} \bigoplus_{i \in I} (M_i \otimes_R N), \qquad (x_i)_{i \in I} \otimes y \longmapsto (x_i \otimes y)_{i \in I},$$

which is uniquely characterized by the stated mapping rule on tensors. In particular, tensor products are compatible with direct sums.

Next we define for two given R-linear maps $\varphi\colon M \longrightarrow M'$ and $\psi\colon N \longrightarrow N'$ their tensor product $\varphi \otimes \psi\colon M \otimes_R N \longrightarrow M' \otimes_R N'$ by $x \otimes y \longmapsto \varphi(x) \otimes \psi(y)$. This is permissible, since $(x, y) \longmapsto \varphi(x) \otimes \psi(y)$ gives rise to an R-bilinear map $M \times N \longrightarrow M' \otimes_R N'$. In particular, we can consider the tensor product $\varphi \otimes \mathrm{id}\colon M \otimes_R N \longrightarrow M' \otimes_R N$ of φ with the identical map on N; for simplicity, it is said that one constructs the tensor product of the map $\varphi\colon M \longrightarrow M'$ with the R-module N. To study the behavior of R-linear maps when taking tensor products with an R-module N, we use the notion of *exact sequence*. Hereby one understands a sequence of R-linear maps

$$M_1 \xrightarrow{\varphi_1} M_2 \xrightarrow{\varphi_2} \ldots \xrightarrow{\varphi_{r-1}} M_r$$

such that $\operatorname{im}\varphi_i = \ker\varphi_{i+1}$ for $i = 1, \ldots, r-2$.

Proposition 5. *Let*

$$M' \xrightarrow{\varphi} M \xrightarrow{\psi} M'' \longrightarrow 0$$

be an exact sequence of R-modules. Then, for every R-module N, the sequence

$$M' \otimes_R N \xrightarrow{\varphi\otimes\mathrm{id}} M \otimes_R N \xrightarrow{\psi\otimes\mathrm{id}} M'' \otimes_R N \longrightarrow 0$$

is exact.

Proof. First observe that $(\psi \otimes \mathrm{id}) \circ (\varphi \otimes \mathrm{id}) = (\psi \circ \varphi) \otimes \mathrm{id} = 0$, since $\psi \circ \varphi = 0$, which shows that $\operatorname{im}(\varphi \otimes \mathrm{id}) \subset \ker(\psi \otimes \mathrm{id})$. Therefore, $\psi \otimes \mathrm{id}$ gives rise to an R-linear map

$$\overline{\psi}\colon (M \otimes_R N)\big/\operatorname{im}(\varphi \otimes \mathrm{id}) \longrightarrow M'' \otimes_R N,$$

and it is enough to show that $\overline{\psi}$ is an isomorphism. To construct an inverse of $\overline{\psi}$ we use the surjectivity of ψ and choose for every element $x'' \in M''$ an element $\iota(x'') \in M$ such that $\psi(\iota(x'')) = x''$; the resulting map $\iota\colon M'' \longrightarrow M$ is meant as a map between *sets* only. Now consider the map

$$\sigma\colon M'' \times N \longrightarrow (M \otimes_R N)\big/\operatorname{im}(\varphi \otimes \mathrm{id}), \qquad (x'', y) \longmapsto \overline{\iota(x'') \otimes y},$$

where $\overline{\iota(x'') \otimes y}$ indicates the residue class of $\iota(x'') \otimes y$ in $(M \otimes_R N)/\operatorname{im}(\varphi \otimes \mathrm{id})$. We claim that σ is an R-bilinear map, the linearity in the second argument being clear. To justify the linearity in the first argument, it is enough to show that the element $\overline{\iota(x'') \otimes y}$ is independent of the choice of the preimage $\iota(x'') \in M$ of $x'' \in M''$. To check such an independence, consider two preimages $x_1, x_2 \in M$ of x''. Then we have $x_1 - x_2 \in \operatorname{im}\varphi$, say $x_1 - x_2 = \varphi(x')$, since the sequence $M' \to M \to M'' \to 0$ is exact. However, this implies

$$\overline{x_1 \otimes y} - \overline{x_2 \otimes y} = \overline{\varphi(x') \otimes y} = \overline{(\varphi \otimes \mathrm{id})(x' \otimes y)} = 0,$$

as claimed. Therefore, the above map σ is R-bilinear, and it is seen that the induced R-linear map $M'' \otimes_R N \longrightarrow (M \otimes_R N)/\operatorname{im}(\varphi \otimes \mathrm{id})$ is an inverse of $\overline{\psi}$. □

Given R-modules M, N, and a submodule $M' \subset M$, we can use Proposition 5 to make tensor products of type $(M/M') \otimes_R N$ more explicit. From the canonical exact sequence

$$M' \xrightarrow{\varphi} M \xrightarrow{\psi} M/M' \longrightarrow 0$$

we get the exact sequence

$$M' \otimes_R N \xrightarrow{\varphi\otimes\mathrm{id}} M \otimes_R N \xrightarrow{\psi\otimes\mathrm{id}} (M/M') \otimes_R N \longrightarrow 0,$$

which yields an isomorphism

$$(M/M') \otimes_R N \xrightarrow{\sim} (M \otimes_R N)/\operatorname{im}(\varphi \otimes \mathrm{id}), \qquad \overline{x} \otimes y \longmapsto \overline{x \otimes y}.$$

However, it should be observed that even if φ is injective, this does not necessarily imply injectivity for the tensor product map $\varphi \otimes \mathrm{id}$. As a consequence, we cannot generally use $\varphi \otimes \mathrm{id}$ to view $M' \otimes_R N$ as a submodule of $M \otimes_R N$. For example, look at the inclusion $2\mathbb{Z} \hookrightarrow \mathbb{Z}$ and consider its tensor product with $\mathbb{Z}/2\mathbb{Z}$ over $\mathbb{Z}$, i.e., the induced map $2\mathbb{Z} \otimes_{\mathbb{Z}} (\mathbb{Z}/2\mathbb{Z}) \longrightarrow \mathbb{Z} \otimes_{\mathbb{Z}} (\mathbb{Z}/2\mathbb{Z})$. This is zero, since in $\mathbb{Z} \otimes_{\mathbb{Z}} (\mathbb{Z}/2\mathbb{Z})$, all tensors of type $2a \otimes \overline{b}$ can be written as $a \otimes 2\overline{b}$ and hence are zero. On the other hand, $2\mathbb{Z} \otimes_{\mathbb{Z}} (\mathbb{Z}/2\mathbb{Z}) \simeq \mathbb{Z} \otimes_{\mathbb{Z}} (\mathbb{Z}/2\mathbb{Z}) \simeq \mathbb{Z}/2\mathbb{Z}$ is nonzero.

An R-module N is called *flat* if for every injection of R-modules $M' \hookrightarrow M$ the tensor product map $M' \otimes_R N \longrightarrow M \otimes_R N$ is injective as well. This is equivalent to the condition that exact sequences of type $0 \to M' \to M \to M'' \to 0$ remain exact under taking the tensor product with N over R. For example, we can conclude the following from Remarks 3 and 4:

Remark 6. *Free R-modules are flat. In particular, every vector space over a field is flat.*

Next we want to explain the process of coefficient extension for R-modules. Let $f\colon R \longrightarrow R'$ be a ring homomorphism and view R' as an R-module with respect to f. Then we can construct for any R-module M the tensor product $M \otimes_R R'$, which, by its definition, is an R-module. We claim that the R-module structure extends canonically to an R'-module structure. Indeed, define the product of an element $a \in R'$ with a tensor $(x \otimes b) \in M \otimes_R R'$ by $x \otimes (ab)$. To see that we get a well-defined multiplication by a on $M \otimes_R R'$, observe that the map

$$M \times R' \longrightarrow M \otimes_R R', \qquad (x, b) \longmapsto x \otimes (ab),$$

is R-bilinear and hence gives rise to an R-linear map

$$M \otimes_R R' \longrightarrow M \otimes_R R', \qquad x \otimes b \longmapsto x \otimes (ab).$$

Furthermore, using the calculation rules of tensors, it is immediately clear that this multiplication satisfies the conditions necessary for an R'-module structure on $M \otimes_R R'$. We say that the R'-module $M \otimes_R R'$ is obtained from M by *coefficient extension*. Also it is easily seen that this definition of coefficient extension coincides with the one given in Section 4.11, where we restricted ourselves to vector spaces over fields; see also Exercise 1 below.

Remark 7. *Let $R \longrightarrow R' \longrightarrow R''$ be ring homomorphisms, and M an R-module. Then there is a canonical isomorphism of R''-modules*

$$(M \otimes_R R') \otimes_{R'} R'' \xrightarrow{\sim} M \otimes_R R'', \qquad (x \otimes a') \otimes a'' \longmapsto x \otimes (a'a''),$$

which is uniquely characterized by the stated mapping rule on tensors.

Proof. Viewing $R' \longrightarrow R''$ as an R-linear map, it induces an R'-linear map $\sigma\colon M \otimes_R R' \longrightarrow M \otimes_R R''$ by taking the tensor product with M. Furthermore,

$$(M \otimes_R R') \times R'' \longrightarrow M \otimes_R R'', \qquad (x, a'') \longmapsto a'' \cdot \sigma(x),$$

is a well-defined R'-bilinear map and therefore gives rise to an R'-linear map $(M \otimes_R R') \otimes_{R'} R'' \longrightarrow M \otimes_R R''$, the one considered in the assertion. This map is R''-linear, as is easily checked on tensors. To construct an inverse of it, look at the R-bilinear map

$$M \times R'' \longrightarrow (M \otimes_R R') \otimes_{R'} R'', \qquad (x, a'') \longmapsto (x \otimes 1) \otimes a'',$$

as well as at the corresponding map $M \otimes_R R'' \longrightarrow (M \otimes_R R') \otimes_{R'} R''$. □

We will use the process of coefficient extension especially for ring homomorphisms of type $R \longrightarrow R_S$, where $S \subset R$ is a multiplicative system; here R_S is the localization of R by S, i.e., $R_S = S^{-1}R$ in the notation of Section 2.7. Therefore, given an R-module M, we can always consider $M \otimes_R R_S$ as an R_S-module. On the other hand, we can just as well construct from M an R_S-module by applying the process of localization to modules instead of rings. Indeed, consider the set of all fractions $\frac{x}{s}$ for $x \in M$, $s \in S$, and identify $\frac{x}{s}$ in each case with another fraction $\frac{x'}{s'}$ if there exists an element $s'' \in S$ such that $s''(s'x - sx') = 0$. Then, using the usual rules of fractional arithmetic, the resulting set is an R_S-module. It is denoted by M_S.

Proposition 8. *Let $S \subset R$ be a multiplicative system.*

(i) *The canonical map $R \longrightarrow R_S$ is flat, i.e., R_S is a flat R-module under this map.*

(ii) *For every R-module M, there is a canonical isomorphism of R-modules, resp. R_S-modules,*

$$M \otimes_R R_S \xrightarrow{\sim} M_S, \qquad x \otimes \frac{a}{s} \longmapsto \frac{ax}{s},$$

which is uniquely characterized by the stated mapping rule on tensors.

Proof. We start with assertion (ii). The map

$$M \times R_S \longrightarrow M_S, \qquad \left(x, \frac{a}{s}\right) \longmapsto \frac{ax}{s},$$

is well defined, as is easily checked, and R-bilinear. Hence, it gives rise to a unique R-linear map $\varphi\colon M \otimes_R R_S \longrightarrow M_S$ with mapping rule as stated in (ii). In addition, the mapping rule shows that φ is even R_S-linear. On the other hand, it is easily checked that

$$\psi\colon M_S \longrightarrow M \otimes_R R_S, \qquad \frac{x}{s} \longmapsto x \otimes \frac{1}{s},$$

is a well-defined R-linear map that is an inverse of φ. Consequently, φ is an isomorphism.

Now assertion (i) is easy to obtain. Let $\sigma\colon M' \longrightarrow M$ be an injection of R-modules. Then, applying (ii), it is enough to show that the natural map $\sigma_S\colon M'_S \longrightarrow M_S$, $\frac{x}{s} \longmapsto \frac{\sigma(x)}{s}$, induced from σ, is injective. Therefore, look at an element $\frac{x}{s}$ in M'_S whose image in M_S is zero. By the definition of M_S, there is an element $s'' \in S$ such that $\sigma(s''x) = s''\sigma(x) = 0$. However, the injectivity of σ implies $s''x = 0$ and therefore $\frac{x}{s} = 0$, i.e., σ_S is injective. □

Finally, fixing two ring homomorphisms $f\colon R \longrightarrow R'$ and $g\colon R \longrightarrow R''$, let us look at the tensor product $R' \otimes_R R''$. In order to avoid mentioning the homomorphisms f and g explicitly, we prefer to view R' and R'' as *R-algebras*; see Section 3.3. Observe that $R' \otimes_R R''$ is from the left an R'-module and from the right an R''-module. Furthermore, we want to show that $R' \otimes_R R''$ is even an R-algebra. For this we introduce on $R' \otimes_R R''$ a ring multiplication via the rule

$$(a \otimes b) \cdot (c \otimes d) = (ac) \otimes (bd).$$

To check that we get a well-defined multiplication on $R' \otimes_R R''$, consider for an arbitrary element $z = \sum_{i=1}^{r} c_i \otimes d_i \in R' \otimes_R R''$ the map

$$R' \times R'' \longrightarrow R' \otimes_R R'', \qquad (a, b) \longmapsto a \cdot z \cdot b = \sum_{i=1}^{r} (ac_i) \otimes (bd_i).$$

This map is well defined and R-bilinear, since we know already that $R' \otimes_R R''$ is an R'-module, as well as an R''-module. Therefore, we obtain the "multiplication" by z as an R-linear map

$$R' \otimes_R R'' \longrightarrow R' \otimes_R R'', \qquad a \otimes b \longmapsto a \cdot z \cdot b.$$

Varying z over $R' \otimes_R R''$, we get a map

$$(R' \otimes_R R'') \times (R' \otimes_R R'') \longrightarrow R' \otimes_R R'',$$

which is characterized by the rule

$$(a \otimes b, c \otimes d) \longmapsto (ac) \otimes (bd).$$

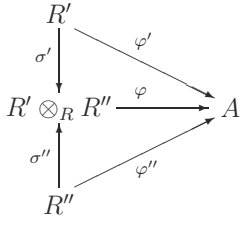
In addition, it admits the properties of a ring multiplication, due to the bilinearity property of tensors. Then, to equip the tensor product $R' \otimes_R R''$ with the structure of an R-algebra, we use the ring homomorphism $R \longrightarrow R' \otimes_R R''$ given by $a \longmapsto (a \cdot 1) \otimes 1 = 1 \otimes (a \cdot 1)$.

The tensor product $R' \otimes_R R''$ of two R-algebras R' and R'' comes equipped with two canonical R-algebra homomorphisms:

$$\begin{aligned} \sigma' \colon R' &\longrightarrow R' \otimes_R R'', \qquad a' \longmapsto a' \otimes 1, \\ \sigma'' \colon R'' &\longrightarrow R' \otimes_R R'', \qquad a'' \longmapsto 1 \otimes a''. \end{aligned}$$

Together, they provide a unique characterization of the tensor product $R' \otimes_R R''$ as an R-algebra, as we will see in the next lemma.

Lemma 9. *The above maps* $\sigma' \colon R' \longrightarrow R' \otimes_R R''$, $\sigma'' \colon R'' \longrightarrow R' \otimes_R R''$ *admit the following universal property: Given two* R*-algebra homomorphisms* $\varphi' \colon R' \longrightarrow A$ *and* $\varphi'' \colon R'' \longrightarrow A$ *to an* R*-algebra* A*, there is a unique* R*-algebra homomorphism* $\varphi \colon R' \otimes_R R'' \longrightarrow A$ *such that the diagram*



is commutative. Furthermore, φ *is characterized by* $a' \otimes a'' \longmapsto \varphi'(a') \cdot \varphi''(a'')$.

If the homomorphisms φ', φ'' *admit the same universal property as do* σ', σ''*, then* φ *is an isomorphism. In particular, the tensor product* $R' \otimes_R R''$ *as an* R*-algebra is uniquely characterized by the above universal property.*

Proof. To show that φ is unique, look at a tensor $a' \otimes a'' \in R' \otimes_R R''$. Then we have

$$\varphi(a' \otimes a'') = \varphi\big((a' \otimes 1) \cdot (1 \otimes a'')\big) = \varphi(a' \otimes 1) \cdot \varphi(1 \otimes a'') = \varphi'(a') \cdot \varphi''(a'').$$

Therefore, φ is unique on all tensors in $R' \otimes_R R''$ and hence on all of $R' \otimes_R R''$. On the other hand, we can consider the map

$$R' \times R'' \longrightarrow A, \qquad (a', a'') \longmapsto \varphi'(a') \cdot \varphi''(a'').$$

It is R-bilinear and hence induces an R-linear map $\varphi \colon R' \otimes_R R'' \longrightarrow A$. That φ is an R-algebra homomorphism with the required properties can readily be checked. □

Proposition 10. *Let R' be an R-algebra and $\mathfrak{X}$ a system of variables, as well as $\mathfrak{a} \subset R[\mathfrak{X}]$ an ideal. Then there are canonical isomorphisms*

$$R[\mathfrak{X}] \otimes_R R' \xrightarrow{\sim} R'[\mathfrak{X}], \qquad f \otimes a' \longmapsto a'f,$$
$$\big(R[\mathfrak{X}]/\mathfrak{a}\big) \otimes_R R' \xrightarrow{\sim} R'[\mathfrak{X}]/\mathfrak{a}R'[\mathfrak{X}], \qquad \overline{f} \otimes a' \longmapsto \overline{a'f},$$

which are uniquely characterized by the stated mapping rules on tensors.

Proof. The canonical R-algebra homomorphisms $\varphi' \colon R[\mathfrak{X}] \longrightarrow R'[\mathfrak{X}]$ and $\varphi'' \colon R' \longrightarrow R'[\mathfrak{X}]$ give rise, due to Lemma 9, to an R-algebra homomorphism

$$\varphi \colon R[\mathfrak{X}] \otimes_R R' \longrightarrow R'[\mathfrak{X}], \qquad f \otimes a' \longmapsto a'f.$$

On the other hand, $R' \longrightarrow R[\mathfrak{X}] \otimes_R R'$, $a' \longmapsto 1 \otimes a'$, can be extended via $\mathfrak{X} \longmapsto \mathfrak{X} \otimes 1$ to yield a ring homomorphism $\psi \colon R'[\mathfrak{X}] \longrightarrow R[\mathfrak{X}] \otimes_R R'$; cf. 2.5/5. One can check without problems that ψ and φ are inverse to each other. Thus, φ is an isomorphism, as claimed. The second isomorphism mentioned in the assertion follows from the first one with the help of Proposition 5. □

To end the present section, we want to apply the preceding results on tensor products of algebras to the special case of field extensions. In more precise terms, we want to determine the tensor product $L \otimes_K K'$ for certain field extensions L/K and K'/K. First of all, let us point out that $L \otimes_K K'$ is a nonzero K-algebra containing L and K' as subalgebras. Indeed, the canonical maps

$$L \simeq L \otimes_K K \longrightarrow L \otimes_K K', \qquad K' \simeq K \otimes_K K' \longrightarrow L \otimes_K K'$$

are injective, due to the flatness of L/K and K'/K.

Remark 11. *Let K'/K be a field extension and $f \in K[X]$ a polynomial in one variable X. Then*

$$\big(K[X]/fK[X]\big) \otimes_K K' \simeq K'[X]/fK'[X].$$

Furthermore, if $f = p_1^{\nu_1} \dots p_r^{\nu_r}$ is a factorization in $K'[X]$ with pairwise nonassociated prime polynomials $p_i \in K'[X]$, we get

$$\big(K[X]/fK[X]\big) \otimes_K K' \simeq \prod_{i=1}^{r} K'[X]/p_i^{\nu_i}K'[X].$$

Proof. Use Proposition 10 in conjunction with the Chinese remainder theorem 2.4/14. □

If L/K is a simple algebraic field extension, say $L = K(a)$ with minimal polynomial $f \in K[X]$ of a, and if K'/K is any field extension, then we see in the setting of Remark 11 that

$$K(a) \otimes_K K' \simeq K'[X]/fK'[X] \simeq \prod_{i=1}^{r} K'[X]/p_i^{\nu_i} K'[X]$$

is a field if and only if f is irreducible over K'. However, in general $K(a) \otimes_K K'$ will contain nontrivial zero divisors and even nontrivial nilpotent elements, i.e., elements $z \neq 0$ such that there is an exponent $n \in \mathbb{N}$ satisfying $z^n = 0$. Indeed, $K(a) \otimes_K K'$ contains such nilpotent elements if and only if at least one of the above exponents ν_i is greater than 1. In particular, f cannot be separable then.

Remark 12. *Let L/K be a purely transcendental field extension and $\mathfrak{X}$ a transcendence basis generating L over K, i.e., such that $L = K(\mathfrak{X})$. Then, for every field extension K'/K, there are canonical homomorphisms*

$$L \otimes_K K' \xrightarrow{\sim} K'[\mathfrak{X}]_S \hookrightarrow K'(\mathfrak{X}),$$

where $K'[\mathfrak{X}]_S$ is the localization of the polynomial ring $K'[\mathfrak{X}]$ by the multiplicative system $S = K[\mathfrak{X}] - \{0\}$. In particular, $L \otimes_K K'$ is an integral domain.

Proof. We view L as the localization $K[\mathfrak{X}]_S$ of the polynomial ring $K[\mathfrak{X}]$ by the multiplicative system $S = K[\mathfrak{X}] - \{0\}$. Then Proposition 10 yields

$$K[\mathfrak{X}] \otimes_K K' \simeq K'[\mathfrak{X}]$$

and, in conjunction with Remark 7 and Proposition 8,

$$\begin{aligned} L \otimes_K K' &\simeq K[\mathfrak{X}]_S \otimes_K K' \\ &\simeq K[\mathfrak{X}]_S \otimes_{K[\mathfrak{X}]} \left(K[\mathfrak{X}] \otimes_K K'\right) \\ &\simeq K[\mathfrak{X}]_S \otimes_{K[\mathfrak{X}]} K'[\mathfrak{X}] \\ &\simeq K'[\mathfrak{X}]_S. \end{aligned}$$

Therefore, $L \otimes_K K'$ is a subring of the field of fractions of $K'[\mathfrak{X}]$, and as such, an integral domain. □

We want to explain another property that quite often allows one to reduce problems on tensor products to the case in which certain finiteness conditions are given. It concerns the compatibility of tensor products with direct limits, see Exercise 8 below, a property that we formulate for simplicity only in a special case at this place.

Lemma 13. *Let A and A' be algebras over a field K and observe for subalgebras $A_0 \subset A$ and $A'_0 \subset A'$ that $A_0 \otimes_K A'_0$ is canonically a subalgebra in $A \otimes_K A'$. Furthermore, let $(A_i)_{i\in I}$, $(A'_j)_{j\in J}$ be directed systems of subalgebras in A, resp. A', such that $A = \bigcup_{i\in I} A_i$ and $A' = \bigcup_{j\in J} A'_j$. Then $(A_i \otimes_K A'_j)_{i\in I, j\in J}$ is a directed system of subalgebras in $A \otimes_K A'$ such that*

$$A \otimes_K A' = \bigcup_{i\in I, j\in J} (A_i \otimes_K A'_j).$$

(*Recall that the system* $(A_i)_{i\in I}$ *is called* directed *if for* $i, i' \in I$ *there is always an index* $k \in I$ *such that* $A_i \cup A_{i'} \subset A_k$, *likewise for all remaining systems*; *see also* 4.2.)

Proof. By the flatness of K-algebras, the inclusions $A_0 \hookrightarrow A$ and $A'_0 \hookrightarrow A'$ give rise to injections

$$A_0 \otimes_K A'_0 \hookrightarrow A_0 \otimes_K A' \hookrightarrow A \otimes_K A'.$$

In particular, the tensor products $A_i \otimes_K A'_j$ can canonically be viewed as subalgebras of $A \otimes_K A'$, for all $i \in I$, $j \in J$. Now consider an element $z \in A \otimes_K A'$ and write it as a finite sum $z = \sum_{\rho=1}^{r} x_\rho \otimes y_\rho$, where $x_\rho \in \bigcup_{i\in I} A_i$ and $y_\rho \in \bigcup_{j\in J} A'_j$. Since $(A_i)_{i\in I}$ and $(A'_j)_{j\in J}$ are directed, there are indices $i \in I$, $j \in J$ such that $x_1, \ldots, x_r \in A_i$ and $y_1, \ldots, y_r \in A'_j$. Thus, we get $z \in A_i \otimes_K A'_j$. □

As an example, we can consider for two field extensions L/K and L'/K the directed systems $(L_i)_{i\in I}$, resp. $(L'_j)_{j\in J}$, of all subfields $L_i \subset L$, resp. $L'_j \subset L'$, that are *finitely* generated over K. Then we obtain $L \otimes_K L' = \bigcup_{i\in I, j\in J}(L_i \otimes_K L'_j)$ from Lemma 13. Now, if we want to prove a certain property for $L \otimes_K L'$, for instance to be an integral domain, we see that $L \otimes_K L'$ does not contain nontrivial zero divisors if and only if all $L_i \otimes_K L'_j$ admit this property. In this way, we can often reduce considerations about general field extensions to the case of finitely generated field extensions. In the next section we will make use of this possibility on several occasions.

Exercises

Let R always be a ring.

1. *Consider the tensor product* $M \otimes_R R'$ *of an* R*-module* M *and an* R*-algebra* R'. *Then show that* $M \otimes_R R'$, *as an* R'*-module together with the* R*-linear map* $\tau\colon M \longrightarrow M \otimes_R R'$, $x \longmapsto x \otimes 1$, *is uniquely characterized by the following universal property*: *For every* R*-linear map* $\Phi\colon M \longrightarrow E$ *to an* R'*-module* E, *there is a unique* R'*-linear map* $\varphi\colon M \otimes_R R' \longrightarrow E$ *such that* $\Phi = \varphi \circ \tau$.

2. *Prove the existence of the tensor product* $R' \otimes_R R''$ *of two* R*-algebras in a direct way, by constructing an* R*-algebra* T *admitting the universal property of Lemma* 9.

3. Let R' be an R-algebra. Show for a flat R-module M that $M \otimes_R R'$ is a flat R'-module.

4. Let M be an R-module. Show:
   (i) If M is flat and $a \in R$ is not a zero divisor in R, then every equation $ax = 0$ for $x \in M$ implies $x = 0$.
   (ii) If R is a principal ideal domain, then M is a flat R-module if and only if M is *torsion-free*, in the sense that an equation $ax = 0$ for elements $a \in R$, $x \in M$ always implies $a = 0$ or $x = 0$.

5. Consider two R-algebras R', R'', as well as two ideals $\mathfrak{a}' \subset R'$, $\mathfrak{a}'' \subset R''$, and show that $(R'/\mathfrak{a}') \otimes_R (R''/\mathfrak{a}'') \simeq (R' \otimes_R R'')/(\mathfrak{a}', \mathfrak{a}'')$. Here $(\mathfrak{a}', \mathfrak{a}'')$ indicates the ideal in

$R' \otimes_R R''$ that is generated by the images $\sigma'(\mathfrak{a}')$ and $\sigma''(\mathfrak{a}'')$ with respect to the canonical R-algebra homomorphisms $\sigma' \colon R' \longrightarrow R' \otimes_R R''$, $\sigma'' \colon R'' \longrightarrow R' \otimes_R R''$.

6. For a normal algebraic field extension L/K in characteristic $p > 0$, consider the separable closure K_s, as well as the purely inseparable closure K_i of K in L; cf. 3.7/4 and 3.7/5. Show that the canonical map $K_s \otimes_K K_i \longrightarrow L$, $a \otimes b \longmapsto ab$, is an isomorphism.

7. Let L/K and K'/K be finitely generated field extensions, where L/K is transcendental of transcendence degree > 0. Show that $L \otimes_K K'$ is a field if and only if the extension K'/K is algebraic.

8. Let $(M_i)_{i \in I}$ and $(N_i)_{i \in I}$ be two inductive systems of R-modules (see Section 4.2). Show that $(M_i \otimes_R N_i)_{i \in I}$ is naturally an inductive system of R-modules, and that there is a canonical isomorphism of R-modules

$$(\varinjlim M_i) \otimes_R (\varinjlim N_i) \xrightarrow{\sim} \varinjlim (M_i \otimes_R N_i).$$

7.3 Separable, Primary, and Regular Extensions*

In the present section we study certain classes of not necessarily algebraic field extensions whose defining properties can be characterized in terms of tensor products. We start with separable field extensions, recalling first of all the definition of the *radical* of a ring R. It is denoted by $\operatorname{rad} R$ and consists of all elements $z \in R$ such that there is an exponent $n \in \mathbb{N}$ satisfying $z^n = 0$. The ring R is called *reduced* if $\operatorname{rad} R = 0$.

Remark 1. *The following conditions are equivalent for an* algebraic *field extension* L/K:

(i) *The extension* L/K *is separable in the sense of Definition* 3.6/3.

(ii) *The tensor product* $L \otimes_K K'$ *is reduced for every field extension* K'/K.

Proof. First, consider the case that L/K is separable. Then, applying 7.2/13, we may assume that the extension L/K is finitely generated and hence of finite degree. By the primitive element theorem 3.6/12, the extension L/K is simple, say $L = K(a)$ for some element $a \in L$, and it follows from 7.2/11 and the separability of a over K that $L \otimes_K K'$ is reduced for every field extension K'/K. This shows that (i) implies (ii). Conversely, assuming (ii), let K' be an algebraic closure of K and fix an element $a \in L$. Since K'/K is flat, the inclusion $K(a) \hookrightarrow L$ leads to an inclusion $K(a) \otimes_K K' \hookrightarrow L \otimes_K K'$, and we see that $K(a) \otimes_K K'$ is reduced. But then the minimal polynomial of a over K can have only simple zeros in K', due to 7.2/11, and it follows that a is separable over K. Letting a vary over L, it follows that L/K is separable. □

Since condition (ii) in Remark 1 is meaningful also for nonalgebraic extensions L/K, we can define the separability for arbitrary field extensions as follows:

Definition 2. *A field extension L/K is called* separable *if for arbitrary field extensions K'/K the tensor product $L \otimes_K K'$ is reduced.*

Remark 3. *Every purely transcendental field extension L/K is separable.*

Proof. Use 7.2/12. □

Next we want to list some simple properties of separable field extensions.

Proposition 4. *Let M/K be a field extension.*

(i) *If M/K is separable and L is an intermediate field of M/K, then the extension L/K is separable as well.*[2]

(ii) *M/K is separable if and only if L/K is separable for every intermediate field L of M/K that is finitely generated over K.*

(iii) *If for an intermediate field L of M/K the extensions M/L and L/K are separable, then M/K is separable as well.*

Proof. Consider an intermediate field L of M/K and assume that M/K is separable. Then, for any field extension K'/K, the inclusion $L \hookrightarrow M$ gives rise to an inclusion $L \otimes_K K' \hookrightarrow M \otimes_K K'$, since K'/K is flat. In particular, $L \otimes_K K'$ is reduced if $M \otimes_K K'$ is reduced. Thus, the separability of M/K implies that of L/K. Moreover, we conclude from 7.2/13 that M/K is separable if and only if L/K is separable for all intermediate fields L that are finitely generated over K. This verifies assertions (i) and (ii).

