Galois Theory for Beginners

A Historical Perspective

STUDENT MATHEMATICAL LIBRARY
Volume 35

Galois Theory for Beginners

A Historical Perspective

Jörg Bewersdorff

Translated by David Kramer

AMS
American Mathematical Society
Providence, Rhode Island

Originally published in the German language by
Friedr. Vieweg & Sohn Verlag, 65189 Wiesbaden, Germany,
as "Jörg Bewersdorff: Algebra für Einsteiger. 2. Auflage (2^{nd} edition)".

Translated by David Kramer with additions and corrections
by the author.

2000 *Mathematics Subject Classification.* Primary 12–01;
Secondary 12F10.

For additional information and updates on this book, visit
www.ams.org/bookpages/stml-35

Library of Congress Cataloging-in-Publication Data

Bewersdorff, Jörg.
[Algebra für Einsteiger. English]
Galois theory for beginners : a historical perspective / Jörg Bewersdorff ;
translated by David Kramer.
p. cm. — (Student mathematical library, ISSN 1520-9121 ; v. 35)
Includes index.
ISBN-13: 978-0-8218-3817-4 (acid-free paper)
1. Galois theory. 2. Polynomials. I. Title.

QA214.B49 2006
512′.32—dc22 2006048423

10 9 8 7 6 5 4 3 2 1 11 10 09 08 07 06

Contents

Preface to the English Edition vii

Prefaces to the German Editions ix

Chapter 1. Cubic Equations 1

Chapter 2. Casus Irreducibilis: The Birth of the Complex Numbers 9

Chapter 3. Biquadratic Equations 23

Chapter 4. Equations of Degree n and Their Properties 27

The Fundamental Theorem of Algebra: Plausibility and Proof 32

Chapter 5. The Search for Additional Solution Formulas 37

Permutations 42

The Fundamental Theorem on Symmetric Polynomials 47

Ruffini and the General Equation of Fifth Degree 50

Chapter 6. Equations That Can Be Reduced in Degree 55

The Decomposition of Integer Polynomials 57

Eisenstein's Irreducibility Criterion 60

Chapter 7. The Construction of Regular Polygons 63

Constructions with Straightedge and Compass 69

The Classical Construction Problems 74

Chapter 8. The Solution of Equations of the Fifth Degree 81

The Transformations of Tschirnhaus and of Bring and Jerrard 89

Chapter 9. The Galois Group of an Equation 93

Computing the Galois Group 114

A Quick Course in Calculating with Polynomials 119

Chapter 10. Algebraic Structures and Galois Theory 125

Groups and Fields 130

The Fundamental Theorem of Galois Theory: An Example 144

Artin's Version of the Fundamental Theorem of Galois Theory 149

The Unsolvability of the Classical Construction Problems 161

Epilogue 165

Index 177

Preface to the English Edition

This book is a translation of the second edition of my German book *Algebra für Einsteiger: Von der Gleichungsauflösung zur Galois-Theorie*, Vieweg, 2004. The original German edition has been expanded by the addition of exercises. The goal of the book is described in the original preface. In a few words it can be sketched as follows: Galois theory is presented in the most elementary way, following the historical evolution. The main focus is always the classical application to algebraic equations and their solutions by radicals. I am grateful to David Kramer, who did more than translate the present book, having also offered several suggestions for improvements. My thanks are also directed to Ulrike Schmickler-Hirzebruch, of Vieweg, who first proposed a translation to the American Mathematical Society, and to Edward Dunne, of the AMS, for managing the translation.

Jörg Bewersdorff

Translator's Note

I wish to express my appreciation to Jörg Bewersdorff for his helpful collaboration on the translation and to the following individuals at the American Mathematical Society: Edward Dunne for entrusting

me with this project, Barbara Beeton for her friendly and intelligent TEXnical support, and Arlene O'Sean for her careful copyediting of the translation.

David Kramer

Prefaces to the German Editions

Math is like love; a simple idea, but it can get complicated.

—R. Drabek

Preface to the First German Edition

The subject of this book is the history of a classical problem in algebra. We will recount the search for formulas describing the solutions of polynomial equations in one unknown and how a succession of failures led finally to knowledge of a quite unexpected sort, and indeed, of fundamental importance in mathematics.

Let us look briefly at the object that enticed many of the world's best mathematicians over a period of three centuries. Perhaps, dear reader, you recall from your school days quadratic equations of the form

$$x^2 - 6x + 1 = 0$$

as well as the "quadratic formula"

$$x_{1,2} = -\frac{p}{2} \pm \sqrt{\frac{p^2}{4} - q}$$

for the solution of the "general" quadratic equation

$$x^2 + px + q = 0.$$

If we apply this formula to our example, we obtain the two solutions

$$x_1 = 3 + 2\sqrt{2} \quad \text{and} \quad x_2 = 3 - 2\sqrt{2}.$$

If you are interested in a numerical solution, you can pull out your handy pocket calculator (or perhaps you know how to compute square roots by hand) and obtain the decimal representations $x_1 = 5.828427\ldots$ and $x_2 = 0.171572\ldots$. You could also use your calculator to verify that these values are in fact solutions to the original equation. A skeptic who wished to verify that the solutions derived from the formula are the exact solutions would have to substitute the expressions containing the square roots into the equation and demonstrate that the quadratic polynomial $x^2 - 6x + 1 = 0$ actually vanishes—that is, assumes the value zero—at the values $x = x_1$ and $x = x_2$.

The Solution of Equations of Higher Degree. It has long been known how to solve cubic equations such as

$$x^3 - 3x^2 - 3x - 1 = 0$$

by means of a formula similar to the quadratic formula. Indeed, such formulas were first published in 1545 by Cardano (1501–1676) in his book *Ars Magna*. However, they are quite complicated, and have little use for numerical calculation. In an age of practically unlimited computing power, we can do without such explicit formulas in practical applications, since it suffices completely to determine the solutions by means of numeric algorithms. Indeed, for every such equation in a single variable there exist approximation methods that iteratively, that is, step by step, compute the desired solution more and more precisely. Such a procedure is run until the solution has reached an accuracy suitable for the given application.

However such iterative numeric procedures are unsuitable when not only the numerical value of a solution is sought, such as $x_1 = 3.847322\ldots$ in the previous example, but the "exact" value

$$x_1 = 1 + \sqrt[3]{2} + \sqrt[3]{4}.$$

It is not only that such an algebraic representation possesses a certain aesthetic quality, but in addition, a numeric solution is insufficient if

one hopes to derive mathematical knowledge and principles from the solution of the equation. Let us hypothesize, for example, based on numeric calculation, the following identities:

$$\sqrt[3]{\sqrt[3]{2}-1} = \frac{1}{3}\left(\sqrt[3]{3} - \sqrt[3]{6} + \sqrt[3]{12}\right),$$

$$e^{\pi\sqrt{163}} = 262537412640768744,$$

and

$$2\cos\frac{2\pi}{17} = -\frac{1}{8} + \frac{1}{8}\sqrt{17} + \frac{1}{8}\sqrt{34-2\sqrt{17}} + \frac{1}{4}\sqrt{17+3\sqrt{17}-\sqrt{34-\sqrt{17}}-2\sqrt{34+2\sqrt{17}}}.$$

Without going into detail, it seems plausible that behind such identities, if indeed they are correct, lie some mathematical laws. A direct check to determine whether they are in fact correct or are merely the result of chance numeric approximation would be difficult.[1]

But back to Cardano. In addition to the solution for cubic equations, Cardano published in his *Ars Magna* a general formula for quartic equations, that is, equations of the fourth degree, also known as biquadratic equations. Using such formulas, the equation

$$x^4 - 8x + 6 = 0$$

[1] I will reveal that only the first and third identities are correct. The first was discovered by the Indian mathematician Ramanujan (1887–1920) and can be easily checked. The third, which will be discussed in Chapter 7, contains within it a proof that the regular heptadecagon (seventeen-sided polygon) can be constructed with straightedge and compass.

The second equation is not exact. The actual value of the right-hand side is

$$262537412640768743.999999999992501\ldots.$$

However, this approximate identity is more than mere chance. It is based on some deep number-theoretic relationships. For more on this, see Philip J. Davies, Are there coincidences in mathematics? *American Mathematical Monthly* 88 (1981), pp. 311–320.

can be shown to have the solution

$$x_1 = \frac{\sqrt{2}}{2}\left(\sqrt{\sqrt[3]{4+2\sqrt{2}}+\sqrt[3]{4-2\sqrt{2}}} + \sqrt{-\sqrt[3]{4+2\sqrt{2}}-\sqrt[3]{4-\sqrt{2}}+2\sqrt{2\sqrt[3]{3+2\sqrt{2}}+2\sqrt[3]{3-2\sqrt{3}}-2}}\right).$$

With the almost simultaneous discovery of formulas for solving third- and fourth-degree equations came the inevitable problem of finding similar formulas for equations of higher degree. To accomplish this, the techniques that were used for the cubic and quartic equations were systematized, already in Cardano's time, so that they could be applied to equations of the fifth degree. But after three hundred years of failure, mathematicians began to suspect that perhaps there were no such formulas after all.

This question was resolved in 1826 by Niels Henrik Abel (1802–1829), who showed that there cannot exist general solution formulas for equations of the fifth and higher degree that involve only the usual arithmetic operations and extraction of roots. One says that such equations cannot be *solved in radicals*. The heart of Abel's proof is that for the intermediate values that would appear in a hypothetically existing formula, one could prove corresponding symmetries among the various solutions of the equation that would lead to a contradiction.

Galois Theory. A generalization of Abel's approach, which was applicable to all polynomial equations, was found a few years later by the twenty-year-old Évariste Galois (1811–1832). He wrote down the results of his researches of the previous few months on the evening before he was killed in a duel. In these writings are criteria that allow one to investigate any particular equation and determine whether it can be solved in radicals. For example, the solutions to the equation

$$x^5 - x - 1 = 0$$

cannot be so expressed, while the equation

$$x^5 + 15x - 44 = 0$$

has the solution

$$x_1 = \sqrt[5]{-1+\sqrt{2}} + \sqrt[5]{3+2\sqrt{2}} + \sqrt[5]{3-2\sqrt{2}} + \sqrt[5]{-1-\sqrt{2}}.$$

Of much greater significance than such solutions is the method that Galois discovered, which was unorthodox, indeed revolutionary, at the time, but today is quite usual in mathematics. What Galois did was to establish a relationship between two completely different types of mathematical objects and their properties. In this way he was able to read off the properties of one of these objects, namely the solvability of a given equation and the steps in its solution, from those of the corresponding object.

But it was not only the principle of this approach that benefited future mathematics. In addition, the class of mathematical objects that Galois created for the indirect investigation of polynomial equations became an important mathematical object in its own right, one with many important applications. This class, together with similar objects, today forms the foundation of modern algebra, and other subdisciplines of mathematics have also progressed along analogous paths.

The object created by Galois that corresponds to a given equation, called today the *Galois group*, can be defined on the basis of relations between the solutions of the equation in the form of identities such as $x_1^2 = x_2 + 2$. Concretely, the Galois group consists of renumberings of the solutions. Such a renumbering belongs to the Galois group precisely if every relationship is transformed by this renumbering into an already existing relationship. Thus for the case of the relation $x_1^2 = x_2 + 2$ in our example, the renumbering corresponding to exchanging the two solutions x_1 and x_2 belongs to the Galois group only if the identity $x_2^2 = x_1 + 2$ is satisfied. Finally, every renumbering belonging to the Galois group corresponds to a symmetry among the solutions of the equation. Moreover, the Galois group can be determined without knowledge of the solutions.

The Galois group can be described by a finite table that is elementary but not particularly elegant. Such a table is called a *group table*, and it can be looked upon as a sort of multiplication table, in

which each entry is the result of operating on two elements of the Galois group in succession. An example is shown in Figure 0.1. What is significant about the Galois group, and its corresponding group table, is that it always contains the information about whether, and if so, how, the underlying equation can be solved in radicals. To be sure, the proof of this in a concrete application can be quite involved; nevertheless, it can always be accomplished in a finite number of steps according to a fixed algorithm.

	A	B	C	D	E	F	G	H	I	J
A	A	B	C	D	E	F	G	H	I	J
B	B	C	D	E	A	J	F	G	H	I
C	C	D	E	A	B	I	J	F	G	H
D	D	E	A	B	C	H	I	J	F	G
E	E	A	B	C	D	G	H	I	J	F
F	F	G	H	I	J	A	B	C	D	E
G	G	H	I	J	F	E	A	B	C	D
H	H	I	J	F	G	D	E	A	B	C
I	I	J	F	G	H	C	D	E	A	B
J	J	F	G	H	I	B	C	D	E	A

Figure 0.1. The Galois group of the equation $x^5 - 5x + 12$ is represented as a table by means of which the solvability in radicals can be determined by purely combinatorial means. This equation will be considered in detail in Section 9.17. Equations of the fifth degree that are not solvable in radicals have tables of size 60×60 or 120×120.

Today, Galois's ideas are described in textbooks in a very abstract setting. Using the class of algebraic objects that we previously mentioned, it became possible at the beginning of the twentieth century to reformulate what has come to be called Galois theory, and indeed in such a way that the problem itself can be posed in terms of such objects. More precisely, the properties of equations and their solution can be characterized in terms of associated sets of numbers whose common characteristic is that they are closed under the four basic arithmetic operations. These sets of numbers are called *fields*.

Thus starting with a given equation

$$x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0,$$

one forms the smallest set of numbers that contains all quantities, such as

$$\frac{a_2}{a_0} - a_1^2 + a_0,$$

that can be obtained from the coefficients of the equation using successive basic arithmetic operations. Then one obtains an enlarged set of numbers that is of particular use in studying the given equation by allowing in one's calculations, in addition to the coefficients of the equation, the solutions $x_1, x_2, \ldots$. This set is therefore formed of all numbers that can be obtained from expressions of the form, for example,

$$\frac{a_0}{a_2}x_1^2 - a_2 x_2 + a_1.$$

If it now possible to represent the solutions of the given equation by nested expressions involving radicals, then one can obtain additional fields of numbers by allowing in addition to the coefficients some of these nested radicals. Thus every solution of an equation corresponds to a series of nested fields of numbers, and these can be found, according to the main theorem of Galois theory, by analysis of the Galois group. Thus by an analysis of the Galois group alone, one can answer the question whether the solutions of an equation can be expressed in radicals.

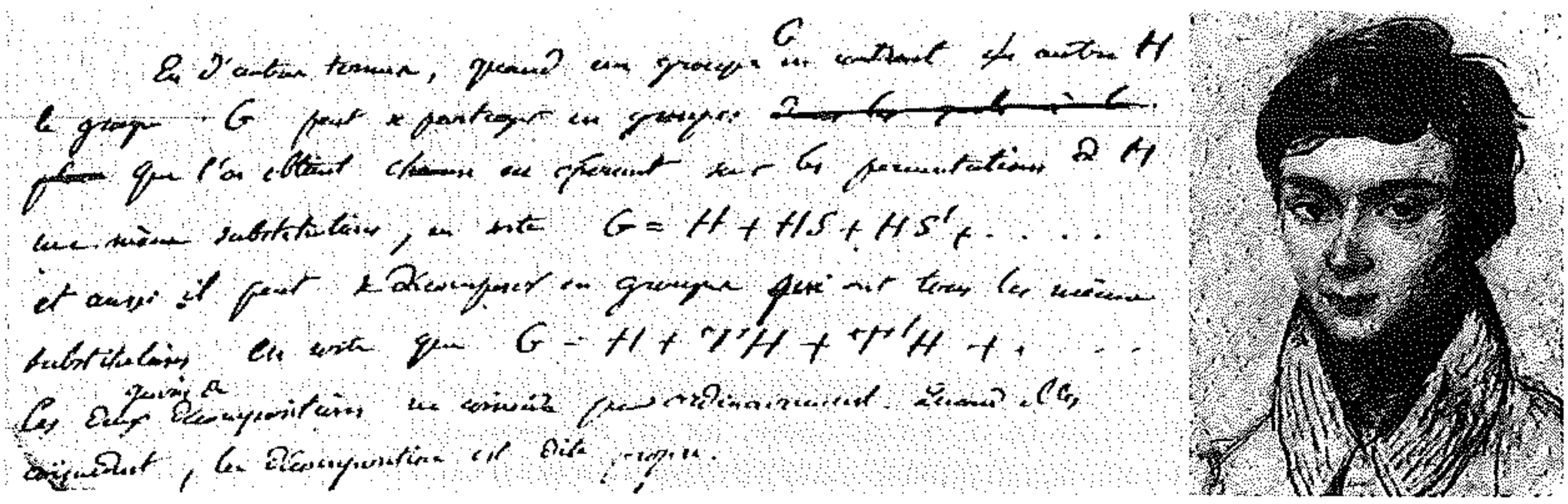


Figure 0.2. Évariste Galois and a fragment from his last letter. In this passage he describes how a group G can be decomposed with the help of the subgroup H. See Section 10.4.

This abstraction achieved at the beginning of the twentieth century and today basically unchanged marks both the end of a historical process during which interest in the problem that we have described has shifted in focus: For Cardano and his contemporaries the main problem was to find concrete solutions to explicit problems using procedures of general applicability. But soon the point of view shifted and the focus was on the important properties of the equations. Beginning with Galois, but in full force only after the turn of the twentieth century, the focus shifted drastically. Now abstract classes of objects such as groups and fields became the basis for the formulation of a host of problems, including those that inspired the creation of these objects in the first place.[2]

About This Book. In order to reach as wide an audience as possible (assumed is only general knowledge obtained from college courses in mathematics), no attempt has been made to achieve the level of generality, precision, and completeness that are the hallmarks of mathematical textbooks. The focus will be rather on ideas, concepts, and techniques, which will be presented only insofar as they are applicable to some concrete application and make further reading in the extensive literature possible. In such a presentation, complicated proofs have no place. However, proofs are without doubt the backbone of any serious engagement with mathematics. In the spirit of compromise, difficult proofs, except those in the last chapter, are set off from the main text so that gaps in the logic can be avoided without the flow of the narrative being interrupted.

Considerable emphasis is placed on the historical development of the subject, especially since the development of modern mathematics in recent centuries is much less well known than that of the natural sciences, and also because it can be very interesting to be able to give a time-lapse view of false starts and important discoveries.

[2]In particular, many important applications have been found in modern information theory, in particular in cryptography, as in, for example, the public key codes realized in 1978. In these asymmetric encryption procedures, the key for encoding is made public without creating the risk of unauthorized decoding. The mathematical basis for such public key encryption algorithms as RSA and ElGamal is computations carried out in special algebraic objects with a very large—but finite—number of elements (precisely, the objects are residue class rings and elliptic curves defined over finite fields). An introduction to this subject can be obtained from Johannes Buchmann, *Introduction to Cryptography*, Springer, 2004.

And furthermore, a presentation that follows the historical development has the advantage of making many mathematical abstractions seem the natural consequence of individual investigations, so that one never gets the impression of starting with an unmotivated definition somehow descended from heaven in a completely arbitrary manner. At the same time, we are able to leave out a great deal of material that would be necessary to include in a work seeking great generality. However, we must mention a significant drawback to our approach: Many complicated calculations will be necessary, even if they are of an elementary nature, whose results would be more simply derived from a qualitative point of view on the basis of general principles.

In order to make this book as distinct as possible from mathematical textbooks, I have chosen the same style of presentation as in my book *Luck, Logic, and White Lies*. Every chapter begins with a simple, usually more or less rhetorical, question that gives the reader an idea of the nature and level of difficulty of the chapter ahead, even if the chapter usually goes far beyond simply answering the question posed. This structure should also offer the more mathematically sophisticated reader, for whom the overview offered here will often be too superficial and incomplete, a quick way of determining which parts of the book are of particular interest, after which the references to the literature will indicate a path of additional reading.

The topics of the individual chapters are too closely woven together to make it possible to read the chapters independently of one another. Nevertheless, the reader who is interested in only a particular aspect of the subject is encouraged to plunge directly into the relevant chapter. Even if one then encounters a reference to another chapter, at least the details of the calculations carried out there will be unnecessary for an understanding of the following chapters. Of course, the beginning of every chapter offers the opportunity to start over if the details of the previous chapter became too difficult.

The reader who wishes to keep the very abstract passages at a greater distance might adhere to the following plan:

- In Chapters 1 through 6 the proofs in the set-off sections may be skipped.

- For understanding the following chapters, the only part of Chapter 7 that is necessary is the first part, which deals with the regular heptadecagon (17-gon).
- Chapter 8 can be omitted entirely.
- In Chapter 9 the set-off sections at the end of the chapter may be skipped.
- Chapter 10 and the epilogue may also be omitted.

Readers who wish to follow a typical "Algebra I" course should place Chapters 9 and 10, which deal with Galois theory, as well as the epilogue, at the center of their reading. For a deep understanding of the subject the following are of particular importance: the main theorem on symmetric polynomials (Chapter 5), the factorization of polynomials (Chapter 6), and the ideas around cyclotomy (the division of the circle) (Chapter 7). How much relative attention should be given to the remaining chapters depends on the reader's interests and prior knowledge.

Following the historical development of the subject, the presentation on the solvability of equations is divided into three parts:

- Classical methods of solution, based on more or less complicated equivalent reformulations of equations, were used historically for deriving the general formulas for quadratic, cubic, and quartic equations (Chapters 1 through 3).
- Systematic investigation of the discovered solution formulas becomes possible when one expresses the intermediate results of the individual calculational steps in terms of the totality of the solutions being sought (Chapters 4 and 5). This leads to the solution of equations in special forms, namely, those that are less complex than those in the general form in that they exhibit particular relationships among the solutions that can be formulated as polynomial identities. In addition to equations that can be broken down into equations of lower degree (Chapter 6), the so-called cyclotomic equations $x^n - 1 = 0$ are examples of such less-complex equations (Chapter 7). Finally, in this part should be included the attempt, described in Chapter 8, at finding a

general solution formula for fifth-degree equations, the result of which is a formula that works only in special cases.

- Based on systematic attempts at finding solution formulas, we finally arrive at the limits of solvability of equations in radicals. These limits, as recognized and investigated by Abel and Galois, are dealt with, aside from a brief preview in Chapter 5, in Chapters 9 and 10. The focus here is on Galois groups.

With the investigation of Galois groups we reach a level of difficulty well beyond that of the first chapters. Therefore, two different presentations are given. In Chapter 9 a relatively elementary overview is given, supplemented by numerous examples, in which the scope of the concepts introduced is reduced as much as possible. The resulting holes are filled in Chapter 10, which leads to the main theorem of Galois theory, which involves the mathematical objects called fields referred to earlier, which are closed under the four basic arithmetic operations. The discussion of these objects will be limited to those aspects relevant to Galois theory.

The reader who wishes to deepen his or her understanding of Galois theory beyond what is contained in this book can move on to any textbook on modern algebra. One might mention as representatives of these books the two classics *Algebra*, by Bartel Leendert van der Waerden (1903–1996), and *Galois theory*, by Emil Artin (1898–1962), whose first editions appeared in 1930 and 1948. But conversely, the present book can be seen as an extension of the usual algebra textbooks in the direction of providing examples and historical motivation.

Acknowledgments

I would like to thank all those who shared in the creation of this book: I received considerable advice about errors and infelicities from Jürgen Behrndt, Rudolf Ketter, and Franz Lemmermeyer. Thanks to their help I was able to reduce the number of errors considerably, though of course the errors that remain are my fault entirely. I thank Vieweg-Verlag and its program director Ulrike Schmickler-Hirzebruch for having accepted this book for publication. Finally, I thank my

wife, Claudia, without whose often tried patience this book could not have been written.

Preface to the Second German Edition

The pleasant circumstance that this book's first edition sold out in only two years gives me the opportunity to expand the bibliography and to correct some errors spotted by several alert readers, particularly Daniel Adler, Ulrich Brosa, Kurt Ewald, Volker Kern, Ralf Krawczyk, and Heinz Lüneburg.

Preface to the Third German Edition

I again have alert readers to thank for the discovery of errors: Erwin Hartmann, Alfred Moser, and David Kramer, who is also the translator of the English-language edition. Finally, I have fulfilled my frequently mentioned desire to provide the book with a set of exercises for the reader.

The Author's Coordinates. Readers are encouraged to report errors or infelicities via e-mail to `mail@bewersdorff-online.de`. Questions will also be answered to the extent possible. Additions and corrections will be published on my website: `http://www.bewersdorff-online.de`. The AMS will also maintain a web page for this book. The URL can be found on the back cover.

Jörg Bewersdorff

Chapter 1

Cubic Equations

Find a number that when added to its cube root yields 6.

1.1 Problems like the one given above have "entertained" generations of schoolchildren. Such problems are at least several hundred years old. They appear as the first thirty problems that were posed to Niccolò Fontana (1499 or 1500–1557), better known as Tartaglia (the stutterer), in a mathematical competition. His challenger was Antonio Fior (1506–?), to whom Tartaglia also posed thirty problems.[1]

As usual, the path to a solution begins with finding an equation that represents the problem. In our example, with x representing the cube root in question, we obtain the equation

$$x^3 + x - 6 = 0.$$

But how are we to solve it? Quadratic equations can always be solved by "completing the square." Then one simply takes the square root and out pops the solution. That is, in the general case of a quadratic equation

$$x^2 + px + q = 0,$$

[1] A complete listing of the thirty problems set by Fior can be found in Renato Acampora, *"Die Cartelli di matematica disfida." Der Streit zwischen Nicolò Tartaglia und Ludovico Ferrari*, Institut für die Geschichte der Naturwissenschaften (Reihe Algorismus, 35), Munich, 2000, pp. 41–44. See also Friedrich Katscher, *Die kubischen Gleichungen bei Nicolo Tartaglia: die relevanten Textstellen aus seinen "Quesiti et inventioni diverse" auf deutsch übersetzt und kommentiert*, Vienna, 2001.

the quantity $\left(\frac{p}{2}\right)^2$ is added to both sides, and the q is moved to the other side of the equation, yielding

$$x^2 + px + \left(\frac{p}{2}\right)^2 = \left(\frac{p}{2}\right)^2 - q.$$

Now the left-hand side of the equation can be represented as a square:

$$\left(x + \frac{p}{2}\right)^2 = \left(\frac{p}{2}\right)^2 - q.$$

And by taking square roots, one obtains the general solution formula

$$x_{1,2} = -\frac{p}{2} \pm \sqrt{\frac{p^2}{4} - q}.$$

Note the following important property of quadratic equations: if one forms the negative sum of the two solutions and their product, one obtains the coefficients of the original equation, namely,

$$x_1 + x_2 = -p \quad \text{and} \quad x_1 x_2 = q.$$

Such completions of the square, in the form of geometric manipulations, were known already to the Babylonians around 1700 BCE. Quadratic equations were treated systematically in the works of the Baghdad scholar al-Khwarizmi (ca. 780–850), which were later translated into Latin, inspiring mathematical progress in Europe for centuries. His name is the origin of the word *algorithm*. Moreover, the word *algebra* is derived from the title of one of his works, *al-Jabr*.