To justify assertion (iii), assume that M/L and L/K are separable. Furthermore, choose an arbitrary field extension K'/K. Then $R = L \otimes_K K'$ is nonzero and reduced. We need as an auxiliary result the fact that the zero ideal in R is an intersection of prime ideals. To prove this, consider for an element $s \neq 0$ in R the multiplicative system $S = \{s^0, s^1, \ldots\}$. Since R is reduced, S does not contain 0. Now, proceeding as in the proof of 3.4/6, we can use Zorn's lemma 14.5 to construct an ideal $\mathfrak{p} \subset R$ that is maximal with respect to the condition that $\mathfrak{p} \cap S = \emptyset$. It follows that $\mathfrak{p}$ is a prime ideal. Thus, for each element $s \neq 0$ in R, there is a prime ideal $\mathfrak{p} \subset R$ such that $s \notin \mathfrak{p}$. Consequently, the zero ideal in R is an intersection of prime ideals, say $0 = \bigcap_{j \in J} \mathfrak{p}_j$.

Let Q_j be the field of fractions of $R/\mathfrak{p}_j$ for all $j \in J$. Then the canonical homomorphisms $R \longrightarrow R/\mathfrak{p}_j \hookrightarrow Q_j$ give rise to an injection $R \hookrightarrow \prod_{j \in J} Q_j$ and, due to the flatness of M/L, to an injection $M \otimes_L R \hookrightarrow M \otimes_L \prod_{j \in J} Q_j$. Now we use the fact that the map

$$(*) \qquad M \otimes_L \prod_{j \in J} Q_j \longrightarrow \prod_{j \in J} (M \otimes_L Q_j), \qquad x \otimes (q_j)_{j \in J} \longmapsto (x \otimes q_j)_{j \in J},$$

is injective, a result that we will prove further below. Since M/L is separable, we know that all tensor products $M \otimes_L Q_j$ are reduced. Therefore, $M \otimes_L R$ is reduced as well, and using the isomorphism

[2] See Exercise 1 for the fact that M/L is not necessarily separable if M/K is separable.

$$M \otimes_L R = M \otimes_L (L \otimes_K K') \xrightarrow{\sim} M \otimes_K K'$$

of 7.2/7, the same is true for $M \otimes_K K'$. But this asserts that M/K is separable, as desired.

It remains to show that the map $(*)$ above is injective. To achieve this we fix an L-vector space basis $(y_i)_{i\in I}$ of M. Since tensor products commute with direct sums, see 7.2/4, we can write each element $z \in M \otimes_L \prod_{j\in J} Q_j$ as a sum $z = \sum_{i\in I} y_i \otimes (q_{ij})_{j\in J}$ with unique elements $q_{ij} \in Q_j$, where almost all terms of such a sum are trivial. In particular, only for finitely many indices $i \in I$ can there exist indices $j \in J$ such that $q_{ij} \neq 0$. In a similar way, we write elements in $\prod_{j\in J}(M \otimes_L Q_j)$ uniquely as families of type $(\sum_{i\in I} y_i \otimes q_{ij})_{j\in J}$. All members of such a family are sums admitting only finitely many nonzero terms. In particular, for each $j \in J$ there are at most finitely many indices $i \in I$ such that $q_{ij} \neq 0$, although all in all, we can have $q_{ij} \neq 0$ for infinitely many indices $i \in I$ and certain indices $j \in J$ depending on i. Since the map $(*)$ associates to an element of type $\sum_{i\in I} y_i \otimes (q_{ij})_{j\in J}$ the element $(\sum_{i\in I} y_i \otimes q_{ij})_{j\in J}$, we see that $(*)$ is always injective, but not surjective in general. □

Definition 5. *A field extension L/K is called* separably generated *if there is a transcendence basis $\mathfrak{X}$ of L/K such that L is separable (algebraic) over $K(\mathfrak{X})$. In this case, $\mathfrak{X}$ is called a* separating transcendence basis *of L/K.*

Since every field extension L/K admits a transcendence basis, cf. 7.1/3, we see for $\operatorname{char} K = 0$ that L/K is automatically separably generated. Furthermore, we can directly conclude the following corollary from Proposition 4 (iii) in conjunction with Remark 3:

Corollary 6. *Every separably generated field extension L/K, in particular every field extension in characteristic 0, is separable.*

Our next objective is to show for finitely generated field extensions a converse of this assertion. To do this, recall for a field K of characteristic $p > 0$ that we can construct the field $K^{p^{-i}}$ of all p^ith roots of elements in K. In this way, we get a chain of inclusions

$$K = K^{p^{-0}} \subset K^{p^{-1}} \subset K^{p^{-2}} \subset \ldots,$$

where $K^{p^{-\infty}} = \bigcup_{i=0}^{\infty} K^{p^{-i}}$ equals the purely inseparable closure of K, a field that is perfect and purely inseparable over K; cf. Exercise 6 of Section 3.7.

Proposition 7. *For a field K of characteristic $p > 0$ and an extension field L of K, the following conditions are equivalent*:

(i) *L/K is separable.*

(ii) *$L \otimes_K K^{p^{-\infty}}$ is reduced.*

(iii) *For every finite extension K'/K such that $K' \subset K^{p^{-1}}$, the tensor product $L \otimes_K K'$ is reduced.*

(iv) *If $a_1, \dots, a_r \in L$ are linearly independent over K, then so are $a_1^p, \dots, a_r^p$.*

(v) *Every subfield $L' \subset L$ that is finitely generated over K is separably generated over K.*

If L/K is finitely generated, say $L = K(x_1, \dots, x_n)$, and separable over K, then the system of elements x_i can be reduced to a separating transcendence basis of L/K.

Proof. The implication (i) $\Longrightarrow$ (ii) is trivial. Furthermore, the implication (ii) $\Longrightarrow$ (iii) follows from the flatness of L/K, since every K' in the situation of (iii) is a subfield of $K^{p^{-\infty}}$, so that $L \otimes_K K'$ is a subring of $L \otimes_K K^{p^{-\infty}}$.

To verify the implication (iii) $\Longrightarrow$ (iv), consider elements $a_1, \dots, a_r \in L$ that are linearly independent over K, as well as elements $c_1, \dots, c_r \in K$ such that $\sum_{i=1}^r c_i a_i^p = 0$. Then we can extract the pth root $c_i^{p^{-1}} \in K^{p^{-1}}$ of c_i, $i = 1, \dots, r$, and consider the field $K' = K(c_1^{p^{-1}}, \dots, c_r^{p^{-1}}) \subset K^{p^{-1}}$, which is finite over K. Now look at the element $z = \sum_{i=1}^r a_i \otimes c_i^{p^{-1}} \in L \otimes_K K'$. Since

$$z^p = \sum_{i=1}^r a_i^p \otimes c_i = \sum_{i=1}^r (c_i a_i^p) \otimes 1 = \Big(\sum_{i=1}^r c_i a_i^p \Big) \otimes 1 = 0$$

and since $L \otimes_K K'$ is reduced, we get $z = 0$. On the other hand, we claim that the tensors $a_1 \otimes 1, \dots, a_r \otimes 1 \in L \otimes_K K'$ are linearly independent over K'. Indeed, $(\bigoplus_{i=1}^r K a_i) \otimes_K K'$ is a subvector space of $L \otimes_K K'$, due to the flatness of K'/K, and we have

$$\Big(\bigoplus_{i=1}^r K a_i \Big) \otimes_K K' \xrightarrow{\sim} \bigoplus_{i=1}^r (K a_i \otimes_K K')$$

by 7.2/4. Therefore, $z = 0$ implies that all coefficients $c_i^{p^{-1}}$ and hence all c_i are zero. This shows that $a_1^p, \dots, a_r^p$ are linearly independent over K.

Next assume condition (iv). To derive (v), it is enough to look at the case that L/K is a finitely generated field extension, say $L = K(x_1, \dots, x_n)$. We show by induction on n that L/K is separably generated. The base case $n = 0$ is trivial. Therefore, let $n > 0$ and assume that $x_1, \dots, x_t$ is a maximal subsystem of $x_1, \dots, x_n$ that is algebraically independent over K. Then $t \le n$, and $x_1, \dots, x_t$ is a transcendence basis of L/K. Since nothing has to be shown for $t = n$, assume $t < n$. Furthermore, let $f \in K[X_1, \dots, X_{t+1}]$ be a nontrivial polynomial of minimal total degree d such that $f(x_1, \dots, x_{t+1}) = 0$. If f is even a polynomial in $X_1^p, \dots, X_{t+1}^p$, then f is of type $f = \sum_{\nu \in I} c_\nu (X^p)^\nu$ for coefficients $c_\nu \in K$ and a finite index set $I \subset \mathbb{N}^{t+1}$, where we can assume $c_\nu \neq 0$ for all $\nu \in I$. Hence, the pth powers $(x_1^{\nu_1})^p \dots (x_{t+1}^{\nu_{t+1}})^p$, $\nu \in I$, are linearly dependent over K, and the same holds for the monomials $x_1^{\nu_1} \dots x_{t+1}^{\nu_{t+1}}$ by (iv). In this way, we obtain a relation $g(x_1, \dots, x_{t+1}) = 0$ for a nontrivial polynomial $g \in K[X_1, \dots, X_{t+1}]$, whose total degree is $< d$. However, this is excluded by the choice of d, and it follows that $f \notin K[X_1^p, \dots, X_{t+1}^p]$. As a consequence, there is a variable X_i such that f is not a polynomial in X_i^p. In particular, $h = f(x_1, \dots, x_{i-1}, X_i, x_{i+1}, \dots, x_{t+1})$ is a nontrivial polynomial in X_i with coefficients in $K[x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_{t+1}]$,

admitting x_i as a zero and having a derivative that is not identically zero. Therefore, L is algebraic over $K(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{t+1})$. In addition, since $\operatorname{transdeg}_K(L) = t$, we see that $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{t+1}$ constitute a transcendence basis of L/K and hence are algebraically independent over K. Now observe that the minimality of the total degree of f implies that h, as a polynomial with coefficients in $K[x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{t+1}]$, is irreducible and primitive. Furthermore, the polynomial ring in X_i over this coefficient domain is a unique factorization domain by 2.7/3. Therefore, h is prime and hence also prime in $K(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{t+1})[X_i]$; cf. 2.7/7. Since in addition, the derivative of h is nonzero, h is separable by 3.6/1. Therefore, x_i is separable algebraic over $K(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{t+1})$, and in particular, separable algebraic over $K(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$, which in turn, is separably generated over K by the induction hypothesis. Therefore, all in all, $L = K(x_1, \ldots, x_n)$ is separably generated over K. This finishes the verification of the implication (iv) $\Longrightarrow$ (v). In particular, our argument shows for $L = K(x_1, \ldots, x_n)$ that the system of the x_i can be reduced to a separating transcendence basis of L/K.

Finally, the implication (v) $\Longrightarrow$ (i) follows from Proposition 4 (ii) and Corollary 6. □

If K is a perfect field in the setting of Proposition 7, then K does not admit nontrivial purely inseparable field extensions and hence satisfies $K = K^{p^{-\infty}}$. Therefore, using Corollary 6 we can state the following corollary:

Corollary 8. *Every field extension L/K of a perfect field K is separable.*

Next we want to consider two further classes of field extensions, namely primary and regular field extensions, where primary extensions can be viewed as a generalization of purely inseparable algebraic extensions; for example, use the characterization of primary extensions given in Proposition 13 below. A ring R is called *irreducible* if its radical $\operatorname{rad} R$ is a prime ideal.

Definition 9. *A field extension L/K is called* primary (*resp.* regular) *if for every field extension K'/K the tensor product $L \otimes_K K'$ is irreducible* (*resp. an integral domain*).[3] *Thus, an extension L/K is regular if and only if it is separable and primary.*[4]

We can easily read from 7.2/11 that purely inseparable simple field extensions L/K in characteristic $p > 0$ are examples of primary extensions. Indeed, if $L = K(a)$, and if $f = X^{p^r} - c \in K[X]$ is the minimal polynomial of a over K, we get $L \otimes_K K' \simeq K'[X]/(f)$. Over an algebraic closure $\overline{K'}$ of K', we can write $f = (X - a)^{p^r}$, where we have identified a with the corresponding zero of f in $\overline{K'}$. Then, since $\operatorname{rad}(\overline{K'}[X]/(f)) = (X - a)/(f)$, it follows that

[3] In the literature, a field extension L/K is usually said to be primary if K is separably closed in L. This condition is equivalent to the one given here; see Proposition 13.

[4] Use the fact that a ring is an integral domain if its zero ideal is prime.

$\operatorname{rad}(L \otimes_K \overline{K}')$ is prime. Since the inclusion $K' \hookrightarrow \overline{K}'$ gives rise to an injection $L \otimes_K K' \hookrightarrow L \otimes_K \overline{K}'$ by the flatness of L/K, and since $\operatorname{rad}(L \otimes_K \overline{K}')$ is prime as we have seen, the same is true for its intersection with $L \otimes_K K'$, which equals the radical of $L \otimes_K K'$. As a consequence, L/K is primary.

Similarly as we did for separable field extensions, let us derive some elementary properties for primary and regular field extensions.

Remark 10. *Every purely transcendental field extension L/K is regular, and hence in particular, primary.*

Proof. Use 7.2/12. □

Proposition 11. *Let M/K be a field extension.*

(i) *If M/K is primary (resp. regular), the extension L/K is primary (resp. regular) for every intermediate field L of M/K.*

(ii) *M/K is primary (resp. regular) if and only if the extension L/K is primary (resp. regular) for all intermediate fields L of M/K that are finitely generated over K.*

(iii) *If for an intermediate field L of M/K the extensions M/L and L/K are primary (resp. regular), the same is true for M/K.*

Proof. It is enough to consider primary extensions. Indeed, recalling the fact that a field extension is regular if and only if it is separable and primary, we can use Proposition 4 to settle the separable part. If L is an intermediate field of M/K and K' an arbitrary extension field of K, the inclusion $L \hookrightarrow M$ gives rise to an injection $L \otimes_K K' \hookrightarrow M \otimes_K K'$, due to the flatness of K'/K. Next observe for a ring extension $R \subset R'$ and a prime ideal $\mathfrak{p}' \subset R'$ that the intersection $R \cap \mathfrak{p}'$ is a prime ideal in R, and that $\operatorname{rad} R = R \cap \operatorname{rad} R'$. Thereby we see that L/K is primary if the same is true for M/K. On the other hand, if L/K is primary for every intermediate field L of M/K that is finitely generated over K, we can read from 7.2/13 that M/K is primary. Thus, assertions (i) and (ii) are clear.

To verify (iii) consider primary extensions M/L and L/K, as well as an extension K'/K. Then $R = (L \otimes_K K')/\operatorname{rad}(L \otimes_K K')$ is an integral domain, and we can consider its field of fractions Q. Now look at the following sequence of homomorphisms:

$$M \otimes_K K' \xrightarrow{\sim} M \otimes_L (L \otimes_K K') \xrightarrow{\varphi} M \otimes_L R \xrightarrow{\psi} M \otimes_L Q.$$

The first map is the isomorphism of 7.2/7, while the two subsequent ones are constructed from the canonical maps $L \otimes_K K' \longrightarrow R \hookrightarrow Q$ by taking the tensor product with M over L. Using 7.2/5 in conjunction with the flatness of M/L, we can identify $\ker \varphi$ with the tensor product $M \otimes_L \operatorname{rad}(L \otimes_K K')$, which consists of nilpotent elements. Furthermore, ψ is injective. To see that $\operatorname{rad}(M \otimes_K K')$ is a prime ideal, consider elements $a, b \in M \otimes_L (L \otimes_K K')$ whose product ab is nilpotent. Then $(\psi \circ \varphi)(ab) = (\psi \circ \varphi)(a) \cdot (\psi \circ \varphi)(b)$ is nilpotent in $M \otimes_L Q$. Since

M/L is primary, one of the two factors, say $(\psi \circ \varphi)(a)$, must be nilpotent. But then, since $\ker \psi \circ \varphi = \ker \varphi$ consists of nilpotent elements, a must be nilpotent itself. Thus, $\operatorname{rad}(M \otimes_K K')$ is prime. □

In the following we want to show that a field extension L/K is already primary, resp. regular, if the tensor product $L \otimes_K K'$ is irreducible, resp. an integral domain, for all *algebraic* field extensions K'/K. To prove this, we need the following key ingredient:

Lemma 12. *A tensor product $A \otimes_K A'$ of two algebras A and A' over an* algebraically closed *field K is an integral domain if and only if A and A' are integral domains.*

Proof. First assume that $A \otimes_K A'$ is an integral domain. Then $A \otimes_K A'$ is nonzero, and the same is true for A and A'. Therefore, the structural maps $K \longrightarrow A$, $K \longrightarrow A'$ are injective. Due to the flatness of A and A' over K, the tensor product maps

$$A \simeq A \otimes_K K \longrightarrow A \otimes_K A', \qquad A' \simeq K \otimes_K A' \longrightarrow A \otimes_K A'$$

are injective as well, and we see that A and A' are integral domains.

To prove the converse, we apply the geometric methods of Section 3.9. In particular, we use Hilbert's Nullstellensatz 3.9/4. Therefore, assume that A and A' are integral domains. Relying on 7.2/13, we may assume that A and A' are finitely generated K-algebras, say of type

$$A \simeq K[X]/\mathfrak{p}, \qquad A' \simeq K[Y]/\mathfrak{q},$$

for variables $X = (X_1, \ldots, X_r)$, $Y = (Y_1, \ldots, Y_s)$ and prime ideals $\mathfrak{p}$, $\mathfrak{q}$. Furthermore, we know from 7.2/10 that there is a canonical isomorphism

$$\big(K[X]/\mathfrak{p}\big) \otimes_K \big(K[Y]/\mathfrak{q}\big) \xrightarrow{\sim} K[X,Y]/(\mathfrak{p},\mathfrak{q}), \qquad \overline{f} \otimes \overline{g} \longmapsto \overline{fg}.$$

Now let $U = V(\mathfrak{p}) \subset K^r$ and $U' = V(\mathfrak{q}) \subset K^s$ be the algebraic subsets of K^r, resp. K^s, that correspond to $\mathfrak{p}$ and $\mathfrak{q}$. Then $U \times U' = V(\mathfrak{p},\mathfrak{q})$, which means that $U \times U'$ coincides with the algebraic set given by the ideal $(\mathfrak{p},\mathfrak{q}) \subset K[X,Y]$. Since all polynomials in $\mathfrak{p}$ vanish on U, the substitution homomorphism $K[X] \longrightarrow K$, $f \longmapsto f(x)$, factors for $x \in U$ through $A \simeq K[X]/\mathfrak{p}$ and hence yields a substitution homomorphism $A \longrightarrow K$. In particular, as indicated at the end of Section 3.9, we can view the elements of A as "functions" on U. Furthermore, due to Hilbert's Nullstellensatz 3.9/4, a function $f \in A$ vanishes on all of U if and only if $f \in \operatorname{rad} A$, i.e., if and only if f is nilpotent. However, in our case, $\mathfrak{p}$ is a prime ideal, and hence $A = K[X]/\mathfrak{p}$ is an integral domain, so that $f(U) = 0$ is equivalent to $f = 0$. In a similar way we consider the elements of A' as functions on U', as well as the elements of $A \otimes_K A'$ as functions on $U \times U'$.

In a first step we want to show that $A \otimes_K A'$ is reduced, which means for an element $g \in A \otimes_K A'$ that $g(U \times U') = 0$ implies $g = 0$. To do this we need to

consider for points $x \in U$ the tensor product of the substitution homomorphism $A \longrightarrow K$, $a \longmapsto a(x)$, with A', i.e., the map

$$\sigma_x \colon A \otimes_K A' \longrightarrow A', \qquad \sum a_i \otimes a'_i \longmapsto \sum a_i(x) \cdot a'_i,$$

as well as later on the analogous map

$$\tau_y \colon A \otimes_K A' \longrightarrow A, \qquad \sum a_i \otimes a'_i \longmapsto \sum a_i \cdot a'_i(y),$$

for points $y \in U'$. Choosing a K-basis $(e'_i)_{i \in I}$ of A', we can apply 7.2/4 and write every $g \in A \otimes_K A'$ as a sum $g = \sum_{i \in I} g_i \otimes e'_i$ with unique elements $g_i \in A$. Now, after these preparations, look at a nilpotent element $g \in A \otimes_K A'$, say $g = \sum_{i \in I} g_i \otimes e'_i$. Then g vanishes on $U \times U'$. Hence, the same is true for the functions $\sigma_x(g) = \sum_{i \in I} g_i(x) \cdot e'_i$ on U', where $x \in U$. Since A' is reduced, we get $g_i(x) = 0$ for all $x \in U$. Furthermore, since A is reduced as well, this means that $g_i = 0$ for all $i \in I$, and hence $g = 0$. Therefore, $A \otimes_K A'$ is reduced.

In a similar way we can show that $A \otimes_K A'$ is even an integral domain. Indeed, look at elements $f, g \in A \otimes_K A'$, $f \neq 0$, such that $f \cdot g = 0$, and write $g = \sum_{i \in I} g_i \otimes e'_i$ with unique elements $g_i \in A$ again. Since

$$\sigma_x(f) \cdot \sum_{i \in I} g_i(x) \cdot e'_i = \sigma_x(f) \cdot \sigma_x(g) = \sigma_x(fg) = 0,$$

and since A' is an integral domain, we get $\sigma_x(g) = 0$, and hence $g_i(x) = 0$ for all $i \in I$ at those points $x \in U$ where $\sigma_x(f) \neq 0$, i.e., at all points $x \in U$ such that there exists $y \in U'$ satisfying $f(x, y) \neq 0$. Therefore, $f \cdot (g_i \otimes 1)$ vanishes for all $i \in I$ on $U \times U'$, thus implying $f \cdot (g_i \otimes 1) = 0$, as we have seen before. Furthermore, we look at the equation

$$\tau_y(f) \cdot g_i = \tau_y\big(f \cdot (g_i \otimes 1)\big) = 0$$

for $y \in U'$. Since $f \neq 0$, there exist points $(x, y) \in U \times U'$ such that $f(x, y) \neq 0$ and hence such that $\tau_y(f) \neq 0$. However, since A is an integral domain, this implies $g_i = 0$ for all $i \in I$, and hence $g = 0$. □

Proposition 13. *The following conditions are equivalent for a field extension L/K:*

(i) *L/K is primary.*

(ii) *$L \otimes_K K'$ is irreducible for every finite separable extension K'/K.*

(iii) *K is separably closed in L, i.e., every element $a \in L$ that is separable algebraic over K is already contained in K.*

Proof. The implication (i) $\Longrightarrow$ (ii) is trivial. Therefore, assume condition (ii) and let $a \in L$ be separable algebraic over K, with corresponding minimal polynomial $f \in K[X]$ of a over K. This polynomial factorizes over L into a product of irreducible factors, say $f = f_1 \dots f_r$, and there cannot be multiple prime factors, since a, and hence f, are separable. Writing $K' = K(a)$, we get

$$L \otimes_K K' \simeq \prod_{i=1}^{r} L[X]/(f_i),$$

from 7.2/11, thus implying that $L \otimes_K K'$ is a finite product of fields. In particular, the radical $\mathrm{rad}(L \otimes_K K')$ is zero. On the other hand, this radical is prime by our assumption. Therefore, we must have $r = 1$, and it follows that f is irreducible in $L[X]$. Since $a \in L$ is a zero of f by the definition of f, there is a factorization of type $f = (X - a) \cdot g$ in $L[X]$. But f being irreducible implies $g = 1$, and therefore $a \in K$. This shows that K is separably closed in L.

Now assume condition (iii). To show that $L \otimes_K K'$ is irreducible for arbitrary field extensions K'/K, let us first consider the case that K'/K is a finite separable extension. Then K'/K is a simple extension by the primitive element theorem 3.6/12, say $K' = K(a)$. Let $f \in K[X]$ be the minimal polynomial of a over K. It is irreducible over K, but also over L. Indeed, if $f = g \cdot h$ is a factorization into monic polynomials $g, h \in L[X]$, then the coefficients of g and h are separable algebraic over K, since they belong to the splitting field of f over K. Since K is separably closed in L, we get $g, h \in K[X]$. Furthermore, the irreducibility of f over K implies $g = 1$ or $h = 1$ and hence that f is irreducible over L. But then, using the isomorphism $L \otimes_K K' \simeq L[X]/(f)$ of 7.2/11, we see that $L \otimes_K K'$ is a field.

In a next step, we consider a finite separable extension K'/K as before, and a finite purely inseparable field extension K''/K', assuming that we are in positive characteristic. Immediately following Definition 9 we showed that purely inseparable simple extensions are primary. Therefore, we can conclude from Proposition 11 (iii) that K''/K' is primary as well. Thus, the tensor product $L \otimes_K K'' \simeq (L \otimes_K K') \otimes_{K'} K''$ is irreducible, and we see that $L \otimes_K K''$ is irreducible for all finite extensions K''/K. But then, if $\overline{K}$ is an algebraic closure of K, it follows that $L \otimes_K \overline{K}$ is irreducible as well. Indeed, using 7.2/13, we can interpret the radical $\mathrm{rad}(L \otimes_K \overline{K})$ as the union of all radicals $\mathrm{rad}(L \otimes_K K'')$, where K''/K varies over the finite extensions that are contained in $\overline{K}$.

Using Lemma 12, it is now easy to see that $L \otimes_K K'$ is irreducible for arbitrary extensions K'/K, and therefore that L/K is primary. Indeed, choose an algebraic closure $\overline{K'}$ of K' and consider the injection $L \otimes_K K' \hookrightarrow L \otimes_K \overline{K'}$ induced from $K' \hookrightarrow \overline{K'}$. It is enough to show that $L \otimes_K \overline{K'}$ is irreducible. We have just seen that $L \otimes_K \overline{K}$ is irreducible if $\overline{K}$ is the algebraic closure of K in $\overline{K'}$. Since the tensor product

$$\big((L \otimes_K \overline{K})/\,\mathrm{rad}(L \otimes_K \overline{K})\big) \otimes_{\overline{K}} \overline{K'}$$

is an integral domain by Lemma 12, it is seen similarly as in the proof of Proposition 11 (iii) that $L \otimes_K \overline{K'}$ is irreducible. □

Combining the characterizations we have proved for separable and for primary field extensions, we can derive a corresponding characterization for regular field extensions.

Proposition 14. *The following conditions are equivalent for a field extension L/K:*

(i) *L/K is regular.*
(ii) *$L \otimes_K K'$ is an integral domain for every finite extension K'/K.*
(iii) *L/K is separable and K is algebraically closed in L.*

Proof. A ring R is an integral domain if and only if the zero ideal $0 \subset R$ is prime. This is equivalent to the fact that the radical $\operatorname{rad} R$ is both prime and zero. This justifies the equivalence of (i) and (ii) if we use Propositions 7 and 13.

To derive the equivalence of (i) and (iii), start with a regular extension L/K. Then the algebraic closure of K in L is regular over K as well, due to Proposition 11 (i). Therefore, it is enough to consider the case that L/K is algebraic. But in this case, Remark 1 and Proposition 13 imply immediately $L = K$ and therefore (iii). On the other hand, if condition (iii) is given, we can deduce (i) from Proposition 13 again. □

Finally, let us indicate a geometric application of the results we have obtained in the present section. Let K be a field and $\overline{K}$ an algebraic closure. Working in the setting of Section 3.9, consider an algebraic subset $U \subset \overline{K}^n$ of $\overline{K}^n$ that is defined over K and *irreducible* over K, in the sense that the corresponding ideal $\mathfrak{p} = I_K(U) \subset K[X_1, \ldots, X_n]$ is prime; see also the geometric interpretation of irreducibility in Exercise 4 of Section 3.9. Then we can view U just as well as an algebraic subset of $\overline{K}^n$ that is defined over $\overline{K}$. Let $I_{\overline{K}}(U)$ be its corresponding ideal in $\overline{K}[X_1, \ldots, X_n]$, which by Hilbert's Nullstellensatz 3.9/4 satisfies $I_{\overline{K}}(U) = \operatorname{rad}(\mathfrak{p}\overline{K}[X_1, \ldots, X_n])$. The algebraic set U is called *geometrically reduced* if $I_{\overline{K}}(U) = \mathfrak{p}\overline{K}[X_1, \ldots, X_n]$, i.e., if the ideal $\mathfrak{p}\overline{K}[X_1, \ldots, X_n]$ is reduced. Furthermore, U is called *geometrically irreducible* if $I_{\overline{K}}(U) = \operatorname{rad}(\mathfrak{p}\overline{K}[X_1, \ldots, X_n])$ is prime, hence if U is irreducible as an algebraic set defined over $\overline{K}$. In Exercise 4 below we show that U is geometrically reduced (resp. geometrically irreducible, resp. geometrically reduced and geometrically irreducible) if and only if, writing Q for the field of fractions of $K[X_1, \ldots, X_n]/\mathfrak{p}$, the extension Q/K is separable (resp. primary, resp. regular).

Exercises

1. *Consider field extensions $K \subset L \subset M$, where M/K is separable (resp. primary, resp. regular). Then the extension L/K is separable (resp. primary, resp. regular), as we have seen. Is a corresponding assertion also valid for the extension M/L?*

2. *A field extensionL/K is primary if and only if K is separably closed in L. Is it possible to characterize separable extensions L/K for $p = \operatorname{char} K > 0$ in a similar way, for example by requiring that $a^p \in K$ for $a \in L$ implies $a \in K$ or that the algebraic closure of K in L is separable over K?*

3. *Give an example of a separable field extension that is not separably generated.*

4. Let $K[X]$ be the polynomial ring over a field K in finitely many variables $X_1, \ldots, X_n$. Furthermore, consider a prime ideal $\mathfrak{p} \subset K[X]$, the corresponding field of fractions $Q = Q(K[X]/\mathfrak{p})$, as well as an algebraic closure $\overline{K}$ of K. Show:
   (i) The extension Q/K is separable if and only if the ideal $\mathfrak{p}\overline{K}[X]$ is reduced in $\overline{K}[X]$.
   (ii) The extension Q/K is primary if and only if the ideal $\mathrm{rad}(\mathfrak{p}\overline{K}[X])$ is prime in $\overline{K}[X]$.
   (iii) The extension Q/K is regular if and only if the ideal $\mathfrak{p}\overline{K}[X]$ is prime in $\overline{K}[X]$.
5. For a field K and an algebraic closure $\overline{K}$ show that an extension L/K is regular if and only if $L \otimes_K \overline{K}$ is a field.
6. For a perfect field K, consider two reduced K-algebras A, A'. Show that their tensor product $A \otimes_K A'$ is reduced as well.
7. Let K be a field of characteristic $p > 0$. A system $x = (x_1, \ldots, x_n)$ of elements in $K^{p^{-1}}$ is called *p-free* over K if the extension $K(x)/K$ cannot be generated by fewer than n elements. Show:
   (i) n elements $x_1, \ldots, x_n \in K^{p^{-1}}$ are p-free over K if and only if the canonical map $K[X_1, \ldots, X_n]/(X_1^p - x_1^p, \ldots, X_n^p - x_n^p) \longrightarrow K(x)$ is an isomorphism.
   (ii) A field extension L/K is separable if and only if the following condition holds: If $x_1, \ldots, x_n \in K$ are p-free over K^p, then these elements are p-free over L^p as well.

7.4 Differential Calculus*

The objective of the present section is to characterize separable field extensions in terms of differential calculus. However, the methods we will use are not based on infinitesimal limits as is customary in analysis. In fact, they are of purely algebraic nature and find their natural continuation in the study of so-called étale and smooth morphisms in algebraic geometry; see, for example, [3], Chap. 8. In the following, let R always be a ring.

Definition 1. *An* R-derivation *of an R-algebra A to an A-module M is given by an R-linear map $\delta\colon A \longrightarrow M$ satisfying the "product rule"*

$$\delta(fg) = f \cdot \delta(g) + g \cdot \delta(f), \qquad f, g \in A.$$

Without further specification, a derivation *is meant as a $\mathbb{Z}$-derivation.*

For elements $r \in R$ we have always $\delta(r \cdot 1) = 0$. Furthermore, from the product rule one easily derives the "quotient rule"

$$\delta\Big(\frac{f}{g}\Big) = \frac{g\delta(f) - f\delta(g)}{g^2}$$

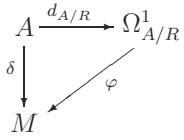
for elements $f, g \in A$, where g is a unit. The set of all R-derivations $\delta\colon A \longrightarrow M$ yields an A-module, which is denoted by $\mathrm{Der}_R(A, M)$, or by $\mathrm{Der}(A, M)$ if $R = \mathbb{Z}$. For example, if $A = R[X]$ is the polynomial ring in one variable over R, then the usual formalism of differentiating polynomials

$$\frac{d}{dX}\colon R[X] \longrightarrow R[X], \qquad f(X) \longmapsto f'(X),$$

gives rise to a well-defined R-derivation of $R[X]$ to itself. Since an R-derivation $\delta\colon R[X] \longrightarrow R[X]$ is already uniquely determined by the element $\delta(X)$ due to the product rule, we see that $\mathrm{Der}_R(R[X], R[X])$ is a free $R[X]$-module, generated by the derivation $\frac{d}{dX}$.

Proposition 2. *Let A be an R-algebra. Then there exists an A-module $\Omega^1_{A/R}$ together with an R-derivation $d_{A/R}\colon A \longrightarrow \Omega^1_{A/R}$ such that $(\Omega^1_{A/R}, d_{A/R})$ admits the following universal property*:

For every R-derivation $\delta\colon A \longrightarrow M$ to an A-module M, there is a unique A-linear map $\varphi\colon \Omega^1_{A/R} \longrightarrow M$ satisfying $\delta = \varphi \circ d_{A/R}$, in other words, making the diagram



commutative. The pair $(\Omega^1_{A/R}, d_{A/R})$ is uniquely determined by this property, up to canonical isomorphism. Furthermore, $(\Omega^1_{A/R}, d_{A/R})$, or just $\Omega^1_{A/R}$, is referred to as the module of relative differential forms (*of degree* 1) *of A over R.*

Proof. We start with the case $A = R[\mathfrak{X}]$, for a system $\mathfrak{X}$ of (arbitrarily many) variables X_i, $i \in I$. Then define $\Omega^1_{A/R} = A^{(I)}$ as the free A-module generated by I. Writing dX_i for the basis element of $\Omega^1_{A/R}$ that is given by $i \in I$, we get $\Omega^1_{A/R} = \bigoplus_{i\in I} A \cdot dX_i$. Now, taking the partial derivatives of elements $f \in A$ with respect to the X_i in a formal sense, it is easily seen that

$$d_{A/R}\colon A \longrightarrow \Omega^1_{A/R}, \qquad f \longmapsto \sum_{i\in I} \frac{\partial f}{\partial X_i} dX_i,$$

is an R-derivation satisfying $d_{A/R}(X_i) = dX_i$, and further, that $(\Omega^1_{A/R}, d_{A/R})$ satisfies the universal property of a module of relative differential forms of A over R. Indeed, if $\delta\colon A \longrightarrow M$ is an R-derivation to an arbitrary A-module M, define an A-linear map $\varphi\colon \Omega^1_{A/R} \longrightarrow M$ by $\varphi(dX_i) = \delta(X_i)$ for $i \in I$. Then $\varphi \circ d_{A/R}$ is an R-derivation from A to M that coincides with δ on the variables X_i, $i \in I$. Using the A-linearity and the product rule, we get

$$\delta(f) = \sum_{i\in I} \frac{\partial f}{\partial X_i}\delta(X_i) = \sum_{i\in I} \frac{\partial f}{\partial X_i}\varphi(dX_i) = \varphi \circ d_{A/R}(f)$$

for all $f \in A$, thus implying $\delta = \varphi \circ d_{A/R}$. In particular, since $\varphi(dX_i) = \delta(X_i)$ by the latter relation, φ is unique.

In the general case we may assume A to be of type $R[\mathfrak{X}]/\mathfrak{a}$, for a system of variables $\mathfrak{X}$ and an ideal $\mathfrak{a} \subset R[\mathfrak{X}]$. Therefore, it is enough to prove the following lemma:

Lemma 3. *For an R-algebra A and an ideal $\mathfrak{a} \subset A$, write $B = A/\mathfrak{a}$. Furthermore, let $(\Omega^1_{A/R}, d_{A/R})$ be the module of relative differential forms of A over R. Then*

$$\Omega = \Omega^1_{A/R} \big/ \big(\mathfrak{a}\Omega^1_{A/R} + A d_{A/R}(\mathfrak{a})\big),$$

together with the R-linear map $d\colon B \longrightarrow \Omega$ induced from $d_{A/R}\colon A \longrightarrow \Omega^1_{A/R}$, is the module of relative differential forms of B over R.

Proof. First observe that Ω is indeed a B-module. Further, since $d_{A/R}$ admits the properties of an R-derivation, the same is true for d. To establish the universal property for d, consider an R-derivation $\overline{\delta}\colon B \longrightarrow M$ to a B-module M. Then the composition $\delta = \overline{\delta} \circ \pi$ of $\overline{\delta}$ with the projection $\pi\colon A \longrightarrow A/\mathfrak{a} = B$ is an R-derivation from A to M, when viewing M as an A-module. Furthermore, the universal property of $d_{A/R}\colon A \longrightarrow \Omega^1_{A/R}$ ensures that δ factors uniquely through an A-linear map $\varphi\colon \Omega^1_{A/R} \longrightarrow M$. Since we have $\delta(\mathfrak{a}) = 0$ and since M is a module over B, we see that necessarily, $\varphi(\mathfrak{a}\Omega^1_{A/R} + A d_{A/R}(\mathfrak{a})) = 0$. Therefore, φ gives rise to a B-linear map $\overline{\varphi}\colon \Omega \longrightarrow M$ satisfying $\overline{\delta} = \overline{\varphi} \circ d$. That $\overline{\varphi}$ is uniquely determined by this equation follows from the uniqueness of φ. This settles the proof of Lemma 3 and thereby also the proof of Proposition 2. □

In particular, the preceding proposition says that the map $\varphi \longmapsto \varphi \circ d_{A/R}$ gives rise to an A-linear bijection

$$\mathrm{Hom}_A\big(\Omega^1_{A/R}, M\big) \longrightarrow \mathrm{Der}_R\big(A, M\big)$$

between A-module homomorphisms $\Omega^1_{A/R} \longrightarrow M$ and R-derivations from A to M. Furthermore, we conclude from its universal property that $\Omega^1_{A/R}$ is generated by all differentials of elements in A, i.e., by all elements of type $d_{A/R}(f)$, $f \in A$. In more precise terms, the argument given in the proof of Proposition 2 yields the following:

Proposition 4. *Let A be an R-algebra and $x = (x_i)_{i\in I}$ a system of elements in A such that $A = R[x]$. Then:*

(i) *$(d_{A/R}(x_i))_{i\in I}$ is an A-generating system of $\Omega^1_{A/R}$.*

(ii) *If $x = (x_i)_{i\in I}$ is algebraically independent over R, then $(d_{A/R}(x_i))_{i\in I}$ is a basis of $\Omega^1_{A/R}$; in particular, $\Omega^1_{A/R}$ is then free.*

Next we want to show that every homomorphism of R-algebras $\tau\colon A \longrightarrow B$ gives rise to a canonical exact sequence of B-modules

$$\Omega^1_{A/R} \otimes_A B \xrightarrow{\alpha} \Omega^1_{B/R} \xrightarrow{\beta} \Omega^1_{B/A} \longrightarrow 0.$$

To define the map α, look at the composition of $\tau\colon A \longrightarrow B$ with the R-derivation $d_{B/R}\colon B \longrightarrow \Omega^1_{B/R}$. Viewing $\Omega^1_{B/R}$ as an A-module under τ, we see that $d_{B/R} \circ \tau$ is an R-derivation of A. By the definition of $\Omega^1_{A/R}$, this derivation factors through an A-linear map

$$\Omega^1_{A/R} \longrightarrow \Omega^1_{B/R}, \qquad d_{A/R}(f) \longmapsto d_{B/R}\big(\tau(f)\big),$$

which, in turn, gives rise to a B-linear map

$$\alpha\colon \Omega^1_{A/R} \otimes_A B \longrightarrow \Omega^1_{B/R}, \qquad d_{A/R}(f) \otimes b \longmapsto b \cdot d_{B/R}\big(\tau(f)\big).$$

Finally, to set up β, observe that every A-derivation of B can be viewed as an R-derivation of B. Thus, by the universal property of $\Omega^1_{B/R}$, there is a well-defined map

$$\beta\colon \Omega^1_{B/R} \longrightarrow \Omega^1_{B/A}, \qquad d_{B/R}(g) \longmapsto d_{B/A}(g).$$

Proposition 5. *For every homomorphism of R-algebras $\tau\colon A \longrightarrow B$, the sequence*

$$\Omega^1_{A/R} \otimes_A B \xrightarrow{\alpha} \Omega^1_{B/R} \xrightarrow{\beta} \Omega^1_{B/A} \longrightarrow 0$$

given by $d_{A/R}(f) \otimes b \overset{\alpha}{\longmapsto} b \cdot d_{B/R}(\tau(f))$, $d_{B/R}(g) \overset{\beta}{\longmapsto} d_{B/A}(g)$ is exact.

Proof. Since $\Omega^1_{B/A}$ is generated by the elements of type $d_{B/A}(g)$, $g \in B$, and since $\beta(d_{B/R}(g)) = d_{B/A}(g)$, we see that β is surjective. Furthermore, we have $\beta \circ \alpha = 0$, which implies $\operatorname{im}\alpha \subset \ker\beta$. To prove $\operatorname{im}\alpha = \ker\beta$, it is enough to show that $\Omega^1_{B/R}/\operatorname{im}\alpha$ together with the map $d\colon B \longrightarrow \Omega^1_{B/R}/\operatorname{im}\alpha$ induced by $d_{B/R}$ admits the properties of the module of relative differential forms of B over A. To justify this, consider the following commutative diagram:

$$\begin{array}{ccccc} A & \xrightarrow{d_{A/R}\otimes 1} & \Omega^1_{A/R} \otimes_A B & & \\ {\scriptstyle\tau}\downarrow & & \downarrow{\scriptstyle\alpha} & & \\ d\colon B & \xrightarrow{d_{B/R}} & \Omega^1_{B/R} & \longrightarrow & \Omega^1_{B/R}/\operatorname{im}\alpha \end{array}$$

First, $d\colon B \longrightarrow \Omega^1_{B/R}/\operatorname{im}\alpha$ is an A-derivation, since it is an R-derivation by its definition and since we have $d_{B/R}(\tau(f)) \in \operatorname{im}\alpha$ for all $f \in A$. Now if $\delta\colon B \longrightarrow M$ is an A-derivation from B to a B-module M, then it is an R-derivation as well. Thus, there exists a unique B-linear map $\varphi\colon \Omega^1_{B/R} \longrightarrow M$ such that $\delta = \varphi \circ d_{B/R}$. Since δ is an A-derivation, we get $\delta \circ \tau = 0$ and hence $\varphi \circ \alpha = 0$. However, this means that φ factors through a B-linear map $\overline{\varphi}\colon \Omega^1_{B/R}/\operatorname{im}\alpha \longrightarrow M$. We have $\delta = \overline{\varphi} \circ d$ by construction, and $\overline{\varphi}$ is uniquely determined by this equation. □

Let us evaluate the exact sequence of Proposition 5 in a special case.

Proposition 6. *For an R-algebra A and a multiplicative system $S \subset A$, consider the canonical map $\tau\colon A \longrightarrow A_S$ from A to its localization by S. Then the corresponding map*

$$\alpha\colon \Omega^1_{A/R} \otimes_A A_S \longrightarrow \Omega^1_{A_S/R}, \qquad d_{A/R}(f) \otimes a \longmapsto a \cdot d_{A_S/R}\big(\tau(f)\big),$$

is bijective. In particular, we get $\Omega^1_{A_S/A} = 0$.

Proof. The relation $\Omega^1_{A_S/A} = 0$ is easily derived from the bijectivity of α; use Proposition 5, or look at the case $R = A$ and use $\Omega^1_{A/A} = 0$. Therefore, it remains only to show that α is bijective. To do this we identify $\Omega^1_{A/R} \otimes_A A_S$ with the A_S-module $(\Omega^1_{A/R})_S$, see 7.2/8, and show that $(\Omega^1_{A/R})_S$ together with the map

$$d\colon A_S \longrightarrow (\Omega^1_{A/R})_S, \qquad \frac{f}{s} \longmapsto \frac{s d_{A/R}(f) - f d_{A/R}(s)}{s^2},$$

admits the universal property of the module of relative differential forms of A_S over R. First, we have to justify that d is well defined. Indeed, if $\frac{f}{s} = \frac{f'}{s'}$ for elements $f, f' \in A$ and $s, s' \in S$, then there is an element $s'' \in S$ such that $s''(s'f - sf') = 0$ holds in A. This implies

$$(s'f - sf') \cdot d_{A/R}(s'') + s'' \cdot d_{A/R}(s'f - sf') = 0,$$

and we see, multiplying by s'', that $d_{A/R}(s'f - sf')$ vanishes as an element of $(\Omega^1_{A/R})_S$. In other words, we have $s'\delta(f) - s\delta(f') = f'\delta(s) - f\delta(s')$, writing δ for the composition of $d_{A/R}$ with the canonical map $\Omega^1_{A/R} \longrightarrow (\Omega^1_{A/R})_S$. That the map $d : A_S \longrightarrow (\Omega^1_{A/R})_S$ is well defined follows then from the following computation:

$$\begin{aligned}
& s'^2\big(s\delta(f) - f\delta(s)\big) - s^2\big(s'\delta(f') - f'\delta(s')\big) \\
&= ss'\big(s'\delta(f) - s\delta(f')\big) - s'^2 f\delta(s) + s^2 f'\delta(s') \\
&= ss'\big(f'\delta(s) - f\delta(s')\big) - s'^2 f\delta(s) + s^2 f'\delta(s') \\
&= s'(sf' - s'f)\delta(s) + s(sf' - s'f)\delta(s') \\
&= 0.
\end{aligned}$$

One can check now in a straightforward manner that d is a derivation. However, we will skip this step. Finally, to test the universal property, consider an R-derivation $\delta\colon A_S \longrightarrow M$ to an A_S-module M. Then $\delta \circ \tau$ is an R-derivation from A to M, and there exists an A-linear map $\varphi\colon \Omega^1_{A/R} \longrightarrow M$ such that $\delta \circ \tau = \varphi \circ d_{A/R}$. Using A_S-linear extension, φ gives rise to an A_S-linear map $\varphi_S\colon (\Omega^1_{A/R})_S \longrightarrow M$ satisfying $\delta = \varphi_S \circ d$, and it is easily seen that φ_S is uniquely determined by this equation. □

In the following we want to apply the theory of differential forms to *field extensions*. Looking at some special cases, we will determine the module of

relative differential forms $\Omega^1_{L/K}$ that is attached to a field extension L/K. In principle, this can be done using Lemma 3 and Proposition 6. However, from a technical point of view, it is often easier to determine instead of $\Omega^1_{L/K}$ its dual space, namely the L-vector space $\mathrm{Der}_K(L,L) \simeq \mathrm{Hom}_L(\Omega^1_{L/K}, L)$. There is a canonical injection

$$\Omega^1_{L/K} \hookrightarrow \mathrm{Hom}_L\big(\mathrm{Der}_K(L,L), L\big), \qquad d_{L/K}(x) \longmapsto \big(\delta \longmapsto \delta(x)\big),$$

in terms of L-vector spaces, which is bijective if one of the spaces $\Omega^1_{L/K}$ and $\mathrm{Der}_K(L,L)$ is of finite dimension over L.

Proposition 7. *Let L/K be a field extension and $x = (x_j)_{j\in J}$ a system of generators. Choosing a system of variables $\mathfrak{X} = (X_j)_{j\in J}$, look at the K-homomorphism $\pi\colon K[\mathfrak{X}] \longrightarrow L$ given by $X_j \longmapsto x_j$, and let $(f_i)_{i\in I}$ be a system of generators of the ideal $\ker\pi$. Furthermore, consider a derivation $\delta\colon K \longrightarrow V$ to an L-vector space V, as well as a system $(v_j)_{j\in J}$ of elements in V. Then the following conditions are equivalent*:

(i) *δ extends to a derivation $\delta'\colon L \longrightarrow V$ satisfying $\delta'(x_j) = v_j$ for all $j \in J$.*

(ii) *The relations*

$$f_i^\delta(x) + \sum_{j\in J} \frac{\partial f_i}{\partial X_j}(x)\cdot v_j = 0, \qquad i \in I,$$

are valid in V, where f^δ for $f \in K[\mathfrak{X}]$ indicates the "polynomial" that is obtained in $V[\mathfrak{X}] := V \otimes_K K[\mathfrak{X}]$ by applying δ to the coefficients of f, i.e., $f^\delta = \sum_\nu \delta(c_\nu)\mathfrak{X}^\nu$ for $f = \sum_\nu c_\nu \mathfrak{X}^\nu$.

If there exists an extension as in (i), *it is unique.*

Proof. Assuming condition (i), we obtain for polynomials $f = \sum_\nu c_\nu \mathfrak{X}^\nu \in K[\mathfrak{X}]$ the equation

$$\delta'\big(f(x)\big) = \sum_\nu \delta(c_\nu)x^\nu + \sum_\nu c_\nu \delta'(x^\nu) = f^\delta(x) + \sum_{j\in J} \frac{\partial f}{\partial X_j}(x)\cdot v_j,$$

i.e., δ', as an extension of δ, is uniquely determined on $K[x]$ by the equations $\delta'(x_j) = v_j$, $j \in J$. Furthermore, using the quotient rule

$$\delta'\Big(\frac{a}{b}\Big) = \frac{b\delta'(a) - a\delta'(b)}{b^2}$$

for elements $a, b \in K[x]$, $b \neq 0$, the uniqueness assertion of δ' extends to $K(x)$; alternatively, one can use Proposition 6. Also we see that the equations of (ii) are valid because $f_i(x)$ vanishes for all $i \in I$.

Now assume condition (ii) and define a derivation $\hat\delta\colon K[\mathfrak{X}] \longrightarrow V$ by

$$\hat\delta(f) = f^\delta(x) + \sum_{j\in J} \frac{\partial f}{\partial X_j}(x)\cdot v_j.$$

Viewing V as a $K[\mathfrak{X}]$-module under the map $\pi\colon K[\mathfrak{X}] \longrightarrow L$, it is easily verified that we indeed get a derivation. It satisfies $\hat{\delta}(f_i) = 0$ for all $i \in I$, due to the equations in (ii). Furthermore, the product rule implies $\hat{\delta}(gf_i) = 0$ for arbitrary elements $g \in K[\mathfrak{X}]$. In this way, $\hat{\delta}$ vanishes on the ideal generated by $(f_i)_{i\in I}$ in $K[\mathfrak{X}]$ and hence on the kernel of the map $\pi\colon K[\mathfrak{X}] \longrightarrow L$, $\mathfrak{X} \longmapsto x$. Therefore, $\hat{\delta}$ induces a derivation $\overline{\delta}\colon K[x] \longrightarrow V$ extending δ. Finally, we can use the quotient rule or Proposition 6 in order to extend $\overline{\delta}$ to a derivation $\delta'\colon K(x) \longrightarrow V$. □

The result of Proposition 7, which we have just proved, provides a useful tool for computing $\Omega^1_{L/K}$ and $\mathrm{Der}_K(L, L)$ for field extensions L/K, since K-derivations $L \longrightarrow L$ may be viewed as extensions of the trivial derivation $K \longrightarrow L$. However, in most cases one will subdivide an extension L/K by suitable intermediate fields, say $K \subset L' \subset L$, and start by examining the K-derivations of L'. Subsequently one needs to know how to extend K-derivations on L' to K-derivations on L so that after all, one obtains information on the K-derivations of L. This is a typical scenario for applying Proposition 7. Alternatively, one can use the exact sequence of Proposition 5 for a chain of field extensions $K \subset L' \subset L$. Here one would like the map $\alpha\colon \Omega^1_{L'/K} \otimes_{L'} L \longrightarrow \Omega^1_{L/K}$ to be injective, which, however, is not the case in general. One can show that the injectivity of the map α is equivalent to the condition that every K-derivation $L' \longrightarrow L$ admits an extension to a K-derivation $L \longrightarrow L$; cf. Exercise 3.

Next we want to rephrase the assertion of Proposition 7 to yield assertions on modules of differential forms.

Corollary 8. *Let L/K be a purely transcendental field extension that is generated by a transcendence basis $(x_j)_{j\in J}$. Then $(d_{L/K}(x_j))_{j\in J}$ is a basis of the L-vector space $\Omega^1_{L/K}$.*

Proof. Use Propositions 4 and 6. Alternatively, one can rely on the assertion of Proposition 7, at least if the transcendence basis $(x_j)_{j\in J}$ is finite. □

Corollary 9. *Let L/K be a separable algebraic field extension. Then, for every derivation $\delta\colon K \longrightarrow V$ to an L-vector space V, there exists a unique extension as a derivation $\delta'\colon L \longrightarrow V$. In particular, we get $\Omega^1_{L/K} = 0$.*

Proof. Let $\delta\colon K \longrightarrow V$ be a derivation to an L-vector space V, and let L' be an intermediate field of L/K such that L'/K is finite. As we know, L'/K is simple by the primitive element theorem 3.6/12, say $L' = K(x)$ for some $x \in L$. Let $f \in K[X]$ be the minimal polynomial of x over K. Then, given $v \in V$, the condition of Proposition 7 for extending δ to a derivation $\delta'\colon K(x) \longrightarrow V$ such that $\delta'(x) = v$ reads

$$f^\delta(x) + f'(x) \cdot v = 0.$$

Since f is separable, the derivative f' of f is a nonzero polynomial. Furthermore, we have $f'(x) \neq 0$, since the degree of f' is smaller than the degree of the minimal polynomial f of x. Therefore, v is uniquely determined by the preceding equation, and we see that δ extends uniquely to a derivation $\delta' \colon L' \longrightarrow V$.

From this we can easily conclude that δ extends uniquely to a derivation $\delta' \colon L \longrightarrow V$. Indeed, for every intermediate field L' of L/K that is finite over K, we can extend δ to a derivation $\delta' \colon L' \longrightarrow V$, as just seen. Since every such extension is uniquely determined by δ and since L can be exhausted by subfields of type L', it follows that there is a unique extension of δ to a derivation $L \longrightarrow V$.

In particular, the trivial derivation $K \longrightarrow L$ extends to the trivial derivation $L \longrightarrow L$ only, which implies $\mathrm{Der}_K(L, L) = 0$ and hence $\Omega^1_{L/K} = 0$. □

In contrast to the assertion of the preceding corollary, the extension of derivations, for example in the setting of Proposition 7, poses problems if the extension L/K is not separable.

Corollary 10. *Let K be a field of characteristic $p > 0$ and L/K a purely inseparable field extension of degree p, say $L = K(x)$ with minimal polynomial of x over K given by $f = X^p - c \in K[X]$. Furthermore, let $\delta \colon K \longrightarrow V$ be a derivation to an L-vector space V. Then*:

(i) *$\delta(c) = 0$ if there exists a derivation $\delta' \colon L \longrightarrow V$ extending δ.*

(ii) *Conversely, if $\delta(c) = 0$, there exists for every $v \in V$ a unique derivation $\delta' \colon L \longrightarrow V$ that extends δ and satisfies $\delta'(x) = v$. In particular, $d_{L/K}(x)$ is an L-basis of $\Omega^1_{L/K}$.*

Proof. According to Proposition 7, the derivation δ admits an extension to a derivation $\delta' \colon L \longrightarrow V$ satisfying $\delta'(x) = v$ if and only if the equation

$$-\delta(c) + px^{p-1} \cdot v = 0$$

is satisfied, hence if and only if $\delta(c) = 0$, where in the latter case, the value of $\delta'(x) = v$ can be selected without any restriction. Therefore, $\mathrm{Der}_K(L, L)$ is of dimension 1 over L, and the same is true for $\Omega^1_{L/K}$. Furthermore, $d_{L/K}(x)$ is a basis of $\Omega^1_{L/K}$. □

Now we are able to derive the desired characterization of separable field extensions, at least in the case of finitely generated extensions.

Theorem 11. *Consider a field extension L/K that is generated by r elements, say $L = K(y_1, \ldots, y_r)$. Then*

$$\mathrm{transdeg}_K L \leq \dim_L \Omega^1_{L/K} \leq r.$$

Furthermore, $\mathrm{transdeg}_K L = \dim_L \Omega^1_{L/K}$ is equivalent to the fact that L/K is separable.

Corollary 12. *A finitely generated field extension L/K is separable algebraic if and only if $\Omega^1_{L/K} = 0$.*

Corollary 13. *Let L/K be a separable finitely generated field extension. Then the following conditions are equivalent for elements $x_1, \ldots, x_n \in L$:*

(i) *$x_1, \ldots, x_n$ form a separating transcendence basis of L/K.*

(ii) *$d_{L/K}(x_1), \ldots, d_{L/K}(x_n)$ form an L-basis of $\Omega^1_{L/K}$.*

The assertion of Corollary 12 is a special case of Theorem 11. Therefore, it has been called a "corollary." However, from a proof-theoretic point of view, it serves as a preparatory lemma that will be used during the proof of Theorem 11.

Accordingly, we start with the *proof of Corollary* 12. If L/K is separable algebraic, we have $\Omega^1_{L/K} = 0$ by Corollary 9. Conversely, assume $\Omega^1_{L/K} = 0$, which is equivalent to $\mathrm{Der}_K(L, L) = 0$. Choosing a transcendence basis $x_1, \ldots, x_n$ of L/K, the field L is a finite algebraic extension of $K(x_1, \ldots, x_n)$. If this extension is even separable, we read from Corollaries 8 and 9 that $\mathrm{Der}_K(L, L)$ is of dimension n over L. Therefore, $n = 0$, and L/K is separable algebraic.

On the other hand, if the extension $K(x_1, \ldots, x_n) \subset L$ is not separable for $p = \mathrm{char}\, K > 0$, then there is an intermediate field L' of L/K such that L/L' is purely inseparable of degree p. Furthermore, there exists a nontrivial L'-derivation $L \longrightarrow L$ by Corollary 10, and hence in particular a nontrivial K-derivation $L \longrightarrow L$. However, this is in contradiction to $\mathrm{Der}_K(L, L) = 0$, so that L must be separable over $K(x_1, \ldots, x_n)$. This settles the assertion of Corollary 12. □

Now we turn to the *proof of Theorem* 11. It follows from Propositions 4 and 6 that $\Omega^1_{L/K}$ is generated by the elements $d_{L/K}(y_1), \ldots, d_{L/K}(y_r)$. Therefore, we get $\dim_L \Omega^1_{L/K} \leq r$. Now choose elements $x_1, \ldots, x_n \in L$ such that the differential forms $d_{L/K}(x_1), \ldots, d_{L/K}(x_n)$ yield a basis of $\Omega^1_{L/K}$. Furthermore, let $L' = K(x_1, \ldots, x_n)$. Then, looking at the exact sequence

$$\Omega^1_{L'/K} \otimes_{L'} L \xrightarrow{\alpha} \Omega^1_{L/K} \xrightarrow{\beta} \Omega^1_{L/L'} \longrightarrow 0$$

of Proposition 5, the map α is surjective, which shows that $\Omega^1_{L/L'} = 0$. Therefore, as we have seen, the extension L/L' is separable algebraic, and we get

$$\mathrm{transdeg}_K L = \mathrm{transdeg}_K L' \leq n = \dim_L \Omega^1_{L/K}.$$

If the above inequality is an equality, then $x_1, \ldots, x_n$ are algebraically independent over K, and we see that the extension L/K is separably generated and in particular separable; cf. 7.3/6. Conversely, if L/K is a finitely generated separable field extension of transcendence degree n, then L/K is separably generated by 7.3/7. Furthermore, Corollaries 8 and 9 show that $\mathrm{Der}_K(L, L)$ and hence $\Omega^1_{L/K}$ are of dimension n over L. □

Finally, to achieve the *proof of Corollary* 13, let $L' = K(x_1, \ldots, x_n)$ and consider the exact sequence

$$\Omega^1_{L'/K} \otimes_{L'} L \xrightarrow{\alpha} \Omega^1_{L/K} \xrightarrow{\beta} \Omega^1_{L/L'} \longrightarrow 0$$

of Proposition 5. Furthermore, assume that $x_1, \ldots, x_n$ form a separating transcendence basis of L/K. Then we conclude that $\Omega^1_{L/L'} = 0$ from Corollary 12 or Corollary 9, and the map α is surjective. But it is even bijective, since $\dim_L(\Omega^1_{L'/K} \otimes_{L'} L) = n$ by Corollary 8, as well as $\dim \Omega^1_{L/K} = n$ by Theorem 11. Since $d_{L'/K}(x_1), \ldots, d_{L'/K}(x_n)$ form a basis in $\Omega^1_{L'/K}$, the same is true for the images in $\Omega^1_{L/K}$, since α is bijective.

Conversely, if $d_{L/K}(x_1), \ldots, d_{L/K}(x_n)$ form a basis of $\Omega^1_{L/K}$, we conclude as in the proof of Theorem 11 that $x_1, \ldots, x_n$ are a separating transcendence basis of L/K. □

The assertion of Corollary 13 shows again, using Proposition 4 in conjunction with Proposition 6, that for a separable finitely generated field extension L/K, every generating system can be reduced to a separating transcendence basis.