From the modern point of view, al-Khwarizmi's method of handling quadratic equations is quite cumbersome. All statements and proofs are expressed in words, without algebraic symbols, which had not yet been invented. Furthermore, all the argumentation is of a geometric nature. And finally, since negative numbers had not been discovered—and no wonder, given the geometric context—al-Khwarizmi had to distinguish various types of equations, which today we would notate as $x^2 = px$, $x^2 = q$, $x^2 + q = px$, $x^2 + px = q$, and $x^2 = px + q$, and since we have no difficulty accepting coefficients that are less than or equal zero, we can easily reduce all these to a single type.

Figure 1.1 gives an impression of the method of argumentation used by al-Khwarizmi. From the figure one can see that the desired side length x of the inner square can be calculated from area $q =$

$x^2 + px$ of the hatched region using a calculation corresponding to the formula

$$x = \sqrt{q + 4\left(\frac{p}{4}\right)^2} - \frac{p}{2}.$$

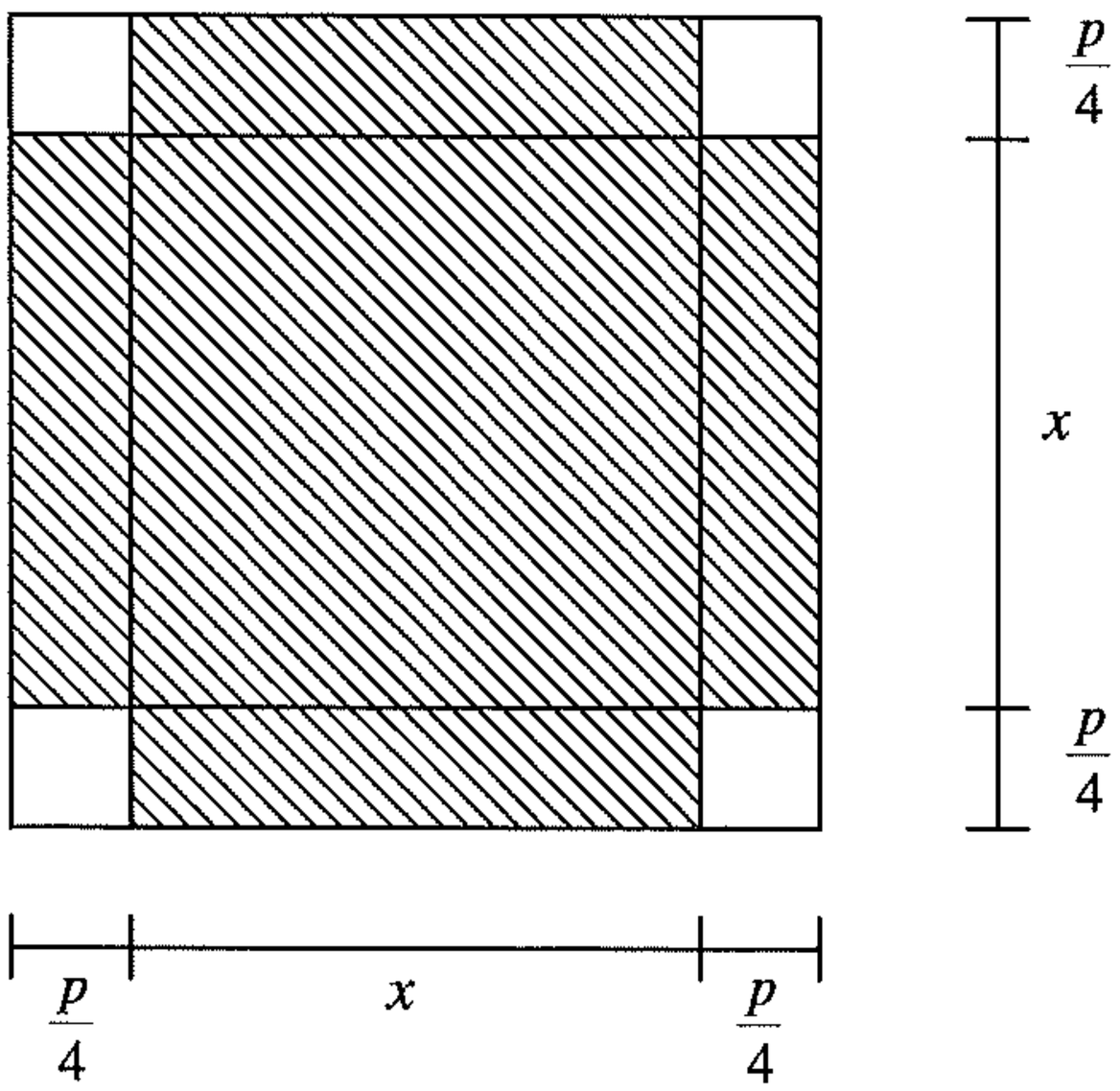


Figure 1.1. Al-Khwarizmi's treatment of the quadratic equation $x^2+px = q$. The hatched area corresponds to the constant term q.

1.2 Following this digression on quadratic equations, we shall turn our attention again to the question posed by Fior. When Fior posed his problems, mathematics had made little progress in the seven hundred years since al-Khwarizmi. European mathematics was marked by its static adherence to the knowledge gained in earlier periods of activity, notably from the Arabs and the ancient Greeks. Arabic numerals had been introduced into Europe, which served the need of the mathematical calculations required in trade. Such calculations were generally carried out by professional masters of calculation.

Although mathematics was used in many commercial applications, negative numbers remained unknown. Moreover, mathematical notation was slow to develop. Thus, for example, in the fifteenth century the notation *R3 V31 m R16* was used for the expression

$$\sqrt[3]{31 - \sqrt{16}}.$$

One of the first great advances beyond Arabic mathematics was taken by Scipione del Ferro (1465?–1526), who taught at the University of Bologna at the beginning of the sixteenth century. He was the first, as reported some decades later by Cardano in his book *Ars Magna*,[2] to solve general cubic equations, namely those of the type $x^3 + px^2 = q$. Without making his method public, del Ferro revealed it to his student Antonio Fior. At this time, Niccolò Fontana, alias Tartaglia, was also working on solving cubic equations. Tartaglia, who was working as a master calculator in Venice, was one of the best mathematicians in Italy.[3] And in fact, he figured out how to solve cubic equations of the type $x^3 + px^2 = q$. However, his method seems to have been less a general solution algorithm than a way of setting up special equations whose solutions could be easily found.[4]

At the competition mentioned at the beginning of the chapter, Fior posed thirty problems of the type $x^3 + px = q$ for Tartaglia to solve, while conversely, Tartaglia posed thirty somewhat atypical problems, including cubic equations of the type $x^3 + px = q$. At first, neither contestant cold solve any of the problems that he had been posed. But shortly before the end of the competition, on 13 February 1535, Tartaglia figured out how to solve equations of the form $x^3 + px = q$. Like del Ferro and Fior, he kept his method of solution a secret.

And now there enters upon the stage the man whose name is associated today with the solution formula. Girolamo Cardano, known today more for his discovery of what is called the Cardano wave and Cardano suspension and by profession actually a physician, managed to convince Tartaglia to reveal to him his formula, under the guarantee—according to Tartaglia afterward—that he would keep it

---

[2]Girolamo Cardano, *The Great Art or the Rules of Algebra*, the English translation of the 1545 edition with additions from the editions of 1570 and 1663 (Cambridge, Massachusetts, 1968); see the beginning of Chapter 1 and Chapter 11.

[3]An idea of the accomplishments of a master calculator, and in particular of the person of Tartaglia, can be found in the historical novel *Der Rechenmeister*, by Dieter Jörgensen, Berlin, 1999. A substantial part of the novel deals with the discovery of the solution formula for cubic equations and the resulting conflict.

[4]See the work cited by Renato Acampora, pp. 32–34. On the other hand, based on the fact that Tartaglia is known to have studied the work of Archimedes, Phillip Schultz speculates (Tartaglia, Archimedes and cubic equations, *Australian Mathematical Society Gazette* 11 (1984), pp. 81–84) that Tartaglia could have used a geometric method in which he determined the intersection point of the parabola $y = x^2$ and the hyperbola $y = -q/(x + p)$.

secret. Nonetheless, Cardano published the solution procedure in his *Ars Magna*, a book describing the current state of algebraic knowledge.[5]

The solution of cubic equations is based on the cubic binomial formula

$$(u+v)^3 = 3uv(u+v) + \left(u^3+v^3\right),$$

which Cardano was able to derive in an analogous manner to the geometric method used by al-Khwarizmi for the quadratic equation, though in this case the argument of course used three-dimensional figures and volumes (see Figure 1.2). However, this identity can also be interpreted as a cubic equation, where the sum $u+v$ yields a solution x of the cubic equation

$$x^3 + px + q = 0$$

if the conditions

$$3uv = -p,$$

$$u^3 + v^3 = -q,$$

are satisfied. The cubic equation $x^3 + px + q = 0$ can then be solved if one can find suitable quantities u and v. But that is a relatively simple task. Since both the sum and product of the quantities u^3 and v^3 are known, one can solve the quadratic equation

$$w^2 + qw - \left(\frac{p}{3}\right)^3 = 0$$

to obtain u^3 and v^3 as the two solutions

$$-\frac{q}{2} \pm \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3},$$

so that u and v can be determined from the two equations

$$u = \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}, \quad v = \sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}.$$

---

[5]The result of Cardano's alleged breaking of his word led to a great falling out between Tartaglia and Cardano. It is thanks to the publications around this dispute (see the works referred to in an earlier footnote) that we have knowledge of the history leading up to the *Ars Magna*.

Finally, the desired solution x of the cubic equation $x^3 + px + q = 0$ is obtained from *Cardano's formula*

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

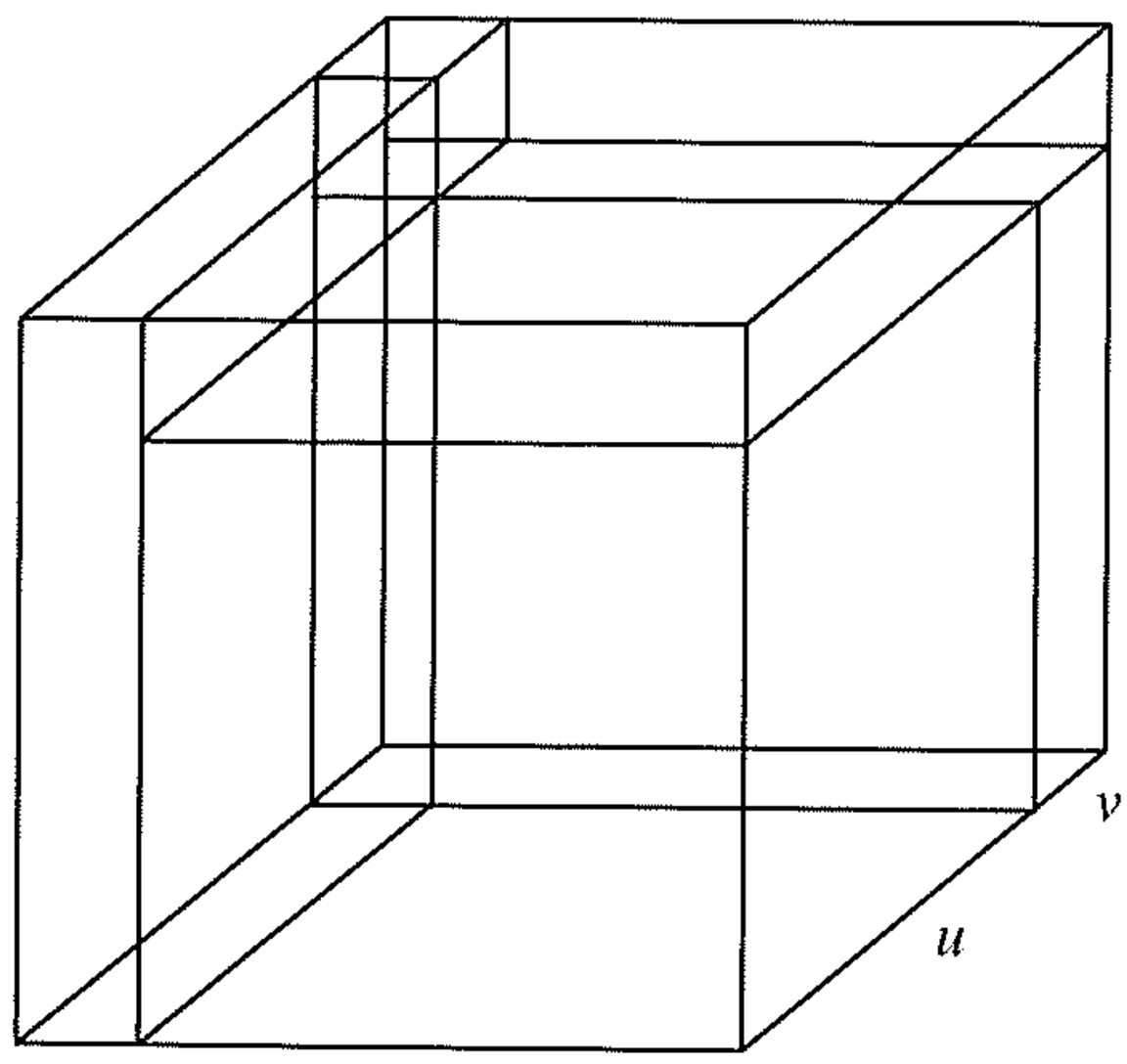


Figure 1.2. Depicted here is the geometric basis of the binomial equation $(u+v)^3 = 3uv(u+v) + (u^3+v^3)$, similar to Cardano's presentation in his *Ars Magna*. The large cube can be decomposed into two subcubes and three rectangular parallelepipeds, all with side lengths u, v, and $u+v$.

If we apply this result to our problem $x^3 + x - 6 = 0$, we obtain

$$x = \sqrt[3]{3 + \frac{2}{3}\sqrt{\frac{61}{3}}} + \sqrt[3]{3 - \frac{2}{3}\sqrt{\frac{61}{3}}},$$

whose decimal value is approximately 1.634365.

1.3 In his *Ars Magna* Cardano also solved cubic equations involving quadratic terms.[6] We have already seen, in the introduction, an

[6] *Ars Magna*, Chapter XXIII.

example of such an equation with a quadratic term:

$$x^3 - 3x^2 - 3x - 1 = 0.$$

To solve such equations, Cardano transformed them using a generally applicable procedure into equations of the form $y^3 + py + q = 0$. Starting with a cubic equation in the general form

$$x^3 + ax^2 + bx + c = 0,$$

the transformation consists in adding the summand $\frac{a}{3}$ to the desired solution x, which allows the quadratic and cubic terms to be combined:

$$x^3 + ax^2 = \left(x + \frac{a}{3}\right)^3 - \frac{a^2}{3}x - \frac{a^3}{27} = \left(x + \frac{a}{3}\right)^3 - \frac{a^2}{3}\left(x + \frac{a}{3}\right) + \frac{2}{27}a^3.$$

To obtain the complete transformation of the coefficients of this equation, one replaces every occurrence of x in the equation via the *substitution*

$$x = y - \frac{a}{3},$$

obtaining, after collecting terms in like powers of y, the identity

$$x^3 + ax^2 + bx + c = y^3 + py + q,$$

with

$$p = -\frac{1}{3}a^2 + b,$$

$$q = \frac{2}{27}a^3 - \frac{1}{3}ab + c.$$

Once one has solved the reduced cubic equation $y^3 + py + q = 0$ with Cardano's formula, the solution of the original equation can be obtained with the transformation $x = y - \frac{a}{3}$. In the concrete example $x^3 - 3x^2 - 3x - 1 = 0$, the transformation $x = y + 1$ leads to the equation

$$y^3 - 6y - 6 = 0,$$

whose solution

$$y = \sqrt[3]{2} + \sqrt[3]{4}$$

obtained from Cardano's formula leads to the following solution of the original equation:

$$x = 1 + \sqrt[3]{2} + \sqrt[3]{4}.$$

In addition to the progress in calculation evidenced in Cardano's *Ars Magna*, two fundamental developments are to be found that would aid in the future development of mathematics, namely the extension of the set of numbers to include first negative quantities and then complex numbers. Cardano did not in fact use negative numbers in the *Ars Magna*, which would have allowed him to solve various types of cubic equations such as $x^3 + px = q$ and $x^3 = px + q$ as a single type. But he did show a greater openness to negative numbers by listing in addition to the "true" solutions to an equation the negative solutions, which he called "false" solutions. For Cardano, a "false" solution corresponded to a "true" solution of another equation, namely one with x replaced by $-x$. For example, for Cardano -4 is a false solution to the equation $x^3 + 16 = 12x$, while 4 is a true solution of the equation $x^3 = 12x + 16$.[7]

Exercises

(1) Find a solution to the cubic equation

$$x^3 + 6x^2 + 9x - 2 = 0.$$

(2) The cubic equation

$$x^3 + 6x - 20 = 0$$

has 2 as a solution. How is this solution given by Cardano's formula?

[7] *Ars Magna*, Chapter I.

Chapter 2

Casus Irreducibilis: The Birth of the Complex Numbers

If you try to solve the cubic equation $x^3 = 8x + 3$ using Cardano's formula, the formula appears to fail. But this does not indicate that the equation has no solution, for clearly $x = 3$ is a solution.

2.1 Like the problem that introduced Chapter 1, the above equation may also be considered "classical," since it comes from Cardano's book *Ars Magna*.[1] However, Cardano, who simply gives 3 as a solution and then calculates two additional solutions, does not go more deeply into the difficulties presented by his formula, though they can hardly have been unknown to him.[2]

Let us look at the details. From the coefficients $p = -8$ and $q = -3$ of the equation, one obtains not the expected solution $x = 3$ but the complicated expression

$$x = \sqrt[3]{\frac{3}{2} + \frac{19}{6}\sqrt{-\frac{5}{3}}} + \sqrt[3]{\frac{3}{2} - \frac{19}{6}\sqrt{-\frac{5}{3}}},$$

[1] Chapter XIII.

[2] Cardano mentions the problem in a 1539 letter to Tartaglia, thus six years before the publication of *Ars Magna*. See the work by Acampora cited earlier, pp. 62–63. Moreover, because negative numbers were not frequently used, such situations could not arise in equations of the type $x^3 + px = q$.

whose simplification would have been practically out of the question in Cardano's time on account of the square roots of negative numbers, even though at another place in the *Ars Magna* Cardano makes some tentative calculations with such square roots of negative numbers when he seeks to solve the problem of finding two numbers whose sum is 10 and whose product is 40. Of course, Cardano knew how to find the solutions, if in fact they existed, namely, as the solutions of the quadratic equation

$$x^2 - 10x + 40 = 0,$$

with the result that one obtains the following two numbers:

$$5 + \sqrt{-15} \quad \text{and} \quad 5 - \sqrt{-15}.$$

Although both of the numbers found hardly represent what in Cardano's time would have been thought of as numbers, Cardano dared to carry out the calculation

$$\left(5 + \sqrt{-15}\right)\left(5 - \sqrt{-15}\right) = 25 + 15 = 40,$$

thereby performing the first known calculation with what we today call complex numbers.[3]

It is such calculations, using the rules of arithmetic to achieve meaningful results, that doubtless motivated mathematicians following Cardano to allow square roots of negative numbers, at first as intermediate values in a calculation, and later as mathematical objects in their own right, for which there developed an independent interest. An important role in this extension of the set of allowable numbers was played by the so-called *casus irreducibilis*, which denoted the case in the solution of cubic equations in which within Cardano's formula there appeared square roots of negative numbers. In particular, this situation occurs in the case of reduced cubic equations of the form

$$x^3 + px + q = 0$$

when the radicand of the square root is negative:

$$\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3 < 0.$$

[3] *Ars Magna*, Chapter XXXVII, Rule II.

Is it possible to calculate with such roots of negative numbers and end up with the correct results? The first steps in this direction were taken by Rafael Bombelli (1526–1572) in his book *L'Algebra*, published in the year of his death. There he solved the equation

$$x^3 = 15x + 4,$$

by courageously calculating with the radical expression

$$x = \sqrt[3]{2 + \sqrt{-121}} + \sqrt[3]{2 - \sqrt{-121}} = \sqrt[3]{2 + 11\sqrt{-1}} + \sqrt[3]{2 - 11\sqrt{-1}}$$

produced by Cardano's formula. Finally, with the known solution $x = 4$ staring him in the face, he obtained values for the two cubic roots, namely, by calculating

$$\left(2 + \sqrt{-1}\right)^3 = 8 + 12\sqrt{-1} - 6 - \sqrt{-1} = 2 + 11\sqrt{-1},$$

$$\left(2 - \sqrt{-1}\right)^3 = 8 - 12\sqrt{-1} - 6 + \sqrt{-1} = 2 - 11\sqrt{-1},$$

thereby obtaining the desired $x = 2 + \sqrt{-1} + 2 - \sqrt{-1} = 4$.

It thus seemed that the complicated expression involving the square roots of negative numbers was equal to 4. Bombelli commented thus: "An extravagant thought, according to many. I myself was for a long time of the same opinion. The matter seemed to rest more on sophistry than on truth, but I searched until I found a proof."[4] These daring calculations had provided an explanation for an already known result, just as in the history of mathematics it is certain that similar developments occurred when negative numbers were first viewed as a useful addition to the set of permissible numbers. In comparison to the negative numbers, the square roots of negative numbers introduced a greater level of abstraction, since there is no obvious analogy in our everyday experience, in contrast to the negative balance in a bank account, which can be represented by a negative number. Thus it took almost another two hundred years before the objects hesitantly introduced by Bombelli were accepted into general mathematical use, under the name *complex numbers*. What was needed was a description of their fundamental properties so that there would be no doubt as to how they could be used. This could occur only when the more or less philosophical question, "what actually are complex numbers?"

[4]Quoted by Moritz Cantor, *Vorlesungen über Geschichte der Mathematik*, Berlin, 1900–1908, Band 2, p. 625.

was put aside in favor of *defining* them on the basis of their properties. The first decisive steps in this direction were taken in 1797 by Caspar Wessel (1745–1818).

However, Wessel's formal definition by no means eliminated all doubt, not least because his writings were not widely circulated. Thus it was for almost another half century that these *imaginary* or *impossible* numbers eked out an existence similar to that of the then still current *infinitesimals*, those infinitely small quantities of mathematical analysis (calculus): Mathematicians could use them effectively and elegantly to obtain "correct" results quickly, results that then might well be obtained by another method without use of the suspect intermediate steps. Thus even the great Carl Friedrich Gauss (1777–1855), on the occasion of his 1796 formulation of the fundamental theorem of algebra, which uses complex numbers decisively, wrote thus:

> I will carry out my proof without the use of imaginary quantities, although I could well have allowed myself that luxury employed by all modern analysts.[5]

Perhaps Gauss hit upon the essence of the reservations of many—even Leibniz (1646–1716) had spoken in 1702 of a "wonder of analysis, a monstrosity of the human imagination"—thirty-six years later, after he had meanwhile been often compelled to mention the "metaphysics" of complex numbers:

> The difficulties that one believes to surround the theory of imaginary quantities have their basis in large measure in the less than optimal nomenclature. If one had ... called positive numbers "direct," negative numbers "inverse," and imaginary numbers "lateral," there would have been simplicity instead of confusion, clarity instead of darkness.

This remark of Gauss should be understood generally to indicate that mathematical definitions should be viewed in isolation and freed from later interpretation and nomenclature that reflects such interpretations: A mathematical object "lives" only because of its freedom

[5]This and following quotations are taken from Herbert Pieper, *Die komplexen Zahlen*, Frankfurt/M., 1999. The last chapter of this book offers an extensive presentation of the history of complex numbers.

from contradiction. It is "created" when a use for it is conceived, and "tended" as long as it continues to serve a purpose.

2.2 The set of complex numbers includes by definition all pairs (a, b) whose coordinates a and b are real numbers. Geometrically, the set of complex numbers can be viewed as a plane, in analogy with the number line representing the set of real numbers. With the idea that a pair of numbers (a, b) is to interpreted as

$$a + b\sqrt{-1},$$

one *defines* the mathematical operations on complex numbers as follows:

$$(a, b) + (c, d) = (a + c, b + d),$$
$$(a, b) \times (c, d) = (ac - bd, bc + ad).$$

The inverse operations are explained in reference to so-called inverse elements. That is, subtraction is defined as the addition of a negated value, and division as multiplication by the multiplicative inverse. The inverse elements are defined thus:

$$-(a, b) = (-a, -b)$$
$$(a, b)^{-1} = \left(\frac{a}{a^2 + b^2}, \frac{-b}{a^2 + b^2}\right).$$

Of course, in the last definition it is assumed that $(a, b) \neq (0, 0)$.

Indeed, these definitions accomplish the desired goal, since except for failing to have an order relation, the complex numbers have all the familiar laws of operation of the real numbers. We shall list these laws, which the reader can easily verify from the definitions, so that we will have some common notation and nomenclature:

- All the relations among real numbers, such as the commutative, associative, and distributive law properties, continue to hold. Furthermore, the zero $(0, 0)$ and the one $(1, 0)$ have the familiar properties of serving as identity elements in addition and multiplication. Finally, subtraction and division are in fact the inverse operations of addition and multiplication.[6]

[6] Taken all together, a set with two operations satisfying these conditions is called a *field*. We shall return to this concept in Chapters 9 and 10.

- The subset of complex numbers of the form $(a, 0)$ behaves under these operations like the set of real numbers, and it can be identified with them, just as the set of fractions with denominator 1 can be identified with the integers. The complex numbers can therefore be looked on as an extension of the real numbers. For simplicity, we will often write a complex number of the form $(a, 0)$ simply as a, and for a complex number (a, b) we call a the *real part*.
- We have $(0, 1) \times (0, 1) = (0, -1) \times (0, -1) = (-1, 0)$, a result that corresponds to the real number -1. Therefore, the two complex numbers $(0, 1)$ and $(0, -1)$ can be interpreted as square roots of -1. The number $(0, 1)$ is given the special notation i, called the imaginary unit. In a complex number (a, b), we call b the *imaginary part*.
- We have the equation $(a, b) \times (a, -b) = a^2 + b^2$, where $(a, -b)$ is called the *conjugate* of the complex number (a, b). It is denoted by $\overline{(a, b)}$. We call $\sqrt{a^2 + b^2}$ the *absolute value* or *modulus* of the number (a, b). Within the *complex plane*, as the geometric representation of the complex numbers is called, the modulus of a complex number represents the distance of the number from the origin.[7] An example is displayed in Figure 2.1. Finally, complex conjugation (taking the conjugate of a complex number) possesses the following property: $\overline{((a, b) \times (c, d))} = \overline{(a, b)} \times \overline{(c, d)}$.

All these properties together make us certain that with pairs (a, b) of the form $(a, b) = (a, 0) + (b, 0) \times (0, 1) = a + bi$, where $i^2 = 1$, we have in fact defined the set of mathematical objects of the form $a + b\sqrt{-1}$. This resulting extension of the real numbers was achieved without using the previously undefined expression $\sqrt{-1}$, whose use, moreover, is not always completely unproblematic, since it can easily lead to erroneous calculations such as $\sqrt{-1}\sqrt{-1} = \sqrt{(-1)(-1)} = \sqrt{1} = 1$.

Another, very important, example of the complex numbers that will be of use to us in what follows is closely related to the geometric representation of the complex numbers. We first note that every

[7] The definition of the distance between two complex numbers as the modulus of their difference makes possible the creation of a function theory, or complex analysis, on the complex numbers, whereby notions such as convergence, continuity, derivative, and integral are defined with properties similar to those of classical analysis.

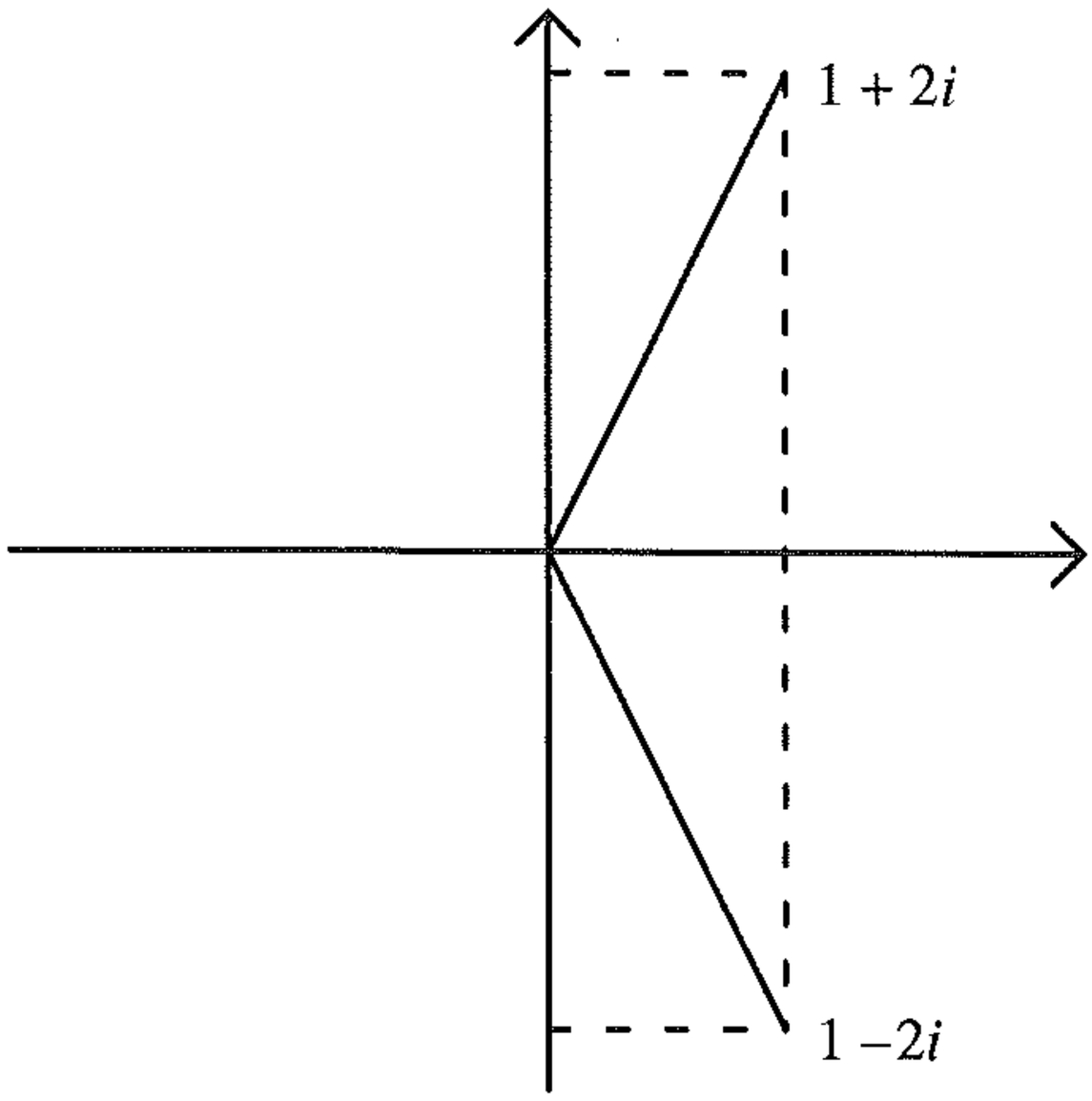


Figure 2.1. The complex plane with the number $1 + 2i$ and its conjugate $1 - 2i$. The modulus of both numbers is $\sqrt{5}$.

complex number located on the unit circle, that is, the circle of radius 1 with center at the origin, can be represented in terms of the trigonometric functions sine and cosine. To be precise, such complex numbers have a representation

$$\cos\phi + i\sin\phi,$$

where ϕ is the angle that runs counterclockwise from the positive horizontal axis (i.e., the positive real axis) to the line from the origin to the complex number in question (see Figure 2.2). If we now multiply together two such numbers lying on the unit circle, we can do so merely by adding their angles together. A proof of this follows easily from the addition laws for sine and cosine:

$$\begin{aligned}(\cos\phi + i\sin\phi)(\cos\psi + i\sin\psi)&\\ &= (\cos\phi\cos\psi - \sin\phi\sin\psi) + i(\cos\phi\sin\psi + \sin\phi\cos\psi)\\ &= \cos(\phi+\psi) + i\sin(\phi+\psi).\end{aligned}$$

If we now introduce the modulus of a complex number (which is 1 for the complex numbers lying on the unit circle), we can generalize this result to all nonzero complex numbers. Note first that if a complex number has nonzero (hence positive) modulus m, we

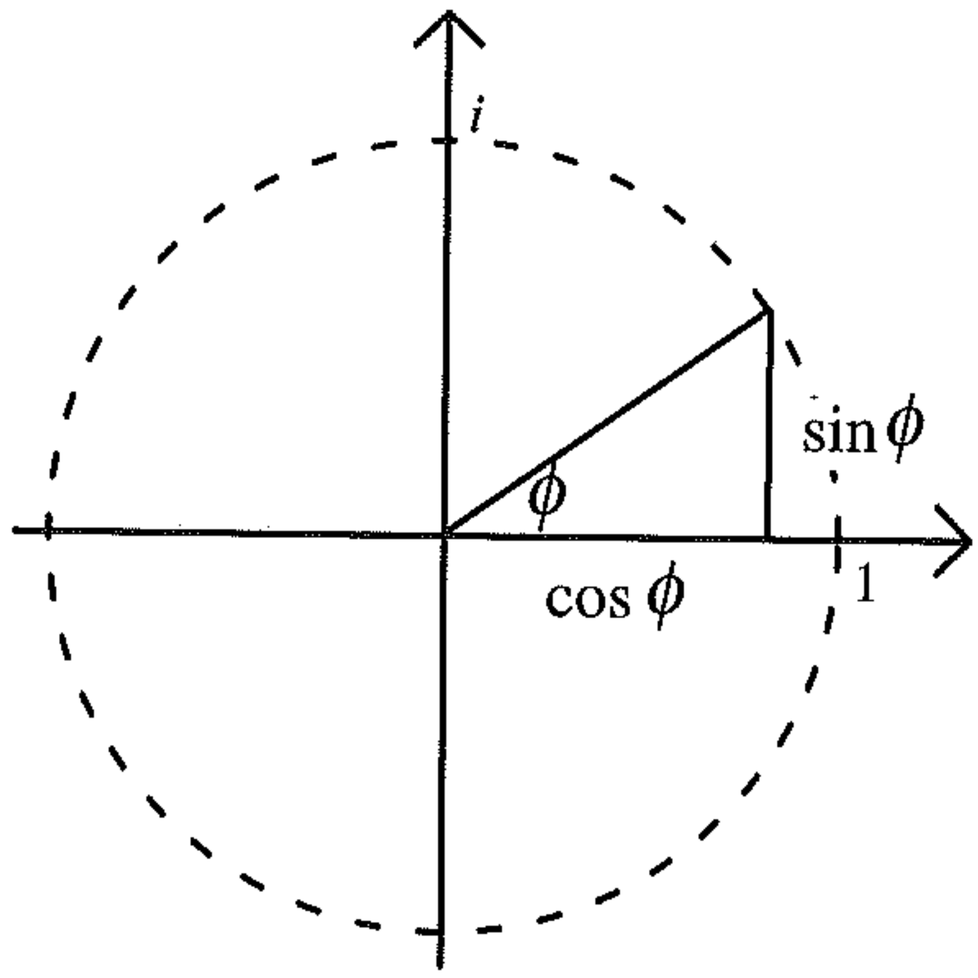


Figure 2.2. Representation of a complex number of the form $\cos\phi + i\sin\phi$ located on the unit circle.

can set $s = \ln m$ (the natural logarithm) and write $m = e^s$. Thus a complex number z with angle ϕ and modulus e^s can be written $z = e^s(\cos\phi + i\sin\phi)$. We then see[8] that

$$e^s(\cos\phi + i\sin\phi) \times e^t(\cos\psi + i\sin\psi) = e^{s+t}\big(\cos(\phi+\psi) + i\sin(\phi+\psi)\big).$$

The special case of raising a complex number to a power goes under the name *de Moivre's formula*, even though its namesake Abraham de Moivre (1667–1754) never formulated it explicitly:

$$\big(e^s(\cos\phi + i\sin\phi)\big)^n = \big(e^s\big)^n\big(\cos(n\phi) + i\sin(n\phi)\big).$$

2.3 Before we return to the *casus irreducibilis*, we would like to apply the knowledge we have just gained to the equation

$$x^3 - 1 = 0.$$

In the field of real numbers it is clear that $x_1 = 1$ is the only solution. If we move to the field of complex numbers, then de Moivre's formula suggests that the equation must have two additional solutions, both of which lie on the unit circle (see Figure 2.3) and form angles with the positive real axis of $\frac{2\pi}{3}$ and $\frac{4\pi}{3}$, so that the three solutions form

[8] The reason for the validity of this equation will become clear when the power series for sine, cosine, and exponential functions are extended to the complex numbers, which was first accomplished in 1748 by Leonhard Euler (1707–1783). One can then see that for arbitrary complex numbers $x + iy$, one has the identity $e^{x+iy} = e^x(\cos y + i\sin y)$.

an equilateral triangle. It is thus apparent that the two additional solutions are

$$x_2 = -\frac{1}{2} + \frac{\sqrt{3}}{2}i \quad \text{and} \quad x_3 = -\frac{1}{2} - \frac{\sqrt{3}}{2}i.$$

These three solutions are called third *roots of unity*. Equations of the form $x^n - 1 = 0$, which form the topic of Chapter 7, will be called *cyclotomic equations*[9] because of their geometric significance.

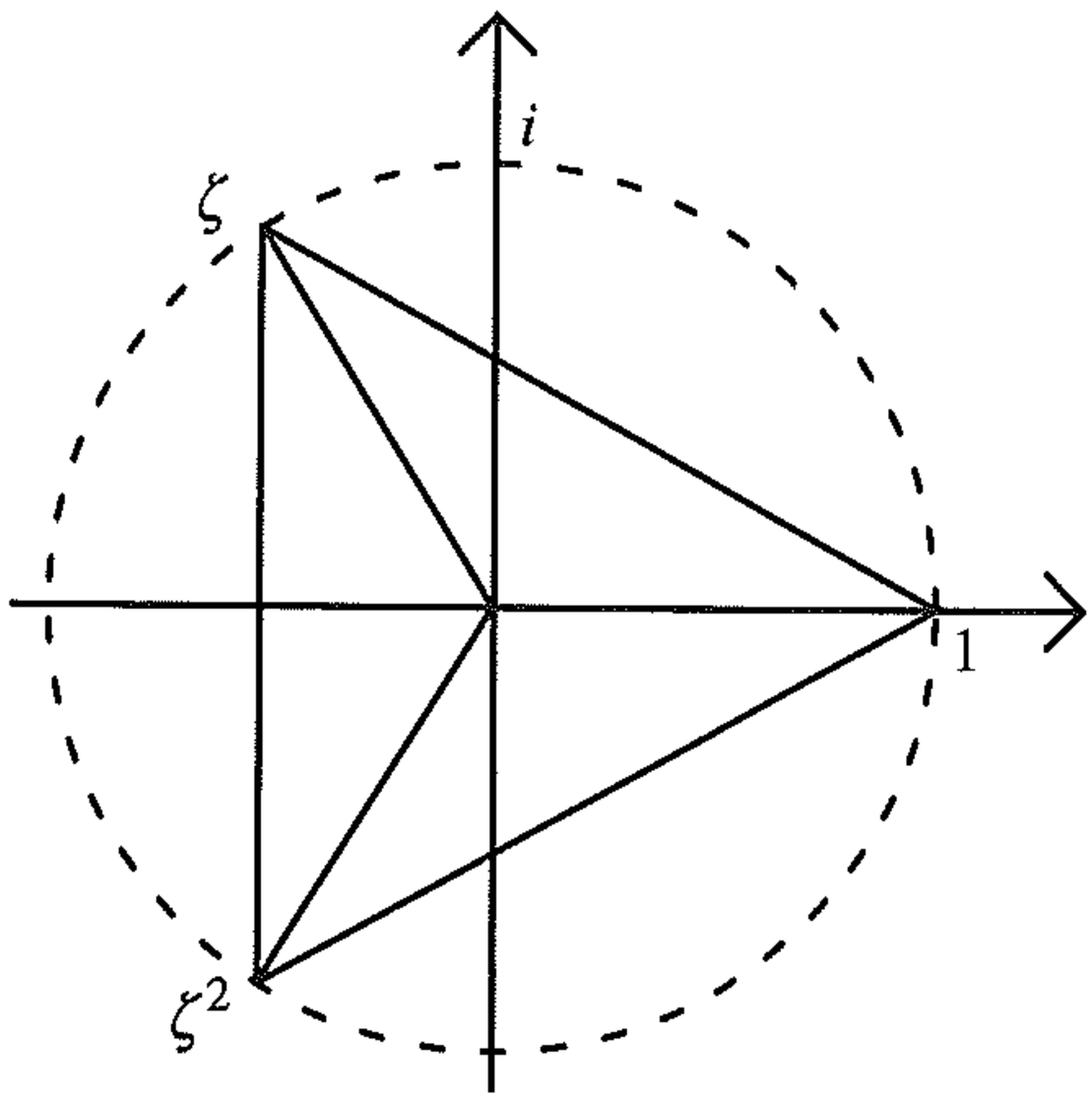


Figure 2.3. The three solutions 1, ζ, and ζ^2 of the cyclotomic equation $x^3 - 1 = 0$.

2.4 The cube roots of unity are of significance for the general cubic equation in that with them one can extend Cardano's formula so that three roots are always obtainable. We begin by defining

$$\zeta = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$$

and recognizing that the two equations underlying Cardano's formula,

$$3uv = -p,$$
$$u^3 + v^3 = -q,$$

have as solutions, in addition to the pair (u, v) introduced in Chapter 1, the solutions $(\zeta u, \zeta^2 v)$ and $(\zeta^2 u, \zeta v)$, so that altogether, one

[9] From the Greek *kuklos*, circle, and *-tomia*, cutting.

obtains the following three solutions to the reduced cubic equation $x^3 + px + q = 0$:

$$x_1 = \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}},$$

$$x_2 = \zeta \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}} + \zeta^2 \sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}},$$

$$x_3 = \zeta^2 \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}} + \zeta \sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}.$$

This formula holds in general, since no restrictions on the solutions were assumed in its derivation. However, in the case of *casus irreducibilis*, one must note that for the two complex numbers u^3 and v^3, which are conjugate to each other, pairs of cube roots u and v are selected that are also conjugate to each other; only in this way can the two equations that determine u and v be satisfied. Not only have we shown that the calculations carried out by Bombelli are justified, but furthermore, we see that in the general case, the three solutions x_1, x_2, and x_3 are all real, since

$$\overline{x_j} = \overline{\zeta^{j-1}u + \zeta^{-(j-1)}v} = \zeta^{-(j-1)}\overline{u} + \zeta^{j-1}\overline{v} = \zeta^{-(j-1)}v + \zeta^{j-1}u = x_j,$$

for $j = 1, 2, 3$. That is, in employing Cardano's formula, it is necessary to calculate with complex numbers even in the case that all three solutions are real and distinct.

For the problem $x^3 = 8x + 3$, which appears at the beginning of Cardano's *Ars Magna*, one obtains the solution

$$\begin{aligned} x_1 &= \sqrt[3]{\frac{3}{2} + i\frac{19}{6}\sqrt{\frac{5}{3}}} + \sqrt[3]{\frac{3}{2} - i\frac{19}{6}\sqrt{\frac{5}{3}}} \\ &= \frac{1}{2}\left(3 + i\sqrt{\frac{5}{3}}\right) + \frac{1}{2}\left(3 - i\sqrt{\frac{5}{3}}\right) \\ &= 3. \end{aligned}$$

The other two solutions, which Cardano knew, are

$$x_2 = \frac{1}{4}\left(-1+i\sqrt{3}\right)\left(3+i\sqrt{\frac{5}{3}}\right)+\frac{1}{4}\left(-1-i\sqrt{3}\right)\left(3-i\sqrt{\frac{5}{3}}\right)$$
$$= \frac{1}{2}\left(-3-\sqrt{5}\right)$$

and

$$x_3 = \frac{1}{4}\left(-1-i\sqrt{3}\right)\left(3+i\sqrt{\frac{5}{3}}\right)+\frac{1}{4}\left(-1+i\sqrt{3}\right)\left(3-i\sqrt{\frac{5}{3}}\right)$$
$$= \frac{1}{2}\left(-3+\sqrt{5}\right).$$

So, was all this effort worth it? In any case, the extension of the underlying field of numbers to the complex numbers has converted the solution algorithm into a unified process. Moreover, the extension to the complex numbers has removed the uncertainty that we might obtain incorrect results in calculating with nonreal intermediate results. However, in actual calculations there is still one problem: Our procedure does not provide an effective way of simplifying expressions of the type

$$\sqrt[3]{\frac{3}{2}+i\frac{19}{6}\sqrt{\frac{5}{3}}}$$

or at least to approximate them numerically. At least the latter option is relatively easy to accomplish if one begins with a number in polar coordinates.[10] In the case of *casus irreducibilis*, that is, in the case

$$\left(\frac{q}{2}\right)^2+\left(\frac{p}{3}\right)^3<0,$$

where the coefficient p must be negative, the two numbers u^3 and v^3 are complex conjugates of the form

$$-\frac{q}{2}\pm i\sqrt{-\left(\frac{q}{2}\right)^2-\left(\frac{p}{3}\right)^3}.$$

[10] On the other hand, the former problem, that of simplifying cube roots of complex numbers, that is, to express the real and imaginary parts separately with expressions involving roots, cannot be completely resolved: If a cubic equation with rational coefficients has three real solutions, none of which is rational, such as, for example, $x^3-6x+2=0$, then there is no expression for the roots involving nested radicals whose intermediate values are all real. See B. L. van der Waerden, *Algebra*, vol. I, Section 64, Springer, 2003.

The modulus of each of these numbers is

$$\sqrt{\left(\frac{q}{2}\right)^2 - \left(\frac{q}{2}\right)^2 - \left(\frac{p}{3}\right)^3} = \sqrt{\left(-\frac{p}{3}\right)^3} = \left(\sqrt{-\frac{p}{3}}\right)^3.$$

The angle made by the two numbers with the positive real axis can be obtained from the quotient of the real part of the number and its modulus. To be precise, the "upper" angle is give by

$$\phi = \arccos\left(\frac{-\frac{q}{2}}{\left(\sqrt{-\frac{p}{3}}\right)^3}\right) = \arccos\left(\frac{3q}{2p\sqrt{-\frac{p}{3}}}\right),$$

so that we obtain the following formula for the three solutions of the reduced cubic equation $x^3 + px + q = 0$:

$$x_{j+1} = 2\sqrt{-\frac{p}{3}}\cos\left(\frac{1}{3}\phi + j\frac{2\pi}{3}\right), \quad j = 0, 1, 2.$$

Such a solution based on trigonometric functions actually has nothing to do with algebra. However, the question of the solutions of a cubic equation is certainly an algebraic problem. Furthermore, the solution method of *casus irreducibilis* is excellently suited for learning how to work with complex numbers.

The formulas we have just described were first discovered in 1591 by François Viète (1540–1603). They were published posthumously in 1615. However, Viète did not use complex numbers in his derivation, using instead the triple-angle formula for the cosine:

$$\cos 3\psi = 4\cos^3\psi - 3\cos\psi.$$

Then, using an equation of the form

$$y^3 - \frac{3}{4}y - \frac{1}{4}\cos 3\psi = 0,$$

a solution can be found using the relation $y = \cos\psi$. To solve a cubic equation of the reduced form $x^3 + px + q = 0$, one first makes the transformation $x = sy$, where the parameter s is chosen such that the resulting equation

$$y^3 + \frac{p}{s^2}y + \frac{q}{s^3} = 0$$

has the desired form, which is the case for $s = 2\sqrt{-\frac{p}{3}}$:

$$y^3 - \frac{3}{4}y - \frac{3q}{8p}\frac{1}{\sqrt{-p/3}} = 0.$$

One then obtains a solution of the reduced cubic $x^3 + px + q = 0$ by starting with an angle ψ such that

$$\cos 3\psi = -\frac{3q}{2p}\frac{1}{\sqrt{-p/3}}$$

and making the substitution

$$x = 2\sqrt{-\frac{p}{3}}\cos\psi,$$

where the procedure works only if the conditions

$$p < 0 \quad \text{and} \quad \left|\frac{3q}{2p}\frac{1}{\sqrt{-p/3}}\right| \leq 1$$

are satisfied. The second inequality is equivalent to

$$\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3 \leq 0.$$

Literature on the History of Complex Numbers

Paul J. Nahin, *An Imaginary Tale: The Story of* $\sqrt{-1}$, Princeton, 1998.

Lutz Führer, *Kubische Gleichungen und die widerwillige Entdeckung der komplexen Zahlen*, Praxis der Mathematik, 43 (2001), pp. 57–67.

Exercises

(1) Derive the formulas for the real and imaginary parts of the square root of a complex number $a + bi$. Also attempt to find analogous formulas for the real and imaginary parts of the cube root of a complex number. Explain the problem that arises.

(2) Show that the complex conjugate of a solution of a polynomial equation with real coefficients is also a solution.

(3) Which of the three following complex numbers are roots of unity?

$$\frac{3}{11}\sqrt{7} + i\frac{2}{5}\sqrt{3}, \quad \frac{1}{2}\sqrt{2-\sqrt{3}} - i\frac{1}{2}\sqrt{2+\sqrt{3}}, \quad \frac{5}{7}\sqrt{6} + i\frac{2}{7}\sqrt{6}.$$

Chapter 3

Biquadratic Equations

We seek a solution of the equation $x^4 + 6x^2 + 36 = 60x$.

3.1 The problem for this chapter is also classical, coming as well from Cardano's *Ars Magna* (Chapter XXXIX, problem V). However, such problems caused difficulties for Cardano, because they offered no geometric interpretation. Thus he mentions in the forward to his book, "While *positio* is associated with a line, *quadratum* with a surface, and *cubum* with a solid, it would be foolish to attempt to extrapolate. Nature does not permit it."