Exercises

1. *Is the condition $\Omega^1_{L/K} = 0$ for an arbitrary field extension L/K equivalent to the fact that L/K is separable algebraic?*
2. Consider a field extension L/K in characteristic 0. Show that every derivation $K \longrightarrow V$ to an L-vector space V extends to a derivation $L \longrightarrow V$.
3. For field extensions $R \subset K \subset L$, consider the map
$$\alpha\colon \Omega^1_{K/R} \otimes_K L \longrightarrow \Omega^1_{L/R}, \qquad d_{K/R}(x) \otimes a \longmapsto a \cdot d_{L/R}(x).$$
Show that α is injective if and only if every R-derivation $K \longrightarrow L$ extends to an R-derivation $L \longrightarrow L$.
4. Let L/K be a finitely generated field extension, say $L = K(x_1, \ldots, x_n)$. Assume that the kernel of the K-homomorphism $K[X_1, \ldots, X_n] \longrightarrow L$ mapping X_i to x_i for all i is generated by polynomials $f_1, \ldots, f_r \in K[X_1, \ldots, X_n]$ that satisfy the condition
$$\operatorname{rank}\left(\frac{\partial f_i}{\partial X_j}(x)\right)_{\substack{i=1,\ldots,r \\ j=1,\ldots,n}} = r.$$
Show that L/K is a separable extension of transcendence degree $n - r$.
5. Let L/K be a field extension in characteristic $p > 0$ such that $L^p \subset K$. Furthermore, let $(x_i)_{i \in I}$ be a *p-basis* of L/K, i.e., a p-free system (cf. Exercise 7 of Section 7.3) generating the extension L/K, and let $\delta\colon K \longrightarrow V$ be a derivation into an L-vector space V. Show for $c_i = x_i^p$:
   (i) If there exists a derivation $\delta'\colon L \longrightarrow V$ extending δ, then $\delta(c_i) = 0$ for all $i \in I$.
   (ii) Conversely, if $\delta(c_i) = 0$ for all $i \in I$, then for every system $(v_i)_{i \in I}$ of elements in V, there is a unique extension $\delta'\colon L \longrightarrow V$ of δ satisfying $\delta'(x_i) = v_i$ for all i.
   (iii) The differential forms $d_{L/K}(x_i)$, $i \in I$, constitute an L-basis of $\Omega^1_{L/K}$.

6. Show that a field extension L/K is separable if and only if every derivation $K \longrightarrow L$ extends to a derivation $L \longrightarrow L$. *Hint*: Use Exercise 2, as well as Exercise 5 for characteristic $p > 0$ in conjunction with the characterization of separable extensions in Exercise 7 of Section 7.3.

Appendix

Solutions to Exercises

Each section of the book ends with a list of specially adapted exercises, some of them printed in *italics*. The latter are particularly suited for digesting the material presented in the corresponding sections. Different from the remaining exercises of more classical type, they concern problems that could just as well be addressed in conversational discussions. Only for these exercises do we give explanations and hints for their solutions below.

1.1, Exercise 1. Clearly, conditions (ii) and (iii) of 1.1/1 imply conditions (ii′) and (iii′) of 1.1/2. Conversely, let G be a set with an associative law of composition such that there is a left neutral element $e \in G$, as well as for every element $a \in G$ a left inverse $b \in G$. We show first that b is also a right inverse of a. Indeed, let $ba = e$. Then b admits a left inverse c such that $cb = e$. However, this implies

$$ab = eab = cbab = cb = e.$$

Hence, if b is a left inverse of a, it is a right inverse of a as well, so that condition 1.1/1(iii) is met. It remains to show that the left neutral element $e \in G$ is right neutral as well. Let $a \in G$. If $b \in G$ is a left inverse of a, it is also a right inverse of a, as we have seen, and we get

$$ae = aba = ea = a,$$

justifying condition 1.1/1(ii).

1.1, Exercise 2. We show that there cannot exist a group isomorphism between $\mathbb{Q}$ and $\mathbb{Q}_{>0}$, due to incompatible properties of the two groups. Given $x \in \mathbb{Q}$, there is always a rational number $y \in \mathbb{Q}$ such that $x = y + y$, namely $y = \frac{1}{2}x$. However, the corresponding assertion, that for every $x \in \mathbb{Q}_{>0}$ there is an element $y \in \mathbb{Q}_{>0}$ such that $x = y \cdot y$, is false. Indeed, for $x = 2$ there does not exist, as one knows, a rational number y whose square equals 2. This is proved by relying on the unique prime factorization of integers. Now, if there exists an isomorphism of groups $\varphi\colon \mathbb{Q} \longrightarrow \mathbb{Q}_{>0}$, the surjectivity of φ shows that there exists an element $a \in \mathbb{Q}$ satisfying $\varphi(a) = 2$. But then, writing $b = \frac{1}{2}a$, we would get $\varphi(b)^2 = \varphi(2b) = \varphi(a) = 2$, contradicting the fact that 2 does not admit a rational square root.

1.2, Exercise 1. Since H is of index 2 in G, we see that G is the union of two disjoint left cosets of H, namely of H itself and of another one denoted by H' that necessarily coincides with the complement of H in G. The same argument works for right cosets of H as well, and it follows that H' is both a left and a right coset of H. Now let $a \in G$. If $a \in H$, we get $aH = Ha$ for trivial reasons. On the other hand, if $a \notin H$, then the two cosets aH and Ha are both different from H and hence must coincide with H'. Therefore, also in this case we get $aH = Ha$, so that H is normal in G.

To see that a subgroup of index 3 is not necessarily normal, let us look at the symmetric group $\mathfrak{S}_3$. Let $\sigma \in \mathfrak{S}_3$ be the permutation that interchanges the numbers 1 and 2, leaving 3 fixed. Then $H := \{\mathrm{id}, \sigma\} \subset \mathfrak{S}_3$ is a subgroup of order 2 and hence a subgroup of index 3 by the theorem of Lagrange 1.2/3, since $\operatorname{ord} \mathfrak{S}_3 = 6$. Now let $\tau \in \mathfrak{S}_3$ be the permutation that leaves 1 fixed and interchanges 2 and 3. Then $\tau \circ \sigma \circ \tau^{-1}$ interchanges the numbers 1 and 3, leaving 2 fixed, and therefore does not belong to H. Thus, we get $\tau H \neq H\tau$, and H cannot be normal in $\mathfrak{S}_3$.

1.2, Exercise 2. To begin with, let us assume that N is a subgroup in G, not necessarily normal. Then $\tau_g \colon G \longrightarrow G$, $a \longmapsto ga$, the left translation by an element $g \in G$, maps left cosets of N onto sets of the same type. Thereby τ_g gives rise to a map $\overline{\tau}_g \colon X \longrightarrow X$ that is characterized by $aN \longmapsto gaN$. Since $gaN = ga'N$ for $a, a' \in G$ implies $aN = a'N$, we conclude that $\overline{\tau}_g$ is injective. On the other hand, $\overline{\tau}_g$ is surjective as well, since τ_g is surjective. Thus, $\overline{\tau}_g$ is bijective and thereby satisfies $\overline{\tau}_g \in S(X)$. Furthermore, $g \longmapsto \overline{\tau}_g$ defines a map $\varphi \colon G \longrightarrow S(X)$, which is even a group homomorphism, due to the relation $\tau_{gg'} = \tau_g \circ \tau_{g'}$ for $g, g' \in G$. Let us look at the kernel of φ. Indeed, an element $g \in G$ is contained in $\ker \varphi$ if and only if $\overline{\tau}_g \colon X \longrightarrow X$ is the identity map, i.e., if and only if $gaN = aN$ for all $a \in G$. This equation is equivalent to $ga \in aN$, resp. $g \in aNa^{-1}$, so that we obtain $\ker \varphi = \bigcap_{a \in G} aNa^{-1}$. Now, if we assume that N is normal in G, this implies $aNa^{-1} = N$ for all $a \in G$ and therefore $\ker \varphi = N$. Thus, writing $\overline{G} = \varphi(G)$, we have constructed for every normal subgroup $N \subset G$ a group $\overline{G}$ together with a surjective group homomorphism $p \colon G \longrightarrow \overline{G}$ satisfying $\ker p = N$.

We could view $\overline{G}$ as "the" factor group of G modulo N. In particular, this makes sense if we understand by $\overline{G}$ the specific subgroup $\varphi(G) \subset S(G)$ constructed above. However, a slightly more general point of view is more appropriate, namely to refer to an arbitrary pair $(\overline{G}, p)$, where $p \colon G \longrightarrow \overline{G}$ is a surjective group homomorphism with kernel $\ker p = N$, as a "factor group" of G modulo N. For such a homomorphism $p \colon G \longrightarrow \overline{G}$ we can prove the fundamental theorem on homomorphisms 1.2/6 in the same way as we did in Section 1.2 for the explicitly constructed surjective group homomorphism $\pi \colon G \longrightarrow G/N$. As a consequence, all "factor groups" $(\overline{G}, p)$ are canonically isomorphic, in particular, isomorphic to the explicitly constructed "factor group" $(G/N, \pi)$.

1.3, Exercise 1. First observe that the law of composition "$\circ$" is commutative. To verify the associativity, consider elements $a, b, c \in G_m$. According to the definition of "$\circ$", there are integers $q, q' \in \mathbb{Z}$ satisfying

$$a + b = qm + (a \circ b), \qquad (a \circ b) + c = q'm + \big((a \circ b) \circ c\big),$$

which yields

$$a + b + c = (q + q')m + \big((a \circ b) \circ c\big).$$

Therefore, $(a \circ b) \circ c$ is the remainder of $a+b+c$ on division by m. In the same way we see that also $a \circ (b \circ c)$ equals the remainder of $a + b + c$ on division by m. Hence, we get $(a \circ b) \circ c = a \circ (b \circ c)$, and the law of composition "$\circ$" is associative. The remaining axioms are easily checked. 0 serves as a zero element of the law "$\circ$", and $m - a$, for $a \in G_m$, $a \neq 0$, is an inverse of a. Thus, G_m is a commutative group.

To exhibit an isomorphism between G_m and $\mathbb{Z}/m\mathbb{Z}$, consider the bijection $\iota\colon G_m \longrightarrow \mathbb{Z}/m\mathbb{Z}$, $a \longmapsto a + m\mathbb{Z}$. Since the integers $a \circ b$ and $a + b$ for $a, b \in G_m$ can differ only by a multiple of m, we get $(a \circ b) + m\mathbb{Z} = (a + b) + m\mathbb{Z}$ and hence $\iota(a \circ b) = \iota(a) + \iota(b)$. It follows that ι is a group isomorphism.

1.3, Exercise 2. Look at the epimorphism $\pi\colon \mathbb{Z} \longrightarrow \mathbb{Z}/m\mathbb{Z}$ that is given by $a \longmapsto a + m\mathbb{Z}$, and observe that the preimage $\pi^{-1}(\overline{H})$ of any subgroup $\overline{H} \subset \mathbb{Z}/m\mathbb{Z}$ yields a subgroup of $\mathbb{Z}$ containing $m\mathbb{Z}$. On the other hand, the image $\pi(H)$ of any subgroup $H \subset \mathbb{Z}$ is a subgroup in $\mathbb{Z}/m\mathbb{Z}$. In this way, it is easily seen that $\overline{H} \longmapsto \pi^{-1}(\overline{H})$ gives rise to a bijection between the subgroups $\overline{H} \subset \mathbb{Z}/m\mathbb{Z}$ and the subgroups $H \subset \mathbb{Z}$ containing $m\mathbb{Z}$.

Let us determine all subgroups $H \subset \mathbb{Z}$ containing $m\mathbb{Z}$. If H is such a subgroup, it is cyclic by 1.3/4, say $H = d\mathbb{Z}$. Furthermore, the inclusion $m\mathbb{Z} \subset d\mathbb{Z}$ shows that that m admits a factorization $m = cd$ for some $c \in \mathbb{Z}$, so that d divides m. On the other hand, every divisor d of m leads to an inclusion $m\mathbb{Z} \subset d\mathbb{Z}$. Therefore, the subgroups in $\mathbb{Z}$ that contain $m\mathbb{Z}$ correspond bijectively to the groups of type $d\mathbb{Z}$, where d divides m. Since the generating element d of such a subgroup $d\mathbb{Z} \subset \mathbb{Z}$ is unique up to sign, these groups correspond bijectively to the positive divisors of m.

Now, to specify all subgroups of $\mathbb{Z}/m\mathbb{Z}$, we have only to apply the epimorphism π to the subgroups $d\mathbb{Z} \subset \mathbb{Z}$ just considered, where d varies over the positive divisors of m. Since $d\mathbb{Z}$ is cyclic with generating element d, its image $\pi(d\mathbb{Z})$ is cyclic as well, generated by $\pi(d) = d + m\mathbb{Z}$. The order of the latter group equals $\frac{m}{d}$, so that the index of $\pi(d\mathbb{Z})$ in $\mathbb{Z}/m\mathbb{Z}$ is d; cf. 1.2/3. Therefore, we can state that for every positive divisor d of m, there is a unique subgroup $\overline{H} \subset \mathbb{Z}/m\mathbb{Z}$ of index d, namely the cyclic subgroup generated by $d + m\mathbb{Z}$, and there are no further subgroups in $\mathbb{Z}/m\mathbb{Z}$. Relying on the fact that every cyclic group of order m is isomorphic to $\mathbb{Z}/m\mathbb{Z}$, we can also state more generally that for every cyclic group of order m, there exists a unique subgroup of index d for every positive divisor d of m and, in view of 1.2/3, a unique subgroup of order d.

Finally, let us point out that the preceding result can also be obtained in a direct way, without looking at corresponding subgroups in $\mathbb{Z}$, if we are willing to use the *ideal-theoretic* characterization of the greatest common divisor of integers in $\mathbb{Z}$. A basic step of proof consists in showing that a given subgroup $H \subset \mathbb{Z}/m\mathbb{Z}$ is generated by the residue class $\overline{d}$ of some divisor d of m. To justify

this, consider elements $a_1, \ldots, a_r \in \mathbb{Z}$, whose residue classes $\overline{a}_1, \ldots, \overline{a}_r \in \mathbb{Z}/m\mathbb{Z}$ generate the group H. If d is the greatest common divisor of $a_1, \ldots, a_r, m$, then there is an equation of type $d = c_1a_1 + \ldots + c_ra_r + cm$ for suitable coefficients $c_1, \ldots, c_r, c \in \mathbb{Z}$; see, for example, 2.4/13. From this we can conclude that H is already generated by the residue class $\overline{d}$ of d.

2.1, Exercise 1. We obtain $0 \cdot a + 0 \cdot a = (0 + 0) \cdot a = 0 \cdot a$ and therefore $0 \cdot a = 0$ for all $a \in R$ by applying the distributive law. In a similar way, we get $a \cdot b + (-a) \cdot b = (a + (-a)) \cdot b = 0 \cdot b = 0$. Hence, $(-a) \cdot b$ is an inverse of $a \cdot b$ with respect to addition, and we see that $(-a) \cdot b = -(a \cdot b)$.

2.1, Exercise 2. The construction of polynomial rings $R[X]$ in 2.1 did not require R to be commutative. Therefore, we can consider the polynomial ring $R[X]$ over an arbitrary not necessarily commutative ring R. Observe that the resulting multiplication on $R[X]$ satisfies $aX = Xa$ for all elements $a \in R$, i.e., that X commutes with the elements of R. In addition, if $R \subset R'$ is an extension of not necessarily commutative rings, we can evaluate polynomials of $R[X]$ at elements $x \in R'$ as usual, by substituting the variable X by x. The addition on $R[X]$ is compatible with such a process, i.e., we get $(f+g)(x) = f(x)+g(x)$ for $f, g \in R[X]$ and $x \in R'$. However, as a rule, the corresponding multiplicative equation $(f \cdot g)(x) = f(x) \cdot g(x)$ is valid only if x commutes with the elements of R, in the sense that $ax = xa$ for all $a \in R$. For example, thinking of the trivial extension $R \subset R$ and taking $x \in R$, this is the reason why polynomial rings, as introduced in 2.1, should be considered only over a *commutative* coefficient ring R. On the other hand, if we want to evaluate polynomials of $R[X]$ on a ring R' strictly extending R, it is not really necessary for R' to be commutative. As indicated before, it is enough that the elements of R commute with those of R'.

2.2, Exercise 1. From $\mathfrak{a} = \sum_{i=1}^{m} Ra_i$ and $\mathfrak{b} = \sum_{j=1}^{n} Rb_j$ we get immediately $\mathfrak{a}+\mathfrak{b} = \sum_{i=1}^{m} Ra_i + \sum_{j=1}^{n} Rb_j$ and thereby see that $a_1, \ldots, a_m, b_1, \ldots, b_n$ generate the ideal $\mathfrak{a} + \mathfrak{b}$. Next we want to show that the elements a_ib_j, $i = 1, \ldots, m$, $j = 1, \ldots, n$, give rise to a system of generators of $\mathfrak{a} \cdot \mathfrak{b}$. Let $\mathfrak{q}$ be the ideal that is generated by these elements. Since $a_ib_j \in \mathfrak{a} \cdot \mathfrak{b}$ for all i and j, we get $\mathfrak{q} \subset \mathfrak{a} \cdot \mathfrak{b}$. To verify the reverse inclusion consider an element $z \in \mathfrak{a} \cdot \mathfrak{b}$. Then z is a finite sum of type $z = \sum_\lambda \alpha_\lambda\beta_\lambda$ for suitable elements $\alpha_\lambda \in \mathfrak{a}$, $\beta_\lambda \in \mathfrak{b}$, and there are elements $c_{\lambda i}, d_{\lambda j} \in R$ such that $\alpha_\lambda = \sum_{i=1}^{m} c_{\lambda i}a_i$ as well as $\beta_\lambda = \sum_{j=1}^{n} d_{\lambda j}b_j$. However, this implies $\alpha_\lambda\beta_\lambda = \sum_{i,j} c_{\lambda i}d_{\lambda j}a_ib_j \in \mathfrak{q}$ and hence $z \in \mathfrak{q}$. Therefore, $\mathfrak{q} = \mathfrak{a} \cdot \mathfrak{b}$, and the elements a_ib_j generate the ideal $\mathfrak{a} \cdot \mathfrak{b}$.

Concerning the ideal $\mathfrak{a} \cap \mathfrak{b}$, it is more complicated to construct a system of generators from the elements a_i and b_j. Let us consider the case $R = \mathbb{Z}$ as an example. Then $\mathfrak{a}$ is generated by the greatest common divisor a of the elements a_i, and likewise $\mathfrak{b}$ by the greatest common divisor b of all b_j; for example, assertions of this type are proved in 2.4/13. Furthermore, the ideal $\mathfrak{a} \cap \mathfrak{b}$ is generated by the least common multiple of a and b; cf. 2.4/13 again. However, such a description of a generating element of $\mathfrak{a} \cap \mathfrak{b}$ is valid only for principal ideal domains. In more general settings it is totally unclear how to arrive at a system of generators for $\mathfrak{a} \cap \mathfrak{b}$.

2.2, Exercise 2. Let $\mathfrak{a}, \mathfrak{b}$ be ideals of a ring R. We claim that $\mathfrak{a} \cup \mathfrak{b}$ is an ideal in R if and only if $\mathfrak{a} \subset \mathfrak{b}$ or $\mathfrak{b} \subset \mathfrak{a}$. If one of these inclusions holds, say $\mathfrak{a} \subset \mathfrak{b}$, then $\mathfrak{a} \cup \mathfrak{b} = \mathfrak{b}$ is clearly an ideal in R. On the other hand, if $\mathfrak{a} \subsetneq \mathfrak{b}$ and $\mathfrak{b} \subsetneq \mathfrak{a}$, then there exists an element $a \in \mathfrak{a}$ that is not contained in $\mathfrak{b}$, as well as an element $b \in \mathfrak{b}$ that is not contained in $\mathfrak{a}$. This implies that $a + b$ is contained neither in $\mathfrak{a}$ nor in $\mathfrak{b}$. Consequently, $\mathfrak{a} \cup \mathfrak{b}$ is not closed with respect to addition and hence cannot be an ideal in R. This justifies our claim.

It is not easy to decide for a family $(\mathfrak{a}_i)_{i \in I}$ of ideals in R consisting of more than two elements whether the union $\mathfrak{a} = \bigcup_{i \in I} \mathfrak{a}_i$ is also an ideal. Certainly, $\mathfrak{a}$ is closed under the multiplication by elements of R, as well as closed under the formation of inverses with respect to addition. Therefore, it remains to check whether $\mathfrak{a}$ is closed under addition, i.e., whether $a, b \in \mathfrak{a}$ always implies $a+b \in \mathfrak{a}$. For example, a sufficient condition for this is that for any two indices $i, j \in I$ and elements $a \in \mathfrak{a}_i$, $b \in \mathfrak{a}_j$, there is always an index $k \in I$ such that $a, b \in \mathfrak{a}_k$. In particular, the union of an ascending chain of ideals $\mathfrak{a}_1 \subset \mathfrak{a}_2 \subset \ldots$ is again an ideal.

2.2, Exercise 3. It is quite easy to determine the ideals in K^2. We claim that except for 0, $K \times 0$, $0 \times K$, K^2, there are no further ideals in K^2. Indeed, consider an ideal $\mathfrak{a} \subset K^2$. If $\mathfrak{a}$ contains an element (a, b), where $a \neq 0 \neq b$, then $(1, 1) = (a^{-1}, b^{-1})(a, b) \in \mathfrak{a}$, i.e., $\mathfrak{a}$ contains the unit element of K^2, and we have $\mathfrak{a} = K^2$. Next, if $\mathfrak{a}$ does not contain an element (a, b) such that $a \neq 0 \neq b$, then $\mathfrak{a}$ consists only of elements of type $(a, 0)$ or $(0, b)$. Since $(a, 0) + (0, b) = (a, b)$, we see that nontrivial elements of type $(a, 0)$ and $(0, b)$ cannot be present in $\mathfrak{a}$ at the same time. Therefore, let us assume that all elements of $\mathfrak{a}$ are of type $(a, 0)$. Then $\mathfrak{a}$ is either the zero ideal, or there is an element $(a, 0)$ in $\mathfrak{a}$ such that $a \neq 0$. In the latter case we get $(1, 0) = (a^{-1}, 1)(a, 0) \in \mathfrak{a}$ and therefore $\mathfrak{a} = K \times 0$.

In particular, we can see that all ideals are subvector spaces in K^2. There is a general reason for this. Indeed, considering the so-called diagonal embedding $K \longrightarrow K^2$, $a \longmapsto (a, a)$, we may identify K with its image Δ in K^2 and thereby view $K = \Delta$ as a subring of K^2. Then the product av in the sense of K^2 as a K-vector space coincides for elements $a \in K$ and $v \in K^2$ with the product av of the ring multiplication in K^2. Since ideals are closed under multiplication by elements of K^2, we see once again that every ideal in K^2 is a K-subvector space in K^2. Since the same is true for every subring of K^2 containing the diagonal Δ, we conclude by reasons of dimension that there cannot exist such subrings, except for Δ and K^2 themselves. Also we see that Δ is an example of a subvector space in K^2 that does not admit the properties of an ideal. Furthermore, Δ is the only subvector space that is at the same time a subring in K^2.

If K consists of more than two elements, one can show that K^2 contains further subvector spaces beyond the ones already mentioned. In general, there will exist further proper subrings in K^2 besides Δ, in particular, subrings that are not contained in Δ.

2.3, Exercise 1. Certainly, the image $\varphi(\mathfrak{a})$ of an ideal $\mathfrak{a} \subset R$ is a subgroup of R'. But it is not an ideal in general, since $\varphi(\mathfrak{a})$ does not need to be closed

with respect to multiplication by elements of R'. For example, consider the ring homomorphism $\mathbb{Z} \hookrightarrow \mathbb{Q}$ and observe for $m \geq 1$ that $m\mathbb{Z}$ is an ideal in $\mathbb{Z}$, but not in $\mathbb{Q}$. Indeed, $\mathbb{Q}$ is a field and therefore contains only the trivial ideals. The situation is different if we assume $\varphi\colon R \longrightarrow R'$ to be *surjective*. In such a case the image $\varphi(\mathfrak{a})$ of an ideal $\mathfrak{a} \subset R$ is always an ideal in R'. To verify that $\varphi(\mathfrak{a})$ is closed with respect to multiplication by elements of R', consider elements $r' \in R'$, $a' \in \varphi(\mathfrak{a})$, as well as preimages $r \in R$, $a \in \mathfrak{a}$. Then we have $ra \in \mathfrak{a}$ and hence $r'a' = \varphi(ra) \in \varphi(\mathfrak{a})$. Let us examine the circumstances under which the ideal $\varphi(\mathfrak{a})$ is prime or maximal in R'. To do this, look at the composition $\psi\colon R \longrightarrow R' \longrightarrow R'/\varphi(\mathfrak{a})$ of φ with the canonical projection $R' \longrightarrow R'/\varphi(\mathfrak{a})$, of course under the assumption that φ is surjective. As a composition of surjective ring homomorphisms, ψ is again a surjective ring homomorphism. Its kernel equals $\mathfrak{a} + \ker\varphi$, so that $R'/\varphi(\mathfrak{a})$ is isomorphic to $R/(\mathfrak{a}+\ker\varphi)$ by 2.3/5. Therefore, using 2.3/8, it follows that $\varphi(\mathfrak{a})$ is prime (resp. maximal) if and only if $R'/\varphi(\mathfrak{a})$ is an integral domain (resp. a field), i.e., if and only if $\mathfrak{a} + \ker\varphi$ is prime (resp. maximal) in R. In particular, the image $\varphi(\mathfrak{a})$ of a prime (resp. maximal) ideal $\mathfrak{a}$ is prime (resp. maximal) if $\mathfrak{a}$ contains $\ker\varphi$.

Next we look at the preimage $\mathfrak{a} = \varphi^{-1}(\mathfrak{a}')$ of an ideal $\mathfrak{a}' \subset R'$, for an arbitrary ring homomorphism $\varphi\colon R \longrightarrow R'$. It is easily verified that $\mathfrak{a}$ is an ideal in R. Indeed, to show that $\mathfrak{a}$ is closed under multiplication by R, consider elements $r \in R$ and $a \in \mathfrak{a}$. Then we have $\varphi(ra) = \varphi(r)\varphi(a) \in \mathfrak{a}'$ and hence $ra \in \varphi^{-1}(\mathfrak{a}') = \mathfrak{a}$. To deal with the question whether $\mathfrak{a}$ is prime or maximal in R, consider the composition $\psi\colon R \longrightarrow R' \longrightarrow R'/\mathfrak{a}'$ again, which now satisfies $\ker\psi = \mathfrak{a}$. We conclude from 2.3/4 that ψ induces an injective homomorphism $\overline{\psi}\colon R/\mathfrak{a} \longrightarrow R'/\mathfrak{a}'$. Now if $\mathfrak{a}'$ is prime in R', we see from 2.3/8 that $R'/\mathfrak{a}'$ is an integral domain. Then the same is true for $R/\mathfrak{a}$, and $\mathfrak{a}$ is a prime ideal in R. However, the same argument does not work for maximal ideals instead of prime ideals, since the map $\overline{\psi}$ does not need to be surjective. Indeed, the preimage $\mathfrak{a} \subset R$ of a maximal ideal $\mathfrak{a}' \subset R'$ is not necessarily maximal. Just consider the inclusion map $\mathbb{Z} \hookrightarrow \mathbb{Q}$, as well as the ideal $\mathfrak{a}' = 0$. It is maximal in $\mathbb{Q}$, but its preimage $\mathfrak{a} = 0$ is not maximal in $\mathbb{Z}$.

In particular, if φ is surjective, the argument given above shows that the ideals (resp. prime ideals, resp. maximal ideals) in R' correspond bijectively to the ideals (resp. prime ideals, resp. maximal ideals) in R that contain $\ker\varphi$.

2.3, Exercise 2. We claim that $\ker\varphi_x$ equals the principal ideal $(X - x)$ that is generated by $X - x$. Clearly, we have $X - x \in \ker\varphi_x$. On the other hand, given any element $f \in \ker\varphi_x$, we may apply Euclidean division 2.1/4 and write $f = q(X - x) + r$ for a polynomial $r \in R[X]$ of degree < 1, i.e., for a constant $r \in R$. However, since $\varphi_x(r) = \varphi_x(f) = 0$, this means that $r = 0$ and therefore $f \in (X - x)$. Thus, $\ker\varphi_x = (X - x)$.

Using the surjectivity of φ_x, the fundamental theorem 2.3/5 on homomorphisms yields an isomorphism $R[X]/\ker\varphi_x \xrightarrow{\sim} R$. Furthermore, 2.3/8 shows that $\ker\varphi_x$ is prime if and only if R is an integral domain, and maximal if and only if R is a field.

2.4, Exercise 1. Let R be a ring. We want to show that the polynomial ring $R[X]$ is a principal ideal domain if and only if R is a field. The condition is sufficient, as we have seen in 2.4/3. To show that it is necessary as well, assume that $R[X]$ is a principal ideal domain. In particular, $R[X]$ is then an integral domain, and the same is true for R. Furthermore, observe that the element X is irreducible in $R[X]$. Indeed, consider a factorization $X = fg$ for polynomials $f, g \in R[X]$. Then 2.1/2 yields $\deg f + \deg g = 1$, say $\deg f = 0$ and $\deg g = 1$. Therefore, the polynomial f is constant, i.e., is given by an element in R, and the product of f with the coefficient of degree 1 in g equals 1, due to the equation $X = fg$. However, this shows that f is a unit in R and $R[X]$, so that X is irreducible in $R[X]$.

Now consider the substitution homomorphism $\varphi\colon R[X] \longrightarrow R$, $h \longmapsto h(0)$. It is surjective and satisfies $\ker\varphi = (X)$. Therefore, it gives rise to an isomorphism $R[X]/(X) \simeq R$, due to the fundamental theorem on homomorphisms 2.3/5. Since X is irreducible, we conclude from 2.4/6 that the ideal (X) is maximal in $R[X]$. But then $R[X]/(X) \simeq R$ is a field by 2.3/8.

2.4, Exercise 2. Let R be a unique factorization domain. If the ideal generated by two elements $x, y \in R$ is always principal in R, then, by an inductive argument, every *finitely* generated ideal in R is principal. Furthermore, using the fact that R is a unique factorization domain, we can conclude that *every* ideal in R is principal and hence that R is a principal ideal domain. Indeed, if there exists an ideal $\mathfrak{a} \subset R$ that is not finitely generated, one can find in $\mathfrak{a}$ a sequence of elements $a_1, a_2, \ldots$ giving rise to a strictly ascending chain of ideals

$$(a_1) \subsetneq (a_1, a_2) \subsetneq (a_1, a_2, a_3) \subsetneq \ldots \ .$$

Since each member of this chain is finitely generated and hence principal, say generated by some element x_i, we can write the chain just as well as

$$(x_1) \subsetneq (x_2) \subsetneq (x_3) \subsetneq \ldots,$$

where in each case x_{i+1} is a nontrivial divisor of x_i. This implies that the number of prime factors in the prime factorization of x_{i+1} will be strictly smaller than the corresponding number of prime factors for x_i. As a consequence, there cannot exist an infinite chain of the above type. Hence, every ideal of R is finitely generated, and therefore principal, so that R is a principal ideal domain. In particular, we thereby see that the ideal-theoretic characterization of the greatest common divisor is generally valid only in principal ideal domains.