However, thanks to his student Ludovico Ferrari (1522–1569), Cardano was able to describe in the *Ars Magna* the solution to biquadratic (quartic, fourth-degree) equations. In particular, Ferrari was able to transform equations of the form

$$x^4 + px^2 + qx + r = 0$$

by the addition of two terms in powers of x and x^2 in such a way that a perfect square is obtained on both sides of the equation. Deviating only slightly from the method described by Cardano, one most simply adds $2zx^2 + z^2$ to both sides of the equation, where the value for z is to be chosen later, and thereby obtains

$$x^4 + 2zx^2 + z^2 = (2z - p)x^2 - qx + \left(z^2 - r\right).$$

Although the left-hand side of the equation is already in the form of a perfect square, $\left(x^2 + z\right)^2$, such is not necessarily the case for the

right-hand side. However, z can now be suitably chosen, namely, so as to satisfy the condition

$$2\sqrt{2z-p}\sqrt{z^2-r} = -q.$$

Squaring both sides of this condition leads to

$$(2z-p)\left(z^2-r\right) = \frac{q^2}{4}$$

and thus to the cubic equation

$$z^3 - \frac{p}{2}z^2 - rz + \frac{pr}{2} - \frac{q^2}{8} = 0.$$

Once one has determined a solution z to this so-called *cubic resolvent*, the solutions to the original biquadratic equation results from

$$x^2 + z = \pm\left(\sqrt{2z-p}\,x + \sqrt{z^2-r}\right),$$

where each of the two possible signs yields two solutions by virtue of the quadratic formula. Altogether, one therefore obtains the following four solutions:

$$x_{1,2} = \frac{1}{2}\sqrt{2z-p} \pm \sqrt{-\frac{1}{2}z - \frac{1}{4}p + \sqrt{z^2-r}},$$
$$x_{3,4} = \frac{1}{2}\sqrt{2z-p} \pm \sqrt{-\frac{1}{2}z - \frac{1}{4}p - \sqrt{z^2-r}}.$$

It remains to note that in his *Ars Magna*, Cardano illustrates Ferrari's procedure with examples in which the numbers calculated are sometimes incorrect. For the problem posed at the start of this chapter, one obtains the cubic resolvent

$$z^3 - 3z^2 - 36z - 342 = 0,$$

which can be transformed via the substitution $z = y + 1$ into the reduced cubic

$$y^3 - 39y - 380 = 0.$$

Using the resolvent solution

$$z = 1 + \sqrt[3]{190 + 3\sqrt{3767}} + \sqrt[3]{190 - 3\sqrt{3767}},$$

one obtains towers of expressions in radicals that solve the original biquadratic equation.

By now it should be clear that as indicated in the introduction, these algebraic formulas are completely useless if the goal is simply to obtain numerical values, for these can be more quickly and easily computed by means of iteration procedures. Thus for the equation at the beginning of this chapter one obtains the solutions 3.09987... and 0.64440... as well as the pair of complex conjugates $-1.87214... \pm i \cdot 3.81014...$.

Nevertheless, from a mathematical point of view, the algebraic result is impressive. Who would have guessed a priori that cube roots would appear in the solution of a fourth-degree equation? However, if viewed correctly, this result is not as surprising as it might appear at first glance. In fact, we encountered something comparable in the case of cubic equations: Just as Cardano's formula contains square roots in addition to cube roots, a general formula for biquadratic equations must be similarly constituted. Otherwise, the general formula would not be applicable to a special equation like $x^4 - 2x = 0$, with the solution $x_1 = \sqrt[3]{2}$.

3.2 Since Ferrari's formula as presented here can be used only for biquadratic equations in which the variable x does not appear to the third power, we must describe a method of converting the general biquadratic equation of the form

$$x^4 + ax^3 + bx^2 + cx + d = 0$$

into an equation in the reduced form

$$y^4 + py^2 + qy + r = 0.$$

In analogy to the case of the cubic equation, this can be done by replacing the variable x via the substitution

$$x = y - \frac{a}{4},$$

with the result that the two terms in y^3 that arise cancel each other:

$$x^4 + ax^3 + bx^2 + cx + d = y^4 + py^2 + qy + r.$$

Of course, as in the case of the cubic, the coefficients of the reduced equation can be calculated from those of the original equation using polynomial expressions.

Literature on Biquadratic Equations

Ludwig Matthiessen, *Grundzüge der antiken und modernen Algebra der litteralen Gleichungen*, Leipzig, 1896.

Heinrich Dörrie, *Kubische und biquadratische Gleichungen*, Munich, 1948.

Exercises

(1) Determine all four solutions of the equation

$$x^4 - 8x + 6 = 0$$

presented at the beginning of the chapter.

(2) Determine all four solutions of the equation

$$x^4 + 8x^3 + 24x^2 - 112x + 52 = 0.$$

Chapter 4

Equations of Degree n and Their Properties

We seek an equation whose solutions are the numbers 1, 2, 3, 4, *and* 5.

4.1 The success in finding solution procedures for cubic and biquadratic equations led inevitably to the desire to do the same for equations of higher degree. This search included a wish to obtain a better understanding of polynomial equations through systematic study. In this connection, the problem posed at the start of this chapter was stated and solved. It can be found in François Viète's 1591 work *In artem analyticem isagoge.*

In addition to creating a useful symbolic notation, Viète discussed extensively the sorts of transformations of equations that are permissible without changing the solutions. He also found a way of constructing equations that possess given numbers $x_1, x_2, \ldots, x_n$ as solutions. In the case of two given solutions x_1, x_2, one needs only a quadratic equation, namely

$$x^2 - (x_1 + x_2)x + x_1x_2 = 0.$$

In the case of three prescribed solutions x_1, x_2, x_3, the cubic equation

$$x^3 - (x_1 + x_2 + x_3)x^2 + (x_1x_2 + x_1x_3 + x_2x_3)x - x_1x_1x_3 = 0$$

meets the required conditions. Analogously, the four numbers x_1, x_2, x_3, x_4 are solutions of the biquadratic equation

$$\begin{aligned} x^4 &- (x_1 + x_2 + x_3 + x_4)x^3 \\ &+ (x_1x_2 + x_1x_3 + x_2x_3 + x_1x_4 + x_2x_4 + x_3x_4)x^2 \\ &- (x_1x_2x_3 + x_1x_2x_4 + x_1x_3x_4 + x_2x_3x_4)x + x_1x_2x_3x_4 = 0. \end{aligned}$$

Finally, Viète produced an equation whose solutions are the five given numbers x_1, x_2, x_3, x_4, x_5:

$$\begin{aligned} x^5 &- (x_1 + x_2 + x_3 + x_4 + x_5)x^4 \\ &+ (x_1x_2 + x_1x_3 + x_2x_3 + x_1x_4 + x_2x_4 + x_3x_4 + x_2x_5 + x_3x_5 \\ &\quad + x_4x_5)x^3 \\ &- (x_1x_2x_3 + x_1x_2x_4 + x_1x_3x_4 + x_2x_3x_4 + x_1x_2x_5 + x_1x_3x_5 \\ &\quad + x_2x_3x_5 + x_1x_4x_5 + x_2x_4x_5 + x_3x_4x_5)x^2 \\ &+ (x_1x_2x_3x_4 + x_1x_2x_3x_5 + x_1x_2x_4x_5 + x_1x_3x_4x_5 + x_2x_3x_4x_5)x \\ &- x_1x_2x_3x_4x_5 = 0. \end{aligned}$$

Viète's last example solves the problem posed at the beginning of the chapter, for which one obtains the equation[1]

$$x^5 - 15x^4 + 85x^3 - 225x^2 + 274x - 120 = 0.$$

Only the obvious symmetry prevents Viète's formula from being completely confusing. Of course, one may easily check that *Viète's root theorem* (*root* is a frequent synonym for *solution*) is correct by substituting one of the values in for the variable. However, of greater interest is the question of how one obtains such results, including the analogous ones for more than five solutions. This, too, is not difficult, and was described for the first time in 1637 by René Descartes (1596–1650) in his work *La Géometrie*: If an equation is sought whose solutions are the given numbers $x_1, x_2, \ldots, x_n$, then one may simply

---

[1] In Viète's notation, the equation is

$$1QC - 15QQ + 85C - 225Q + 274N, \text{ equatur } 120.$$

A facsimile together with German translation can be found in Henk J. M. Bos, Karin Reich, Der doppelte Auftakt zur frühneuzeitlichen Algebra: Viète und Descartes, in Erhard Scholz (ed.), *Geschichte der Algebra*, Mannheim, 1990, pp. 183–234.

take the equation

$$(x - x_1)(x - x_2) \cdots (x - x_n).$$

In this form, it is obvious that $x_1, x_2, \ldots, x_n$ are solutions and that there are no others. One is required simply to multiply out to obtain an equation in the more familiar form.

In particular, Viète's root theorem explains an observation made already by Cardano in his *Ars Magna* (Chapter I, equation $x^3 + 72 = 11x^2$). He had found three solutions for some equations of the form $x^3 + bx = ax^2 + c$ and observed that the sum of the solutions agrees with the coefficient a of the quadratic term. An explanation of this fact would have been difficult for Cardano to have discovered, since it presupposed the existence of negative numbers to bring the equation into a form in which the right-hand side is equal to zero.

4.2 Descartes also discusses the problem of whether and under what circumstances the left-hand side of an equation of the form

$$x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0$$

can be decomposed into a product of the form $(x - x_1) \cdots (x - x_n)$. If there is such a decomposition into *linear factors*, then clearly the solutions are known. But conversely, says Descartes, each solution provides one step in the factorization of the equation into linear factors. For example, if x_1 is a solution, then the variable x on the left side of the equation can be replaced by $x_1 + (x - x_1)$. If one then develops the powers $(x_1 + (x - x_1))^k$ into powers of x_1 and $(x - x_1)$, then one finds that the term $(x - x_1)$ can be factored out:

$$\begin{aligned} & x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \\ &= (x - x_1)^n + b_{n-1}(x - x_1)^{n-1} + \cdots + b_1(x - x_1) + b_0, \end{aligned}$$

with

$$b_0 = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0.$$

One thus has the desired result:

$$\begin{aligned} & x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \\ &= (x - x_1)\left((x - x_1)^{n-1} + b_{n-1}(x - x_1)^{n-2} + \cdots + b_1\right) \\ &= (x - x_1)\left(x^{n-1} + c_{n-2}x^{n-2} + \cdots + c_0\right). \end{aligned}$$

Finally, we have succeeded in dividing the original polynomial by the factor $(x - x_1)$, where all the coefficients $c_0, c_1, \ldots, c_{n-2}$ of the resulting polynomial can be determined from those of the original equation and the solution x_1 via multiplication and addition.

If one finds additional solutions, the process of breaking off a linear factor can be continued. In an equation of the nth degree, there are at most n linear polynomials that can be thus broken off. Therefore, as Descartes asserted, an nth-degree equation can have at most n solutions.

4.3 If the number of solutions of a polynomial equation of degree n can be at most n, then what is the least number of possible solutions? Here we do not mean the number of *different* solutions, since the equation $x^n = 0$, for example, has only the single solution zero. By "number of solutions" we mean rather the number of linear factors. Thus we are asking for the number of linear factors that can be broken off an nth-degree equation, and what the smallest possible such number is. If a linear factor appears more than once, then the corresponding solution is called a *multiple solution*, and we say that we are counting solutions with *multiplicity*.[2]

The possibility of breaking off factors corresponding to solutions allows us to say something about the minimum number of solutions: If there is no nth-degree equation without any solutions, then a linear factor can be broken off of any polynomial. Furthermore, since the resulting equation, provided that its degree is at least 1, must again have a solution, the process can be continued, and indeed, it can be continued until the polynomial has been entirely decomposed into linear factors. That is, if it can be proved that every nth-degree equation has at least one solution, then there will always be n solutions, where solutions are counted with multiplicity.

Already before Descartes, Albert Girard (1590–1632) had conjectured in 1629 that an equation with complex coefficients always possesses the number of solutions equal to its degree. Despite the efforts of many mathematicians, a complete proof of this conjecture, now called the *fundamental theorem of algebra*, was achieved only in

[2]Cardano dealt with multiple solutions, for example the equation $x^3 + 16 = 2x$ in Chapter I of the *Ars Magna*.

1799, by Carl Friedrich Gauss. This proof justified—at least from an algebraic point of view—the use of complex numbers, since a further enlargement of the set of available numbers is not necessary.

The designation "fundamental theorem of algebra" is to be understood historically. From today's point of view it is somewhat misleading, since in fact, this theorem is fundamentally not algebraic in nature. That is, it is based only tangentially on the properties of the complex numbers related to the four basic arithmetic operations. Of much greater importance are the properties of the complex numbers that are related to distance, that is, properties related to convergence, continuity, and so on. A comparison with a similar result in the realm of the real numbers might make this clearer: The graph of the polynomial $x^3 - 2$ viewed as a function of a real variable runs, in the standard coordinate system, from "down and to the left" to "up and to the right." Therefore, it "must" cross the x axis at least once. That is, the polynomial under investigation, and indeed any polynomial of odd degree, has at least one zero (that is, a value of x that makes the polynomial equal to zero). What appears to be simple and obvious is in fact due to fundamental properties of the real numbers, which find expression in the so-called *intermediate value theorem*. Two properties are crucial:

- A function defined by a polynomial is *continuous*, that is, there are no holes or jumps in its graph; rather, its value changes at every point by less than any prescribed bound, provided the change in x is sufficiently small.
- The set of real numbers has no "holes," such as exist, for example, in the rational numbers. Indeed, for each point on the number line one can find infinitely many rational numbers within any given small distance of the given point. Nonetheless, the process of approximating the point by rational numbers takes one outside the set of rational numbers. For example, if we approximate the square root of 2 by more and more terms of its decimal expansion,

$$1, \quad 1.4, \quad 1.41, \quad 1.414, \quad 1.4142, \quad \ldots,$$

this sequence of rational numbers approaches a number that is not rational, namely $\sqrt{2}$. The point here is that there is no comparable example in the set of real numbers.

A proof of the fundamental theorem of algebra can be given based on the continuity of polynomial functions and the two key properties of the complex numbers, namely, the lack of "holes" (called *completeness*) and the existence of the number i, which satisfies the equation $i^2 = -1$. The following section on the fundamental theorem of algebra contains an argument for the plausibility of the theorem, as well as a sketch of a proof.

The Fundamental Theorem of Algebra: Plausibility and Proof

As stated in the text, it suffices to prove the following theorem: A polynomial with complex coefficients whose degree is at least one has at least one complex zero.

We begin with a plausibility argument, which makes crucial use of the properties of the absolute value function for complex numbers, $|a+bi| = \sqrt{a^2+b^2}$, namely, that for two arbitrary complex numbers z_1 and z_2, the *triangle inequality* holds, that is, $|z_1+z_2| \le |z_1|+|z_2|$, as well as the identity $|z_1 z_2| = |z_1||z_2|$. This has a consequence that for a given polynomial

$$f(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0,$$

with complex coefficients $a_{n-1}, \ldots, a_1, a_0$, the absolute values of the values taken on by the function are approximated by $|z^n|$ for sufficiently large (in absolute value) z. Concretely, for a complex number z with

$$|z| \ge R := 1 + 2\big(|a_{n-1}| + \cdots + |a_1| + |a_0|\big),$$

one has the inequality

$$\begin{aligned} &\left|a_{n-1}z^{n-1} + \cdots + a_1 z + a_0\right| \\ &\le |a_{n-1}|\,\left|z^{n-1}\right| + \cdots + |a_1||z| + |a_0| \\ &\le \big(|a_{n-1}| + \cdots + |a_1| + |a_0|\big)|z|^{n-1} \le \frac{1}{2}|z|^n. \end{aligned}$$

We would now like to consider how the motion of a complex number z about a circle of radius R with center the origin is affected by the polynomial $f(z)$. For the term z^n it is clear, since by de Moivre's formula, one circuit around the circle of radius R is mapped to n circuits around the circle of radius R^n. This circle is depicted in the middle part of Figure

4.1. The remaining terms of the polynomial have little effect on this result for large R, as we have seen, so that we obtain for a function value $f(z)$ an n-fold circuit within an annulus with 0 as center, inner radius $\frac{1}{2}R^n$, and outer radius $\frac{3}{2}R^n$. For example, the middle of Figure 4.1 shows two bounding circles for the remaining terms; the bounding annulus is shown as a dashed line in both the center and right parts of the figure.

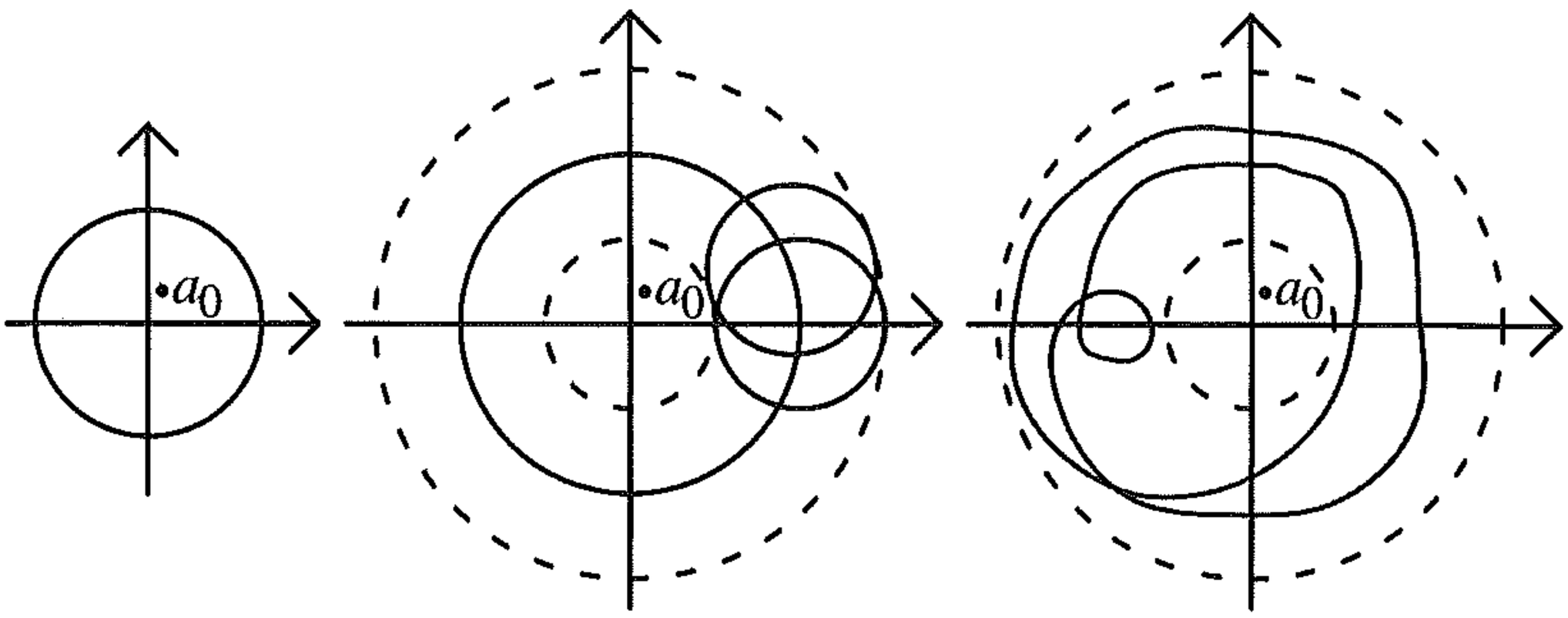


Figure 4.1. A revolution of a circle of sufficiently large radius about the origin (left) is mapped by an nth-degree polynomial to an n-fold revolution within an annulus centered at the origin (right). The middle image shows the curve of the highest power, z^n, and the maximal "perturbation" at two sample points due to the terms of lower degree.

What happens if we vary the radius of the initial circle? Ignoring the details, what is important for us is that "everything is continuous." Since we are only arguing for plausibility, we will now offer a verbal description of the consequences: Independent of the radius of the initial circle that the point z traverses, the image points $f(z)$ always form a closed curve without any discontinuities (holes). Furthermore, any changes experienced by the curve due to a change in the radius can be kept arbitrarily small by keeping the change in radius sufficiently small.

There is one more obvious fact: For zero radius there is but a single image point, namely a_0, which by our construction lies within the inner circle of radius $\frac{1}{2}R^n$.

Now comes, in analogy to the case of a real polynomial of odd degree, the decisive continuity argument: If one gradually contracts the circle on which z is moving toward the origin beginning at radius R, then the image curve, beginning with its n-fold circuit of the origin, contracts to the point a_0. Thus at some point the origin must be crossed, and therefore the polynomial $f(z)$ must contain within the circle of radius R at least one—and indeed, for R sufficiently large, n—complex zero.

Since it is not at all simple to make the previous heuristic argument watertight, we will take a different tack altogether in presenting a formal proof, one taken first in 1815 by Jean Robert Argand (1768–1822) and simplified a few years later by Augustin-Louis Cauchy (1789–1857).

We have already seen that the function values $f(z)$ of the polynomial exceed in absolute value $|f(0)| = |a_0|$ outside a sufficiently large circle. The minimum of the real-valued function $|f(z)|$ is therefore to be found inside the circle, where it is taken, according to a theorem about the extreme values of continuous real-valued functions, at some point, which we call z_0. Developing the polynomial as a function of $z - z_0$ leads to

$$f(z) = b_0 + b_m(z - z_0)^m + b_{m+1}(z - z_0)^{m+1} + \cdots + b_n(z - z_0)^n,$$

where the index $m \geq 1$ is such that $b_m \neq 0$. Furthermore, we may assume $b_0 \neq 0$, since otherwise, we have already found a zero.

We now determine a complex number w, using de Moivre's formula, for example, with the property

$$w^m = -\frac{b_0}{b_m},$$

and then form the argument $z_1 = z_0 + \epsilon w$, where ϵ is a small number, $0 < \epsilon < 1$, to be chosen later. For the associated function value $f(z_1)$, we now obtain

$$\begin{aligned} f(z_1) &= b_0 - b_m\epsilon^m \frac{b_0}{b_m} + b_{m+1}\epsilon^{m+1}w^{m+1} + \cdots + b_n\epsilon^n w^n \\ &= (1 - \epsilon^m)\, b_0 + b_{m+1}\epsilon^{m+1}w^{m+1} + \cdots + b_n\epsilon^n w^n. \end{aligned}$$

This equation allows us to estimate the value $|f(z_1)|$:

$$\begin{aligned} |f(z_1)| &\leq (1 - \epsilon^m)\,|b_0| + \epsilon^{m+1}\left(\left|b_{m+1}w^{m+1}\right| + \cdots + |b_n w^n|\right) \\ &= |b_0|\big(1 - \epsilon^m(1 - \epsilon B)\big), \end{aligned}$$

where $B = \left(\left|b_{m+1}w^{m+1}\right| + \cdots + |b_n w^n|\right)/|b_0|$ depends only on the coefficients $b_0, b_m, \ldots, b_n$ and the choice of the number w. One may now choose ϵ sufficiently small that $1 - \epsilon B$ is positive. Then with such a selection, we have found a smaller value than the assumed minimum: $|f(z_1)| < |b_0| = |f(z_0)|$. The contradiction is eliminated if we give up the assumption $b_0 \neq 0$.

Finally, we observe that the shortest and most beautiful proofs of the fundamental theorem of algebra are based on basic theorems of the theory of functions of a complex variable.

Exercises

(1) Show that the nonreal zeros of a polynomial with real coefficients come in pairs of complex conjugate numbers.

(2) Construct an nth-degree polynomial that takes on the given values $y_1, \ldots, y_n$ at the distinct points $x_1, \ldots, x_n$. Hint: Consider for $j = 1, \ldots, n$ polynomials of the form

$$g_j(x) = \prod_{\substack{i=1,\ldots,n \\ i \neq j}} \frac{x - x_i}{x_j - x_i}$$

at the points $x = x_1, \ldots, x_n$. The polynomial that solves this problem is called the *Lagrange interpolation formula.*

Chapter 5

The Search for Additional Solution Formulas

Is there a common "blueprint" for the solution formulas to equations up to the fourth degree?

5.1 The procedures that Cardano published for solving cubic and biquadratic equations marked the beginning of a historical period in which a variety of attempts were made to find a general formula for solving equations of the fifth degree. In pursuit of this goal, it seemed a good idea to search for similarities in the solution procedures already discovered. In the case of equations of fourth degree, various alternatives to Ferrari's solution method were considered, which with other equivalence transformations and other intermediate results led to the same results.[1]

To be sure, the fundamental theorem of algebra guarantees the existence of n complex roots for an nth-degree equation. However, it offers no clue as to how those solutions can be calculated. Nevertheless, based on the fundamental theorem of algebra, we can reformulate the problem of finding the solutions of an nth-degree equation: Since

[1]The most complete description of such methods is to be found in Ludwig Matthiessen, *Grundzüge der antiken und modernen Algebra der litteralen Gleichungen*, Leipzig, 1896.

in every equation of the form

$$x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 = 0$$

the left side can be decomposed into linear factors

$$x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 = (x - x_1)(x - x_2)\cdots(x - x_n),$$

can the given nth-degree equation be transformed into an equivalent system of equations corresponding to Viète's root theorem? That is, for given complex coefficients $a_{n-1}, \ldots, a_1, a_0$, we seek complex numbers $x_1, x_2, \ldots, x_n$ that satisfy the system of equations

$$\begin{aligned} x_1 + x_2 + \cdots + x_n &= -a_{n-1}, \\ x_1x_2 + x_1x_3 + \cdots + x_{n-1}x_n &= a_{n-2}, \\ &\cdots \\ x_1x_2\cdots x_{n-1}x_n &= (-1)^n a_0. \end{aligned}$$

The symmetric expressions in $x_1, x_2, \ldots, x_n$ appearing on the left sides of the equations are called *elementary symmetric polynomials.* But it is not only the solution of given equations with explicitly known coefficients that can be reinterpreted on the basis of Viète's root theorem. We may certainly view the solutions $x_1, x_2, \ldots, x_n$ as variables, so that the search for a general solution formula corresponds to the problem of determining the variables $x_1, x_2, \ldots, x_n$ from the elementary polynomials $a_{n-1}, \ldots, a_1, a_0$. This interpretation is usually called the *general equation.*

For the case of a quadratic equation, the well-known quadratic formula is given the following interpretation:

$$x_{1,2} = \frac{1}{2}(x_1+x_2) \pm \frac{1}{2}\sqrt{(x_1+x_2)^2 - 4x_1x_2} = \frac{1}{2}(x_1+x_2) \pm \frac{1}{2}(x_1-x_2).$$

It is worth noting that in place of the square root, which is surely the key intermediate value in solving the equation, we have a simple expression in terms of the solutions, namely $(x_1 - x_2)$.

5.2 Similar expressions in terms of the solutions of the general cubic and biquadratic equations can also be found. To be sure, the requisite computations are correspondingly more complicated. We begin with the cubic equation, for whose reduced form $x^3 + px + q = 0$ the three

solutions can be computed with the help of Cardano's formula using the values

$$u = \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}, \quad v = \sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}},$$

namely,

$$x_1 = u + v,$$
$$x_2 = \zeta u + \zeta^2 v,$$
$$x_3 = \zeta^2 u + \zeta v.$$

From these three equations, using the identity $\zeta^2+\zeta+1 = 0$, which is clear from the third-degree cyclotomic equation written $z^3 - 1 = \left(z^2 + z + 1\right)(z - 1)$, one obtains expressions for u and v in terms of the solutions:

$$u = \frac{1}{3}\left(x_1 + \zeta^2 x_2 + \zeta x_3\right),$$
$$v = \frac{1}{3}\left(x_1 + \zeta x_2 + \zeta^2 x_3\right).$$

Again there is a concise expression for the square root appearing in Cardano's formula in terms of the three solutions x_1, x_2, x_3:

$$\begin{aligned}
\sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3} &= \frac{1}{2}\left(u^3 - v^3\right) \\
&= \frac{1}{54}\left(x_1 + \zeta^2 x_2 + \zeta x_3\right)^3 - \frac{1}{54}\left(x_1 + \zeta x_2 + \zeta^2 x_3\right)^3 \\
&= \frac{1}{18}\left(\zeta^2 - \zeta\right) \\
&\quad \times \left(x_1^2 x_2 - x_1 x_2^2 + x_2^2 x_3 - x_2 x_3^2 + x_1 x_3^2 - x_1^2 x_3\right) \\
&= -\frac{1}{18} i\sqrt{3}(x_1 - x_2)(x_2 - x_3)(x_1 - x_3).
\end{aligned}$$

As one can see from the presence of i in the last expression, for the particular case of three real roots the expression under the radical sign is always negative. More significant is the observation that the expression is equal to zero precisely when there is a multiple root. One can also write the analogous product for the general equation of every other degree in which all the differences of pairs of roots appear. Such a product of differences, whose square is called the

discriminant, irrespective of the degree of the equation, is then equal to zero precisely when there is a multiple root.

If we are given a cubic equation with quadratic term

$$x^3 + ax^2 + bx + c = 0,$$

then the solution process begins, as presented in Chapter 1, with the substitution

$$x = y - \frac{a}{3},$$

so that a reduced cubic equation is obtained. The intermediate values u, v, and $\sqrt{(q/2)^2 + (p/3)^3}$ that appear in Cardano's formula can be determined from the solutions of the original equation. All that is necessary is to replace each solution x_j, $j = 1, 2, 3$, in the three formulas just derived with

$$x_j + \frac{1}{3}a = x_j - \frac{1}{3}(x_1 + x_2 + x_3),$$

thereby leaving the three formulas unchanged. The formulas for the three intermediate values u, v, and $\sqrt{(q/2)^2 + (p/3)^3}$ thus hold unchanged for the general cubic equation.

5.3 In Ferrari's procedure for solving the reduced biquadratic equation $x^4 + px^2 + qx + r = 0$, the crucial step is determining a solution z of the cubic resolvent

$$z^3 - \frac{p}{2}z^2 - rz + \frac{pr}{2} + \frac{q^2}{8} = 0,$$

on the basis of which the four solutions can be determined pairwise from two quadratic equations:

$$x^2 \mp \sqrt{2z - p}\, x \mp \sqrt{z^2 - r} + z = 0.$$

Using Viète's root theorem for quadratic equations, we can derive from these two equations the following values for the products of the two pairs of solutions:

$$x_1 x_2 = z + \sqrt{z^2 - r},$$

$$x_3 x_4 = z - \sqrt{z^2 - r}.$$

From this one immediately obtains

$$z = \frac{1}{2}(x_1 x_2 + x_3 x_4).$$

For the sake of completeness we should note that the solution $z - z_1$ of the cubic resolvent corresponds to a possible, but by no means prescribed, numbering of the solutions x_1, x_2, x_3, x_4. Since Ferrari's procedure is a consequence of equivalence transformations, resting on the condition prescribed by the cubic resolvent with respect to the value z, the selection of a different resolvent solution also leads to the correct solutions and therefore can affect only the numbering of the solutions. This has as a consequence that the two other solutions of the resolvent can be determined from the solutions x_1, x_2, x_3, x_4 as follows:

$$z_2 = \frac{1}{2}(x_1x_3 + x_2x_4),$$
$$z_3 = \frac{1}{2}(x_1x_4 + x_2x_3).$$

With this, we can express the square root that appears in Cardano's formula in the solution of the cubic resolvent in terms of the solutions x_1, x_2, x_3, x_4 of the original equation. Up to a constant factor, the square root is equal to $(z_1 - z_2)(z_2 - z_3)(z_1 - z_3)$, where a single factor of this product of differences looks like

$$(z_1 - z_2) = \frac{1}{2}(x_1x_2 + x_3x_4 - x_1x_3 - x_2x_4) = \frac{1}{2}(x_1 - x_4)(x_2 - x_3),$$

with the result that for the complete product of differences we obtain

$$(z_1 - z_2)(z_2 - z_3)(z_1 - z_3)$$
$$= \frac{1}{8}(x_1 - x_4)(x_2 - x_3)(x_1 - x_2)(x_3 - x_4)(x_1 - x_3)(x_2 - x_4).$$

Therefore, up to a constant factor, the discriminant of the original equation coincides with that of the cubic resolvent.

For biquadratic equations that are not in reduced form, one may proceed as in the case of a cubic equation: First, a biquadratic equation

$$x^4 + ax^3 + bx^2 + cx + d = 0$$

with a cubic term is transformed via the substitution

$$x = y - \frac{a}{4}$$

into a reduced biquadratic equation. In order to obtain formulas for the intermediate values that arise in the process in terms of the

solutions of the original equation, we should replace each solution x_j, $j = 1, 2, 3, 4$, in the formulas just derived by

$$x_j + \frac{1}{4}a = x_j - \frac{1}{4}(x_1 + x_2 + x_3 + x_4).$$

The polynomials thus obtained take care of the general biquadratic equation. In particular, we obtain the "first" solution of the cubic resolvent as

$$z_1 = \frac{1}{2}(x_1x_2 + x_3x_4) - \frac{1}{16}(x_1 + x_2 + x_3 + x_4)^2.$$

5.4 What, then, are the similarities among the three methods for solving quadratic, cubic, and biquadratic equations? In all three cases, the crucial intermediate values, that is, the expressions for the roots appearing in the solution formulas, are representable as "simple," that is, polynomial, expressions in the solutions $x_1, x_2, \ldots$. Of course, the actual form of such expressions depends on how the solutions are numbered.

Is the ability to express intermediate values as polynomials in the solutions $x_1, x_2, \ldots$ really as surprising at it might seem at first glance? Since a solution method always starts with the coefficients of the equation, that is, in reference to the general equation with the elementary symmetric polynomials in the solutions, it is actually obvious that all the intermediate values can be expressed in terms of the solutions using the usual arithmetic operations *and* nested roots. However, what is not a priori obvious is that polynomials alone suffice, which was the case for degrees two, three, and four. That is, in the representation of the intermediate values, *no* expression of the form, say,

$$\sqrt{x_1 + x_2^3x_4}$$

may appear.

Permutations

Changing the order of a finite number of objects is called a *permutation*. Since it doesn't matter what the objects are called, one usually names them with the natural numbers $1, 2, \ldots, n$.

It is easy to determine the number of permutations of n objects. It is $n!$, read "n factorial," defined by $n! := 1 \cdot 2 \cdot 3 \cdots n$. To see that this is so, observe that the number 1 can be placed in any one of the n positions, which leaves $n-1$ positions for the number 2, and so on. Altogether, then, the number of arrangements is equal to $n \cdot (n-1) \cdots 3 \cdot 2 \cdot 1$.

One simple way of notating a permutation σ is to list the images of the elements, that is, the locations $\sigma(1), \sigma(2), \ldots, \sigma(n)$ after the exchange. Symbolically, one has

$$\begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ \sigma(1) & \sigma(2) & \sigma(3) & \cdots & \sigma(n) \end{pmatrix},$$

or, more simply,

$$\begin{pmatrix} \sigma(1) & \sigma(2) & \sigma(3) & \cdots & \sigma(n) \end{pmatrix}.$$

In special cases it can be useful to use a more suggestive notation. We will use such a notation for the so-called *cyclic* permutations, in which all the numbers from 1 to n are moved one to the other in some order, writing $1 \to 3 \to 4 \to 2 \to 1$ instead of $\begin{pmatrix} 3 & 1 & 4 & 2 \end{pmatrix}$.

An important property of permutations is that one can execute them one after another to obtain another permutation. As with other mappings and functions, such a process is called *composition* and is frequently denoted by the symbol $\circ$. For example,

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 4 & 2 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix},$$

where the order, as is usual for functions and mappings, is read from right to left.[2] What we have, then, is that the number 1 is sent to position 2, as seen in the second permutation from the left, and then is sent to position 3, as seen in the permutation on the far left.

The collection of all $n!$ permutations together with the operation of composition is called the *symmetric group* and is denoted by S_n. Its identity element is the *identity permutation*, which leaves every element in its original place.

The recognition that all intermediate values of the known solution formulas for the general polynomial equations up to fourth degree are given as polynomials in the solutions $x_1, x_2, \ldots$ is due to Joseph Louis Lagrange (1736–1813). Lagrange, who was active in Berlin for the twenty years beginning in 1766 thanks to Friedrich II, published in 1771 an investigation into general solution theorems for

---

[2] This perhaps unusual order can be explained by the fact that the argument of a function appears on the right, and so one writes $(\sigma \circ \tau)(j) = \sigma(\tau(j))$.

equations of the nth degree. Lagrange's starting point was the systematic investigation of the general solution formulas for equations up to the fourth degree. Since the intermediate values in the solution of equations up to degree four are expressible as polynomials in the solutions $x_1, x_2, \ldots$, it makes sense to search for methods that express arbitrary polynomials in the solutions $x_1, x_2, \ldots$. More precisely, one asks the question how a given polynomial $h(x_1, x_2, \ldots, x_n)$ in the solutions $x_1, x_2, \ldots$ can be determined from the coefficients of the general equation, that is, from the elementary symmetric polynomials. Concretely, how can one find a simple equation for which $h(x_1, x_2, \ldots, x_n)$ is a solution and whose coefficients can be expressed in terms of the elementary symmetric polynomials.

Lagrange recognized that such an equation can always be found via a construction of the form

$$\big(z - h(x_1, x_2, \ldots, x_n)\big)\big(z - h(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)})\big) \cdots = 0,$$

where the product is formed from a suitable selection from among the $n!$ *permuations*, that is, exchanges of the variables' indices $1, 2, \ldots$. Concretely, the permutations σ are chosen such that every possible polynomial that can arise from $h(x_1, \ldots, x_n)$ through permuting the variables $x_1, \ldots, x_n$ appears exactly once in the product. Then, as we shall see, one achieves that the coefficients of the equation arising for the unknown z can be calculated in terms of the coefficients of the general equations, that is, the elementary symmetric polynomials, using basic arithmetic operations. Therefore, for the polynomial $h(x_1, \ldots, x_n)$ we will have obtained the desired equation.

That this all sounds more complicated than it is in reality can be seen by means of an example. The polynomial

$$h(x_1, x_2, x_3, x_4) = \frac{1}{2}(x_1x_2 + x_3x_4) - \frac{1}{16}(x_1 + x_2 + x_3 + x_4)^2$$

appeared earlier when we were investigating the cubic resolvent in Ferrari's procedure for solving a biquadratic equation. Lagrange's

universal construction leads for this example to the equation

$$\left(z-\frac{1}{2}(x_1x_2+x_3x_4)+s\right)\left(z-\frac{1}{2}(x_1x_3+x_2x_4)+s\right)$$
$$\times\left(z-\frac{1}{2}(x_1x_4+x_2x_3)+s\right)=0,$$

where we have used the abbreviation

$$s(x_1,x_2,x_3,x_4)=\frac{1}{16}(x_1+x_2+x_3+x_4)^2.$$

If we now multiply the three linear factors together, we obtain again—but now in a generally applicable way—the cubic resolvent from Chapter 3.

Not only in this special case, but also in general, one finds with Lagrange's construction an equation for the unknown z in which the coefficients are polynomials in the variables $x_1, x_2, \ldots$. Since any permutation of the variables $x_1, x_2, \ldots$ simply rearranges the linear factors, the polynomials that form the coefficients of the equation constructed for the unknown z remain unchanged. Thus all coefficients are *symmetric polynomials* in the variables $x_1, x_2, \ldots$. And such symmetric polynomials, that is, the coefficients of the general equation of nth degree, are always able to be expressed in terms of the elementary symmetric polynomials using addition, subtraction, and multiplication. This can be summarized in the *fundamental theorem of symmetric polynomials*.

Theorem 5.1. *Every symmetric polynomial in $x_1, x_2, \ldots$ is a polynomial in the elementary symmetric polynomials.*

This theorem was first formulated by Lagrange. However, the theorem was apparently known a century earlier by the physicist and inventor of the calculus Isaac Newton (1643–1727), along with a procedure for representing symmetric polynomials, for example, by determining concretely

$$x_1^2+x_2^2+x_3^2+\cdots=(x_1+x_2+x_3+\cdots)^2-2(x_1x_2+x_1x_3+x_2x_3+\cdots)$$

and

$$\begin{aligned} &x_1^2x_2 + x_2^2x_1 + x_1^2x_3 + x_3^2x_1 + x_2^2x_3 + x_3^2x_2 + \cdots \\ &= (x_1 + x_2 + x_3 + \cdots)(x_1x_2 + x_1x_3 + x_2x_3 + \cdots) \\ &\quad - 3(x_1x_2x_3 + x_1x_2x_4 + x_1x_3x_4 + x_2x_3x_4 + \cdots). \end{aligned}$$

How this theorem is proved using a constructive algorithm is explained in the section on the fundamental theorem of symmetric polynomials. There is a specific application of Lagrange's theorem on symmetric polynomials to the discriminant

$$\prod_{i<j}(x_i - x_j)^2 = (x_1 - x_2)^2(x_1 - x_3)^2(x_2 - x_3)^2 \cdots,$$

for which there must be a polynomial expression in the coefficients due to its symmetry for each degree n of the underlying equation.

Another polynomial in the variables $x_1, x_2, \ldots$ that had central importance for Lagrange irrespective of the degree of the equation is what is today called the *Lagrange resolvent*[3]

$$h(x_1, x_2, \ldots, x_n) = x_1 + \zeta x_2 + \zeta^2 x_3 + \cdots + \zeta^{n-1}x_n,$$

where ζ is an nth root of unity. Since

$$\begin{aligned} h(x_1, x_2, \ldots, x_n) &= \zeta \cdot h(x_2, x_3, \ldots, x_1) = \cdots \\ &= \zeta^{n-1} \cdot h(x_n, x_1, \ldots, x_{n-1}) \end{aligned}$$

and thus

$$h(x_1, x_2, \ldots, x_n)^n = h(x_2, x_3, \ldots, x_1)^n = \cdots = h(x_n, x_1, \ldots, x_{n-1})^n,$$

one obtains with Lagrange's universal procedure for $h(x_1, x_2, \ldots, x_n)^n$ a resolvent equation of degree $(n-1)!$ whose coefficients can be expressed in terms of those of the original equation. If this equation were solvable using a general formula, the original equation could be solved, since we have

$$\begin{aligned} x_1 = \frac{1}{n}\big(&x_1 + \cdots + x_n + h(x_1, x_2, x_3, \ldots, x_n) + h(x_1, x_3, x_4, \ldots, x_2) \\ &+ \cdots + h(x_1, x_n, x_2 \ldots, x_{n-1})\big) \end{aligned}$$

[3]However, even before Lagrange, such expressions were used by Bézout (1730–1783) and Euler (1707–1783) in their work on solution formulas for the general nth-degree equation.

and analogous equations for the other solutions. Although for $n \geq 5$ no general solution for Lagrange's resolvent is apparent, his method is successful in some special cases. The first to achieve such success was Alexandre-Théophile Vandermonde (1735–1796), who in 1770, independently of Lagrange, investigated "his" resolvent. We will have more to say about this in Chapter 7.

The Fundamental Theorem on Symmetric Polynomials

We recall Theorem 5.1:

> Every symmetric polynomial in the variables $x_1, x_2, \ldots$ can be expressed as a polynomial in the elementary symmetric polynomials.

A proof of this theorem is most easily accomplished by complete induction. The order in which the induction will be taken will be based on a special ordering of the polynomials, one related to the lexicographic ordering of words in a language.

We define a *monomial* $x_1^{j_1} \cdots x_n^{j_n}$ to be "bigger" than the monomial $x_1^{k_1} \cdots x_n^{k_n}$ if in listing the exponents $j_1, j_2, \ldots$ in order, the first exponent j_s that differs from the corresponding exponent k_s is larger than k_s. For example, according to this definition, the monomial $x_1^2 x_2^5 x_3$ is bigger than the monomial $x_1^2 x_2^4 x_3^2 x_4$, based on the lexicographic order of the two strings "251" and "2421."

The induction step now begins with an arbitrary symmetric polynomial

$$f(x_1, \ldots, x_n) = \sum_{j_1, \ldots, j_n} a_{j_1 \ldots j_n} x_1^{j_1} \cdots x_n^{j_n}$$

whose biggest monomial with nonzero coefficient $a_{m_1 \ldots m_n}$ is the monomial $x_1^{m_1} \cdots x_n^{m_n}$. We assume for the induction hypothesis that the theorem has already been proved for all polynomials whose monomials with nonzero coefficients are all smaller than the monomial $x_1^{m_1} \cdots x_n^{m_n}$. Since $f(x_1, \ldots, x_n)$ is a symmetric polynomial, every monomial $x_1^{m_{\sigma(1)}} \cdots x_n^{m_{\sigma(n)}}$, where σ is any permutation of the numbers $1, 2, \ldots, n$, has the same coefficient $a_{m_1 \ldots m_n}$. It follows that $m_1 \geq m_2 \geq \cdots \geq m_n$, for otherwise, one could use a suitable permutation to find a monomial bigger than $x_1^{m_1} \cdots x_n^{m_n}$ whose coefficient in the polynomial $f(x_1, \ldots, x_n)$ is nonzero.

Now a very special polynomial in the elementary symmetric polynomials is formed:

$$g(x) = a_{m_1 \dots m_n} \Big(\sum_j x_j \Big)^{m_1 - m_2} \Big(\sum_{j<k} x_j x_k \Big)^{m_2 - m_3} \cdots (x_1 x_2 \cdots x_n)^{m_n}.$$

The term with the largest nonzero coefficient is

$$x_1^{m_1 - m_2} (x_1 x_2)^{m_2 - m_3} \cdots (x_1 x_2 \cdots x_n)^{m_n} = x_1^{m_1} x_2^{m_2} \cdots x_n^{m_n},$$

so that the induction hypothesis is valid for the polynomial $f - g$.

The procedure for the induction step can be put to practical purpose if for a given symmetric polynomial, a polynomial in the elementary symmetric polynomials is to be calculated explicitly. After a finite number of steps the procedure ends with the zero polynomial, which is valid both for the formal induction proof and as the starting point for the induction.

Furthermore, the procedure described shows that for symmetric polynomials with integer coefficients one can always find integer polynomials in the elementary symmetric polynomials.

Finally, we note that with symmetric polynomials the representation by elementary symmetric polynomials is unique. This is due to the following *uniqueness theorem*, in whose formulation one can restrict to the case of the zero polynomial (for the "general" situation of two equal polynomials, one looks at their difference).

Theorem 5.2. *A polynomial $f(y_1, \dots, y_n)$ that vanishes at the elementary symmetric polynomials, that is, for which*

$$f\Big(\sum_j x_j, \sum_{j<k} x_j x_k, \dots, x_1 x_2 \cdots x_n \Big) = 0,$$

is itself identically equal to zero.

The proof is by contradiction. Suppose that we have a polynomial $f(y_1, \dots, y_n) \neq 0$. We select from among the monomials $y_1^{m_1} \cdots y_n^{m_n}$ with nonzero coefficient a the monomial that in the arrangement of n-tuples $(m_1 + m_2 + \cdots + m_n, m_2 + \cdots + m_n, \dots, m_n)$ has the "largest" n-tuple in the lexicographic sort order. Such a monomial is uniquely determined, since for two equally large n-tuples, the exponents of the two monomials in question are the same. It follows that

$$\begin{aligned} &f\Big(\sum_j x_j, \sum_{j<k} x_j x_k, \dots, x_1 x_2 \cdots x_n \Big) \\ &\qquad = a x_1^{m_1 + m_2 + \cdots + m_n} x_2^{m_2 + \cdots + m_n} \cdots x_n^{m_n} + g(x_1, \dots, x_n), \end{aligned}$$

where in the polynomial g only monomials appear that are lexicographically smaller than the first monomial. Altogether, then, as a sum, expressed as

a polynomial in the variables $x_1, \ldots, x_n$, we obtain a polynomial different from the zero polynomial. But this contradicts the assumption.

5.5 We are not going to go any more deeply into Langrange's investigations, since aside from his theorem on symmetric polynomials, his results will not be needed in the later chapters of this book. It is worth noting, though, the great influence of Lagrange's work on later mathematicians such as Abel and Galois. It was Lagrange's great accomplishment to have noticed the significance of permuting the solutions of an equation. Lagrange was also certainly the first to have recognized the main difficulties in solving the general equation of fifth degree. Lagrange was able to simplify the universal method for "his" resolvents in the case of the general fifth-degree equation, though it led to a resolvent of the sixth degree to be solved, for which no simplification presented itself. The first attempt to establish the impossibility of a solution of the general fifth-degree equation in terms of nested radicals—called a *solution in radicals*—was undertaken by the Italian Paolo Ruffini (1765–1822), who held chairs in mathematics and medicine at the University of Modena. Although his attempts at a proof are incomplete, his arguments go a long way toward establishing that unlike the general fourth-degree equation, that of the fifth degree can have no solution in radicals for which the intermediate values are polynomials in the variables $x_1, x_2, \ldots$. Lagrange's methods of finding a general solution to the equation of fifth degree could not, then, lead to success. (See the section on Ruffini and the general equation of fifth degree.)

Ruffini's work on the impossibility of solving the general fifth-degree equation in radicals appeared between 1799 and 1813. A complete proof of this impossibility—originally completely independently of Ruffini's work—was given in 1826 by the twenty-four-year-old Danish mathematician Niels Henrik Abel. Abel's proof contains in particular a proof of the assumption made by Ruffini without proof, namely, that if a solution in radicals of the general fifth-degree equation were possible, the steps in the solution could always be arranged in such a way that all intermediate values are polynomials in the variables

$x_1, x_2, \ldots$. Thus intermediate values of the form, say,

$$\sqrt[5]{1 + x_3 + \sqrt{x_1 + x_2^3 x_4}}$$

in a general solution formula with radicals can always be avoided.

Abel's impossibility proof applies only to the general equation of fifth or higher degree, in which the solutions $x_1, x_2, \ldots$ interpreted as variables are to be determined in the sense of a "general formula" in elementary symmetric polynomials. Whether the solutions of special fifth-degree equations such as, for example,

$$x^5 - x - 1 = 0,$$

$$x^5 + 330x - 4170 = 0,$$

can be represented in terms of nested radicals is not answered in Abel's proof. Thus, for example, the solutions of the first of the above equations cannot be so represented, while in fact the second one can. For instance,

$$x_1 = \sqrt[5]{54} + \sqrt[5]{12} + \sqrt[5]{648} - \sqrt[5]{144}.$$

From Abel's posthumous papers, it is known that in 1828, after he had returned to Norway from research trips to Berlin (1825) and Paris (1826), he was working on questions of the solvability in radicals of special equations of the nth degree. Alas, at the time, Abel was seriously ill with tuberculosis. In modest circumstances and without having achieved a position commensurate with his mathematical accomplishments, Abel died in 1829 at the age of twenty-six.[4] The problem of whether and under what circumstances a particular equation is solvable in radicals had to wait several years for a solution by Galois.

Ruffini and the General Equation of Fifth Degree

Ruffini's argument for why there is no general solution consisting of only arithmetic computations and extraction of roots of equations of degree five and higher was incomplete in some details. Moreover, Ruffini's method of

[4]For a biography of Abel, see Arild Stubhaug, *Niels Henrik Abel and His Times*, Springer, 2000 (Norwegian original, 1996).

argument was quite unusual for his time, in which mathematics relied very heavily on concrete calculations. Receiving little attention and recognition from his mathematical peers, he attempted to improve and simplify his argument. In what follows we will look at the central idea of his last attempt at proof in 1813 in a slightly revised form.[5]

Ruffini's research considers the ways in which the variables of a given polynomial can be permuted without changing the polynomial. For example, the polynomial $xy - 3z^2$ remains unchanged when the variables x and y are interchanged, while that is not the case for the remaining four permutations aside from the identity. The following result is the basis of Ruffini's attempted proof:

Theorem 5.3. *For a polynomial $g(x_1, \ldots, x_5)$ in the variables $x_1, \ldots, x_5$, let $f(x_1, \ldots, x_5)$ denote the mth power $g(x_1, \ldots, x_5)^m$, where m is a natural number. Then if the polynomial f satisfies the identities*

$$f(x_1, x_2, x_3, x_4, x_5) = f(x_2, x_3, x_1, x_4, x_5) = f(x_1, x_2, x_4, x_5, x_3)$$

for the corresponding permutations of the variables, then the analogous identities hold for the polynomial g.