The situation is different for the least common multiple v of two elements $x, y \in R$. Indeed, we get $(x) \cap (y) = (v)$, even if R is just a unique factorization domain. This is easy to justify. Since v is a multiple of x and y, we obtain $(x) \cap (y) \supset (v)$. On the other hand, if $a \in (x) \cap (y)$, i.e., if a is a common multiple of x and y, then a is a multiple of v by its definition, and we get $a \in (v)$, resp. $(x) \cap (y) \subset (v)$.

2.5, Exercise 1. Let R be a (commutative) ring and M a not necessarily commutative monoid. Then the polynomial ring $R[M]$ can be constructed as

in 2.5, since the commutativity of M was not really used. However, care is required if the law of composition on M is still written additively. Indeed, if M is not commutative, there are elements $\mu, \nu \in M$ such that $\mu + \nu \neq \nu + \mu$. In particular, the product $X^\mu \cdot X^\nu = X^{\mu+\nu}$ differs from the product $X^\nu \cdot X^\mu = X^{\nu+\mu}$ then, so that $R[M]$ is a ring that is no longer commutative. In accordance with this, one should not limit 2.5/1 to commutative extension rings R' of R. Instead, one can admit more general rings R' whose elements commute with those of R. The assertion of 2.5/1, including its proof, remains valid without changes.

2.5, Exercise 2. The results 2.5/2, 2.5/3, and 2.5/4 remain valid without changes if we replace $R[X_1, \ldots, X_n]$ by the polynomial ring $R[\mathfrak{X}]$ in an arbitrary system $\mathfrak{X} = (X_i)_{i \in I}$ of variables X_i. The argument is that each element of $R[\mathfrak{X}]$ is, in fact, a polynomial in finitely many of the variables X_i. Therefore, it is enough to know the corresponding results for polynomial rings in finitely many variables. For example, consider the assertion of 2.5/4. First we see that $R^* \subset (R[\mathfrak{X}])^*$ as usual. On the other hand, if f is a unit in $R[\mathfrak{X}]$, there is a polynomial $g \in R[\mathfrak{X}]$ such that $fg = 1$. Since both f and g are polynomials in finitely many variables, we can read the equation $fg = 1$ just as well in a subring of type $R[X_{i_1}, \ldots, X_{i_n}] \subset R[\mathfrak{X}]$. In particular, f is a unit in $R[X_{i_1}, \ldots, X_{i_n}]$. However, in this case we know already that f is a unit in R.

Also 2.5/5 can be generalized to systems of arbitrarily many variables: Given a ring homomorphism $\varphi\colon R \longrightarrow R'$ and a system $(x_i)_{i \in I}$ of elements in R', there is a unique ring homomorphism $\Phi\colon R[X_i\,;\, i \in I] \longrightarrow R'$ satisfying $\Phi|_R = \varphi$ and $\Phi(X_i) = x_i$ for all $i \in I$. It is possible to deduce this assertion from 2.5/5, by considering extensions of φ of type $R[X_{i_1}, \ldots, X_{i_n}] \longrightarrow R'$, where $X_{i_j} \longmapsto x_{i_j}$, and using their uniqueness. However, it is more natural to observe that a monoid homomorphism $\mathbb{N}^{(I)} \longrightarrow R'$ is uniquely determined by the images of the elements $e_j = (\delta_{ij})_{i \in I}$, $j \in I$, and that these images are not subject to any restrictions. Then one can apply 2.5/1.

2.5, Exercise 3. We proceed by a repeated application of 2.5/1. To begin with, let $\Phi'\colon R[M] \longrightarrow R[M \times M']$ be the ring homomorphism that is given by the canonical map $R \hookrightarrow R[M \times M']$ in conjunction with the monoid homomorphism $M \longrightarrow R[M \times M']$, $\mu \longmapsto X^{(\mu,0)}$. Furthermore, there is a homomorphism $\Phi\colon R[M][M'] \longrightarrow R[M \times M']$ that extends Φ' and otherwise corresponds to the monoid homomorphism $M' \longrightarrow R[M \times M']$, $\nu \longmapsto X^{(0,\nu)}$. On the other hand, we can define a ring homomorphism $\Psi\colon R[M \times M'] \longrightarrow R[M][M']$ by the canonical map $R \hookrightarrow R[M] \hookrightarrow R[M][M']$ and the monoid homomorphism $M \times M' \longrightarrow R[M][M']$, $(\mu, \nu) \longmapsto X^\mu \cdot X^\nu$. We claim that Φ and Ψ are inverse to each other, i.e., that $\Phi \circ \Psi = \mathrm{id}$ and $\Psi \circ \Phi = \mathrm{id}$. Indeed, $\Phi \circ \Psi$ and the identity map constitute two ring homomorphisms $R[M \times M'] \longrightarrow R[M \times M']$ that extend the canonical map $R \hookrightarrow R[M \times M']$ and satisfy $X^{(\mu,\nu)} \longmapsto X^{(\mu,\nu)}$, thereby corresponding to the monoid homomorphism $M \times M' \longrightarrow R[M \times M']$, $(\mu, \nu) \longmapsto X^{(\mu,\nu)}$. Then the uniqueness assertion in 2.5/1 shows that $\Phi \circ \Psi = \mathrm{id}$. In a similar way one obtains $\Psi \circ \Phi = \mathrm{id}$, first restricted to $R[M]$, and then on all of $R[M][M']$.

2.6, Exercise 1. Let $n \geq 1$. Concluding by induction on n, we write the polynomial $f \in K[X_1, \ldots, X_n]$ under consideration as a polynomial of type $f = \sum_{i=0}^{\infty} f_i X_n^i$, with coefficients given by polynomials $f_i \in K[X_1, \ldots, X_{n-1}]$. Then, for points $x = (x_1, \ldots, x_n) \in K^n$, we get $f(x) = \sum_{i=0}^{\infty} f_i(x') x_n^i$, where $x' = (x_1, \ldots, x_{n-1})$. Assuming $f(x) = 0$ for all $x \in K^n$, the polynomial $\sum_{i=0}^{\infty} f_i(x') X_n^i \in K[X_n]$ vanishes for every $x' \in K^{n-1}$ on all of K. Thus, we can apply 2.6/1 to conclude that $f_i(x') = 0$ for all $i \in \mathbb{N}$ and all $x' \in K^{n-1}$. In particular, using the induction hypothesis in case $n > 1$, we get $f_i = 0$ for all i and therefore $f = 0$.

2.7, Exercise 1. First observe that the image $\varphi(p)$ of a prime element $p \in R$ is again a prime element. Therefore, if $x = p_1 \ldots p_n$ is a prime factorization of an element $x \in R$, then $\varphi(x) = \varphi(p_1) \ldots \varphi(p_n)$ is a prime factorization of its image $\varphi(x)$. In particular, we get $\nu_{\varphi(p)}(\varphi(x)) = \nu_p(x)$ for $x \in R$. Let us show that more generally, the equation $\nu_{\varphi(p)}(\Phi(f)) = \nu_p(f)$ holds for all prime elements $p \in R$ and all polynomials $f \in R[X]$. Considering besides Φ also its inverse Φ^{-1}, it is enough to show for polynomials $f \neq 0$ that $\nu_{\varphi(p)}(\Phi(f)) \geq \nu_p(f)$ for all p. Now, if $\nu_p(f) = r \geq 0$, we can view $\tilde{f} = p^{-r} f$ as a polynomial in $R[X]$. Furthermore, $\Phi(f) \in R[X]$ implies $\nu_{\varphi(p)}(\Phi(\tilde{f})) \geq 0$. Since $\Phi(f) = \Phi(p^r \tilde{f}) = \varphi(p)^r \Phi(\tilde{f})$, we see that $\nu_{\varphi(p)}(\Phi(f)) \geq r = \nu_p(f)$, as desired. Under the condition that $\varphi(p)$ is always associated to p, for example if $\varphi = \Phi|_R = \mathrm{id}$, we get even $\nu_p(\Phi(f)) = \nu_p(f)$ for all prime elements $p \in R$.

A polynomial $f \in R[X]$ is primitive if and only if $\nu_p(f) = 0$ for all prime elements $p \in R$. Since φ is an isomorphism of R, it induces a bijection on the set of all classes of associated prime elements. Therefore, using the equality $\nu_{\varphi(p)}(\Phi(f)) = \nu_p(f)$ for all p, we conclude that $\Phi(f)$ is primitive if and only if f is primitive. The automorphism $\Phi\colon R[X] \longrightarrow R[X]$, $f \longmapsto f(X + a)$, may serve as an example. Thereby we see that a polynomial $f \in R[X]$ is primitive if and only if $f(X + a)$ is primitive.

2.7, Exercise 2. Gauss's lemma establishes for prime elements $p \in R$ and polynomials $f, g \in K[X]$ the formula $\nu_p(fg) = \nu_p(f) + \nu_p(g)$. If $f, g \neq 0$, we can read from it

$$\prod_{p \in P} p^{\nu_p(fg)} = \prod_{p \in P} p^{\nu_p(f)} \cdot \prod_{p \in P} p^{\nu_p(g)},$$

in other words, the formula $a_{fg} = a_f \cdot a_g$ for the content. Conversely, departing from this formula, it implies $\nu_p(a_{fg}) = \nu_p(a_f) + \nu_p(a_g)$ for $p \in P$. Since the content a_h of a polynomial $h \neq 0$ is characterized by the relation $\nu_p(a_h) = \nu_p(h)$, we get again $\nu_p(fg) = \nu_p(f) + \nu_p(g)$. Therefore the assertion of Gauss's lemma is equivalent to the formula $a_{fg} = a_f \cdot a_g$ for $f, g \neq 0$.

2.9, Exercise 1. If $M = T \oplus F$ is a decomposition into a torsion module T and a free module F, then T is unique, since it will coincide with "the" torsion submodule of M. On the other hand, F is not unique, except when $T = 0$ or $T = M$. Indeed, if we modify certain elements of a basis of F by adding arbitrary torsion elements, the new system will generate a free submodule $F' \subset M$ that is different from F but satisfies $T \oplus F' = M$ as well.

Similarly, there is no uniqueness for decompositions of type $M = M' \oplus M''$, where $M' \simeq A/p^r A$ and $M'' \simeq A/p^s A$, for a prime element p. For example, if $r = s = 1$, we can view M as an (A/p)-vector space. Then $M = M' \oplus M''$ is a direct sum decomposition of a 2-dimensional (A/p)-vector space into two 1-dimensional subspaces. However, such a decomposition is never unique.

2.9, Exercise 2. We claim that $\mathbb{Q}$ is a torsion-free $\mathbb{Z}$-module of rank 1 that is not free. Indeed, $\mathbb{Q}$ is of rank 1, and if it were a free $\mathbb{Z}$-module, it would be generated by a single element $x \in \mathbb{Q}$, say $x = \frac{a}{b}$ with relatively prime integers $a, b \in \mathbb{Z}$, $a, b \neq 0$. Then, however, $\frac{a}{2b}$ cannot be contained in $\mathbb{Z}x$.

2.9, Exercise 3. Let K be a field and V a finite-dimensional K-vector space, together with a K-endomorphism $\varphi\colon V \longrightarrow V$. A subvector space $U \subset V$ is called *φ-invariant* if $\varphi(U) \subset U$. It is called *φ-cyclic* if U is φ-invariant and if there is a vector $u \in U$ such that the sequence $u, \varphi(u), \varphi^2(u), \ldots$ gives rise to a K-generating system of U. Furthermore, U is called *φ-irreducible* if U is φ-invariant and cannot be decomposed into a direct sum of two proper φ-invariant subvector spaces. As main results, the theory of canonical forms shows that V decomposes into a direct sum of φ-irreducible subvector spaces and that each φ-irreducible subvector space is φ-cyclic. We will deduce these assertions from 2.9/8; see also [4], 6.3–6.5.

In order to apply the methods of 2.9, we view V as a $K[X]$-module under φ, by defining the multiplication by X on V through the application of φ; see the explanations given in 2.9. Then a $K[X]$-submodule $U \subset V$ is the same as a φ-invariant K-subvector space. Similarly, a $K[X]$-submodule generated by a single element is just a φ-cyclic K-subvector space. In this way, a φ-irreducible subvector space of V is recognized as a $K[X]$-submodule of V that cannot be written as a direct sum of two proper $K[X]$-submodules.

Since V is finitely generated as a K-vector space, the same is true for V as a $K[X]$-module. Furthermore, by reasons of vector space dimension, V is a $K[X]$-torsion module, so that 2.9/8 can be applied. Fixing a system P of representatives of the prime polynomials in $K[X]$, we obtain a decomposition

$$V \simeq \bigoplus_{p \in P} \bigoplus_{\nu_p=1}^{r_p} K[X]/(p^{n(p,\nu_p)})$$

with unique integers $r_p, n(p, \nu_p) \in \mathbb{N}$, where r_p is zero for almost all $p \in P$. Using the language of vector spaces, this is a decomposition of V into a direct sum of φ-cyclic subvector spaces. Furthermore, one concludes from the uniqueness assertion in 2.9/8 that the occurring subspaces are even φ-irreducible and in addition, that every φ-irreducible subspace is φ-cyclic. This establishes the results on canonical forms mentioned above.

One can now consider special matrices that describe the endomorphism φ with respect to appropriate K-bases of V. To do this, look at a decomposition $V = \bigoplus_{i=1}^{s} V_i$ into φ-irreducible subvector spaces, choose a basis on each V_i, and consider on V the basis that is composed of the chosen bases on the V_i. Then the

corresponding matrix of φ is of "diagonal shape," in the sense that on its "diagonal" there are the matrices corresponding to the endomorphisms $\varphi_i = \varphi|_{V_i}$, with zero entries otherwise. Therefore, it is enough to assume that V is φ-irreducible, say $V = K[X]/(p^n)$ for a prime element $p \in P$. If $\overline{X} \in K[X]/(p^n)$ denotes the residue class of X, then the elements $1, \overline{X}, \overline{X}^2, \ldots, \overline{X}^{m-1}$, for $m = n \cdot (\deg p)$, form a K-basis of V, and the corresponding matrix of φ is determined to be

$$\begin{pmatrix} 0 & 0 & \ldots & 0 & 0 & -c_m \\ 1 & 0 & \ldots & 0 & 0 & -c_{m-1} \\ . & . & \ldots & . & . & . \\ . & . & \ldots & . & . & . \\ 0 & 0 & \ldots & 1 & 0 & -c_2 \\ 0 & 0 & \ldots & 0 & 1 & -c_1 \end{pmatrix},$$

where $p^n = X^m + c_1 X^{m-1} + \ldots + c_m$ is the minimal polynomial of φ. This is the *rational canonical form*, also referred to as the *Frobenius normal form* of φ. Especially, if p is of degree 1, say $p = X - c$, we can take just as well $1, \overline{X} - c, (\overline{X} - c)^2, \ldots, (\overline{X} - c)^{n-1}$ as a K-basis of V. Then the corresponding matrix of φ is of shape

$$\begin{pmatrix} \mathrm{c} & 0 & \ldots & 0 & 0 \\ 1 & \mathrm{c} & \ldots & 0 & 0 \\ . & . & \ldots & . & . \\ 0 & 0 & \ldots & \mathrm{c} & 0 \\ 0 & 0 & \ldots & 1 & \mathrm{c} \end{pmatrix},$$

and we obtain the *Jordan canonical form* of φ.

3.1, Exercise 1. Let $\sigma\colon R \longrightarrow R'$ be a homomorphism between two integral domains R, R' of characteristic p, resp. p'. Then the homomorphisms $\varphi\colon \mathbb{Z} \longrightarrow R$, $n \longmapsto n \cdot 1$, and $\varphi'\colon \mathbb{Z} \longrightarrow R'$, $n \longmapsto n \cdot 1$, satisfy $\ker\varphi = p\mathbb{Z}$, as well as $\ker\varphi' = p'\mathbb{Z}$. Now observe that there is a unique ring homomorphism from $\mathbb{Z}$ to R'. Therefore, we get $\varphi' = \sigma \circ \varphi$ and hence $\ker\varphi \subset \ker\varphi'$, which implies $p' \,|\, p$. Furthermore, we even have $\ker\varphi = \ker\varphi'$, and hence $p = p'$, if σ is injective. In particular, since homomorphisms of fields are always injective, it follows that there cannot exist homomorphisms between fields of different characteristic.

On the other hand, the canonical map $\mathbb{Z} \longrightarrow \mathbb{Z}/p'\mathbb{Z}$ for p' prime yields an example of a homomorphism between integral domains of characteristic 0 and p'. However, this is the only case of "mixed" characteristic that can occur. Indeed, if $\sigma\colon R \longrightarrow R'$ is a homomorphism between integral domains of characteristic p and p', then $p' \,|\, p$, as we have seen. Since p and p' are (positive) prime numbers if they do not vanish, we necessarily get $p = 0$ for $p \neq p'$.

3.2, Exercise 1. We have to consider a field extension L/K, and two elements $a, b \in L$ that are algebraic over K. To show that $a + b$ is algebraic over K, we could try to explicitly construct from the two minimal polynomials of a and b a nontrivial polynomial admitting $a+b$ as a zero. However, experience shows that such a procedure is not really practicable. A superficial reason for it consists in

the fact that in looking at an expression $f(a+b)$ for a polynomial $f \in K[X]$ of degree ≥ 2, the quantities a and b cannot be "separated" in general, for instance by writing $f(a+b)$ as a sum of a polynomial in a and a polynomial in b. Just consider the field extension $\mathbb{C}/\mathbb{Q}$ and the algebraic numbers $a = \sqrt{2}$ and $b = \sqrt{3}$ as an example. The minimal polynomial of a is $X^2 - 2$, and that of b is $X^2 - 3$. Using the method of Exercise 7 in Section 3.2, the minimal polynomial of $a+b$ is determined to be $X^4 - 10X^2 + 1$. This is a polynomial for which no "obvious" relationship to the minimal polynomials $X^2 - 2$ and $X^2 - 3$ is visible.

Thus, to verify the algebraicity of $a + b$, there is no other way than to apply the theory developed in Section 3.2. We know from 3.2/6 that $K(a)/K$ and $K(a,b)/K(a)$ are finite field extensions. Then $K(a,b)/K$ is finite by the multiplicativity formula 3.2/2 and hence algebraic by 3.2/7. In particular, the sum $a + b \in K(a,b)$ is algebraic over K.

3.2, Exercise 2. We have shown in 3.2/7 that every finite field extension is algebraic. On the other hand, the example of the algebraic closure $\overline{\mathbb{Q}}$ of $\mathbb{Q}$ in $\mathbb{C}$ shows that the converse of this result is not true. However, we can say that a field extension L/K is algebraic if and only if there is a family $(L_i)_{i \in I}$ of intermediate fields of L/K such that $L = \bigcup_{i \in I} L_i$ and L_i/K is finite for every i. Indeed, if this condition holds, then there exists for each $a \in L$ an index $i \in I$ such that $a \in L_i$. It follows that a is algebraic over K and hence that L is algebraic over K. On the other hand, if L/K is algebraic, then L is the union of all intermediate fields $K(a)$, where a varies over L. Furthermore, $K(a)/K$ is finite by 3.2/6.

We want to add that the algebraicity of a field extension L/K can also be characterized by the following condition: Every subfield L' of L that is finitely generated over K is finite over K. Indeed, if L/K is algebraic and if L' is a subfield of L that is finitely generated over K, then L'/K is finite by 3.2/9. Hence, the condition is necessary. But it is also sufficient. For every $a \in L$, the field extension $K(a)/K$ is simple and hence finitely generated. If it is finite in each case, it is algebraic by 3.2/7, and we see that L/K is algebraic.

3.2, Exercise 3. Consider an element $a \in \mathbb{C}$ that is not contained in the algebraic closure $\overline{\mathbb{Q}}$ of $\mathbb{Q}$ in $\mathbb{C}$. If a is not transcendental over $\overline{\mathbb{Q}}$, it is algebraic over $\overline{\mathbb{Q}}$. However, then a is algebraic already over $\mathbb{Q}$ by 3.2/12 and hence must belong to $\overline{\mathbb{Q}}$, in contradiction to the choice of a.

3.3, Exercise 1. Assume $A \neq 0$ and hence $B \neq 0$. As usual, given $b \in B$, we consider the ring homomorphism $\varphi\colon A[Y] \longrightarrow B$ that extends the inclusion $A \hookrightarrow B$ and maps Y to b. Since b satisfies an integral equation over A, it follows that $\ker \varphi$ contains certain monic polynomials; choose one with minimal degree among these, say f. Now, if A is a field K, then $f \in K[Y]$ is uniquely determined by b. Indeed, $\ker \varphi$ is a principal ideal in $K[Y]$, and $\ker \varphi$ is generated by f. Since such a generator is unique up to a unit, we see that f is unique up to a unit in $K[Y]$, i.e., up to a constant in K^*. Therefore, assuming f to be monic, it is uniquely determined by b.

In the general case, however, $\ker \varphi$ does not need to be principal in $A[Y]$. In fact, it is most likely that there will exist several monic polynomials of minimal degree in $\ker \varphi$. As a consequence, none of these can be interpreted as being "the" minimal polynomial of b over A. In particular, looking at the example $A = \{\sum c_i X^i \in K[X]\,;\, c_1 = 0\} \subset K[X] = B$ given in the exercise, then

$$Y^2 - X^2, \qquad Y^2 + X^2 Y - (X^3 + X^2)$$

are two different monic polynomials of minimal degree in $A[Y]$ that admit $b := X$ as a zero. But none of these will generate the ideal $\ker \varphi$.

3.4, Exercise 1. We assume that the polynomial $f \in \mathbb{Q}[X]$ is irreducible; otherwise, we could replace f by an irreducible factor. Then we know from Kronecker's construction, Proposition 3.4/1, that $\mathbb{Q}[X]/(f)$ can be viewed as an extension field of $\mathbb{Q}$ and that the residue class $\overline{X}$ of the variable X is a zero of f. Thereby we have enlarged $\mathbb{Q}$ in a minimal way, so to speak, with the only intent to get hold of a zero of f, without relating it to the real or complex numbers. Studying the same problem in analysis, one constructs first of all from $\mathbb{Q}$ the field $\mathbb{R}$ of real numbers in terms of topological arguments, and subsequently the field $\mathbb{C}$ of complex numbers. Only thereafter is one concerned with zeros of polynomials in these special fields. The construction of such zeros uses in a fundamental way approximation and limit techniques, since the defining properties of $\mathbb{R}$ and $\mathbb{C}$ must come into play, in particular the completeness. Finally, if a zero $a \in \mathbb{C}$ of f has been exhibited, the homomorphism $\mathbb{Q} \longrightarrow \mathbb{C}$ can be extended by 3.4/8 to a homomorphism $\mathbb{Q}[X]/(f) \longrightarrow \mathbb{C}$, mapping $\overline{X}$ to a.

3.4, Exercise 2. In order to apply Zorn's lemma one needs a partially ordered *set*. However, in general the "collection" of all algebraic extensions of K fails to admit the properties of a set.

On the other hand, the proposed argument can be remedied in a certain sense if some set-theoretic precautions are respected. Indeed, consider the power set P of K, and view K as a subset of P via the map $K \longrightarrow P$, $a \longmapsto \{a\}$. Let M be the set of all pairs (L, κ) consisting of a set L, where $K \subset L \subset P$, and a field structure κ on L extending the given field structure on K, and defining L as an algebraic extension of K. Then M is partially ordered by writing $(L, \kappa) \leq (L', \kappa')$ if $L \subset L'$ and if κ' restricts on L to κ. As usual, the argument on unions shows that every totally ordered subset of M admits an upper bound in M. Therefore, we can conclude from Zorn's lemma 3.4/5 that M contains a maximal element. Let us denote such an element by (L_1, κ_1); it constitutes an algebraic extension of K.

We claim that (L_1, κ_1) is already an algebraic closure of K, assuming that K consists of infinitely many elements. To justify this we have to show that (L_1, κ_1) does not admit any nontrivial algebraic extensions. Therefore, consider an algebraic extension E of (L_1, κ_1), which then is algebraic over K as well. Now we use some facts on cardinalities of sets, namely that K and $K[X]$ (for K infinite) are of same cardinality. Thus, they have the same cardinality as L_1 and E, since those fields can be viewed as unions of zero sets of polynomials in

$K[X]$. However, P as the power set of K admits a cardinality that is strictly bigger than that of K, resp. E. Therefore, the same is true for $P - L_1$, and it follows that the inclusion $L_1 \hookrightarrow P$ can be extended to an inclusion $E \hookrightarrow P$. As a result, we get an element $(L_2, \kappa_2) \in M$ such that $(L_1, \kappa_1) \leq (L_2, \kappa_2)$. But then the maximality of (L_1, κ_1) yields $L_1 = L_2$, resp. $(L_1, \kappa_1) = E$, and it follows that (L_1, κ_1) is an algebraic closure of K. For finite fields K, the argument can easily be modified. Just enlarge K to an infinite set K' and define P as the power set of K'.

3.4, Exercise 3. Although two algebraic closures $\overline{K}_1$ and $\overline{K}_2$ of a field K are isomorphic over K, there usually exist several different K-isomorphisms $\overline{K}_1 \xrightarrow{\sim} \overline{K}_2$, i.e., isomorphisms leaving K fixed; cf. 3.4/8 in conjunction with the construction process in the proof of 3.4/9. If we would talk about "the" algebraic closure $\overline{K}$ of K, we would allege an identification of all possible algebraic closures of K. Indeed, given two such closures $\overline{K}_i$ and $\overline{K}_j$, we would have to fix a special isomorphism $\varphi_{ij} \colon \overline{K}_i \xrightarrow{\sim} \overline{K}_j$ such that $\varphi_{ij}|_K = \mathrm{id}_K$, where the compatibility relation $\varphi_{ik} = \varphi_{jk} \circ \varphi_{ij}$ for any three indices i, j, k must be respected. However, since canonical choices for such K-isomorphisms do not exist, except for trivial cases, an identification of all algebraic closures of K is rather problematic.

3.5, Exercise 1. Let L/K be a field extension of degree 2. Choosing an element $a \in L - K$, we get $1 < [K(a) : K] \leq 2$ and hence $[K(a) : K] = 2$, which means that $L = K(a)$. Let $f \in K[X]$ be the minimal polynomial of a over K; it is of degree 2. Then a is a zero of f, and the linear factor $X - a$ divides f in $L[X]$. In particular, f decomposes completely over L into linear factors. Let $a, b \in L$ be the two zeros of f. Since $L = K(a) = K(a, b)$, we see that L is a splitting field of f over K and hence that it is normal over K.

3.5, Exercise 2. We are considering a splitting field L of a nonconstant polynomial $f \in K[X]$. Furthermore, let $g \in K[X]$ be an irreducible polynomial admitting a zero b in L. To see that L contains all possible zeros of g, choose an algebraic closure $\overline{L}$ of L and let $b_1, \ldots, b_r \in \overline{L}$ be the distinct zeros of g. Then there is for each $i = 1, \ldots, r$ a K-homomorphism $\sigma_i \colon K(b) \longrightarrow \overline{L}$ such that $\sigma_i(b) = b_i$, see 3.4/8, and we can extend σ_i by 3.4/9 to a K-homomorphism $\sigma_i' \colon L \longrightarrow \overline{L}$.

It is enough to show that $\sigma_i'(L) \subset L$ for $i = 1, \ldots, r$, since then all zeros $b_1, \ldots, b_r$ of g are contained in L, and hence g decomposes over L into linear factors. Since σ_i' leaves the field K fixed, it maps zeros of f to zeros of f again. However, as L is generated over K by all zeros of f in $\overline{L}$, we get $\sigma_i'(L) \subset L$, as desired.

3.5, Exercise 3. Choose an algebraic closure $\overline{K}$ of L. Then $\overline{K}$ is an algebraic closure of K as well, since L/K is algebraic. Now consider an element $a \in \overline{K}$ with minimal polynomial $f \in K[X]$. Since f is nonconstant and since L is a splitting field of all nonconstant polynomials in $K[X]$, it follows that f decomposes completely over L into linear factors. Thus, $a \in L$ and $L = \overline{K}$, i.e., L is an algebraic closure of K.

3.6, Exercise 1. We proceed similarly as in 3.2, Exercise 1. Since $a \in L$ is separable over K, we read $[K(a) : K]_s = [K(a) : K]$ from 3.6/6. Furthermore, $b \in L$ is separable over K and in particular over $K(a)$, so that we get in the same way $[K(a,b) : K(a)]_s = [K(a,b) : K(a)]$. Then we apply the multiplicativity formulas 3.2/2 and 3.6/7 to conclude that $[K(a,b) : K]_s = [K(a,b) : K]$. To see that $a + b \in K(a,b)$ is separable over K, we can use the implication from (iii) to (i) in 3.6/9. Alternatively, to give a more elementary argument, look at the extensions $K \subset K(a+b) \subset K(a,b)$. Then

$$\begin{aligned}[K(a,b) : K] &= [K(a,b) : K(a+b)] \cdot [K(a+b) : K],\\ [K(a,b) : K]_s &= [K(a,b) : K(a+b)]_s \cdot [K(a+b) : K]_s.\end{aligned}$$

Both terms on the left-hand side coincide. Since the separable degree equals at most the usual degree, see 3.6/6, the corresponding terms on the right-hand side must coincide as well, so that in particular, $[K(a+b) : K]_s = [K(a+b) : K]$. But then, again by 3.6/6, we conclude that $a + b$ is separable over K.

The same argument can be applied to the elements $a - b$, ab, and ab^{-1} if $b \neq 0$. In this way, it follows that the elements of L that are separable over K form an intermediate field of L/K.

3.6, Exercise 2. Given two algebraic closures $\overline{K}_1$ and $\overline{K}_2$ of K, there is a K-isomorphism $\sigma\colon \overline{K}_1 \xrightarrow{\sim} \overline{K}_2$ according to 3.4/10. Let f be a monic polynomial in $K[X]$ and consider its factorizations

$$f = \prod_{i=1}^{m} (X - a_i)^{r_i}, \qquad f = \prod_{i=1}^{n} (X - b_i)^{s_i},$$

into powers of distinct linear factors in $\overline{K}_1[X]$, as well as in $\overline{K}_2[X]$. Then, using the uniqueness of prime factorization, σ transports the first decomposition into the second. In particular, we have $m = n$, and up to a renumbering of the b_i even $\sigma(a_i) = b_i$ for $i = 1, \ldots, m$, as well as $r_i = s_i$. Therefore, f admits multiple zeros in $\overline{K}_1$ if and only if this is the case in $\overline{K}_2$.