A proof begins with the remark that the assumption is equivalent to the identity

$$\begin{aligned} g(x_1, x_2, x_3, x_4, x_5)^m &= g(x_2, x_3, x_1, x_4, x_5)^m \\ &= g(x_1, x_2, x_4, x_5, x_3)^m. \end{aligned}$$

Thus there must exist two mth roots of unity ζ_1 and ζ_2 such that

$$\begin{aligned} g(x_1, x_2, x_3, x_4, x_5) &= \zeta_1\, g(x_2, x_3, x_1, x_4, x_5), \\ g(x_1, x_2, x_3, x_4, x_5) &= \zeta_2\, g(x_1, x_2, x_4, x_5, x_3). \end{aligned}$$

By repeatedly applying the underlying permutation of the variables in the first equation—in both equations three variables are permuted cyclically and the remaining two are unchanged—one obtains the following result:

$$\begin{aligned} g(x_1, x_2, x_3, x_4, x_5) &= \zeta_1\, g(x_2, x_3, x_1, x_4, x_5) = \zeta_1^2\, g(x_3, x_1, x_2, x_4, x_5) \\ &= \zeta_1^3 g\,(x_1, x_2, x_3, x_4, x_5). \end{aligned}$$

This together with the analogous computation for the second permutation implies that

$$\zeta_1^3 = \zeta_2^3 = 1.$$

---

[5]See Raymond G. Ayoub, Paolo Ruffini's contributions to the quintic, *Archive for History Exact Sciences*, 23 (1980), pp. 253–277; Raymond G. Ayoub, On the non-solvability of the general polynomial, *American Mathematical Monthly*, 89 (1982), pp. 307–401; Christian Skau, Gjensen med Abels og Ruffinis bevis for unmuligheten av å løse den generelle n' tegradsligningen algebraisk når n 35, *Nordisk Matematisk Tidskrift* (Normat), 38 (1990), pp. 53–84, 192; Ivo Radloff, Abels Unmöglichkeitsbeweis im Spiegel der modernen Galoistheorie, *Mathematische Semesterberichte*, 45 (1998), pp. 127–139.

Now the two permutations are combined. In particular, the first permutation is followed by the second:

$$g(x_1, x_2, x_3, x_4, x_5) = \zeta_1\, g(x_2, x_3, x_1, x_4, x_5) = \zeta_1\zeta_2\, g(x_2, x_3, x_4, x_5, x_1).$$

If the first permutation is executed twice, then one obtains

$$g(x_1, x_2, x_3, x_4, x_5) = \zeta_1\, g(x_2, x_3, x_1, x_4, x_5) = \zeta_1^2\zeta_2\, g(x_3, x_1, x_4, x_5, x_2).$$

For the two permutations that underlie the last two derived equations, the five variables are permuted cyclically (that is, the two cycles $x_1 \to x_2 \to x_3 \to x_4 \to x_5 \to x_1$ and $x_1 \to x_3 \to x_4 \to x_5 \to x_2 \to x_1$). With analogous argumentation to that for the two three-cycles, one obtains

$$(\zeta_1\zeta_2)^5 = (\zeta_1^2\zeta_2)^5 = 1.$$

From these two identities it follows immediately that $\zeta_1^5 = 1$ and then, using the previously obtained equations, $\zeta_1 = (\zeta_1^3)^2\,(\zeta_1^5)^{-1} = 1$. Building on this, we also obtain $\zeta_2^5 = 1$ and $\zeta_2 = (\zeta_2^3)^2\,(\zeta_2^5)^{-1} = 1$, from which finally the asserted identities for the polynomial g follow.

With this proven property of polynomials in five variables, it is at once plausible that a solution formula for the general equation of fifth degree cannot exist, at least not in the manner of the formulas for equations of lower degree. That is, as described by Lagrange, those formulas are obtained by beginning with the elementary symmetric polynomials and through them determining step by step polynomials $g_1, g_2, \ldots$ in the variables $x_1, x_2, \ldots$, so that for each of them a power is found that can be determined from the polynomials obtained in the previous steps using the four elementary operations. The jth step therefore has the form

$$g_j(x_1, x_2, \ldots)^{m_j} = f_j(x_1, x_2, \ldots),$$

where the function f_j is expressed in terms only of the elementary symmetric polynomials and the polynomials $g_1, g_2, \ldots, g_{j-1}$ determined in the previous steps. If the given general equation is of degree five (or greater), then Ruffini's argument can be applied inductively to assert that every polynomial g_j must satisfy the condition

$$g_j(x_1, x_2, x_3, x_4, x_5) = g_j(x_2, x_3, x_1, x_4, x_5) = g_j(x_1, x_2, x_4, x_5, x_3).$$

None of the steps can therefore lead to the polynomial of the last solution step, such as, for example, $g(x_1, x_2, \dots) = x_1$.[6]

Exercises

(1) For a given cubic equation

$$x^3 + ax^2 + bx + c = 0,$$

determine the cubic equation whose solutions are the squares of the given equation.

(2) Show that the solution of the general biquadratic equation

$$x^4 + ax^3 + bx^2 + cx + d = 0$$

can be obtained directly, that is, without transforming it into a reduced biquadratic equation (without a third power) by constructing a cubic equation for the resolvent

$$z = x_1x_2 + x_3x_4$$

in order to calculate the solutions of the biquadratic equation from the resolvent z.

(3) Carry out the calculations implied in the previous exercise for the resolvent

$$z = (x_1 + x_2)(x_3 + x_4).$$

[6]The question natually arises as to the points at which Ruffini's argument is incomplete and how these deficits can be fixed. We have already mentioned that Abel offered a proof that a solution in radicals of the general equation, if one exists, can always be obtained in such a way that each intermediate step corresponds to a polynomial in the solutions. A reproduction of Abel's proof with commentary can be found in Peter Pesic, *Abel's Proof. An essay on the sources and meaning of mathematical unsolvability*, Cambridge, MA, 2003, pp. 155–174. As an alternative to Abel's argument, it is also possible to extend permutations to formal expressions containing nested radicals, such as

$$\sqrt[5]{1 + x_3 + \sqrt{x_1 + x_2^3x_4}}.$$

One can find a complete proof based on this approach in John Stillwell, Galois theory for beginners, *American Mathematical Monthly*, 101 (1994), pp. 22–27. We will not go into this further, preferring to highlight Galois's more general approach.

(4) For two polynomials with decompositions into linear factors

$$f(X) = (X - x_1) \cdots (X - x_n),$$
$$g(X) = (X - y_1) \cdots (X - y_m),$$

one defines the *resultant* by

$$R(f, g) = \prod_{i=1}^{n} \prod_{j=1}^{m} (x_i - y_j).$$

Clearly, the resultant is identically zero precisely when the two polynomials have a root in common. Show that the resultant can be formally derived from the coefficients of the polynomial. Give an explicit formula for the case $n = m = 2$.

(5) A permutation is called *cyclic* if it permutes k of the n numbers $1, 2, \ldots, n$ cyclically and leaves the remaining $n - k$ numbers in place. A cycle that exchanges exactly two numbers is known as a *transposition*. Prove the following:
   (a) Every permutation is the product of cycles.
   (b) Every cycle is the product of transpositions.
   (c) Every permutation is the product of transpositions.
   (d) Every permutation is the product of transpositions that exchange the number 1 with some other number.
   (e) Every permutation is the product of transpositions that exchange two adjacent numbers j and $j + 1$.

Chapter 6

Equations That Can Be Reduced in Degree

In contrast to the equation $x^5 - 2x^4 - 4x^3 + 2x^2 + 11x + 4 = 0$, *whose solutions can be determined from a quadratic equation and a cubic equation with integer coefficients, nothing is comparable for the equation* $2x^5 + 6x^2 + 3 = 0$. *What is the basis of this difference, and how can it be recognized?*

6.1 The previous chapters dealt with techniques for finding general solution formulas for equations of particular degrees. Now, with Abel's proof of the impossibility of finding a formula for the solution in radicals of the general equation of fifth or higher degree, we shall restrict our attention to special equations of fifth and higher degree.

The first equation presented at the beginning of this chapter is an example of an equation that though not decomposable into linear factors, can be decomposed into factors whose degrees are greater than 1. Since

$$x^5 - 2x^4 - 4x^3 + 2x^2 + 11x + 4 = \left(x^3 - 3x - 4\right)\left(x^2 - 2x - 1\right),$$

three of the five solutions can be determined from the cubic equation

$$\left(x^3 - 3x - 4\right) = 0,$$

and the remaining two solutions form the quadratic equation

$$x^2 - 2x - 1 = 0.$$

Using the methods described in the previous chapters, one obtains the solutions

$$x_{1,2,3} = \zeta \sqrt[3]{2+\sqrt{3}} + \zeta^2 \sqrt[3]{2-\sqrt{3}} \quad \text{with} \quad \zeta^3 = 1,$$
$$x_{4,5} = 1 \pm \sqrt{2}.$$

For the second equation, there is no corresponding decomposition of the polynomial $2x^5 + 6x^2 + 3$ into two polynomials with rational coefficients. How such a negative assertion can be justified and how decompositions can be found when they exist is the topic of this chapter. In contrast to the previous chapters, in which concrete computations stood in the foreground, here we will be rather more concerned with a qualitative point of view relating to properties of polynomials with rational or integer coefficients. The proofs are not too long and difficult, but in comparison to what has gone before, they are based on a different type of argumentation.

The basis of applications is the following proof, due to Carl Friedrich Gauss:

Theorem 6.1. *Let $g(x)$ and $h(x)$ be two polynomials whose leading coefficient (that is, the coefficient of the highest power of x) is 1 (such polynomials are called* monic*), all of whose coefficients are rational, and whose product $g(x)h(x)$ is a polynomial with integer coefficients. Then all the coefficients of the original polynomials g and h must be integers.*

This theorem can be seen as a drastic generalization of the well-known fact that the square root of 2 is irrational. Namely, the polynomial $x^2 - 2$ does not admit a decomposition into linear factors with rational coefficients, since such coefficients would have to be integers, which is obviously impossible. Moreover, the argumentation in the section on the decomposition of polynomials with integer coefficients has a certain relationship with that in the classical proof that the square root of 2 is irrational. The basis of both is a detailed check of divisibility relations, where the assumption of such a relation leads to a contradiction.

The Decomposition of Integer Polynomials

Recall Theorem 6.1:

> If $g(x)$ and $h(x)$ are two monic polynomials with rational coefficients whose product $g(x)h(x)$ has all integer coefficients, then the two polynomials $g(x)$ and $h(x)$ have integer coefficients.

The proof of this theorem, which goes back to Gauss, begins with multiplying through by denominators in $g(x)$ and $h(x)$. In particular, we determine two integers a and b of minimal size for which the two polynomials $a \cdot g(x)$ and $b \cdot g(x)$ have all their coefficients integers. We denote these coefficients by $c_0, c_1, \ldots$ and $d_0, d_1, \ldots$. We now investigate the product $ab \cdot g(x)h(x)$.

We will now show by obtaining a contradiction that there is no prime number p that divides all the coefficients of the product $ab \cdot g(x)h(x)$. This is how we shall prove the assertion: since the product $g(x)h(x)$ has integer coefficients, the nonexistence of the asserted prime p implies at once that $ab = 1$, and therefore $a = b = 1$, so that given how a and b were chosen, it follows that the given polynomials $g(x)$ and $h(x)$ cannot have had any denominators to get rid of, and so their coefficients must have been integers to begin with.

Let us therefore assume that there exists a prime number p that divides all coefficients of the product polynomial $ab \cdot g(x)h(x)$. In reference to this prime number p we consider two cases:

Case 1. We first consider the case in which neither the polynomial $a \cdot g(x)$ nor $b \cdot h(x)$ contains only coefficients divisible by p. Thus we may find the smallest indices j and k such that neither c_j nor d_k is divisible by p. The coefficient of the term x^{j+k} of the polynomial $ab \cdot g(x)h(x)$, which can be expressed as the sum

$$c_j d_k + c_{j-1} d_{k+1} + \cdots + c_{j+1} d_{k-1} + \cdots,$$

is seen not to be divisible by p, in contradiction to our assumption, given the choice of j and k, since the first summand cannot be divisible by p, while all the other summands must be divisible by p.

Case 2. In this case, we assume that all the coefficients of either $a \cdot g(x)$ or $b \cdot h(x)$ are divisible by p. Without loss of generality we may assume that this holds for $a \cdot g(x)$. Since in the polynomial $g(x)$ the leading coefficient is 1, it must be the case that a is divisible by p. In particular, $a > 1$. But that delivers our desired contradiction, since the polynomial $\frac{a}{b} \cdot g(x)$ must possess integer coefficients, contradicting the minimal choice of a.

If we now apply the theorem to the special case of searching for a linear factor, we see at once that monic polynomials with integer coefficients can have only integer roots among its rational roots. Since such roots must divide the constant term of the polynomial, one can easily check a finite number of integers as potential roots to determine all rational roots of the polynomials.

6.2 Before we describe how in many special cases the demonstration of the impossibility of a product decomposition into polynomials with rational coefficients can be simplified, we shall look at some examples. We would like to know how a product decomposition can be found if one exists. As an example, let us consider the polynomial $x^5 - 2x^4 - 4x^3 + 2x^2 + 11x + 4$ presented at the beginning of the chapter. If there is a decomposition, then one of the factors must be of first or second degree, and the theorem tells us that we may limit our search to monic polynomials with integer coefficients. We immediately see that there is no linear factor, since there are only six factors of 4 to check, namely $\pm 1, \pm 2, \pm 4$, none of which turns out to be a root of the polynomial. Therefore any decomposition would have to be of the form

$$\begin{aligned} x^5 - 2x^4 - 4x^3 + 2x^2 + 11x + 4 \\ = (x^2 + ax + b)\left(x^3 - (a+2)x^2 + cx + \frac{4}{b}\right), \end{aligned}$$

where a and c must be integers and b must be one of the six factors of 4. One can obtain further restrictions on the coefficients by evaluating the fifth-degree polynomial at integer arguments. For example, at $x = 2$, the polynomial has the value 2, so that, for instance, the hypothetical quadratic factor must divide 2 when $x = 2$. Therefore, the expression $4 + 2a + b$ must be one of the four factors of 2, namely one of $\pm 1, \pm 2$. Already these two restrictions allow us to conclude that in all, "only" $6 \cdot 4$ possibilities for a and b need to be tried.[1]

[1] A completely different approach is possible if one determines the five complex roots of the polynomial to be factored using numerical approximation. Then one need only check which possible selection of linear factors produces a polynomial with integer coefficients, which can be done without the possibility of rounding error by multiplying out any polynomials thus obtained. Those in a hurry can try a computer algebra system. Thus the Mathematica command Factor$[x^5 - 2x^4 - 4x^3 + 2x^2 + 11x + 4]$ immediately yields the result $\left(-1 - 2x + x^2\right)\left(-4 - 3x + x^3\right)$.

With such restrictions on the coefficients one can of course also reach negative conclusions, that a given polynomial with integer coefficients cannot be expressed as the product of two polynomials of lower degree. Such polynomials are called *irreducible* over the rational numbers.

6.3 As promised, for a proof of irreducibility there frequently exist easier ways that make decisive use of divisibility relations. On the polynomial $2x^5 + 6x^2 + 3$ introduced at the beginning of the chapter we can use the so-called *Eisenstein irreducibility criterion*, named for the mathematician Ferdinand Gotthold Max Eisenstein (1823–1852), who proved this theorem in 1850, independently of a proof given four years earlier by Theodor Schönemann.

Theorem 6.2. *Suppose we are given a polynomial* $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$ *with integer coefficients that satisfies the following conditions for some prime number* p*:*

- $a_{n-1}, \ldots, a_1, a_0$ *are divisible by* p*.*
- a_0 *is not divisible by* p^2*.*

Then the polynomial $f(x)$ *is irreducible over the rational numbers.*

The proof of the Eisenstein irreducibility criterion is not very difficult. It can be found in the appropriately named section of this chapter.

There is no trick to showing that the polynomial $2x^5 + 6x^2 + 3$ is irreducible using the Eisenstein criterion. (Recall that *irreducible* means that the polynomial cannot be decomposed nontrivially into two polynomials of lower degree with rational coefficients.) To obtain a monic polynomial, we begin with the polynomial multiplied by 16, which can be written $(2x)^5 + 24(2x)^2 + 48$. With respect to the Eisenstein criterion for the prime $p = 3$, the polynomial $y^5 + 24y^2 + 48$ is irreducible over the rational numbers. Therefore, the polynomial $2x^5 + 6x^2 + 3$ is also irreducible, since a decomposition would immediately carry over to one for $y^5 + 24y^2 + 48$.

6.4 An important application of the Eisenstein irreducibility criterion is to the cyclotomic equation $x^n - 1 = 0$. Since the linear factor

$(x-1)$ can be removed, the cyclotomic equation is never irreducible for $n > 1$. However, for a prime exponent n, the linear factor $(x-1)$ is the only possible decomposition into polynomials with rational coefficients. In other words, if you factor out the linear factor, what is left is irreducible. That is, it can be shown that the polynomial

$$\frac{x^n-1}{x-1} = x^{n-1} + x^{n-2} + \cdots + x^2 + x + 1$$

is irreducible if n is prime. To prove this, make the substitution $x = y+1$, obtaining, using the binomial theorem,

$$\begin{aligned}\frac{(y+1)^n-1}{y} &= y^{n-1} + \binom{n}{n-1}y^{n-2} + \cdots + \binom{n}{3}y^2 + \binom{n}{2}y + \binom{n}{1} \\ &= y^{n-1} + \sum_{j=1}^{n-2} \frac{n\cdots(n-j)}{1\cdot 2\cdots(j+1)}y^j + n.\end{aligned}$$

As is well known, all the binomial coefficients are integers. Furthermore, the last representation given shows that all the binomial coefficients that appear are divisible by n, since the prime factor n appearing in the numerator is not canceled by anything in the denominator. We can then apply Eisenstein's criterion for the prime n to the polynomial $(x^n-1)/(x-1)$, whence this polynomial is seen to be irreducible over the rational numbers.

Eisenstein's Irreducibility Criterion

Recall Theorem 6.2:

> Suppose we are given a polynomial $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$ with integer coefficients that satisfies the following conditions for some prime number p:
>
> - $a_{n-1}, \ldots, a_1, a_0$ are divisible by p.
> - a_0 is not divisible by p^2.
>
> Then the polynomial $f(x)$ is irreducible over the rational numbers.

The proof can be carried out indirectly once again; that is, we assume the opposite of what we are trying to prove and arrive at a contradiction. We thus assume that we have a decomposition $f(x) = g(x)h(x)$, where $g(x)$

and $h(x)$ are monic polynomials with rational coefficients:

$$g(x) = c_r x^r + c_{r-1} x^{r-1} + \cdots + c_0,$$
$$h(x) = d_s x^s + d_{s-1} x^{s-1} + \cdots + d_0,$$

with $c_r = d_s = 1$. It is of course assumed that the degrees r and s are each at least equal to 1.

From the previous theorem, all the coefficients $c_r, c_{r-1}, \ldots, c_0$ and $d_s, d_{s-1}, \ldots, d_0$ must be integers. Since the product $a_0 = c_0 d_0$ is divisible by the prime number p, but not p^2, exactly one of the coefficients c_0 and d_0 must be divisible by p. Let us assume that it is c_0. Thus the coefficient d_0 is not divisible by p. Since $c_r = 1$, we can find the smallest index j for which c_j is not divisible by p. For the corresponding coefficients a_j of the polynomial $f(x)$ we have the formula

$$a_j = c_j d_0 + c_{j-1} d_1 + \cdots + c_0 d_j,$$

where the first summand is not divisible by p, while all of the others are divisible by p. Thus a_j is not divisible by p, which on account of $j \leq r < n$ is a contradiction.

Exercises

(1) Find a factorization of the polynomial

$$x^6 + 9x^5 + 19x^4 - 4x^3 + 5x^2 - 13x - 3$$

over the rational numbers into irreducible factors.

(2) Show that the polynomial

$$x^6 + 4x^5 - 2x^4 + x^3 - 3x^2 + 5x + 1$$

is irreducible over the rational numbers.

Chapter 7

The Construction of Regular Polygons

With the words, "With concentrated thought ... in the morning ... (before I got out of bed)," Carl Friedrich Gauss describes the circumstances surrounding his discovery in the year 1796 that the regular seventeen-sided polygon can be constructed using straightedge and compass. How could Gauss have managed to consider the possibility of a geometric construction as an exercise of pure imagination?

7.1 The discovery described above by the eighteen-year-old Gauss on March 29, 1796, marks the beginning of a life in mathematics whose scope and significance have seldom been equaled.[1] Gauss himself described in a literary journal his discovery regarding the regular heptadecagon (seventeen-gon) thus:

> It is known to every beginner in geometry that various regular polygons, namely the triangle, pentagon, fifteen-gon, and those obtainable by doubling the number of sides, have been known to be constructible since the time of Euclid, and it would appear that since that time, mathematicians have convinced themselves that the field of elementary geometry was unable to yield further results; at least I

[1]The chronology of Gauss's discoveries is extraordinarily well documented in his mathematical diaries. The first entry reads, "fundamentals on which the division of the circle is based and indeed its divisibility into seventeen parts, etc." See C. F. Gauss, *Mathematisches Tagebuch, 1796–1814*, Ostwalds Klassiker Nr. 256, Leipzig, 1976.

know of no successful attempt at extending geometry's reach in this direction.

All the more, it seems to me that note should be taken of the discovery that in addition to those regular polygons, a host of others are amenable to geometric construction, for example, the heptadecagon . . . .

Geometric constructions with straightedge and compass, generally of triangles from three given data, are a residue of classical mathematics still a part of the standard school curriculum. The significance of such exercises is less their practical application than, aside from being part of a tradition that stretches back to antiquity, to aid the student in developing logical habits of thought. Construction with straightedge (unmarked ruler) and compass is limited to prescribed elementary operations that allow the construction of certain points, starting with two points separated by a distance of unit length. Thus given a set of points that have been thus constructed, the following can be additionally constructed:

- Draw a circle whose midpoint is a point that has been constructed and whose radius is the distance between two constructed points.
- Draw a straight line between two constructed points.
- Every intersection of circles and lines drawn in the previous two steps is considered a constructed point.

At first glance, there seems to be no connection between such geometric constructions and equations in one variable. However, as we have seen in Chapter 2, the nth roots of unity in the complex plane, that is, the n solutions of the equation $x^n - 1 = 0$, are the vertices of a regular n-gon, and indeed with the unit circle as circumscribing circle. Consider Figure 7.1. If starting at the point $1 = (1, 0)$ we can show that the next point of the n-gon in the counterclockwise direction, namely $\zeta = \cos\left(\frac{2\pi}{n}\right) + i \sin\left(\frac{2\pi}{n}\right)$, can be constructed with straightedge and compass, then we will have succeeded in proving the regular n-gon to be constructible.

Gauss, who was well acquainted with the geometric interpretation of complex numbers as points in the plane—indeed, in his honor one

sometimes speaks of the Gaussian plane—was able to solve cyclotomic equations in radicals. In order to find suitable intermediate values, he first ordered the nth roots of unity in a particular way, motivated by his knowledge of the divisibility properties of the integers.

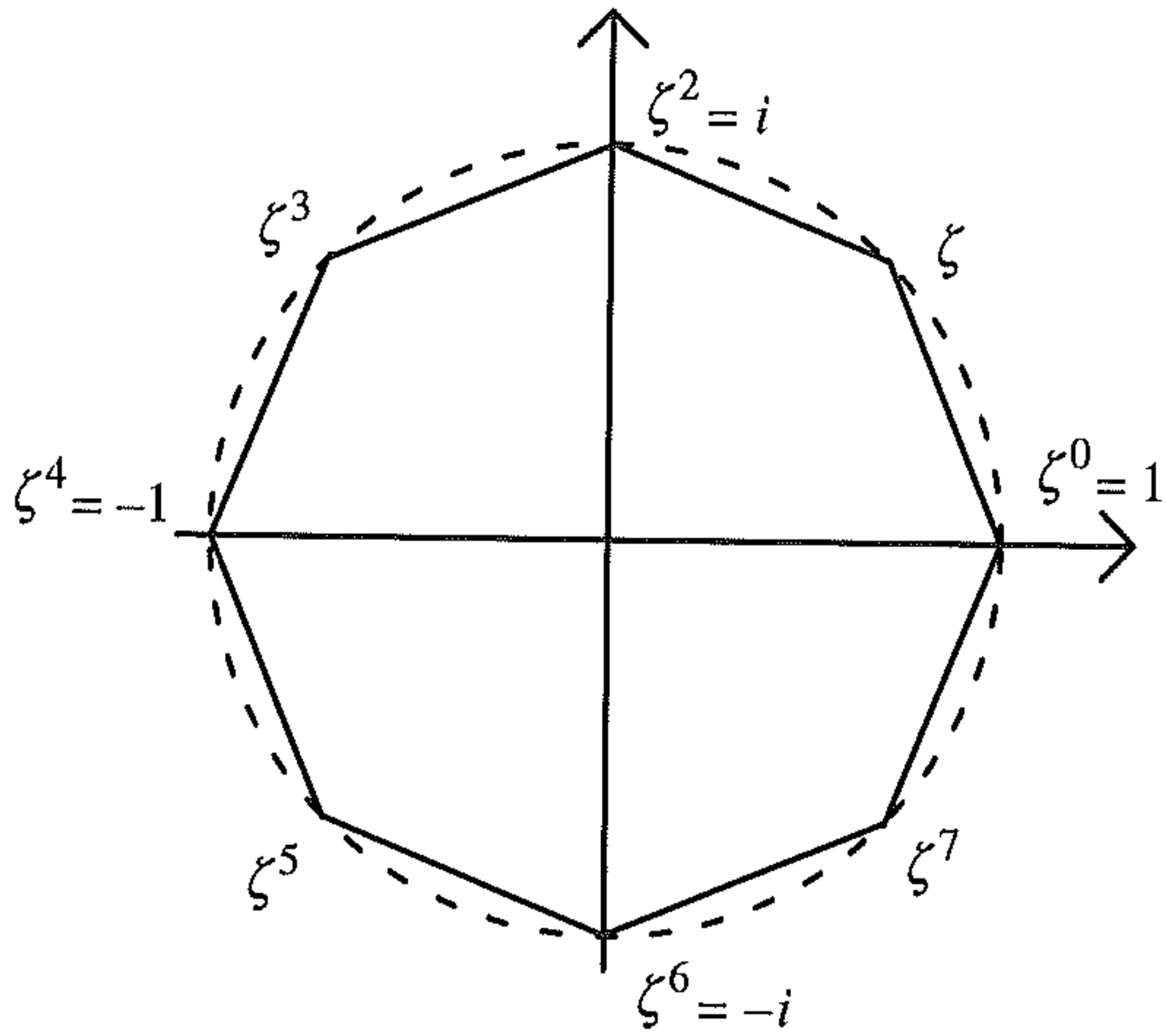


Figure 7.1. The solutions of the cyclotomic equation $x^8 - 1 = 0$ form a regular octagon. All eight eighth roots of unity can be expressed as powers $1, \zeta, \zeta^2, \ldots, \zeta^7$ of the *primitive* root $\zeta = \cos\left(\frac{2\pi}{8}\right) + i \sin\left(\frac{2\pi}{8}\right)$.