3.6, Exercise 3. Let L/K be a finite separable field extension; we are interested only in the case that K consists of infinitely many elements. Using recursion, we can restrict ourselves to the case $L = K(a,b)$. Let f and g be the minimal polynomials of a and b over K, and let L' be a splitting field of f, g over L. Then L' is also a splitting field of f, g over K, and in fact a normal closure of L/K; cf. 3.5/7. In the proof of 3.6/12 we considered all K-homomorphisms $\sigma_1, \ldots, \sigma_n$ from L into an algebraic closure $\overline{K}$ of K. However, since we can assume $L' \subset \overline{K}$, it follows in this case from 3.5/4 that the images of the σ_i are already contained in L'. In other words, it is enough to choose a normal closure L'/K of L/K and to consider all K-homomorphisms $\sigma_1, \ldots, \sigma_n$ from L to L'. Then, looking for an element $c \in K$ such that $\sigma_i(a + cb) \neq \sigma_j(a + cb)$ for all $i \neq j$, we get $K(a,b) = K(a + cb)$.

3.7, Exercise 1. Assume that the elements $a, b \in L$ are purely inseparable over K. This means by 3.7/2 that there exist equations $a^{p^m} = c$ and $b^{p^n} = d$ for

some elements $c, d \in K$. Taking a suitable p-power of one of these equations, we may assume $m = n$. Then the binomial formula 3.1/3 yields the relation $(a+b)^{p^m} = a^{p^m} + b^{p^m} = c + d$. Furthermore, since $(ab)^{p^m} = cd$, this shows, using 3.7/2 again, that $a+b$ and ab are purely inseparable over K. Alternatively, we can conclude similarly as in Exercise 1 of Section 3.2, or as in Exercise 1 of Section 3.6.

3.7, Exercise 2. Purely inseparable field extensions L/K are characterized by the equation $[L : K]_s = 1$. Alternatively, we could write this condition as $[L : K]_i = [L : K]$, however, only in cases in which the degree $[L : K]$ is *finite*. Thus, if we want to work with the inseparable degree instead of the separable degree, we always have to restrict ourselves to finitely generated extensions, similarly as we did when discussing separable extensions in Section 3.6.

3.7, Exercise 3. Let $K(a)/K$ be a simple algebraic field extension with minimal polynomial $f \in K[X]$ of a over K. As in 3.6/2, we can find a polynomial $g \in K[X]$ and a maximal exponent r such that $f(X) = g(X^{p^r})$. Then g is separable, and furthermore, it coincides with the minimal polynomial of a^{p^r} over K. In particular, $K(a)/K(a^{p^r})$ is purely inseparable and $K(a^{p^r})/K$ is separable.

3.8, Exercise 1. Fields of characteristic 0 are perfect (3.6/4). The same is true for finite fields, or more generally, for fields that are algebraic over a finite field (3.8/4). Thus, in order to construct an example of a nonseparable field extension, we must start out from an infinite field K of characteristic $p > 0$ that is not algebraic over its prime subfield $\mathbb{F}_p$. The simplest example of such a field is the function field $K = \mathbb{F}_p(t)$ in a variable t. Adjoining to K a pth root of t yields a nontrivial purely inseparable field extension of K. Applying the Frobenius homomorphism this extension can be identified with $\mathbb{F}_p(t)/\mathbb{F}_p(t^p)$.

3.8, Exercise 2. If $\mathbb{F}$ is a finite field of characteristic $p > 0$ with $q = p^n$ elements, then $\mathbb{F}$ is a splitting field of the polynomial $X^q - X$ over $\mathbb{F}_p$. More precisely, $\mathbb{F}$ consists of the q zeros of this polynomial. Therefore, if $\mathbb{F}$ is a subfield of a field L, it is uniquely characterized by the number of its elements.

3.9, Exercise 1. In the beginning of Section 3.9 we did not use the fact that the zeros of the corresponding polynomials are considered in the affine n-space of an *algebraically closed* field. In this way, the assertion of 3.9/1 remains valid if $\overline{K}$ is replaced by K and $V(\cdot)$ by $V_K(\cdot)$. Also we can read from 3.9/2 that algebraic sets of type $V_K(E)$ can always be defined through finitely many polynomials in $K[X]$ and hence are of type $V_K(f_1, \ldots, f_r)$. Furthermore, in 3.9/3 we obtain the relation $V_K(I(U)) = U$ for subsets $U \subset K^n$ of type $U = V_K(\mathfrak{a})$ and ideals $\mathfrak{a} \subset K[X]$. However, the equation $I(V(\mathfrak{a})) = \mathfrak{a}$ for reduced ideals $\mathfrak{a} \subset K[X]$, which is, so to speak, the essence of Hilbert's Nullstellensatz 3.9/4, does not remain valid. Just consider for $K = \mathbb{R}$ and $n = 1$ the ideal $\mathfrak{a} = (X^2+1) \subset \mathbb{R}[X]$. Then $V_{\mathbb{R}}(\mathfrak{a}) = \emptyset$ and hence $I(V_{\mathbb{R}}(\mathfrak{a})) = \mathbb{R}[X] \neq \mathfrak{a}$. Therefore, in the setting of Hilbert's Nullstellensatz, one cannot abandon the condition that the zeros are considered in the affine n-space of an algebraically closed field.

4.1, Exercise 1. If L/K is a finite Galois extension, then by the fundamental theorem of Galois theory 4.1/6, the intermediate fields of L/K correspond bijectively to the subgroups of the Galois group $\mathrm{Gal}(L/K)$. This fact was used in 4.1/8 to see that a finite separable field extension admits only finitely many intermediate fields, a result that does not extend to nonseparable (finite) extensions. Since the intermediate fields of L/K are characterized as the fixed fields of the subgroups of $\mathrm{Gal}(L/K)$, see 4.1/6 again, such fields can often be explicitly computed if the Galois automorphisms and the group structure of $\mathrm{Gal}(L/K)$ are sufficiently well known. Related to this fact is another aspect of Galois theory. To specify a finite Galois extension L/K is the same as to give a field L together with a finite group G of automorphisms on L. Indeed, from L/K we get the Galois group $G = \mathrm{Gal}(L/K)$ as a finite group of automorphisms on L, and from L together with such a group G we can rediscover K as the fixed field L^G with respect to G; cf. 4.1/4 and 4.1/6. We will further deepen this point of view in Section 4.11, on Galois descent.

4.1, Exercise 2. Let L/K be a quasi-Galois field extension with automorphism group $G = \mathrm{Aut}_K(L)$. Then L/L^G is a Galois extension with Galois group G, according to 4.1/5 (i). Furthermore, $L^G = K_i$ is (for $\mathrm{char}\, K > 0$) the maximal purely inseparable extension of K in L; cf. 3.7/5 and 4.1/5 (iii). Therefore, the assertion of 4.1/6 can be generalized by stating that the subgroups of G correspond, in the manner of 4.1/6, bijectively to those intermediate fields of L/K that contain the maximal purely inseparable extension K_i as a subfield.

4.1, Exercise 3. If L/K is a Galois extension with $\mathrm{Gal}(L/K) = \mathrm{Aut}_K(L)$ as Galois group, then we conclude from 4.1/5 (ii) that K is the fixed field of L with respect to the automorphism group $\mathrm{Aut}_K(L)$. The converse of this follows from 4.1/4.

4.2, Exercise 1. If L/K is an arbitrary Galois extension, then L can be viewed as the union over the system $(L_i)_{i\in I}$ of all intermediate fields of L/K that are finite and Galois over K; see the beginning of Section 4.2. It follows that an element $a \in L$, say $a \in L_i$, is invariant under a subgroup $H \subset \mathrm{Gal}(L/K)$ if and only if a is invariant under the image $H_i = f_i(H)$ of H with respect to the restriction map $f_i\colon \mathrm{Gal}(L/K) \longrightarrow \mathrm{Gal}(L_i/K)$. In other words, we have $L^H \cap L_i = L_i^{H_i}$ and hence $L^H = \bigcup_{i\in I} L_i^{H_i}$. On the other hand, if we consider an intermediate field E of L/K, we can look at the subgroup $H = \mathrm{Gal}(L/E)$ of $\mathrm{Gal}(L/K)$ corresponding to E. Then $H = \bigcap_{i\in I} f_i^{-1}(\mathrm{Gal}(L_i/L_i \cap E))$, as well as $f_i(H) = \mathrm{Gal}(L_i/L_i \cap E)$, where the latter equality is obtained with the help of the extension argument 3.4/9. In terms of these formulas, the Galois theory of L/K is, so to speak, reduced to the Galois theories of the extensions L_i/K. Making use of the fundamental theorem 4.1/6 for these extensions, we obtain for an intermediate field E of L/K with Galois group $H = \mathrm{Gal}(L/E)$ immediately $L^H \cap L_i = E \cap L_i$ and hence $L^H = E$, since $f_i(H) = \mathrm{Gal}(L_i/L_i \cap E)$. On the other hand, departing from a subgroup $H \subset \mathrm{Gal}(L/K)$ and considering its fixed field L^H, we get $\mathrm{Gal}(L/L^H) = \bigcap_{i\in I} f_i^{-1}(f_i(H))$, a group that contains H and is different from H in general. In this respect the general version 4.2/3 of the

fundamental theorem of Galois theory differs from the version 4.1/6 for finite Galois extensions.

4.2, Exercise 2. As we have just explained, the Galois theory of a Galois extension L/K is characterized by the Galois theories of the extensions L_i/K, $i \in I$, where $(L_i)_{i\in I}$ is the system of all intermediate fields of L/K that are finite and Galois over K. In the light of this fact it is natural to identify a Galois automorphism $\sigma\colon L \longrightarrow L$ with the system of its restrictions $(\sigma|_{L_i})_{i\in I}$. Implementing this point of view in a consistent way, we arrive at the interpretation of $\mathrm{Gal}(L/K)$ as a projective limit of the Galois groups $\mathrm{Gal}(L_i/K)$, hence of $\mathrm{Gal}(L/K)$ as a profinite group. Thereby $\mathrm{Gal}(L/K)$ carries a natural topology, namely the one induced from the discrete topologies on the groups $\mathrm{Gal}(L_i/K)$. As we have seen in 4.2/3, resp. 4.2/4, this topology is able to characterize those subgroups in $\mathrm{Gal}(L/K)$ that occur as Galois groups $\mathrm{Gal}(L/E)$ for intermediate fields E of L/K; indeed, these are precisely the closed subgroups of $\mathrm{Gal}(L/K)$. On the other hand, if for an infinite Galois extension L/K, the corresponding Galois group $\mathrm{Gal}(L/K)$ is known merely as an abstract group, without any indications regarding the corresponding topology, this is only of minor importance for characterizing the Galois theory of L/K. For the study of infinite Galois groups $\mathrm{Gal}(L/K)$, there is the choice of either introducing their topology in a direct way, as in 4.2/1, or relying on the formalism of projective limits. Usually the latter bears advantages for explicit computations; see, for example, 4.2/11.

4.3, Exercise 1. Every group G can be viewed as a subgroup of the group of its bijective self-maps $G \longrightarrow G$, by identifying an element $a \in G$ with its corresponding left translation $\tau_a\colon G \longrightarrow G$, $g \longmapsto ag$. Therefore, to solve our problem, it is enough to show that every subgroup G of a permutation group $\mathfrak{S}_n$ can occur as a Galois group. In fact, this is easy to achieve. Consider the rational function field $L = k(T_1, \ldots, T_n)$ in n variables $T_1, \ldots, T_n$ over a coefficient field k. Similarly as we did when considering the generic equation of degree n, we can view G as a subgroup of the isomorphism group of L, by interpreting the elements of G as permutations of the variables $T_1, \ldots, T_n$. Then L/L^G is a Galois extension with Galois group G by 4.1/4. However, much more difficult and partially still open is the question whether a given finite group can be realized as the Galois group of an extension $L/\mathbb{Q}$.

4.4, Exercise 1. By its characterization given in 4.4/3, the discriminant Δ_f of a monic polynomial f with coefficients in a ring R provides a certain measure for the distance between the zeros of f, even if they become available only after extending R. To compute Δ_f, we could extend R until f factorizes completely into linear polynomials, and then try to calculate the product over the squares of the differences of the zeros of f. However, such a procedure is not very practicable. Just think of the difficulties that occur in looking for explicit factorizations of polynomials with coefficients in $\mathbb{Q}, \mathbb{R}$, or $\mathbb{C}$. Instead, one carries out a calculation in a "universal" setting and shows that the result can be transported via suitable ring homomorphisms to all other situations. Indeed, one considers the polynomial $f = \prod_{i=1}^{n}(X - T_i)$ in the variable X over the

coefficient ring $\mathbb{Z}[T_1, \ldots, T_n]$. At this point, the fundamental theorem on symmetric polynomials 4.4/1 enables us to write the discriminant Δ_f as an integer polynomial in the elementary symmetric polynomials $s_1, \ldots, s_n$, which occur as the coefficients of f. The resulting solution for Δ_f can then be transported via ring homomorphisms to any other coefficient domain. In this way, one obtains a formula for Δ_f that is valid independently of the coefficient rings under consideration.

If we restrict the setting of the fundamental theorem on symmetric polynomials 4.4/1 to polynomials over a field K, then the generic polynomial $f = \prod_{i=1}^{n}(X - T_i)$ has to be seen over the coefficient ring $K[T_1, \ldots, T_n]$. As a consequence, Δ_f is obtained as a polynomial in $s_1, \ldots, s_n$, however, now with coefficients that are known only to belong to K. This is why for variable characteristic of the fields under consideration, there is no chance any longer, to relate the different formulas for Δ_f to each other.

4.5, Exercise 1. Let $\Phi_n = g_1 \ldots g_r$ be the prime factorization of the cyclotomic polynomial $\Phi_n \in K[X]$ and observe that the factors $g_1, \ldots, g_r$ are distinct, since Φ_n is separable. Since the zeros of the g_i coincide with the primitive nth roots of unity, we can view every g_i as the minimal polynomial over K of a primitive nth root of unity. Each of these roots generates the same extension field of K, namely $K(\zeta)$, and we see that $\deg g_i = [K(\zeta) : K] = s$ for all i. Since Φ_n is of degree $\varphi(n)$, we get $r = \varphi(n)/s$, as claimed.

4.5, Exercise 2. For $m, n \in \mathbb{N} - \{0\}$, consider primitive mth and nth roots of unity $\zeta_m, \zeta_n \in \overline{\mathbb{Q}}$. Then we get $[\mathbb{Q}(\zeta_m, \zeta_n) : \mathbb{Q}(\zeta_m)] \leq \deg \Phi_n = \varphi(n)$, since ζ_n is a zero of Φ_n. Furthermore, we see that Φ_n is irreducible over $\mathbb{Q}(\zeta_m)$ if and only if $[\mathbb{Q}(\zeta_m, \zeta_n) : \mathbb{Q}(\zeta_m)] = \varphi(n)$, i.e., using $[\mathbb{Q}(\zeta_m) : \mathbb{Q}] = \varphi(m)$, if and only if $[\mathbb{Q}(\zeta_m, \zeta_n) : \mathbb{Q}] = \varphi(m) \cdot \varphi(n)$. To look more closely at the latter relation, we compute the degree of $\mathbb{Q}(\zeta_m, \zeta_n)/\mathbb{Q}$. Let $k = \operatorname{lcm}(m, n)$. Then $\mathbb{Q}(\zeta_m, \zeta_n)$ contains a primitive kth root of unity ζ according to 3.6/13, so that we get $\mathbb{Q}(\zeta_m, \zeta_n) = \mathbb{Q}(\zeta)$ and hence $[\mathbb{Q}(\zeta_m, \zeta_n) : \mathbb{Q}] = \varphi(k)$.

Our considerations show that Φ_n is irreducible over $\mathbb{Q}(\zeta_m)$ if and only if the equation $\varphi(\operatorname{lcm}(m, n)) = \varphi(m) \cdot \varphi(n)$ holds. Now choose decompositions $m = m_0 m'$ and $n = n_0 n'$ such that $\operatorname{lcm}(m, n) = m_0 n_0$ and $\gcd(m_0, n_0) = 1$, as in 3.6/13. Then we get

$$\varphi\big(\operatorname{lcm}(m, n)\big) = \varphi(m_0) \cdot \varphi(n_0) \leq \varphi(m) \cdot \varphi(n)$$

from 4.5/4, where equality holds if and only if $\varphi(m_0) = \varphi(m)$ and $\varphi(n_0) = \varphi(n)$. Further, we read from the explicit formula in 4.5/4 (iii) that $\varphi(m_0) = \varphi(m)$ is equivalent to $m' \in \{1, 2\}$. The corresponding fact holds for the decomposition $n = n_0 n'$, and we see that Φ_n is irreducible over $\mathbb{Q}(\zeta_m)$ if and only if $\gcd(m, n) \in \{1, 2\}$.

4.6, Exercise 1. Assume $\mathbb{F} = \mathbb{F}_q$, where q is a power of a prime number p. Then the multiplicative group of $\mathbb{F}_q$ is cyclic of order $q - 1$ by 3.8/5. Therefore, we have to determine all group homomorphisms $G \longrightarrow \mathbb{Z}/(q-1)\mathbb{Z}$. Now let ζ be a generating element of G, where in a first step, we assume that ζ is of

infinite order. Then we can define for every element $a \in \mathbb{Z}/(q-1)\mathbb{Z}$ a unique group homomorphism $G \longrightarrow \mathbb{Z}/(q-1)\mathbb{Z}$ via $\zeta \longmapsto a$. Thus, in this case, there are $q-1$ characters on G with values in $\mathbb{F}^*$.

It remains to consider a cyclic group G of finite order $m > 0$. Indeed, if $G \longrightarrow \mathbb{Z}/(q-1)\mathbb{Z}$ is a homomorphism with image a of ζ, we get $m \cdot a = 0$, and the order of a divides m. On the other hand, we can define for each element $a \in \mathbb{Z}/(q-1)\mathbb{Z}$ whose order divides m a homomorphism $G \longrightarrow \mathbb{Z}/(q-1)\mathbb{Z}$ via $\zeta \longmapsto a$. Therefore, the desired homomorphisms $G \longrightarrow \mathbb{Z}/(q-1)\mathbb{Z}$ correspond bijectively to the elements in $\mathbb{Z}/(q-1)\mathbb{Z}$ whose order divides m. An elementary calculation shows that their number equals $\gcd(m, q-1)$.

4.7, Exercise 1. Viewing L as a K-vector space, the map $\mathrm{tr}_{L/K} \colon L \longrightarrow K$ is a linear functional on L, and its kernel $\ker \mathrm{tr}_{L/K} = \{a \in L \,;\, \mathrm{tr}_{L/K}(a) = 0\}$ is a K-subvector space of L. If L/K is separable, then $\mathrm{tr}_{L/K}$ is nonzero, and hence $\ker \mathrm{tr}_{L/K}$ is an $(n-1)$-dimensional K-subvector space of L. On the other hand, if L/K is not separable, then $\mathrm{tr}_{L/K}$ is the zero map, and we have $\ker \mathrm{tr}_{L/K} = L$.

4.7, Exercise 2. Let $\mathbb{F} = \mathbb{F}_q$, $\mathbb{F}' = \mathbb{F}_{q'}$ for prime powers q and q' satisfying $q' = q^n$, where n is the degree of the extension $\mathbb{F}'/\mathbb{F}$. First we want to show that the norm map $\mathrm{N} \colon \mathbb{F}'^* \longrightarrow \mathbb{F}^*$ is surjective. Taking into account the fact that the Galois group $\mathrm{Gal}(\mathbb{F}'/\mathbb{F})$ is generated by the relative Frobenius homomorphism $a \longmapsto a^q$, the norm of an element $a \in \mathbb{F}'$ is calculated as

$$\mathrm{N}(a) = a \cdot a^q \cdot a^{q^2} \cdot \ldots \cdot a^{q^{n-1}} = a^{\frac{q^n-1}{q-1}}.$$

In particular, we see that $\mathrm{N}(a)^{q-1} = a^{q^n-1} = 1$. Now we use that the group $\mathbb{F}'^*$ is cyclic, generated by some element α of order q^n-1. Then $\mathrm{N}(\alpha) = \alpha^{\frac{q^n-1}{q-1}} \in \mathbb{F}$ is of order $q-1$ and hence is a generating element of the cyclic group $\mathbb{F}^*$. Therefore, since $\mathrm{N} \colon \mathbb{F}'^* \longrightarrow \mathbb{F}^*$ is a group homomorphism, it is surjective. Furthermore, we see that the kernel of N consists of all elements α^r such that $(q-1) \,|\, r$, or in other words, of all elements that are a $(q-1)$th power of an element in $\mathbb{F}'^*$.

4.8, Exercise 1. Let L/K be a finite cyclic Galois extension with a generating element $\sigma \in \mathrm{Gal}(L/K)$. For $b \in L^*$ consider elements $a, a' \in L^*$ such that $b = a\sigma(a)^{-1} = a'\sigma(a')^{-1}$. Then we get $\sigma(a/a') = a/a'$ and hence $a/a' \in K^*$. Conversely, we see from $a/a' \in K^*$ that $a\sigma(a)^{-1} = a'\sigma(a')^{-1}$. As a conclusion, if $b \in L^*$ satisfies $\mathrm{N}_{L/K}(b) = 1$, then the element $a \in L^*$, which satisfies $b = a\sigma(a)^{-1}$ according to 4.8/1, is unique up to a multiplicative constant in K^*. In the same way one shows in the setting of 4.8/4 for every $b \in L$ with trace $\mathrm{tr}_{L/K}(b) = 0$ that the corresponding element $a \in L$ that satisfies $b = a - \sigma(a)$ is unique up to an additive constant in K.

4.8, Exercise 2. The Galois group $\mathrm{Gal}(\mathbb{C}/\mathbb{R})$ is cyclic of order 2, generated by the complex conjugation $\mathbb{C} \longrightarrow \mathbb{C}$, $z \longmapsto \overline{z}$. Thus, we get $\mathrm{N}_{\mathbb{C}/\mathbb{R}}(z) = z\overline{z} = |z|^2$ for $z \in \mathbb{C}$. Now assume $\mathrm{N}_{\mathbb{C}/\mathbb{R}}(z) = 1$, i.e., that z belongs to the unit circle centered at 0. Then Hilbert's Theorem 90 asserts the existence of an element $x \in \mathbb{C}^*$ satisfying $z = x/\overline{x}$, where we may even assume $x\overline{x} = |x|^2 = 1$. This implies $z = x^2$, i.e., that x is a square root of z.

4.9, Exercise 1. We start with a cyclic extension L/K of degree n. Writing $C = L^n \cap K^*$, we have $L = K(C^{1/n})$ by 4.9/3 and $n = [L : K] = (C : K^{*n})$ by 4.9/1. The Galois group $G_C = \mathrm{Gal}(L/K)$ is cyclic of order n by our assumption. Hence, the same is true for $\mathrm{Hom}(C/K^{*n}, U_n)$ by 4.9/3 and for C/K^{*n} by 4.9/2. Now choose an element $c \in C$ whose residue class generates the group C/K^{*n}. Then $L = K(c^{1/n})$, and we see that the extension L/K is obtained by adjoining a zero a of the polynomial $X^n - c \in K[X]$ to K. By reasons of degree, this polynomial is irreducible and therefore coincides with the minimal polynomial of a over K.

Conversely, consider an extension L/K that is obtained by adjoining a zero a of a polynomial of type $X^n - c$ to K, where we may assume $c \in K^*$. Let C be the subgroup that is generated in K^* by c and K^{*n}. Then we can write $L = K(C^{1/n})$. Furthermore, we conclude from 4.9/3 that L/K is an abelian extension with Galois group $\mathrm{Hom}(C/K^{*n}, U_n)$, resp. C/K^{*n}, since the latter group is finite. Now observe that C/K^{*n} is generated by the residue class of c and hence that L/K is cyclic. Since $c^n \in K^{*n}$, the group C/K^{*n} is cyclic of some order d dividing n. In particular, we get $c^d \in K^{*n}$ and hence $a^d \in K$. Then we can see, similarly as before, that $X^d - a^d$ is the minimal polynomial of a over K.

4.9, Exercise 2. The group $C = K^*$ is the largest of all subgroups in K^* containing K^{*n}. Therefore, it follows from 4.9/3 that likewise, $L_n = K(K^{*1/n})$ is the largest abelian extension of K with an exponent dividing n. Since every group homomorphism $K^* \longrightarrow U_n$ is necessarily trivial on K^{*n}, we get $\mathrm{Gal}(L_n/K) = \mathrm{Hom}(K^*, U_n)$, again by 4.9/3.

4.10, Exercise 1. Let us consider the setting of Theorem 4.10/1. We claim that a field extension L/K is cyclic of a degree dividing n if and only if there exists an element $\alpha \in A$ such that $\wp(\alpha) \in A_K$ and $L = K(\alpha)$. The argument is the same as the one used in Exercise 1 of Section 4.9. First, assume that L/K is a cyclic Galois extension of some degree dividing n. Then, according to 4.10/1, we have $L = K(\wp^{-1}(C))$ for $C = \wp(A_L) \cap A_K$, and there is an isomorphism $C/\wp(A_K) \xrightarrow{\sim} \mathrm{Hom}(G_C, \mu_n)$, where G_C is the Galois group of L/K. By our assumption, G_C is cyclic of some degree dividing n. The same is true for $C/\wp(A_K)$, due to 4.9/2, and there is an element $c \in C$ whose residue class generates $C/\wp(A_K)$. Choosing a preimage $\alpha \in \wp^{-1}(c)$, it follows that $\wp^{-1}(C)$ is generated by α and A_K, and we get $L = K(\alpha)$, as claimed.

On the other hand, assume $L = K(\alpha)$ for some element $\alpha \in A$ satisfying $\wp(\alpha) \in A_K$. Then we get $L = K(\wp^{-1}(C))$, where C is generated by $\wp(\alpha)$ and $\wp(A_K)$. In particular, $C/\wp(A_K)$ is cyclic, generated by the residue class of $\wp(\alpha)$, and we see from 4.10/1 that L/K is an abelian extension of some exponent dividing n. Using 4.9/2, it follows that L/K is even cyclic.

4.10, Exercise 2. Since K is a perfect field of characteristic $p > 0$, the Frobenius homomorphism $K \longrightarrow K$ is an isomorphism, and the same is true for the Frobenius operator $F\colon W(K) \longrightarrow W(K)$. Therefore, the relation $V \circ F = p$ of 4.10/7 implies $p \cdot W(K) = V^1 W(K)$.

Concerning assertion (i), recall the formula

$$(\alpha, 0, 0, \ldots) \cdot (\beta, 0, 0, \ldots) = (\alpha \cdot \beta, 0, 0, \ldots)$$

for the multiplication in $W(K)$, which was mentioned in Section 4.10. It implies that the map $K \longrightarrow W(K)$, $\alpha \longmapsto (\alpha, 0, 0, \ldots)$, is multiplicative and hence restricts to a monomorphism of multiplicative groups $K^* \longrightarrow W(K)^*$. On the other hand, there cannot exist a nontrivial map $K \longrightarrow W(K)$ that is additive. Indeed, multiplication by p yields the zero map on K, while up to the Frobenius operator, it equals the Verschiebung operator on $W(K)$.

Next we turn to assertion (ii). We have to show that $W(K)$, together with the projections $W(K) \longrightarrow W(K)/V^n W(K)$, constitutes a projective limit of the projective system

$$W(K)/V^0 W(K) \longleftarrow W(K)/V^1 W(K) \longleftarrow W(K)/V^2 W(K) \longleftarrow \ldots .$$

To justify this we establish the defining universal property of Section 4.2. Indeed, consider a ring R, together with a system $(h_n)_{n\in\mathbb{N}}$ of ring homomorphisms $h_n \colon R \longrightarrow W(K)/V^n W(K)$ that are compatible with all projections

$$W(K)/V^{i+1} W(K) \longrightarrow W(K)/V^i W(K), \qquad i \in \mathbb{N}.$$

Then the maps h_n admit a common unique factorization through $W(K)$, via the map

$$h \colon R \longrightarrow W(K), \qquad x \longmapsto \big(h_1(x)_0, h_2(x)_1, h_3(x)_2, \ldots\big),$$

where $h_{n+1}(x)_n$ is in each case the component of $h_{n+1}(x) \in W(K)/V^{n+1} W(K)$ of index n. That h is even a ring homomorphism follows by a formal argument on projective limits, or by explicitly using the definition of the ring structure on $W(K)$ in terms of the polynomials S_n, P_n. Therefore, the first part of assertion (ii) is clear, while the second one, namely that $W(\mathbb{F}_p)$ coincides with $\mathbb{Z}_p$, is easily derived from 4.10/10.

To verify (iii) look at the canonical projection $W(K) \longrightarrow W_1(K) = K$. It is an epimorphism with kernel $V^1 W(K) = p \cdot W(K)$, and we thereby see that $p \cdot W(K)$ is a maximal ideal in $W(K)$. Furthermore, we claim that this ideal is the only maximal ideal in $W(K)$, in other words, that the group of units $W(K)^*$ coincides with $W(K) - V^1 W(K)$. To justify this, consider an element $a \in W(K) - V^1 W(K)$. In order to show that it is a unit, we can multiply a by a unit of type $(\alpha, 0, 0, \ldots)$ for some $\alpha \in K^*$; cf. (i). In this way, we can assume that a is of type $1 - p \cdot c$ for some element $c \in W(K)$. Now, using the relation $p^r \cdot W(K) = V^r W(K)$, it is easy to see that $b = \sum_{i\in\mathbb{N}} p^i \cdot c^i$ gives rise to a well-defined element in $W(K)$. Indeed, the image of each finite sum $\sum_{i=0}^{s} p^i \cdot c^i$ with respect to the projection $W(K) \longrightarrow W(K)/V^n W(K)$ is independent of s for $s \geq n$. Furthermore, the formula for the geometric series shows that $a \cdot b$ is mapped to the unit element under all projections $W(K) \longrightarrow W(K)/V^n W(K)$, in other words, that the relation $a \cdot b = 1$ holds in $W(K)$.

Therefore, we have recognized $W(K) - p \cdot W(K) = W(K) - V^1W(K)$ as the group of units in $W(K)$. Next observe for every nonzero element $a \in W(K)$ that there is a unique integer $n \in \mathbb{N}$ satisfying $a \in V^nW(K) - V^{n+1}W(K)$. In this way, we can write $a = p^n \cdot a'$ for some element $a' \in W(K) - V^1W(K)$ and hence for some unit $a' \in W(K)^*$. Since $p^n \cdot W(K) = V^n(K)$, it follows that p is not nilpotent and consequently that $W(K)$ is an integral domain. Moreover, every nontrivial ideal $\mathfrak{a} \subset W(K)$ satisfies

$$\mathfrak{a} = (p^n), \quad \text{where} \quad n = \min\{i \in \mathbb{N}\,;\, p^i \in \mathfrak{a}\}.$$

Thus, $W(K)$ is a principal ideal domain. Let us add that principal ideal domains admitting precisely one nontrivial maximal ideal are referred to as *discrete valuation rings*. In particular, $W(K)$ is such a discrete valuation ring.