It seems sensible at this point to order the roots according to their position on the circle, that is, as seen from Figure 7.1, in the order $1, \zeta, \zeta^2, \ldots, \zeta^{n-1}$, where $\zeta = \cos\left(\frac{2\pi}{n}\right) + i \sin\left(\frac{2\pi}{n}\right)$. Gauss, however, realized that it makes sense to list the roots in quite a different order, at least in the case that n is a prime number. Considering that $\zeta^n = 1$, the value of ζ^j depends only on the remainder when j is divided by n. Therefore, we can choose any order of the possible remainders on division by n. In addition to the obvious order $1, 2, \ldots, n-1$, it is possible, in the case of a prime number n, to obtain all the nonzero remainders $1, 2, \ldots, n-1$ not only by repeated addition of 1, but by repeated multiplication by a suitable number g.[2] This leads to

[2] A proof of this fact can be found in the epilogue to this book.

an ordering $g^0, g^1, g^2, \ldots, g^{n-2}$. The remainder obtained when g is divided by n is called a *primitive root modulo n.*[3]

In the case of $n = 17$, for example, one can choose $g = 3$. Indeed, starting with $g^0 = 1$, after $g^1 = 3^1 = 3$, $g^2 = 9$, one has $g^3 = 3^3 = 27 \equiv 10 \mod 17$. Then comes $3^4 \equiv 3 \cdot 10 = 30 \equiv 13$. Altogether, one obtains the order

$$3,\ 9,\ 10,\ 13,\ 5,\ 15,\ 11,\ 16,\ 14,\ 8,\ 7,\ 4,\ 12,\ 2,\ 6,\ 1.$$

Since the list ends with $g^{16} \equiv 1$, we could continue with $g^{17} \equiv 3$, $g^{18} \equiv 9$, and so on ad infinitum.

In the case of a regular heptadecagon (17-gon), the resulting list of roots of unity takes the form

$$\zeta^1,\ \zeta^3,\ \zeta^9,\ \zeta^{10},\ \zeta^{13},\ \zeta^5,\ \zeta^{15},\ \zeta^{11},\ \zeta^{16},\ \zeta^{14},\ \zeta^8,\ \zeta^7,\ \zeta^4,\ \zeta^{12},\ \zeta^2,\ \zeta^6.$$

The purpose of this is to form partial sums of the roots of unity, called *periods*, which allow for a step-by-step calculation of the roots of unity. One begins with the two periods containing the roots of unity that stand in odd, respectively even, positions. These are called the *eight-member periods*:

$$\eta_0 = \zeta^1 + \zeta^9 + \zeta^{13} + \zeta^{15} + \zeta^{16} + \zeta^8 + \zeta^4 + \zeta^2,$$
$$\eta_1 = \zeta^3 + \zeta^{10} + \zeta^5 + \zeta^{11} + \zeta^{14} + \zeta^7 + \zeta^{12} + \zeta^6.$$

Next one considers the four periods containing the roots whose positions differ by 4 in the list. These sums of four roots of unity are called *four-member periods*:

$$\mu_0 = \zeta^1 + \zeta^{13} + \zeta^{16} + \zeta^4,$$
$$\mu_1 = \zeta^3 + \zeta^5 + \zeta^{14} + \zeta^{12},$$
$$\mu_2 = \zeta^9 + \zeta^{15} + \zeta^8 + \zeta^2,$$
$$\mu_3 = \zeta^{10} + \zeta^{11} + \zeta^7 + \zeta^6.$$

Finally, we consider the *two-member periods*, which are the sums of roots of unity a distance of eight apart in the original list. For our

---

[3]The expression "modulo n" is generally used to indicate that the identity in question holds only up to a division by n. For example, 12 is "equal" to 46 modulo 17, or in mathematical language, 12 is *congruent* to 46 modulo 17, written $12 \equiv 46 \mod 17$, since both numbers have the same remainder, namely 12, on division by 17. Equivalently, when the difference $46 - 12$ is divided by 17, the remainder is zero.

purposes the following two periods suffice:

$$\beta_0 = \zeta^1 + \zeta^{16},$$
$$\beta_4 = \zeta^{13} + \zeta^4.$$

All of these periods are real and have the further property—and this is what Gauss recognized by "concentrated thought"—obtained by this special construction that every period can be obtained from the next-longer period by a quadratic equation. For this the periods are paired off, so that each sum and each product of the pair can be represented as a sum of periods of double the length. Let us see how this works.

The calculation begins with the two eight-member periods η_0 and η_1. Their sum is not too difficult to calculate:

$$\eta_0 + \eta_1 = \zeta^1 + \zeta^2 + \cdots + \zeta^{16} = \left(1 + \zeta^1 + \zeta^2 + \cdots + \zeta^{16}\right) - 1 = -1,$$

where we note that the sum of all nth roots of unity is always zero, which follows at once algebraically from Viète's root theorem applied to the cyclotomic equation, while geometrically, the origin is clearly the center of mass of the n vertices. In contrast, determining the sixty-four products in $\eta_0\eta_1$ is tedious. After great but elementary effort, one obtains that $\eta_1\eta_1 = -4$. Therefore, the two eight-member periods can be calculated as solutions of the quadratic equation

$$y^2 + y - 4 = 0,$$

yielding

$$\eta_{0,1} = -\frac{1}{2} \pm \frac{1}{2}\sqrt{17}.$$

Now from the two eight-member periods η_0 and η_1, the four four-member periods $\mu_0, \mu_1, \mu_2, \mu_3$ can be calculated, though we shall omit the gory details:

$$\mu_0 + \mu_2 = \eta_0,$$
$$\mu_0\mu_2 = \zeta^1 + \zeta^2 + \cdots + \zeta^{16} = -1,$$
$$\mu_1 + \mu_3 = \eta_1,$$
$$\mu_1\mu_3 = -1.$$

These four identities lead to the following two quadratic equations, which make possible the calculation of the four-member periods:

$$y^2 - \eta_0 y - 1 = 0,$$
$$z^2 = \eta_1 z - 1 = 0.$$

The two solutions of the first equation are $y_1 = \mu_0$ and $y_2 = \mu_2$, while those of the second equation are $z_1 = \mu_1$ and $z_2 = \mu_3$.

Finally, we can compute the two two-member periods β_1 and β_4. Again, the key is the calculation of their sum and product:

$$\beta_0 + \beta_4 = (\zeta^1 + \zeta^{16}) + (\zeta^{13} + \zeta^4) = \mu_0,$$
$$\beta_0\beta_4 = (\zeta^1 + \zeta^{16})(\zeta^{13} + \zeta^4) = \zeta^{14} + \zeta^5 + \zeta^{12} + \zeta^3 = \mu_1.$$

From this we obtain the quadratic equation

$$y^2 = \mu_0 y + \mu_1 = 0,$$

whose solutions are the two two-member periods $y_1 = \beta_0$ and $y_2 = \beta_4$.

If one wishes, one may now calculate the seventeenth root of unity ζ from the quadratic equation

$$y^2 = \beta_0 y + 1 = 0,$$

whose two solutions are $y_1 = \zeta^1$ and $y_2 = \zeta^{16}$. However, in a geometric construction, this quadratic equation does not need to be brought into play, since the regular heptadecagon can be constructed using a segment of length $\beta_0 = 2\cos\left(\frac{2\pi}{17}\right)$.

If one solves the quadratic equations we have obtained one after the other and chooses the solutions in an order based on numerical approximations, then one obtains as end result the identity suggested in the introduction:

$$\beta_0 = 2\cos\frac{2\pi}{17} = -\frac{1}{8} + \frac{1}{8}\sqrt{17} + \frac{1}{8}\sqrt{34 - 2\sqrt{17}}$$
$$+ \frac{1}{4}\sqrt{17 + 3\sqrt{17} - \sqrt{34 - 2\sqrt{17}} - 2\sqrt{34 + 2\sqrt{17}}}.$$

This expression in square roots not only shows at once that the regular heptadecagon is constructible, but also indicates how such a

construction can be carried out.[4]. The reason is that the constructibility of a point with straightedge and compass is equivalent to the point being expressible by rational numbers, the four basic arithmetic operations, and the taking of square roots. For more on this, see the section on constructions with straightedge and compass.

Constructions with Straightedge and Compass

Using the system of Cartesian coordinates, the geometric question of what point can be constructed with straightedge and compass can be translated into a purely algebraic problem. We have the following theorem:

Theorem 7.1. *Given the "primitive" unit measure from the point* $(0,0)$ *to the point* $(1,0)$*, a point in the plane can be constructed with straightedge and compass if and only if its two coordinates can be expressed in rational numbers and nested square roots using the four basic arithmetic operations of addition, subtraction, multiplication, and division.*

We begin with the observation that a point with such coordinates can indeed be constructed with straightedge and compass. (Such a point is said to be *constructible*.) In particular, we shall show that the four basic operations and the extraction of square roots lead to constructible points. The three left-hand drawings in Figure 7.2 show how beginning with constructed lengths a and b and the unit length 1, one can construct the lengths $a+b$, $a-b$, ab, and $\frac{a}{b}$. Addition and subtraction are easily carried out by transferring one segment to the other using the compass. Multiplication and division are realized by constructing the parallel lines indicated in gray. The laws of proportions in similar triangles guarantee the results shown.

Taking a square root is accomplished using the laws of proportion in similar right triangles. In the picture on the right in Figure 7.2, all three right triangles (the two smaller triangles form a larger triangle inscribed in the semicircle of diameter $1+a$) are similar. Observe that $\sqrt{a}$ satisfies the relation $\frac{1}{\sqrt{a}} = \frac{\sqrt{a}}{a}$.

The converse statement is also not difficult to prove. To do so, we must analyze the operations of construction with straightedge and compass in terms of their effect on the coordinates of newly constructed points.

René Descartes was the first to make significant use of the idea of formulating geometric problems algebraically. Thus on the first pages of

[4]An explicit description of such a construction can be found in Ian Stewart, Gauss, *Scientific American*, 237, no. 7, pp. 122–131 as well as in Heinrich Tietze, *Famous Problems of Mathematics: Solved and Unsolved Mathematical Problems, from Antiquity to Modern Times*, New York, Graylock Press, 1965.

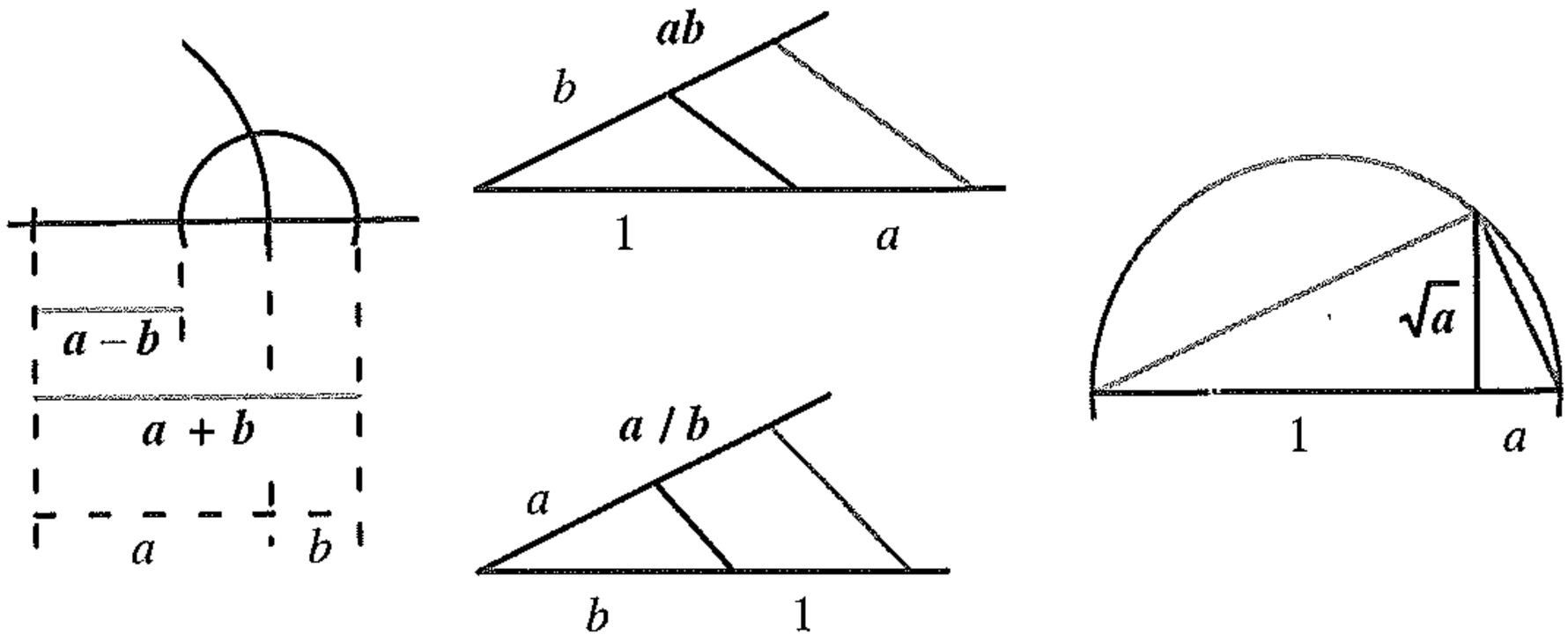


Figure 7.2. How the basic arithmetic operations and extraction of square roots are accomplished with straightedge and compass.

his work *La Géometrie* of 1637 one finds figures corresponding to those of Figure 7.2.[5] Descartes was also the first to interpret geometric products and powers in a way other than surface areas and volumes, which made possible a wide-ranging use of fourth and higher powers. Descartes's contribution is commemorated in the term *Cartesian coordinates* (from the Latinized name *Cartesius*).

7.2 Gauss did not have to carry out such explicit computations to convince himself that there is a method for constructing the regular heptadecagon. It sufficed to realize that using the periods it is possible to calculate a seventeenth root of unity using successive quadratic equations. Finally, what is crucial is that for each period, another period of the same length can be found, so that the sum and product can be computed in terms of periods of double the length. We are now going to consider this in somewhat greater detail in order to see what other regular polygons can be constructed with straightedge and compass. Unfortunately, some rather complicated calculations will be necessary, which, however, are not necessary for an understanding of the later chapters and may therefore be skipped.

As we have already described, the stepwise solution discovered by Gauss of the cyclotomic equation $x^n - 1 = 0$, where n is a prime

---

[5] See Henk J. M. Bos, Karin Reich, Algebra: Viète und Descartes, in: Erhard Scholz (ed.), *Geschichte der Algebra*, Mannheim, 1990, pp. 183–234.

number, employs a primitive root modulo n, that is, an integer g for which the list $g^1, g^2, \ldots, g^{n-1}$, when divided by n, results in a complete set of the numbers $1, 2, \ldots, n-1$. For every factorization of the form $ef = n-1$, one can define, for each power ζ^k of the root of unity $\zeta = \cos\left(\frac{2\pi}{n}\right) + i\sin\left(\frac{2\pi}{n}\right)$, the f-member periods

$$P_f\left(\zeta^k\right) = \zeta^k + \zeta^{kg^e} + \zeta^{kg^{2e}} + \cdots + \zeta^{kg^{(f-1)e}}.$$

Aside from the special case $k = 0, \pm n, \ldots$, for which all periods are equal to $P_f(1) = f$, on account of

$$P_f\left(\zeta^k\right) = P_f\left(\zeta^{kg^e}\right) = \cdots = P_f\left(\zeta^{kg^{(f-1)e}}\right),$$

at most e of the f-member periods can be distinct:

$$P_f(\zeta), \quad P_f\left(\zeta^g\right), \quad P_f\left(\zeta^{g^2}\right), \quad \ldots, \quad P_f\left(\zeta^{g^{e-1}}\right).$$

The property to be proved relates to the product of two f-member periods. Such a product can always be expressed as the sum of f-member periods. To see this, observe that

$$P_f\left(\zeta^j\right) P_f\left(\zeta^k\right) = \left(\sum_{p=0}^{f-1} \zeta^{jg^{pe}}\right)\left(\sum_{q=0}^{f-1} \zeta^{kg^{qe}}\right) = \sum_{p=0}^{f-1}\sum_{q=0}^{f-1} \zeta^{jg^{pe}+kg^{qe}}.$$

If the summation index of the inner sum is transformed by $q = p + r$, then we obtain, as desired,

$$P_f\left(\zeta^j\right) P_f\left(\zeta^k\right) = \sum_{p=0}^{f-1}\sum_{r=0}^{f-1} \zeta^{(j+kg^{re})g^{pe}} = \sum_{r=0}^{f-1}\sum_{p=0}^{f-1} \zeta^{(j+kg^{re})g^{pe}}$$

$$= \sum_{r=0}^{f-1} P_f\left(\zeta^{j+kg^{re}}\right).$$

For the special case $j = k$, we obtain

$$P_f\left(\zeta^j\right)^2 = \sum_{q=0}^{f-1} P_f\left(z^{j+g^{qe}}\right) = \sum_{q=0}^{f-1} P_f\left(\zeta^{j(1+g^{qe})}\right).$$

If the number e is even, then, as in the case of $n = 17$, the f-member periods can be calculated using quadratic equations from the $2f$-member periods. We clearly have the identity

$$P_f\left(\zeta^k\right) + P_f\left(\zeta^{kg^{e/2}}\right) = P_{2f}\left(\zeta^k\right).$$

To see that the associated product also corresponds to a sum of $2f$-member periods, it suffices to show that the sum of the two squares possesses such a representation:

$$P_f\left(\zeta^k\right)^2 + P_f\left(\zeta^{kg^{e/2}}\right)^2 = \sum_{q=0}^{f-1}\left(P_f\left(\zeta^{k(1+g^{qe})}\right) + P_f\left(\zeta^{kg^{e/2}(1+g^{qe})}\right)\right)$$

$$= \sum_{q=0}^{f-1} P_{2f}\left(\zeta^{k(1+g^{qe})}\right).$$

That this general formula is useful would be immediately clear to anyone who has painfully worked out the sixty-four terms of the product $\eta_0\eta_1$ in our earlier investigation of the regular heptadecagon. In comparison to the explicit calculation, one obtains the result using the formula just derived much more quickly:

$$\eta_0^2 + \eta_1^2 = P_8(\zeta)^2 + P_8\left(\zeta^3\right)^2 = \sum_{q=0}^{7} P_{16}\left(\zeta^{1+3^{2q}}\right) = 1\cdot 16 + 7\cdot(-1) = 9.$$

Here it is only for the summation index $q = 4$ that a summand different from -1 appears, namely, $P_{16}(1) = 16$. Consequently, one obtains, as desired,

$$\eta_0\eta_1 = \frac{1}{2}\left(\left(\eta_0 + \eta_1\right)^2 - \left(\eta_0^2 + \eta_1^2\right)\right) = \frac{1}{2}(1 - 9) = -4.$$

In general, the formula for the sum of the period squares shows that Gauss's method for solving the cyclotomic equation $x^n - 1 = 0$ always leads to a sequence of quadratic equations if n is a prime number of the form $n = 2^s + 1$. As of today, there are known only five such primes, called *Fermat primes*,[6] namely 3, 5, 17, 257, and 65537. Finally, it is not difficult to show that a regular n-gon is always constructible when n has only Fermat primes to the first power as its

[6]Since

$$\left(1 - 2^j + 2^{2j} - 2^{3j} + \cdots \pm 2^{(k-1)j}\right) = \frac{(-1)^{k+1}2^{jk} + 1}{2^j + 1},$$

the number 2^{jk} is always composite for an odd number k. Therefore, a number of the form $2^s + 1$ can be prime only if the exponent s is a power of 2. However, $2^{32} + 1$ is not a prime, since it has 641 as a factor. Further details on numbers of the form $2^s + 1$ can be found, for example, in Paulo Ribenboim, *The Book of Prime Number Records*, New York, 1988, 2. VI.

odd prime divisors.[7] And the converse holds as well. Therefore, a regular n-gon is constructible if and only if the prime decomposition of n consists, aside from a possible power of 2, of only Fermat primes to the first power. Thus n-gons can be constructed for the following values of n: $2, 3, 4, 5, 6, 8, 10, 12, 15, 16, 17, 20, 24, 30, 32, 34, 40, \ldots$.

We note finally that an explicit derivation of the quadratic equations in the constructions of regular 257- and 65537-gons is not that difficult using a computer.[8] In both cases, 3 can again be used as a primitive root. Whether it then is worthwhile to derive an explicit construction method on the basis of the obtained quadratic equations is something greatly to be doubted. However, such was actually done in the nineteenth century.[9]

7.3 As a sort of acknowledgment of the construction of the regular pentagon, known since antiquity, we shall derive the construction algebraically here. Starting with the fifth root of unity $\zeta =$

---

[7]If m and n are relatively prime, then there exist—as can be calculated using the Euclidean algorithm—two integers a and b that satisfy the equation $an + bm = 1$. Since

$$a \cdot \frac{2\pi}{m} + b \cdot \frac{2\pi}{n} = \frac{2\pi}{nm},$$

the division of the circle into mn parts can be done given the divisions into m and n parts.

[8]Today it is hardly imaginable that the problem of computing the period products for the cyclotomic equation of degree 257 was the motivation for the author in 1975 to write his first computer program and to learn a computer language, in this case ALGOL 60. Since there was no direct access to a computer, the program was written down on paper and given to someone for input. Indeed, at the first run the desired indices of the periods that appear in the sum were computed correctly. This was much more intersting than the usual oral exam for a high-school diploma.

[9]F. J. Richelot, De resolutione algebraica aequationis $x^{257} = 1$, sive de divisione circuli per bisectionam anguli septies repetitam in partes 257 inter se aequales commentatio coronata, *Journal für die Reine und Angewandte Mathematik*, IX (1832), pp. 12–26, 146–161, 209–230, 337–356. Christian Gottlieb, The simple and straightforward construction of the regular 257-gon, *The Mathematical Intelligencer*, 21/1 (1999), pp. 31–37. Johann Gustav Hermes (1846–1912), of Lingen, is said, as reported by Felix Klein (*Vorträge über ausgewählte Fragen der Elementargeometrie*, Leipzig, 1895, p. 13), to have derived over the course of ten years a construction method for the regular 65537-gon. An overview of this work of over two hundred pages, completed in 1889 and deposited at the University of Göttingen, is given by J. Hermes, Ueber die Teilung des Kreises in 65537 gleiche Teile, *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen*, Math.-Phys. Klasse, 3 (1894), pp. 170–186. Three photographs of the work can be found in Hans-Wolfgang Henn, *Elementare Geometrie und Algebra*, Wiesbaden, 2003, pp. 33–34.

$\cos\left(\frac{2\pi}{5}\right) + i\sin\left(\frac{2\pi}{5}\right)$, we construct the two periods

$$\eta_0 = \zeta^1 + \zeta^4,$$
$$\eta_1 = \zeta^2 + \zeta^3.$$

On account of $\eta_0\eta_1 = -1$ and $\eta_0 + \eta_1 = -1$, one obtains these two periods from the quadratic equation

$$y^2 + y - 1 = 0.$$

This leads to

$$\cos\frac{2\pi}{5} = \Re\zeta = \frac{1}{2}\eta_0 = -\frac{1}{4} + \frac{1}{4}\sqrt{5},$$

where $\Re z$ denotes the real part of the complex number z, from which the construction can immediately be made. The four fifth roots of unity other than 1 are as follows:

$$-\frac{1}{4} + \frac{1}{4}\sqrt{5} + i\frac{1}{4}\sqrt{10 + 2\sqrt{5}},$$
$$-\frac{1}{4} - \frac{1}{4}\sqrt{5} + i\frac{1}{4}\sqrt{10 - 2\sqrt{5}},$$
$$-\frac{1}{4} - \frac{1}{4}\sqrt{5} - i\frac{1}{4}\sqrt{10 - 2\sqrt{5}},$$
$$-\frac{1}{4} + \frac{1}{4}\sqrt{5} - i\frac{1}{4}\sqrt{10 + 2\sqrt{5}}.$$

The Classical Construction Problems

The three famous problems of classical antiquity that remained unsolved into modern times are the *squaring of the circle*, the *doubling of the cube*, and the *trisection of an angle* using only straightedge and compass.