4.11, Exercise 1. Choose a K-vector space basis $(a_i)_{i\in I}$ of A. Then $(a_i \otimes 1)_{i\in I}$ is a K'-vector space basis of $A \otimes_K K'$, and every element of $A \otimes_K K'$ can be written as a sum $\sum_{i\in I} a_i \otimes c_i$ with unique coefficients $c_i \in K'$, where $c_i = 0$ for almost all indices $i \in I$. Now, to introduce the multiplication by some element $\sum_{j\in I} a_j \otimes c'_j$ on $A \otimes_K K'$, consider in a first step the (right) multiplication by a term $a_j \otimes c'_j$:

$$\varphi_{a_j,c'_j}\colon A \otimes_K K' \longrightarrow A \otimes_K K', \qquad \sum_{i\in I} a_i \otimes c_i \longmapsto \sum_{i\in I} a_ia_j \otimes c_ic'_j.$$

Then the multiplication by $\sum_{j\in I} a_j \otimes c'_j$ can be defined as the sum of the maps φ_{a_j,c'_j}. In this way, we get a map

$$(A \otimes_K K') \times (A \otimes_K K') \longrightarrow A \otimes_K K'$$

that is characterized by the mapping rule $(a \otimes c, a' \otimes c') \longmapsto aa' \otimes cc'$. Using this rule, the properties of a ring multiplication can be read directly from the corresponding properties of A and K'. Furthermore, $A \otimes_K K'$ is a K'-algebra via the ring homomorphism $K' \longrightarrow A \otimes_K K'$, $c \longmapsto 1 \otimes c$.

4.11, Exercise 2. To prove assertion 4.11/4 (i), it is enough to show for every finite-dimensional K-subspace $V_0 \subset V$ that the inclusion map $\lambda\colon V \hookrightarrow V'$ gives rise to a K'-linear map $\lambda'_0\colon K' \otimes_K V_0 \longrightarrow V'$ that is injective. This can easily be justified by induction on $r = \dim_K V_0$. Indeed, nothing has to be shown for $r = 0$. Therefore, let $r > 0$. Then there is a nonzero vector $x \in V_0$, and we can view the K-vector space V_0/Kx as a part of the fixed set corresponding to the action that is induced by f on $V'/K'x$. The canonical K'-linear map $K' \otimes_K (V_0/Kx) \longrightarrow V'/K'x$ is injective by the induction hypothesis, and an easy calculation shows that then $\lambda'_0\colon K' \otimes_K V_0 \longrightarrow V'$ is injective as well.

To verify 4.11/4 (ii), observe that $f_\sigma(\alpha_i v) = \sigma(\alpha_i) f_\sigma(v)$ and hence that

$$\sum_{\sigma\in G} f_\sigma(\alpha_i v) = \sum_{\sigma\in G} \sigma(\alpha_i) f_\sigma(v), \qquad i = 1, \ldots, n.$$

Since the matrix $(\sigma(\alpha_i))_{\sigma\in G, i=1\ldots n} \in (K')^{n\times n}$ is invertible by 4.6/3, the elements $f_\sigma(v)$, $\sigma \in G$, and in particular v, can be written as K'-linear combinations

of the elements $v_i = \sum_{\sigma\in G} f_\sigma(\alpha_i v)$, $i = 1, \dots, n$. All vectors v_i are fixed by the action of G on V' and therefore belong to V. However, this implies that $\lambda'\colon K' \otimes_K V \longrightarrow V'$ is surjective.

5.1, Exercise 1. The H-orbit of an element $g \in G$ under left translation by H, i.e., under the action $H \times G \longrightarrow G$, $(h, g) \longmapsto hg$, equals the right coset Hg. If $\{g_1, \dots, g_r\}$ is a system of representatives of the right cosets of G modulo H, then the orbit equation reads $\operatorname{ord} G = \sum_{i=1}^{r} \operatorname{ord}(Hg_i)$. Furthermore, the number r of right cosets of H coincides with the index $(G : H)$, and all right cosets Hg_i contain precisely $\operatorname{ord} H$ elements. Therefore, the orbit equation yields $\operatorname{ord} G = (G : H) \cdot \operatorname{ord} H$, which coincides with the formula given by the theorem of Lagrange 1.2/3. If we consider right translation instead of left translation by H, more precisely, the action $H \times G \longrightarrow G$, $(h, g) \longmapsto gh^{-1}$, then the corresponding H-orbits are of type gH and hence coincide with the left cosets of H. Also in this case, the corresponding orbit equation yields the formula of 1.2/3.

5.1, Exercise 2. A Galois automorphism $\sigma \in \operatorname{Gal}(L/K)$ leaves a given element $a \in L$ fixed if and only if it leaves the field $K(a)$ fixed. Therefore, the stabilizer group of a satisfies $G_a = \operatorname{Gal}(L/K(a))$. Next, the orbit Ga consists of all elements that are conjugate to a over K in the sense of Galois theory; cf. 4.1. For example, if $f \in K[X]$ is the minimal polynomial of a over K, these are precisely the zeros of f. Indeed, every $\sigma \in \operatorname{Gal}(L/K)$ maps the zero set of f into itself. On the other hand, using the normality of L/K, the polynomial f factors in $L[X]$ completely into linear polynomials (cf. 3.5/4 and 3.5/5). In addition, given a zero $a' \in L$ of f, there is an automorphism $\sigma \in \operatorname{Gal}(L/K)$ such that $\sigma(a) = a'$ (cf. 3.4/8 and 3.4/9). In particular, since L/K is separable, we get $\operatorname{ord} Ga = \deg f = [K(a) : K]$, as well as $\operatorname{ord} G_a = [L : K(a)]$.

5.2, Exercise 1. Let G be a finite abelian group and p a prime number. Then there exists a p-Sylow group $S \subset G$ by Theorem 5.2/6 (i). Since all p-Sylow groups in G are conjugate due to 5.2/6 (ii), and since G is abelian, it follows that S is the only p-Sylow group in G. Thus, applying 5.2/6 (i) again, S is as described in 5.2/2. In this way, Theorem 5.2/6 shows that the elements of G, whose orders are p-powers, form a p-Sylow group in G, a fact that was proved in 5.2/2 by elementary means.

5.2, Exercise 2. If $S \subset G$ is a p-Sylow group, its image $\varphi(S)$ consists of elements whose orders are powers of p. Hence, $\varphi(S)$ is a p-group by 5.2/11, and we conclude from 5.2/6 that there exists a p-Sylow group S' in G' such that $\varphi(S) \subset S'$. If φ is injective, say $\varphi\colon G \hookrightarrow G'$, we get necessarily $S' \cap G = S$, since $S' \cap G$ is a p-group in G containing S. In other words, if $G \subset G'$ is a subgroup, the p-Sylow groups of G are restrictions of (certain) p-Sylow groups in G'. On the other hand, if G is a normal subgroup of G', then every p-Sylow group $S' \subset G'$ restricts to a p-Sylow group $S' \cap G$ in G. Indeed, $S' \cap G$ is a p-group and thereby contained in a p-Sylow subgroup S of G. Then S is a p-group in G', and there is an element $g \in G'$ such that $gSg^{-1} \subset S'$; cf. 5.2/9. Since G is normal in G', we get $gSg^{-1} \subset G$, and it follows that gSg^{-1} is a p-Sylow

group in G, since $\operatorname{ord} S = \operatorname{ord} gSg^{-1}$. However, since gSg^{-1} is contained in the intersection $S' \cap G$ and since this is a p-group, we get $S' \cap G = gSg^{-1}$, i.e., $S' \cap G$ is a p-Sylow group in G.

Furthermore, let us look at the case that $\varphi\colon G \longrightarrow G'$ is surjective. Fixing a p-Sylow group $S \subset G$, we claim that its image $H' = \varphi(S)$ is a p-Sylow group in G'. To justify this, consider the left translation of G as an action on the *set* of left cosets G'/H'. This action is transitive, i.e., admits only a single orbit. If H is the stabilizer group of H' viewed as an element in G'/H', we get $S \subset H$, as well as $\operatorname{ord} G/H = \operatorname{ord} G'/H'$ by the orbit equation. Then $p \nmid \operatorname{ord}(G/S)$ implies $p \nmid \operatorname{ord}(G/H)$, and hence $p \nmid \operatorname{ord}(G'/H')$. In particular, since H' is a p-group, it is already a p-Sylow group in G'. Thus, the image of every p-Sylow group $S \subset G$ is a p-Sylow group in G'. On the other hand, it is easy to see, using the conjugation action, that every p-Sylow group in G' is the image of a p-Sylow group in G.

5.3, Exercise 1. A permutation $\pi \in \mathfrak{S}_n$ is a bijective self-map on the set $\{1, \ldots, n\}$. In other words, π "permutes" the numbers $1, \ldots, n$, i.e., it changes their order to $\pi(1), \ldots, \pi(n)$. For example, if π is a transposition, it exchanges precisely two elements of the sequence $1, \ldots, n$. In this way, it is plausible that the numbers of the sequence $1, \ldots, n$ can be put into any desired order by a repeated process of exchanging only two elements at each step. In fact, this corresponds to the assertion that every $\pi \in \mathfrak{S}_n$ can be written as a product of transpositions.

To give a strict proof, use induction on n, where the base case $n = 1$ is trivial, since $\mathfrak{S}_1$ consists only of the unit element, which can be interpreted as the empty product. Therefore, assume $n > 1$. If there exists an index $i \in \{1, \ldots, n\}$ such that $\pi(i) = i$, then π restricts to a bijective self-map π' on $\{1, \ldots, i-1, i+1, \ldots, n\}$. Relying on the induction hypothesis, π' is a product of transpositions, and the same is true for π. On the other hand, if there is an index $i \in \{1, \ldots, n\}$ such that $\pi(i) \neq i$, then $(i, \pi(i)) \circ \pi$ leaves the element i fixed and hence is a product of transpositions, say $(i, \pi(i)) \circ \pi = \tau_1 \circ \ldots \circ \tau_r$, as seen before. But this implies $\pi = (i, \pi(i)) \circ \tau_1 \circ \ldots \circ \tau_r$ and hence that π is a product of transpositions.

5.3, Exercise 2. Let $\pi \in \mathfrak{S}_p$ be a p-cycle, say $\pi = (1, \ldots, p)$. Then the cyclic group $\langle \pi \rangle$ generated by π is a p-Sylow group in $\mathfrak{S}_p$. Indeed, its order is p, and we have $p \nmid (\mathfrak{S}_p : \langle \pi \rangle)$ as $(\mathfrak{S}_p : \langle \pi \rangle) = (p-1)!$.

5.4, Exercise 1. Assume first that H is a normal subgroup of G. Then the relation $g[a,b]g^{-1} = [gag^{-1}, gbg^{-1}]$, see the proof of 5.4/1, shows that $[G,H]$ is a normal subgroup in G. We claim that $[G,H]$ is the smallest of all normal subgroups $N \subset G$ such that the image of H in G/N is contained in the center of G/N. Indeed, the image of H in $G/[G,H]$ commutes elementwise with all residue classes of elements $g \in G$. Hence, the image of H is contained in the center of $G/[G,H]$. Conversely, if $N \subset G$ is a normal subgroup satisfying this condition, then all commutators $[a,b]$ for $a \in G$ and $b \in H$ belong to N, so that we have $[G,H] \subset N$. Therefore, our claim is justified. On the other hand,

if H is only a subgroup in G, then the above argument still yields the following assertion: If $N \subset G$ is a normal subgroup such that the image of H is contained in the center of G/N, then $[G, H] \subset N$.

6.1, Exercise 1. If $f(x) = 0$ is solvable over K, it will not necessarily be solvable over K_0 as well. For example, there exist algebraic equations $f(x) = 0$ over $\mathbb{Q}$ that are not solvable, as we saw at the end of Section 6.1. Clearly, such an equation will become solvable over a splitting field of f, or over an algebraic closure of $\mathbb{Q}$. On the other hand, if $f(x) = 0$ is solvable over K_0, it is solvable over K as well. Indeed, if L is a splitting field of f over K, and $L_0 \subset L$ a splitting field of f over K_0, then there is a canonical restriction map $\mathrm{Gal}(L/K) \longrightarrow \mathrm{Gal}(L_0/K_0)$ by 3.5/4. This map is injective, since both extensions L/K and L_0/K_0 are generated by the zeros of f. Thus, if $\mathrm{Gal}(L_0/K_0)$ is solvable, then the same is true for $\mathrm{Gal}(L/K)$ by 5.4/8.

6.1, Exercise 2. Assume first that the equation $f(x) = 0$ is metacyclic. Then there exists a chain of fields $K = K_0 \subset K_1 \subset \ldots \subset K_n$ such that the splitting field L of f over K is contained in K_n and K_{i+1}/K_i is in each case a (finite) cyclic, and therefore solvable, Galois extension. Using 6.1/4, this implies that the extensions K_n/K and L/K are solvable as well.

Conversely, consider the case that $f(x) = 0$ is solvable. Then the Galois group $\mathrm{Gal}(L/K)$ is solvable, and we can conclude from 5.4/7 that $\mathrm{Gal}(L/K)$ admits a normal series with cyclic factors. Thus, the fundamental theorem of Galois theory 4.1/6 shows in conjunction with the primitive element theorem 3.6/12 that the equation $f(x) = 0$ is metacyclic. Thereby we see that "metacyclic" is equivalent to "solvable." The remaining equivalence to "solvable by radicals" follows from 6.1/5.

6.2, Exercise 1. According to the theory developed in Section 6.2, consider the chain of field extensions

$$K \subset K\big(\sqrt{\Delta}\big) \subset L' \subset L,$$

where L' is a splitting field of g and L a splitting field of f over K, and where Δ is the common discriminant of f and g. The Galois group $G = \mathrm{Gal}(L/K)$ acts on the zeros $x_1, \ldots, x_4 \in L$ of f and thereby can be viewed as a subgroup of $\mathfrak{S}_4$. As we know already, Δ admits a square root in K if and only if G acts in terms of even permutations on the x_i, in other words, if and only if $G \subset \mathfrak{A}_4$. Therefore, the degree of $K(\sqrt{\Delta})$ over K equals 1 or 2, depending on whether we have $G \subset \mathfrak{A}_4$ or $G \not\subset \mathfrak{A}_4$.

Since f does not have real zeros, the complex conjugation $\mathbb{C} \longrightarrow \mathbb{C}$, $z \longmapsto \overline{z}$, restricts on L to a nontrivial element of the Galois group G. In particular, the zeros of f consist of two pairs of conjugate complex numbers, say where $x_2 = \overline{x}_1$ and $x_4 = \overline{x}_3$. Furthermore, we conclude that the zeros

$$z_1 = (x_1 + x_2)(x_3 + x_4), \quad z_2 = (x_1 + x_3)(x_2 + x_4), \quad z_3 = (x_1 + x_4)(x_2 + x_3)$$

of g satisfy $z_1 \in \mathbb{R}$, as well as $z_2, z_3 \geq 0$, where $z_i \neq 0$ for all i, since g is irreducible. In particular, we get $L' \subset \mathbb{R}$. Furthermore, the degree of L/K is

divisible by 3, using the irreducibility of g again, and it follows that $L'/K(\sqrt{\Delta})$ is of degree 3.

Now we claim that the Galois group $H = \mathrm{Gal}(L/K(\sqrt{\Delta}))$ coincides with $\mathfrak{A}_4$. Of course, we have $H \subset \mathfrak{A}_4$. Furthermore, observe that complex conjugation leaves $L' \subset \mathbb{R}$ fixed and therefore gives rise to a nontrivial element in $\mathrm{Gal}(L/L')$. In particular, the degree $[L : L']$ is divisible by 2 and hence is at least 2. It follows that the order $\operatorname{ord} H = [L : K(\sqrt{\Delta})]$ is at least 6, hence 6 or 12, and it is enough to exclude the case $\operatorname{ord} H = 6$. To do this, assume $\operatorname{ord} H = 6$. Then there exists in H precisely one 3-Sylow group by the Sylow theorems 5.2/6. In addition, the chain $K(\sqrt{\Delta}) \subset L' \subset L$, where necessarily $[L : L'] = 2$, shows that H contains a normal subgroup of order 2, which is a 2-Sylow group in H and, being normal, the only 2-Sylow group in H. But then the proof of 5.2/12 shows that H is cyclic of order 6, in contradiction to the fact that $\mathfrak{S}_4$ contains only elements of order 1, 2, 3, or 4. Thus, $\operatorname{ord} H = 12$, and we can conclude that $\mathrm{Gal}(L/K(\sqrt{\Delta})) = \mathfrak{A}_4$, as claimed. To sum up, we get $\mathrm{Gal}(L/K) = \mathfrak{A}_4$ if Δ is a square in K, as well as $\mathrm{Gal}(L/K) \supsetneq \mathfrak{A}_4$, and therefore $\mathrm{Gal}(L/K) = \mathfrak{S}_4$, if Δ is not a square in K.

Let us add an example illustrating our discussion. Consider the algebraic equation $f(x) = 0$, where

$$f = X^4 + X^2 + X + 1 \in \mathbb{Q}[X].$$

Clearly, f does not admit real zeros and is irreducible. The resolvent cubic is given by

$$g = X^3 - 2X^2 - 3X + 1 \in \mathbb{Q}[X]$$

and is irreducible as well. Furthermore, the discriminant of f and g is

$$\Delta = 144 - 128 - 4 + 16 - 27 + 256 = 257.$$

Since 257 is not a square in $\mathbb{Q}$, we see that the Galois group of the equation $f(x) = 0$ equals $\mathfrak{S}_4$.

6.3, Exercise 1. The properties of real numbers used in the proof of 6.3/1 cannot be justified in terms of purely algebraic methods, for instance, as developed in the present book. This comes as no surprise, since we have assumed the real numbers to be "known," without giving a precise characterization of them. Anyway, the study of real numbers and of real-valued functions is rather a part of analysis than of algebra. Accordingly, we justify the desired properties by means of infinitesimal calculus. Therefore, let $f = X^n + a_1X^{n-1} + \ldots + a_n$ be a polynomial of odd degree in $\mathbb{R}[X]$. Using the factorization

$$f(x) = x^n(1 + a_1x^{-1} + \ldots + a_nx^{-n})$$

for $x \in \mathbb{R}$, $x \neq 0$, we conclude that

$$\lim_{x\to\infty} f(x) = \infty, \qquad \lim_{x\to-\infty} f(x) = -\infty.$$

Then the intermediate value theorem asserts that $f(x)$, as a real-valued continuous function, admits a zero in $\mathbb{R}$. For a similar reason, every $a \in \mathbb{R}$, $a \geq 0$, admits a square root in $\mathbb{R}$. Indeed, consider the function $g(x) = x^2 - a$. By the intermediate value theorem again, it admits a zero in $\mathbb{R}$, since $g(0) \leq 0$, as well as $\lim_{x\to\infty} g(x) = \infty$.

6.4, Exercise 1. Write $K = \mathbb{Q}(M \cup \overline{M})$. Looking at a point $z \in \mathfrak{K}(M)$, we conclude from 6.4/1 in conjunction with the multiplicativity formula 3.2/2 that the degree $[K(z) : K]$ is a power of 2. Conversely, let $z \in \mathbb{C}$ be an element satisfying such a condition. Applying 6.4/1, we get $z \in \mathfrak{K}(M)$, provided the extension $K(z)/K$ is Galois. However, in general we obtain $z \in \mathfrak{K}(M)$ only if z is contained in a Galois extension of K whose degree is a power of 2. Writing L for the field that is generated over K by all conjugates of z, i.e., for the splitting field of the minimal polynomial of z over K, the preceding condition is equivalent to the condition that the degree $[L : K]$ is a power of 2. However, there are examples for which $[K(z) : K]$ is a power of 2 but $[L : K]$ is not. Just observe that there exist irreducible algebraic equations of degree 4 with Galois group $\mathfrak{S}_4$, as we will see below. In particular, we cannot necessarily conclude that $z \in \mathfrak{K}(M)$ from the fact that $[K(z) : K]$ is a power of 2.

To explicitly construct such an example, let $M = \{0, 1\}$ and consider a polynomial of type $f = X^4 - pX - 1 \in \mathbb{Q}[X]$ for a prime number p. Then f is irreducible. Indeed, it is enough to show that f is irreducible as a polynomial in $\mathbb{Z}[X]$; see 2.7/7. This is easily checked by showing that f does not admit a decomposition over $\mathbb{Z}$ into a linear and a cubic polynomial, or into two quadratic polynomials. Now let $\alpha_1, \ldots, \alpha_4 \in \mathbb{C}$ be the zeros of f, and let $L = \mathbb{Q}(\alpha_1, \ldots, \alpha_4)$ be the splitting field of f in $\mathbb{C}$. Then the solution process of quartic equations in Section 6.1 shows that the quantities

$$\beta_1 = (\alpha_1 + \alpha_2)(\alpha_3 + \alpha_4), \quad \beta_2 = (\alpha_1 + \alpha_3)(\alpha_2 + \alpha_4), \quad \beta_3 = (\alpha_1 + \alpha_4)(\alpha_2 + \alpha_3)$$

are the zeros of the resolvent cubic of f, which is given by the polynomial $g = X^3 + 4X + p^2$. Similarly as for f, one checks that g is irreducible over $\mathbb{Q}$. Therefore, L contains elements of degree 3 over $\mathbb{Q}$. Hence, the degree $[L : \mathbb{Q}]$ cannot be a power of 2. In particular, $\alpha_1, \ldots, \alpha_4$ do not belong to $\mathfrak{K}(\{0, 1\})$, although each α_i is of degree 4 over $\mathbb{Q}$. In addition, it is easily seen that the Galois group $\mathrm{Gal}(L/\mathbb{Q})$ coincides with the full permutation group $\mathfrak{S}_4$ when the elements $\sigma \in \mathrm{Gal}(L/\mathbb{Q})$ are viewed as permutations of the zeros $\alpha_1, \ldots, \alpha_4$. Indeed, since the Galois group $\mathrm{Gal}(L/\mathbb{Q})$ is a subgroup of $\mathfrak{S}_4$, its order divides 24. Furthermore, since L contains elements of degree 3 as well as of degree 4 over $\mathbb{Q}$, the order is at least 12. Therefore, we can conclude either that $\mathrm{Gal}(L/\mathbb{Q}) = \mathfrak{S}_4$, or that $\mathrm{Gal}(L/\mathbb{Q})$ is a subgroup of index 2 and hence a normal subgroup in $\mathfrak{S}_4$. In the latter case, we have $\mathrm{Gal}(L/\mathbb{Q}) = \mathfrak{A}_4$, since each normal subgroup of index 2 in $\mathfrak{S}_4$ gives rise to an abelian factor group, and since $[\mathfrak{S}_4, \mathfrak{S}_4] = \mathfrak{A}_4$; cf. 5.4/1 and 5.4/2. Now observe that the discriminant

$$\Delta_g = (\beta_1 - \beta_2)^2(\beta_1 - \beta_3)^2(\beta_2 - \beta_3)^2 = -4 \cdot 4^3 - 27p^4$$

of the resolvent cubic g does not admit a square root in $\mathbb{Q}$; concerning the formula for Δ_g, consult example (2) in 4.3 or the end of Section 4.4. Therefore, not all elements of $\mathrm{Gal}(L/\mathbb{Q})$ will restrict to an even permutation on the $\beta_1, \beta_2, \beta_3$, and it follows then from the definition of the β_i that $\mathrm{Gal}(L/\mathbb{Q})$ cannot consist of even permutations of the $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ only. This implies $\mathrm{Gal}(L/\mathbb{Q}) = \mathfrak{S}_4$, as claimed.

6.4, Exercise 2. As we know, the primitive third root of unity $\zeta_3 = e^{2\pi i/3} \in \mathbb{C}$ belongs to $\mathfrak{K}(\{0,1\})$; cf. 6.4/3. Now if angle trisection were generally possible in terms of compass and straightedge constructions, the primitive ninth root of unity $\zeta_9 = e^{2\pi i/9}$ could be obtained via such constructions. However, such is not the case by 6.4/3, since $\varphi(9) = 6$. Therefore, in general, the problem of angle trisection does not admit a solution in terms of compass and straightedge constructions. This is no surprise, since trisecting an angle φ corresponds to solving the equation $z^3 - e^{i\varphi} = 0$, or if we restrict ourselves to the real part of this equation and use $z\overline{z} = 1$, to solving the equation $4x^3 - 3x - \cos\varphi = 0$. In general, such cubic equations cannot be accessed in terms of compass and straightedge constructions.

7.1, Exercise 1. Let L/K be a field extension and $\mathfrak{X} = (x_i)_{i\in I}$ a transcendence basis. Then the system $\mathfrak{X}$ is algebraically independent over K. Furthermore, we can view $\mathfrak{X}$ as a system of variables, and the subring $K[\mathfrak{X}] \subset L$ as a polynomial ring in the variables x_i. Clearly, this implies that the system $\mathfrak{X}$ is linearly independent over K if we view L as a K-vector space. But on the other hand, it is impossible that $\mathfrak{X}$ generates $K[\mathfrak{X}]$, or even L, as a K-vector space. It is for this reason that a transcendence basis of L/K will never give rise to a K-vector space basis of L.

Nevertheless, there is a strong conceptual analogy between bases of vector spaces and transcendence bases of field extensions. Under this analogy, "linear independence" of a system $\mathfrak{X}$ of elements of a K-vector space V corresponds to "algebraic independence" of a system $\mathfrak{X}$ of elements of a field extension L/K. A basis of V is a linearly independent system $\mathfrak{X} \subset V$ generating V as a K-vector space. Likewise, a transcendence basis of L/K is an algebraically independent system $\mathfrak{X} \subset L$, "generating" L/K in the sense that $L/K(\mathfrak{X})$ is algebraic. Just as for vector space bases, transcendence bases can be characterized as maximal algebraically independent systems (cf. 7.1/3), or alternatively, as minimal "generating systems" in the sense just alluded to. Also the proof of 7.1/5, that every two transcendence bases of L/K are of the same cardinality, remains valid for vector space bases.

But let us add that there are natural limits to the analogy between such bases. For example, every bijection $\mathfrak{X} \longrightarrow \mathfrak{Y}$ between two bases of a vector space V gives rise to a unique K-automorphism of V. The corresponding assertion for transcendence bases of a field extension L/K fails to be true, with respect to both the existence and the uniqueness assertions. Just consider a simple purely transcendental extension $L = K(X)$ of a field K. Each of the elements X and X^2 defines a transcendence basis of L/K, and there is a K-isomorphism $K(X) \longrightarrow K(X^2)$ mapping X to X^2. But this isomorphism does not extend to

a K-automorphism of $K(X)$, since X does not admit a square root in $K(X)$. On the other hand, if L is an algebraic closure of $K(X)$, the identity map on $K(X)$ extends to a K-automorphism on L. But it is not unique, since there are nontrivial $K(X)$-automorphisms of L.

7.1, Exercise 2. First let us show that $\mathbb{C}$ admits automorphisms that do not leave $\mathbb{R}$ fixed. To do this, choose an element $x \in \mathbb{R}$, say $x = \pi$, that is transcendental over $\mathbb{Q}$. Following 7.1/4, the extension $\mathbb{C}/\mathbb{Q}$ admits a transcendence basis $\mathfrak{X}$ such that $x \in \mathfrak{X}$. Since the element $ix \in \mathbb{C}$ is transcendental over $\mathbb{Q}$ as well, there exists a transcendence basis $\mathfrak{Y}$ of $\mathbb{C}/\mathbb{Q}$ such that $ix \in \mathfrak{Y}$. Then $\mathfrak{X}$ and $\mathfrak{Y}$ are of same cardinality by 7.1/5, and there is a bijection $\mathfrak{X} \longrightarrow \mathfrak{Y}$, where we may assume $x \longmapsto ix$. This bijection extends to a $\mathbb{Q}$-isomorphism $\mathbb{Q}(\mathfrak{X}) \xrightarrow{\sim} \mathbb{Q}(\mathfrak{Y})$. Since $\mathbb{C}$ is an algebraically closed field that is algebraic over $\mathbb{Q}(\mathfrak{X})$ and over $\mathbb{Q}(\mathfrak{Y})$, we can view $\mathbb{C}$ as an algebraic closure of both $\mathbb{Q}(\mathfrak{X})$ and $\mathbb{Q}(\mathfrak{Y})$. Therefore,

$$\sigma\colon \mathbb{Q}(\mathfrak{X}) \hookrightarrow \mathbb{C}, \qquad \tau\colon \mathbb{Q}(\mathfrak{X}) \xrightarrow{\sim} \mathbb{Q}(\mathfrak{Y}) \hookrightarrow \mathbb{C}$$

are two algebraic closures of $\mathbb{Q}(\mathfrak{X})$. Then, by 3.4/10, there exists an automorphism $\varphi\colon \mathbb{C} \longrightarrow \mathbb{C}$ satisfying $\tau = \varphi \circ \sigma$. Since we have $\varphi(x) = ix$ by construction, it follows that φ is an automorphism of $\mathbb{C}$ that does not leave $\mathbb{R}$ fixed, which had to be shown. In particular, $\varphi(\mathbb{R})$ is a subfield of $\mathbb{C}$ that is isomorphic to $\mathbb{R}$, but different from $\mathbb{R}$.

We proceed in a similar way to show that $\mathbb{C}$ admits proper subfields isomorphic to itself. Choosing a transcendence basis $\mathfrak{X}$ of $\mathbb{C}/\mathbb{Q}$, we use the fact that $\mathfrak{X}$ consists of infinitely many elements; cf. Exercise 3 of 7.1. Then there exists an injective map $\mathfrak{X} \hookrightarrow \mathfrak{X}$ that is not surjective. To justify this, one may use, as shown in 7.1/7, that $\mathfrak{X}$ is a disjoint union of countably infinite subsets of $\mathfrak{X}$. Then the considered injection $\mathfrak{X} \longrightarrow \mathfrak{X}$ extends to a monomorphism $\iota\colon \mathbb{Q}(\mathfrak{X}) \hookrightarrow \mathbb{Q}(\mathfrak{X})$ such that $\mathbb{Q}(\mathfrak{X})$ is not algebraic over the image of ι. Again we consider the two homomorphisms

$$\sigma\colon \mathbb{Q}(\mathfrak{X}) \hookrightarrow \mathbb{C}, \qquad \tau\colon \mathbb{Q}(\mathfrak{X}) \xhookrightarrow{\iota} \mathbb{Q}(\mathfrak{X}) \hookrightarrow \mathbb{C}.$$

Then $\mathbb{C}$ is an algebraic closure of $\mathbb{Q}(\mathfrak{X})$ with respect to the injection σ, but not with respect to τ. Using 3.4/9, there is a $\mathbb{Q}(\mathfrak{X})$-homomorphism $\varphi\colon \mathbb{C} \hookrightarrow \mathbb{C}$ satisfying $\tau = \varphi \circ \sigma$. Since $\mathbb{C}$ is not algebraic over the image of τ, we can exclude that φ is surjective. Therefore, $\varphi(\mathbb{C})$ is a proper subfield of $\mathbb{C}$ that is isomorphic to $\mathbb{C}$.

7.2, Exercise 1. Let $\Phi\colon M \longrightarrow E$ be an R-linear map to an R'-module E. We have only to show that Φ gives rise to a unique R'-linear map $\varphi\colon M \otimes_R R' \longrightarrow E$ such that $x \otimes 1 \longmapsto \Phi(x)$ for $x \in M$. To justify the existence of φ, consider the R-bilinear map $M \times R' \longrightarrow E$, $(x, a) \longmapsto a\Phi(x)$. By the universal property of tensor products, it induces an R-linear map $\varphi\colon M \otimes_R R' \longrightarrow E$ that is uniquely characterized by $\varphi(x \otimes a) = a\Phi(x)$ for $a \in R'$ and $x \in M$. In particular, it follows that φ, as a map between R'-modules, is even R'-linear. On the other

hand, if $\psi\colon M \otimes_R R' \longrightarrow E$ is another R'-linear map satisfying $\psi(x \otimes 1) = \Phi(x)$ for $x \in M$, then ψ coincides on all tensors of type $x \otimes 1$ with φ. However, this implies $\varphi = \psi$, since these tensors generate $M \otimes_R R'$ as an R'-module.

7.2, Exercise 2. First we look at the case of free polynomial rings $R' = R[\mathfrak{X}]$ and $R'' = R[\mathfrak{Y}]$, with systems of variables $\mathfrak{X}, \mathfrak{Y}$. Then the polynomial ring $R[\mathfrak{X}, \mathfrak{Y}]$, together with the canonical injections $\sigma'\colon R[\mathfrak{X}] \longrightarrow R[\mathfrak{X}, \mathfrak{Y}]$ and $\sigma''\colon R[\mathfrak{Y}] \longrightarrow R[\mathfrak{X}, \mathfrak{Y}]$, admits the universal property of 7.2/9. Indeed, a homomorphism of R-algebras $R[\mathfrak{X}, \mathfrak{Y}] \longrightarrow A$ is uniquely characterized by the images of the members of $\mathfrak{X}$ and $\mathfrak{Y}$, while such images can be chosen arbitrarily. In the general case, R' and R'' can be interpreted as residue class rings of free polynomial rings, say $R' = R[\mathfrak{X}]/\mathfrak{a}$ and $R'' = R[\mathfrak{Y}]/\mathfrak{b}$. Then $R[\mathfrak{X}, \mathfrak{Y}]/(\mathfrak{a}, \mathfrak{b})$, together with the canonical maps $\sigma'\colon R[\mathfrak{X}]/\mathfrak{a} \longrightarrow R[\mathfrak{X}, \mathfrak{Y}]/(\mathfrak{a}, \mathfrak{b})$, as well as $\sigma''\colon R[\mathfrak{Y}]/\mathfrak{b} \longrightarrow R[\mathfrak{X}, \mathfrak{Y}]/(\mathfrak{a}, \mathfrak{b})$, admits the universal property of 7.2/9. Indeed, if $\varphi'\colon R[\mathfrak{X}] \longrightarrow A$ and $\varphi''\colon R[\mathfrak{Y}] \longrightarrow A$ are two R-algebra homomorphisms satisfying $\mathfrak{a} \subset \ker \varphi'$ and $\mathfrak{b} \subset \ker \varphi''$, then the resulting R-algebra homomorphism $\varphi\colon R[\mathfrak{X}, \mathfrak{Y}] \longrightarrow A$ satisfies $(\mathfrak{a}, \mathfrak{b}) \subset \ker \varphi$.

7.3, Exercise 1. In many cases the question has a negative answer. Consider as an example of a regular extension a purely transcendental extension $K(X)/K$ for a variable X. If $\operatorname{char} K = 2$, the extension $K(X)/K(X^2)$ is purely inseparable and therefore not separable, since it is nontrivial. On the other hand, if $\operatorname{char} K \neq 2$, this extension is separable algebraic, but not primary.

7.3, Exercise 2. Also this question has a negative answer. To give an example, choose a field k of characteristic $p > 0$ and consider for variables X, Y, Z the purely transcendental extension $k(X, Y, Z)$, as well as the following subfields:

$$K = k(X^p, Y^p), \qquad L = k(X^p, Y^p, Z)(t), \qquad \text{where} \qquad t = X + YZ.$$

We want to show that the extension L/K is not separable, although K is algebraically closed in L. First observe that the extension L/K decomposes into the purely transcendental extension $K(Z)/K$ and the purely inseparable extension $L/K(Z)$ of degree p; note that $t^p - (X^p + Y^p Z^p) = 0$ is the irreducible equation of t over $K(Z)$. To see that L/K is not separable, consider the elements $t^p, 1^p, Z^p$. These are linearly dependent over K, as follows from the preceding equation. Now, if the extension L/K were separable, then by 7.3/7 (iv), also the elements $t, 1, Z$ would be linearly dependent over K. This implies $t \in K(Z)$, which, however, is not the case. Consequently, L/K cannot be separable.

Thus, it remains to show that K is algebraically closed in L. Therefore, look at an element $a \in L$ that is algebraic over K. Then $a^p \in K(Z)$. However, since every element in $K(Z) - K$ is transcendental over K (cf. 7.1/10), we must have $a^p \in K$, so that $a \in k(X, Y)$. Assume that a does not belong to K. Then $a \notin K(Z)$, which implies $K(Z)(a) = L$, since $[L : K(Z)] = p$. Now observe that $K(Z)(a) = K(a)(Z)$. Thus, to obtain L we may just as well first adjoin the algebraic element a to K and then the transcendental element Z. It follows that the element $t = X + YZ \in L = K(a)(Z)$ can be written as a quotient of two polynomials in $K(a)[Z] \subset k(X, Y)[Z]$, say $X + YZ = f(Z)g(Z)^{-1}$. Canceling

powers of Z on the right-hand side we may assume $g(0) \neq 0$, which implies $X = f(0)g(0)^{-1} \in K(a)$. But then, just as t belongs to $K(a)(Z)$, the same is true for YZ. Therefore, Y belongs to $K(a)$. This shows that $K(a) = k(X, Y)$, which, however, is impossible, since a is only of degree p over $K = k(X^p, Y^p)$. Thus, K is algebraically closed in L, as claimed.

7.3, Exercise 3. Let K be a perfect field of characteristic $p > 0$, for instance take $K = \mathbb{F}_p$, and let X be a variable. Furthermore, consider the purely inseparable closure $L = K(X)^{p^{-\infty}}$ of $K(X)$. Then L/K is of transcendence degree 1. We claim that this extension is separable, but not separably generated. To justify this, observe that L is the union of the ascending chain of fields $K(X)^{p^{-i}} = K(X^{p^{-i}})$, $i \in \mathbb{N}$. Since for each i, the field $K(X^{p^{-i}})$ is purely transcendental over K with transcendence basis $X^{p^{-i}}$, we conclude from 7.2/13 and 7.3/3 that L/K is separable.

Let us assume now that the extension L/K is even separably generated. Then there exists an element $x \in L$ that is transcendental over K such that L is separable algebraic over $K(x)$. Since x is contained in one of the fields $K(X^{p^{-i}})$, there is a chain of fields $K(x) \subset K(X^{p^{-i}}) \subset L$. Now if $L/K(x)$ is separable algebraic, the same is true for $L/K(X^{p^{-i}})$ by 3.6/11. However, in contradiction to this, the element $X^{p^{-i-1}}$ is purely inseparable of degree p over $K(X^{p^{-i}})$. Thus, the extension L/K cannot be separably generated.

7.4, Exercise 1. The characterization of separable extensions L/K through the condition $\Omega^1_{L/K} = 0$ is valid only for finitely generated extensions. For example, if K is a nonperfect field of characteristic $p > 0$ and $L = K^{p^{-\infty}}$ is its purely inseparable closure (or an algebraic closure), then the extension L/K is not separable. On the other hand, since each element of L admits a pth root in L, every derivation on L is trivial. In particular, this implies $\Omega^1_{L/K} = 0$.

Literature

1. Artin, E.: *Foundations of Galois Theory*. New York University lecture notes. New York University, New York 1938.
2. Artin, E.: *Galois Theory*. Notre Dame Mathematical Lectures, Number 2. University of Notre Dame Press, Notre Dame 1942.
3. Bosch, S.: *Algebraic Geometry and Commutative Algebra*. Universitext, Springer, London–Heidelberg–New York–Dordrecht 2013.
4. Bosch, S.: *Lineare Algebra*. Springer, Berlin–Heidelberg–New York 2001, 2003, 2006, 2008, 2014.
5. Bourbaki, N.: *Eléments de Mathématique*, *Algèbre*. Hermann, Paris 1947, ...
6. Grothendieck, A.: *Technique de descente et théorèmes d'existence en géométrie algébrique: I. Généralités. Descente par morphismes fidèlement plats*. Séminaire Bourbaki 12, no. 190, 1959/60.
7. Hasse, H.: *Vorlesungen über Zahlentheorie*, Grundlehren der mathematischen Wissenschaften, Vol. 59. Springer, Berlin–Göttingen–Heidelberg–New York 1964.
8. Hermite, Ch.: Sur la fonction exponentielle. C. R. Acad. Sci. Paris 77 (1873).
9. Hilbert, D.: Die Theorie der algebraischen Zahlkörper. Jahresbericht der Deutschen Mathematikervereinigung 4, 175–546 (1897).
10. Huppert, B.: *Endliche Gruppen I*, Grundlehren der mathematischen Wissenschaften, Vol. 134. Springer, Berlin–Heidelberg–New York 1967.
11. Kiernan, B. M.: The Development of Galois Theory from Lagrange to Artin. Archive for History of Exact Sciences 8, 40–154 (1971/72).
12. Lang, S.: *Algebra*. Addison Wesley, 2nd Edition 1965, 3rd Edition 1993.
13. Lindemann, F.: Über die Zahl π. Math. Ann. 20, 213–225 (1882).
14. Serre, J.-P.: *Corps locaux*. Hermann, Paris 1968.
15. Steinitz, E.: Algebraische Theorie der Körper. Crelles Journal 137, 167–309 (1910).
16. Van der Waerden, B. L.: *Moderne Algebra*. Springer, Berlin 1930/31. Further Editions 1936, 1950, 1955, 1960, 1964, 1966 (from 1955 on called "Algebra").
17. Weber, H.: *Lehrbuch der Algebra*, 2 Vols. Vieweg, Braunschweig 1895/96.

Glossary of Notation

$\mathbb{N}$	natural numbers, including 0
$\mathbb{Z}$	(rational) integers
$\mathbb{Q}, \mathbb{Q}_{>0}$	rational numbers, resp. positive rational numbers
$\mathbb{R}, \mathbb{R}_{>0}$	real numbers, resp. positive real numbers
$\mathbb{C}$	complex numbers
$\mathrm{Map}(X, G)$	set of maps 12
$G^X, G^{(X)}$	G-valued functions on X 12
τ_a	left translation by a 14
aH, Ha	coset of a subgroup H 15
$G/H, H\backslash G$	set of cosets modulo H 16
$(G : H)$	index of a subgroup H 16
$\mathrm{ord}\, G$	order of a group 16
G/N	residue class group modulo a normal subgroup N 17
$\langle x \rangle$	subgroup generated by an element 20
$\mathrm{ord}\, a$	order of an element 21
R^*	group of units of a ring 26
$\mathbb{H}$	Hamiltonian quaternions 27
$R^X, R^{(X)}$	R-valued functions on a set X 27
$R[X]$	polynomial ring in one variable X 28
$\deg f$	degree of a polynomial 29
$R[\![X]\!]$	ring of formal power series 31
$\mathfrak{a} + \mathfrak{b}$	sum of ideals 32
$\mathfrak{a} \cdot \mathfrak{b}$	product of ideals 32
$\mathfrak{a} \cap \mathfrak{b}$	intersection of ideals 32
$(a_1, \ldots, a_n)$	ideal generated by $a_1, \ldots, a_n$ 32
$R/\mathfrak{a}$	residue class ring modulo an ideal $\mathfrak{a}$ 36
$\mathbb{F}_p$	field with p elements 38
$x \equiv y \mod \mathfrak{a}$	congruence 40
$x \,\vert\, y$	x divides y 43
$x \nmid y$	x doesn't divide y 43
$\nu_p(a)$	exponent relative to the prime factor p 47
$\gcd(x_1, \ldots, x_n)$	greatest common divisor of $x_1, \ldots, x_n$ 47
$\mathrm{lcm}(x_1, \ldots, x_n)$	least common multiple of $x_1, \ldots, x_n$ 47
$R[M]$	polynomial ring attached to M 51
$R[X_1, \ldots, X_n]$	polynomial ring in n variables 51
$R[\mathfrak{X}]$	polynomial ring in a family $\mathfrak{X}$ of variables 51
$\deg f$	total degree of a polynomial 55

$R[x]$	smallest subring containing R and x 56
Df, f'	derivative of a polynomial 58
$Q(R)$	field of fractions of an integral domain 60
$K(X), K(\mathfrak{X})$	rational function fields 60
$S^{-1}R, R_S$	localization of a ring R 60
$\nu_p(x), \nu_p(f)$	exponents relative to the prime factor p 61
M/N	residue class module modulo a submodule N 68
$\sum_{i\in I} M_i$	sum of modules 69
$\bigoplus_{i\in I} M_i$	direct sum of modules 69
$\operatorname{rank} M$	rank of a module 69
$S^{-1}M$	localization of a module M 70
$l_A(M)$	length of a module 70
M_{sat}	saturation of a submodule 71
$\operatorname{cont}(x)$	content of an element 72
$\bigwedge^t F$	t-fold exterior power of a free module 75
$\operatorname{char} K$	characteristic of a field 85
L/K	field extension 87
$[L:K]$	degree of a field extension 87
$K(\mathfrak{A})$	field generated over K by a system $\mathfrak{A}$ 90
$K(\alpha_1, \ldots, \alpha_n)$	field generated over K by $\alpha_1, \ldots, \alpha_n$ 90
$\overline{\mathbb{Q}}$	algebraic closure of $\mathbb{Q}$ 93
$A[x_1, \ldots, x_n]$	ring generated over A by $x_1, \ldots, x_n$ 94
$\overline{K}$	algebraic closure of a field 104
f^σ	polynomial transported by σ 105
$\operatorname{Hom}_K(L, \overline{K})$	set of K-homomorphisms $L \longrightarrow \overline{K}$ 113
$[L:K]_s$	separable degree of a field extension 113
$\#H$	number of elements in a set 113
$\mathbb{F}_q$	field consisting of $q = p^n$ elements 124
$V(E), V(\mathfrak{a})$	algebraic set attached to $E, \mathfrak{a}$ 126
$I(U)$	vanishing ideal of U 126
$\operatorname{rad} \mathfrak{a}$	radical of an ideal 128
$\operatorname{Aut}_K(L)$	group of K-automorphisms of L 135
$\operatorname{Gal}(L/K)$	Galois group of L/K 135
L^G	fixed field of G 136
$E \cdot E'$	composite field 139
$\varprojlim_{i\in I} G_i$	projective limit 147
$\varinjlim_{i\in I} G_i$	inductive limit 147
$\mathbb{Z}_\ell$	ring of integral ℓ-adic numbers 150
$s_0, \ldots, s_n$	elementary symmetric polynomials in n variables 158
$\operatorname{lexdeg}(f)$	lexicographic degree of a polynomial 160
Δ_f	discriminant of a polynomial 167
$\operatorname{res}(f, g)$	resultant of two polynomials 168
$\mathrm{N}_{A/R}\big(g(x)\big)$	norm of multiplication with $g(x)$ 170
$\mathrm{tr}_{A/R}(a)$	trace of the multiplication by a 173
$D_{A/R}(x_1, \ldots, x_n)$	discriminant of $x_1, \ldots, x_n$ 173
U_n	group of nth roots of unity 176
$\varphi(n)$	Euler's φ-function 177
Φ_n	nth cyclotomic polynomial 182

$\mathrm{tr}_{L/K}(a)$	trace of an element 189
$\mathrm{N}_{L/K}(a)$	norm of an element 189
$H^1(G, A)$	first cohomology group of G with values in A 195
$W(R)$	Witt ring 211
$W_n(X_0, \ldots, X_n)$	Witt polynomial 211
F	Frobenius operator 218
V	Verschiebung operator 218
$K' \otimes_K V$	tensor product 224
$a \otimes v$	tensor 224
τ_g, τ'_g	translations by g 233
int_g	conjugation by g 233
Gx	orbit of x 234
G_x	stabilizer group of x 234
Z_S	centralizer of S 235
Z, Z_G	center of G 235
N_S	normalizer of S 236
$\mathfrak{S}_n$	permutation group 245
$(x_1, \ldots, x_r)$	cycle 245
$\mathrm{sgn}\,\pi$	sign of a permutation 246
$\mathfrak{A}_n$	alternating group 247
$\mathfrak{V}_4$	Klein four-group 248
$[a, b]$	commutator of two elements 249
$[H, H']$	commutator of two subgroups 249
$D^i G$	ith iterated commutator 250
$\mathfrak{K}(M)$	points constructible with compass and straightedge 276
F_ℓ	ℓth Fermat number 281
$\mathrm{card}\, M$	cardinality of a set 286
$\mathrm{transdeg}_K L$	transcendence degree of a field extension 288
$M \otimes_R N$	tensor product of modules 290
$x \otimes y$	tensor 290
M_S	localization of a module 295
$\mathrm{rad}\, R$	radical of a ring 301
$K^{p^{-\infty}} = \bigcup_{i=0}^{\infty} K^{p^{-i}}$	purely inseparable closure of K 303
$\mathrm{Der}_R(A, M)$	A-module of derivations 312
$(\Omega^1_{A/R}, d_{A/R})$	module of differential forms 312

Index

Abel, N. H., 4
action.
  see group action
ℓ-adic
– absolute value, 152
– numbers, 150, 223
adjunction of an element, 4, 83, 101
d'Alembert, J., 3
R-algebra, 94, 296
algebraic
– closure, 84, 93, 104
– element, 88
– equation, 1–7, 23–25, 83, 133
  – degree, 1
  – generic, 4, 157–160, 232, 260
  – integral, 95
  – irreducible, 4, 24
  – metacyclic, 263
  – solvability, 257
  – solvability by radicals, 2–6, 231–232, 255–272
– set, 126
  – geometrically irreducible, 310
  – geometrically reduced, 310
  – irreducible, 132, 310
algebraically
– dependent, 56
– independent, 56, 284
alternating group, 247
angle trisection, 280, 282
Artin, E., 84, 103, 134, 186, 273
Artin–Schreier
– theorem, 198
– theory, 205, 210
associated element, 33
automorphism, 13, 35
– inner, 14, 233

p-basis, 320
Burnside's lemma, 237

canonical forms of matrices, 68, 81
Cantor, G., 6
Cardano, G., 2
Cardano's formulas, 268
cardinality, 286
Cayley's theorem, 14
center, 235
centralizer, 235
character, 186
– linear independence, 134, 186
characteristic of a field, 85
characteristic polynomial, 188
Chinese remainder theorem, 39, 48
class equation, 236
closed subset, 142
closure
– algebraic, 84, 93, 104
– integral, 100
– purely inseparable, 123
– separable algebraic, 123
– topological, 142
coboundary, 195
cocycle, 195
coefficient extension, 295
cofinal subsystem, 149
cohomology group, 195
commutator, 249
– group, 249
– iterated, 250
compass and straightedge construction, 2, 4, 256, 275–282
– of regular polygon, 256, 280–282
composite field, 139
congruence, 40, 50
congruent, 40

conjugate, 135, 233
conjugation, 233
– action, 233
content, 64, 72
coprime
– elements, 47
– ideals, 39
coset, 15, 16
Cramer's rule, 95, 170
cycle, 245
cyclotomic
– field, 157, 177, 179
– polynomial, 182

Dedekind, R., 6, 23, 25
Dedekind ring, 25
degree
– lexicographic, 160
– of a field extension, 87
– of an element, 91
– of a polynomial, 29, 55
derivation, 58, 311
derivative of a polynomial, 58
descent, 224–229
determinant, 188
differential form, 312
dihedral group, 247, 248
directed
– index set, 147, 149
– system, 300
direct limit, 147
discriminant, 154, 161, 162, 166, 172, 173, 264
division ring, 26
division with remainder, 30
divisor, 43
– greatest common, 47–50
duplication of the cube, 1, 256, 280

Eisenstein's criterion, 65
element
– algebraic, 88
– associated, 33
– greatest, 102
– integral, 96
– inverse, 11
– irreducible, 43
– maximal, 102
– nilpotent, 32, 299
– prime, 43
– purely inseparable, 119
– reducible, 43
– separable, 113
– transcendental, 88
elementary divisors, 71, 75
– constructive method, 76
– theorem, 71
endomorphism, 13, 35
epimorphism, 13, 35
equation.
see algebraic equation
Euclidean
– algorithm, 40, 48–50
– division, 30, 41–42
– domain, 41–43, 48–50
– function, 42
Euler, L., 3, 272
Euler's φ-function, 177
exact sequence, 293
exponent of a group, 200
extension of coefficients, 224

factor
– group, 17
– ring, 36
factorial ring, 46
Fermat, P. de, 7, 21
Fermat's
– last theorem, 7
– little theorem, 21
Fermat prime, 281
Ferrari, L., 2
del Ferro, S., 2, 3
field, 27
– algebraically closed, 101
– finite, 24, 38, 123–126
– generated, 90
– homomorphism, 35
– of fractions, 59
– of rational functions, 60
– perfect, 113, 303, 305
– polynomial, 93
field extension, 87
– algebraic, 6, 83, 89
– degree, 87
– finite, 83, 87
– finitely generated, 91

– Galois, 135.
see also Galois extension
– infinite, 87
– multiplicativity formula, 87, 114
– normal, 85, 108
– primary, 305
– purely inseparable, 119
– purely transcendental, 284
– quasi-Galois, 135
– regular, 305
– separable, 113, 302, 318
– separable degree, 113
– separably generated, 303
– simple, 91
– solvable, 257
– solvable by radicals, 256
fixed field, 121, 136
K-form, 224, 225
formal power series, 31
p-free system, 311
Frobenius
– homomorphism, 86, 125
– relative, 125, 149
– operator, 218
function fields, 60
fundamental theorem
– of algebra, 3, 23, 255, 272
– of finitely generated abelian groups, 79, 238
– of finitely generated modules over principal ideal domains, 78
– of Galois theory, 4, 134, 136, 137, 145
– on homomorphisms
– for groups, 17
– for modules, 68
– for rings, 36
– on symmetric polynomials, 160, 163

Galois, E., 4–6, 255
Galois
– cohomology, 195, 199
– descent, 224–229
– extension, 85, 133, 135
– abelian, 134, 139, 200
– cyclic, 134, 139, 194–199, 231
– of Kummer type, 134
– group, 84, 125, 133, 135
– absolute, 149, 205
– as a topological group, 142–152
– closed subgroup, 138, 145
– of an equation, 153–162
– open subgroup, 146
Gauss, C. F., 3, 4, 256, 280
Gauss
– lemma of, 61
– theorem of, 59, 63
generators
– of an ideal, 33
generic equation, 4, 157–160, 232
– solvability by radicals, 260
greatest
– common divisor, 47–50
– element, 102
group, 11
– abelian, 11
– action, 195, 232
– and Galois descent, 226
– transitive, 153, 235
– alternating, 247
– commutative, 11
– cyclic, 20–22, 118
– direct product, 12
– exponent, 200
– finitely generated, 79
– free cyclic, 20
– linear, 263
– nilpotent, 253
– of functions, 12
– of permutations, 12
– operation, 232.
see also group action
– profinite, 149
– solvable, 232, 251
– symmetric, 12, 245
– the concept, 9
– topological, 144
p-group, 237

Hermite, Ch., 6, 56
Hilbert, D., 194
Hilbert's
– basis theorem, 127
– Nullstellensatz, 129, 310
– theorem 90, 194, 228
– additive version, 197
– cohomological version, 195, 206, 222
homomorphism
– finite, 94

– image, 13, 35
– integral, 96
– kernel, 13, 35
– of fields, 35
– of finite type, 94
– of groups, 13
– of modules, 68
– of monoids, 13
– of rings, 35
G-homomorphism, 207
K-homomorphism, 107

ideal, 25, 32
– finitely generated, 33
– generated by elements, 32
– image under a homomorphism, 41
– maximal, 38
– preimage under a homomorphism, 41
– prime, 38
– product, 32
– reduced, 128
– sum, 32
– trivial, 32
index, 16
inductive limit, 147
inseparable degree, 190
integral
– closure, 100
– domain, 26, 27
– element, 96
– equation, 95
intermediate field, 87
inverse element, 11
irreducibility criterion, 65
irreducible element, 43
isomorphism, 13, 35
– theorem
  – for groups, 18
  – for rings, 37
isotropy group, 234

Jacobson ring, 132

Klein four-group, 248
Kronecker, L., 6
Kronecker's
– construction, 25, 50, 84, 100
– symbol, 52
Krull, W., 134
Kummer, E., 134, 200
Kummer
– extension, 200
– theory, 134
  – general, 205–211, 220–223
  – multiplicative, 141, 200–205

Lagrange, J. L., 3, 4, 6, 10, 272
Lagrange
– resolvent, 266
– theorem of, 16
law of composition, 10
– associative, 10
– commutative, 10
least common multiple, 47–48
left translation, 14, 233
Leibniz, G. W., 3
Lie, S., 9
limit
– direct, 147
– inductive, 147
– projective, 147
Lindemann, F., 6, 56, 280
linearly independent, 69
σ-linear map, 226
linear subspace
– defined over K, 225
Liouville, J., 6
localization, 61, 64

map
– R-bilinear, 290
– continuous, 143
– R-linear, 290
maximal
– element, 102
– ideal, 38
minimal polynomial, 83, 89
module, 68, 162
– basis, 69
– direct sum, 69
– exterior power, 75
– finite, 69
– flat, 294
– free, 69
– free system of generators, 69, 162
– length, 70
– localization, 70, 295
– of fractions, 70

– rank, 69
– sum, 69
– system of generators, 69
– torsion-free, 69, 300
G-module, 205
monic polynomial, 29
monoid, 10
monomial, 52
monomorphism, 13, 35
multiplicative system, 60
multiplicativity formula, 87
– for the separable degree, 114

neutral element, 10, 11
nilpotent element, 32, 299
nilradical, 34
Noetherian ring, 45
Noether normalization, 97
norm, 134, 170, 189
– transitivity formula, 190
normal
– closure, 109
– series, 250
normalizer, 236
number
– algebraic, 283
– irrational, 283
– transcendental, 283

open
– neighborhood, 142
– set, 142
operation.
    see group action
orbit-stabilizer lemma, 235
orbit equation, 235
orbits of an action, 234
order
– lexicographic, 160
– of a group, 16
– of an element, 21
– partial, 102
– total, 102

partial fraction decomposition, 64
permutation, 232, 245
– even, 246
– group, 245
– odd, 246
– sign, 246
– signature, 246
Poincaré series, 195, 222
polynomial, 23–25
– g-adic expansion, 32
– degree, 29
– derivative, 58
– elementary symmetric, 158
– function, 23, 131
– generic, 159
– homogeneous, 55
– in several variables, 51–57
– irreducible, 24
– lexicographic degree, 160
– monic, 29
– primitive, 62
– purely inseparable, 119
– reduction of coefficients, 56
– ring
  – in one variable, 28–31
  – in several variables, 51–57
  – universal property, 52
– separable, 111
– symmetric, 160, 162
– total degree, 55
– zero, 56
prime
– element, 43
– factorization, 25, 42, 44–47
– ideal, 38
– number, 43, 47
– subfield, 85
primitive element, 116
– theorem, 116
principal
– ideal, 32, 33
– ideal domain, 25, 33, 43–45, 48
projective
– limit, 147
– system, 146
purely inseparable, 85
– closure, 123
– element, 119

quaternions, 27

radical, 255
– extension, 134
– of an ideal, 128

– of a ring, 34, 301
rational function, 60, 283
– symmetric, 158
reducible element, 43
reduction test for irreducibility, 65
representative, 16
residue class, 16
– group, 17
– module, 68
– ring, 36
resolvent
– cubic, 270
– Lagrange, 266
resultant, 162, 168
– formal degree, 168
right translation, 14, 233
ring, 26, 28
– dimension, 98
– Euclidean, 41–43
– extension, 26
  – finite, 94
  – integral, 96
  – of finite type, 94
– factorial, 46, 63
– fundamental theorem on homomorphisms, 36
– irreducible, 305
– Noetherian, 45, 51, 127
– of formal power series, 31
– of fractions, 60, 64
– of functions, 27
– of Gaussian integers, 42
– of matrices, 27
– reduced, 301
– topological, 150
ring-theoretic product, 28
root of unity, 134, 176
– primitive, 176
Ruffini, P., 4

saturation of a submodule, 71
Schröder–Bernstein theorem, 286
Schur, I., 184
separable, 85
– algebraic closure, 118, 120, 123
separably closed, 308
skew field, 26
splitting field, 84, 107
squaring the circle, 280
stabilizer group, 234
Steinitz, E., 6, 283
subfield, 85
subgroup, 12
– cyclic, 13, 20
– generated, 20
– normal, 16
– trivial, 13
submodule, 68
submonoid, 12
subring, 26
substitution homomorphism, 35, 56
Sylow, L., 232, 237
p-Sylow subgroup, 238
Sylow theorems, 232, 241
symmetric group, 12, 245
system of representatives, 235

tensor, 224, 290
– product, 224
  – coefficient extension, 295
  – of algebras, 296
  – of fields, 298
  – of modules, 290
topological space, 142
– compact, 144
– Hausdorff, 144
– quasicompact, 144
– totally disconnected, 144
topology
– discrete, 143
– generated, 143
– induced, 143
– product, 143
– restriction, 143
– weakest, 143
torsion
– element, 69
– module, 69
– submodule, 69, 78
trace, 134
– of a linear map, 173, 188
– of an element, 189
– transitivity formula, 190
transcendence, 6, 56, 88, 283, 284
– basis, 283, 285
  – separating, 303
– degree, 288
transposition, 245

unique factorization domain, 46, 59, 63
unique prime factorization, 61
unit, 26
– element, 10, 26
– ideal, 32
universal property, 52
upper bound, 102

valuation ring, discrete, 345
Vandermonde, A. T., 4
vector space homomorphism
– defined over K, 225
Verschiebung operator, 218
Viète, F., 3

Weber, H., 9
Witt, E., 135, 205, 211
Witt
– polynomials, 211
– ring, 211, 218
– vectors, 135, 211–223
  – ghost components, 218
  – of finite length, 219, 220

zero
– divisor, 27
– element, 11, 26
– ideal, 32
– multiplicity of, 57
– of a polynomial, 57
– polynomial, 52
– ring, 26
Zorn's lemma, 102