The problem of squaring the circle asks for the construction of a square whose area is equal to that of a given circle. On the assumption that the circle has area 1, the problem amounts to constructing a segment of length $\sqrt{\pi}$. Since one can construct square roots and squares of lengths, the problem is equivalent to the construction of a segment of length π. Thus the algebraic equivalent of the problem is to represent π in terms of nested square roots, rational numbers, and the four basic arithmetic operations. It was shown by Ferdinand Lindemann (1852–1939) in 1882 that the number π is *transcendent*, that is, that it satisfies no polynomial equation with

rational coefficients. Therefore, π cannot be represented in terms of square roots and hence is not constructible.[10]

The doubling of the cube, that is, the construction of a cube whose volume is twice that of the unit cube, amounts to the construction of a segment of length $\sqrt[3]{2}$. With the methods of Galois theory one can prove relatively easily that $\sqrt[3]{2}$ cannot be expressed in terms of rational numbers and nested square roots. We shall return to this topic in Chapter 10.

The resolution of the problem of angle trisection is of a similar nature. In dealing with the *casus irreducibilis* in Chapter 2 we saw the close relationship between angle trisection and cubic equations. In general, one has the identity

$$\cos^3\frac{\psi}{3} - \frac{3}{4}\cos\frac{\psi}{3} - \frac{1}{4}\cos\psi = 0.$$

The relationship with the problem of doubling the cube is seen more clearly when one brings the sine function into the picture, with

$$\left(\cos\frac{\psi}{3} + i\sin\frac{\psi}{3}\right)^3 = \cos\psi + i\sin\psi.$$

To show that a general trisection algorithm does not exist, it suffices to show that a single angle measure cannot be constructed. Of course, one can easily trisect the full circle of 360 degrees, as well as a right angle and a number of other special angles. However, the angle of 120° cannot be trisected, since otherwise, the regular nonagon would be constructible with straightedge and compass. We shall see more on this topic in Chapter 10.

7.4 The cyclotomic equation $x^n - 1 = 0$ has many interesting algebraic properties even for those values of n for which the regular n-gon is not constructible. Gauss himself recognized, as presented, along with many other results, in his 1801 work *Disquistiones arithmeticae*, that all cyclotomic equations are solvable in radicals. By this is not meant simply a "solution" of the form $x = \sqrt[n]{1}$, since such a symbol allows a number of algebraic interpretations. That is, the symbol offers interpretations that differ greatly as regards the four basic arithmetic operations. For example, the expression $\sqrt[4]{1}$ comprises the four complex numbers $1, -1, i, -i$, of which only i and $-i$ are indistinguishable on the basis of their algebraic properties alone. Thus 1 is uniquely defined as the multiplicative identity of the complex

---

[10] A relatively elementary discussion can be found in the very informative and well illustrated book by Jean-Paul Delahaye, *Le fascinant nombre Pi*, Editions Belin, Paris, 1997, Chapter 9.

numbers, while -1 is the unique additive inverse of 1. In contrast, i and $-i$ are characterized only as the two solutions of the equation $x^2 + 1 = 0$. One could also argue that an expression of the form $\zeta = -\frac{1}{2} + \frac{1}{2}i\sqrt{3}$ also has more than one interpretation. As with i and $-i$, this multiplicity relates only to numbers that possess the identical algebraic properties.

Thus one may conclude that the symbol $\sqrt[n]{a}$ can be used unproblematically only when the equation $x^n - a = 0$ is irreducible and therefore its solutions all possess identical algebraic properties. Therefore, the solution of an equation in radicals can be interpreted as a stepwise reduction to the solution of irreducible equations of the form $x^n - a = 0$.

That the cyclotomic polynomials $x^7 - 1 = 0$ and $x^9 - 1 = 0$ are solvable by radicals can be shown relatively easily by a method due to de Moivre. After removing the linear factor $(x-1)$, one can halve the remaining exponent (to 3 or 4) via the substitution $y = x+x^{-1}$. Then one can use the solution formula for a cubic or biquadratic equation. In detail, with the substitutions into the equations divided by x^3, respectively x^4, we have

$$x^3 + x^2 + x + 1 + x^{-1} + x^{-2} + x^{-3} = 0$$

and

$$x^4 + x^3 + x^2 + x + 1 + x^{-1} + x^{-2} + x^{-3} + x^{-4} = 0.$$

One makes the substitutions

$$\begin{aligned} x^2 + x^{-2} &= y^2 - 2, \\ x^3 + x^{-3} &= y^3 - 3y, \\ x^4 + x^{-4} &= y^4 - 4\left(y^2 - 2\right) - 6 = y^4 - 4y^2 + 2. \end{aligned}$$

After the unknown y is determined from the resulting third- or fourth-degree equation, the desired unknown x can be obtained from the quadratic equation

$$x^2 - yx + 1 = 0.$$

7.5 In the case of the cyclotomic equation of degree 11, namely $x^{11} - 1 = 0$, the process we have just described leads to the fifth-degree equation

$$y^5 + y^4 - 4y^3 - 3y^2 + 3y + 1 = 0,$$

whose five solutions are given by $y_j = 2\cos\frac{2\pi j}{11}$ for $j = 1, 2, 3, 4, 5$. That this equation, and therefore the cyclotomic equation of degree 11, can be solved in radicals was discovered before Gauss, in 1771, by Alexandre Théophile Vandermonde (*Mémoire sur la résolution des équations*). Like Lagrange, Vandermonde attempted to study the solution method for the general equation up to the fourth degree in order to generalize it to degree five. To this end, he used the resolvent, now named for Lagrange (see Chapter 5). Although Vandermonde failed to find a general solution formula—as of course he had to—he recognized that in "special cases in which there are equations among the roots," his "method could serve to solve the given equations without having to use the general solution formulas." In the equations mentioned by Vandermonde that exist among the solutions, there are identities for the periods such as

$$y_1^2 = y_2 + 2, \quad y_2^2 = y_4 + 2, \quad y_3^2 = y_5 + 2, \quad y_4^2 = y_3 + 2,$$
$$y_5^2 = y_1 + 2, \quad y_1y_2 = y_1 + y_3, \quad y_1y_3 = y_2 + y_4, \quad y_2y_3 = y_1 + y_5,$$

with which Vandermonde, without explicit reference, established the entire structure by which the solutions are sorted into the order

$$y_1, \quad y_2, \quad y_3, \quad y_4, \quad y_5,$$

exactly corresponding to the general method discovered by Gauss thirty years later using 2 as a primitive root modulo 11. In this order, for which we shall use the notation

$$\eta_k = P_f\left(\zeta^{2^k}\right) = \zeta^{2^k} + \zeta^{-2^k},$$

and so

$$\eta_0 = y_1, \quad \eta_1 = y_2, \quad \eta_2 = y_4, \quad \eta_3 = y_3, \quad \eta_4 = y_5,$$

Vandermonde was able to determine the fifth power of the Lagrange resolvent, and indeed in the form of a sum of integer multiples of fifth

roots of unity. In particular, for the Lagrange resolvent

$$z(\epsilon) = \eta_0 + \epsilon\eta_1 + \epsilon^2\eta_2 + \epsilon^3\eta_3 + \epsilon^4\eta_4$$

defined for a fifth root of unity $\epsilon = \cos\frac{2\pi k}{5} + i\sin\frac{2\pi k}{5}$ for $k = 1, 2, 3, 4$, one obtains on the one hand, as already discussed in Chapter 5,

$$y_1 = \eta_0 = \frac{1}{5}\left(-1 + \sqrt[5]{z(\epsilon)^5} + \sqrt[5]{z\left(\epsilon^2\right)^5} + \sqrt[5]{z\left(\epsilon^3\right)^5} + \sqrt[5]{z\left(\epsilon^4\right)^5}\right)$$

and, on the other hand, though only after a complicated calculation,

$$z(\epsilon)^5 = 11\left(6\epsilon + 41\epsilon^2 + 16\epsilon^3 + 26\epsilon^4\right).$$

From these last two equations one finally obtains, with the help of the already found square root representation for the fifth roots of unity, a root representation for the two-member period $y_1 = 2\cos\frac{2\pi}{11}$. It remains to note that the identity given for $z(\epsilon)^5$ can be derived in a completely elementary way from the $5^5 = 3125$ summands by sorting, grouping, and simplifying on the basis of the period identity discovered by Vandermonde (together with $\eta_0 + \eta_1 + \cdots + \eta_4 = -1$). Independently of the actual values of the result, one can see relatively easily that such a result can be found in the form of a sum of rational multiples of fifth roots of unity. And this could be the reason why Vandermonde sorted the solutions in the way described.

First of all, the sorting has the effect that every one of the period identities discovered by Vandermonde remains valid when each period η_k is replaced by η_{k+1} (taking into account the fact that the numbering of the periods is defined so that they progress cyclically, that is, $\eta_5 = \eta_0$, $\eta_6 = \eta_1$, etc.). Thus the identity

$$\left(\sum_{j=0}^{4}\epsilon^j\eta_j\right)^5 = \sum_{j=0}^{4}\sum_{k=0}^{4}a_{j,k}\epsilon^j\eta_k + \sum_{j=0}^{4}b_j\epsilon^j,$$

which *obviously* can be derived for $z(\epsilon)^5$ via simplification of the period identities with *any* integers $a_{j,k}$ and b_j, is valid also when every period η_k is replaced by η_{k+1}:

$$\left(\sum_{j=0}^{4}\epsilon^j\eta_{j+1}\right)^5 = \sum_{j=0}^{4}\sum_{k=0}^{4}a_{j,k}\epsilon^j\eta_{k+1} + \sum_{j=0}^{4}b_j\epsilon^j.$$

If the indices of the periods are shifted as well by 2, 3, and 4, then altogether one obtains

$$\left(\sum_{j=0}^{4}\epsilon^j\eta_j\right)^5+\cdots+\left(\sum_{j=0}^{4}\epsilon^j\eta_{j+4}\right)^5$$

$$=\sum_{j=0}^{4}\sum_{k=0}^{4}a_{j,k}\epsilon^j(\eta_k+\cdots+\eta_{k+4})+5\sum_{j=0}^{4}b_j\epsilon^j$$

$$=\sum_{j=0}^{4}\left(5b_j-\sum_{k=0}^{4}a_{j,k}\right)\epsilon^j.$$

Here every fifth summand on the left side of the last equation is equal to $z(\epsilon)^5$, since, for example,

$$\sum_{j=0}^{4}\epsilon^j\eta_{j+1}=\epsilon^{-1}\sum_{j=0}^{4}\epsilon^{j+1}\eta_{j+1}=\epsilon^{-1}z(\epsilon).$$

The left side of the previous equation is therefore equal to $5z(\epsilon)^5$, so that Vandermonde's result of a sum of rational multiples of fifth roots of unity is obvious in hindsight:

$$z(\epsilon)^5=\sum_{j=0}^{4}\left(b_j-\frac{1}{5}\sum_{k=0}^{4}a_{j,k}\right)\epsilon^j.$$

Without going into details, we note that there are ways in which the concrete calculation of $z(\epsilon)^5$ can be greatly simplified over the evaluation of 3125 summands.[11]

Although what we have presented applies specifically to the cyclotomic equaton $x^{11}-1=0$, it is not implausible that these considerations apply to every cyclotomic equation of prime degree. The reasons

[11] See Paul Bachmann, *Die Lehre von der Kreistheilung und ihre Beziehungen zur Zahlentheorie*, Leipzig, 1872 (reprint 1988), pp. 75–98. For the special case of the eleventh-degree cyclotomic equation, this method, which works in general, leads to a product representation

$$z(\epsilon)^5=\frac{z(\epsilon)z(\epsilon)}{z(\epsilon^2)}\cdot\frac{z(\epsilon)z(\epsilon^2)}{z(\epsilon^3)}\cdot\frac{z(\epsilon)z(\epsilon^3)}{z(\epsilon^4)}\cdot\left(z(\epsilon)z(\epsilon^4)\right),$$

where each of the four factors represents a sum of integer multiples of fifth roots of unity; the last is in fact an integer. Here the four products in the numerators can be similarly calculated in the general case, as was done for period products. Moreover, the fact that each of the four factors corresponds to a sum of integer multiples of fifth roots of unity can be shown in the same way as was done for $z(\epsilon)^5$.

that such is indeed the case can be found in Vandermonde's equations, which give general relationships between the periods, and the fact that these identities remain valid when the periods are permuted among themselves, as in replacing the root of unity $\zeta = \cos\left(\frac{2\pi k}{n}\right) + i \sin\left(\frac{2\pi k}{n}\right)$ by ζ^g, where g denotes a primitive root modulo n. This situation may be characterized as follows: In comparison to the solutions of the general equation of a given degree, the algebraic calculation of periods is "simpler," since they satisfy additional equations. This allows, in comparison to Lagrange's general procedure, the construction of simpler resolvents, which, although they do not remain unaltered under a permutation of the periods, do so under every permutation arising from the substitution $\zeta \mapsto \zeta^{g^k}$. This limited invariance suffices here to demonstrate, for example for $z(\epsilon)^5$, the possibility of a representation as the the sum of fifth roots of unity.

For the specific task of solving cyclotomic equations and from that eventually deriving a general construction procedure, what we have learned so far should suffice. However, only a general, up to now recognizable only in outline, principle would be completely satisfying from a mathematical point of view. That is precisely what Galois theory will offer us, after we have developed some additional concepts, namely an explanation whose argumentation will be much more homogeneous. Moreover, in the case of the cyclotomic equations, many of the complicated summation expressions for periods will become unnecessary.

Exercises

(1) Express all the seventeenth roots of unity in terms of square roots.

(2) Generalize the calculation for the case $n = 17$ of the product

$$P_{(n-1)/2}(\zeta) \cdot P_{(n-1)/2}\left(\zeta^g\right)$$

to the case of a general prime number $n \geq 3$. How can the necessary case distinction be most simply characterized?

Chapter 8

The Solution of Equations of the Fifth Degree

We seek the solution of the equation $x^5 = 2625x + 61500$.

8.1 This chapter closely follows a talk given in 1977 by the author at the Philips Contest for Young Scientists and Inventors, "Special equations of the fifth degree that are solvable in radicals." The equation presented above is again a classical example. Already in 1762, Leonhard Euler recognized from his studies of solvability of equations that this equation belongs to a class of fifth-degree equations that can be solved in radicals. Like other mathematicians of his time, Euler had attempted to extend the methods for equations of degree less than five to those of fifth degree. Even the mountain of formulas that resulted could not dampen Euler's optimism, for he wrote,

> One may conjecture with apparent certainty that with the correct approach to this elimination procedure, one would finally arrive at an equation of fourth degree. If the result were an equation of higher degree, then ... [the previously used intermediate value for representing the solutions] would itself contain roots of this degree, and that would seem to be unreasonable.

However, in his actual calculations, Euler had to trim his sails somewhat:

> However, since the large number of expressions makes this task so difficult that one cannot achieve any measure of success, it seems appropriate to develop some special cases that do not lead to such complex formulas.[1]

Euler refers to the intermediate results he used as "such values as shorten the calculations." In reality, Euler has avoided not merely calculational difficulties, but the basic impossibility of a general solution. Nonetheless, in this way he arrives at a large class of fifth-degree equations that can be solved in radicals. Since this class does not contain all solvable fifth-degree equations, we will look here at the work of another mathematician. In 1771, thus at almost the same time as the work of Lagrange and Vandermonde, the Italian mathematician Giovanni Francesco Malfatti (1731–1807) was searching for a general formula for equations of the fifth degree. Malfatti, who later, in 1804, commented critically on Ruffini's first attempts at an unsolvability proof based on his own work and thereby motivated Ruffini to refine his work, succeeded in carrying out extremely complicated calculations of a resolvent of the sixth degree. This did not lead to the original goal of a general solution. However, Malfatti noticed that in the special case in which the sixth-degree resolvent possesses a rational solution, the given fifth-degree equation can be solved. Later, using Galois theory, it could be shown that Malfatti had characterized all equations of the fifth degree that are solvable in radicals (in relation to all irreducible fifth-degree polynomials over the rational numbers).

Malfatti's computations are very complicated, and it is very much worth noting that he continued successfully from the point at which Euler had not been able to progress.[2] To get some idea of Malfatti's method of attack, we will consider his calculation, beginning with the

---

[1]Von der Auflösung der Gleichungen aller Grade, reprinted in: Leonhard Euler, *Drei Abhandlungen über die Auflösung der Gleichungen*, Ostwalds Klassiker Nr. 226, Leipzig, 1928. This quotation and the one following appear on page 45; the equation in the epigraph appears on page 50.

[2]See J. Pierpont, Zur Geschichte der Gleichung V. Grades (bis 1858), *Monatshefte für Mathematik und Physik*, 6 (1895), pp. 15–68. Malfatti's attempts at a solution are described on pages 33 through 36.

equation

$$x^5 + 5ax^3 + 5bx^2 + 5cx + d = 0,$$

only for the case $a = b = 0$, that is, for equations of the type

$$x^5 + 5cx + d = 0.$$

Furthermore, we will assume $cd \neq 0$. We should note further that this does not restrict the generality as much as it seems at first glance. In fact, every equation of degree five can be transformed into an equation of this type using a substitution that eliminates the degree-four term. See the section on the transformations of Tschirnhaus and of Bring and Jerrard.[3]

Malfatti's calculations begin with the assumption, without loss of generality, that the solutions are represented in the form

$$x_{j+1} = -\left(\epsilon^j m + \epsilon^{2j} p + \epsilon^{3j} q + \epsilon^{4j} n\right),$$

for $j = 0, 1, 2, 3, 4$ and with $\epsilon = \cos\left(\frac{2\pi}{5}\right) + i \sin\left(\frac{2\pi}{5}\right)$. This corresponds precisely to the method employed already by Bézout, Euler, Lagrange,[4] and Vandermonde. If one multiplies the five associated linear factors together, then one obtains, along with Euler, the equation

$$\begin{aligned} x^5 &- 5(mn + pq)x^3 + 5\left(m^2 p + n^2 q + mp^2 + nq^2\right) x^2 \\ &- 5\left(m^3 p + n^3 q + mq^3 + np^3 - m^2 n^2 + mnpq - p^2 q^2\right) x \\ &+ m^5 + n^5 + p^5 + q^5 + (mn - pq)\left(mp^2 + nq^2 - m^2 q - n^2 p\right) = 0. \end{aligned}$$

Finally, one must try to determine the unknowns m, n, p, q by comparing the coefficients with the original equation. We will employ the following shorthand:

$$\begin{aligned} y &= pq = -mn, \\ r &= m^2 q + n^2 p = -\left(mp^2 + nq^2\right), \\ v &= m^3 p + n^3 q, \\ w &= mq^3 + np^3. \end{aligned}$$

---

[3]For specific applications, however, it is unfortunate that equations with rational coefficients are not transformed into equations of the same type.

[4]Since $m = -\left(x_1 + \epsilon^4 x_2 + \epsilon^3 x_3 + \epsilon^2 x_4 + \epsilon x_5\right)/5$, etc., at issue here are Langrange resolvents for the values m^5, p^5, q^5, n^5.

The two identities mentioned together with the definition of the quantities y and r already contain the result of comparing coefficients for the powers x^3 and x^2. For the other two powers, comparing coefficients gives the pair of equations

$$c = -v - w + 3y^2,$$
$$d = m^5 + n^5 + p^5 + q^5 + 20ry.$$

To be able to formulate as well the last-introduced identity completely in terms of r, v, w, y, we use the relations

$$\begin{aligned} rv &= \left(m^2q + n^2p\right)\left(m^3p + n^3q\right) \\ &= pq\left(m^5 + n^5\right) + (mn)^2\left(mp^2 + nq^2\right) \\ &= \left(m^5 + n^5\right)y - ry^2, \\ rw &= -\left(mp^2 + nq^2\right)\left(mq^3 + np^3\right) \\ &= -mn\left(p^5 + q^5\right) - (pq)^2\left(m^2q + n^2p\right) \\ &= \left(p^5 + q^5\right)y - ry^2, \end{aligned}$$

thereby obtaining for the pair of equations the new form

$$c = -(v + w) + 3y^2,$$
$$dy = r(v + w) + 22ry^2.$$

A calculation of the four unknown quantities r, v, w, y will be possible only if two additional identities are taken into account:

$$\begin{aligned} vw &= \left(m^3p + n^3q\right)\left(mq^3 + np^3\right) \\ &= pq\left(m^4q^2 + n^4p^2\right) + mn\left(m^2p^4 + n^2q^4\right) \\ &= pq\left(m^2q + n^2p\right)^2 + mn\left(mp^2 + nq^2\right)^2 - 4m^2n^2p^2q^2 \\ &= yr^2 + (-y)(-r)^2 - 4y^4 = -4y^4 \end{aligned}$$

and

$$\begin{aligned} -r^2 &= \left(m^2q + n^2p\right)\left(mp^2 + nq^2\right) \\ &= pq\left(m^3p + n^3q\right) + mn\left(mq^3 + np^3\right) = (v - w)y. \end{aligned}$$

Putting these two identities together, we obtain

$$r^4 = (v - w)^2y^2 = (v + w)^2y^2 - 4vwy^2 = (v + w)^2y^2 + 16y^6.$$

Now, using this equation and the pair of equations previously obtained from comparing coefficients, we may determine the values r, v, w, y. First, we eliminate $v + w$ via

$$v + w = 3y^2 - c,$$

so that the following equations remain:

$$dy = (25y^2 - c)\, r,$$
$$r^4 = 25y^6 - 6cy^4 + c^2y^2.$$

To eliminate the variable r as well, we take the fourth power of the first of these two equations and then substitute the second equation into the result to obtain

$$d^4y^4 = (25y^2 - c)^4 \left(25y^4 - 6cy^2 + c^2\right) y^2.$$

Our exclusion of the special case $cd = 0$ helps us in what follows to avoid some complications: First, we have $y \neq 0$, since otherwise, at least three of the values m, n, p, q would be equal to zero, resulting in $c = 0$. Furthermore, we would also have $25y^2 - c \neq 0$, since otherwise we must have $y = 0$.

From $y \neq 0$, we can now multiply the last equation by $25y^{-2}$. We then substitute $z = 25y^2$, so that a *bicubic resolvent* results, that is, an equation of the sixth degree:

$$(z - c)^4 \left(z^2 - 6cz + 25c^2\right) = d^4z.$$

As we shall see, it is sometimes useful to use the bicubic resolvent in the equivalent form

$$\left(z^3 - 5cz^2 + 15c^2z + 5c^3\right)^2 = \left(d^4 + 256c^5\right) z.$$

Of course, in its general form, the bicubic resolvent cannot be solved in radicals. If it were, then beginning with the variable z, the

values y, r, v, w, m, n, p, q could then be calculated in turn:

$$y = \frac{1}{5}\sqrt{z},$$

$$r = \frac{dy}{25y^2 - c},$$

$$v = \frac{3y^3 - cy - r^2}{2y},$$

$$w = \frac{3y^3 - cy + r^2}{2y},$$

$$m, n = \sqrt[5]{\frac{v + y^2}{2y}r \pm \sqrt{\left(\frac{v + y^2}{2y}r\right)^2 + y^5}},$$

$$p, q = \sqrt[5]{\frac{w + y^2}{2y}r \pm \sqrt{\left(\frac{w + y^2}{2y}r\right)^2 - y^5}}.$$

Each equation comes almost directly from the previously derived identities, in the case of the last two equations with the help of Viète's root theorem. Note that the sign of the unknown y can be chosen arbitrarily, since changing the sign merely exchanges the pairs (p, q) and (m, n). Furthermore, note that the ordering of the variables p, q, m, n is always taken such that the equation $v = m^3p + n^3q$ is satisfied.

8.2 Malfatti himself recognized that the bicubic resolvent that he obtained can be used to solve special equations of the fifth degree in radicals. In particular, this is possible when a rational solution to the bicubic resolvent can be found. Here we shall take as an example the equation in the epigraph to this chapter with the coefficients $c = -525$ and $d = -61500$.

Since the bicubic resolvent is a monic polynomial with integer coefficients, all rational solutions, as demonstrated in Chapter 6, must be integers dividing the number $25c^6$. One obtains additional information from the second representation of the bicubic resolvent: Since $d^4 + 256c^5 = 3780900000^2$ is a square, every rational solution must be the square of an integer. And finally, division by 5^6 shows that $\frac{z}{5}$ is also a solution of an equation with integer coefficients, that is, that z is divisible by 5. Having limited the number of possible integer

solutions to 112, one obtains the solution $z = 5625$. It then turns out that $y = 15$, $r = -150$, $v = -150$, $w = 1350$, and finally, for $j = 0, 1, 2, 3, 4$,

$$x_{j+1} = \epsilon^j \sqrt[5]{75\left(5+4\sqrt{10}\right)} + \epsilon^{2j} \sqrt[5]{225\left(35-11\sqrt{10}\right)}$$
$$+ \epsilon^{3j} \sqrt[5]{225\left(35+11\sqrt{10}\right)} + \epsilon^{4j} \sqrt[5]{75\left(5-4\sqrt{10}\right)}.$$

8.3 Malfatti's attempt at a solution shows a methodology in the finest classical tradition, namely, to solve equations using suitable substitutions and transformations. In hindsight, we see that the success of Malfatti's approach, to the extent that success was possible, is clarified if one expresses the relevant intermediate values as polynomials in the solutions $x_1, \ldots, x_5$. Thus from the two identities

$$p = -\frac{1}{5}\left(x_1 + \epsilon^2 x_2 + \epsilon^4 x_3 + \epsilon x_4 + \epsilon^5 x_5\right)$$

and

$$q = -\frac{1}{5}\left(x_1 + \epsilon^3 x_2 + \epsilon x_3 + \epsilon^4 x_4 + \epsilon^2 x_5\right)$$

one obtains

$$25y = 25pq = \sum_{j=1}^{5} x_j^2 + \left(\epsilon^2+\epsilon^3\right)\left(x_1x_2 + x_2x_3 + x_3x_4 + x_4x_5 + x_5x_1\right)$$
$$+ \left(\epsilon+\epsilon^4\right)\left(x_1x_3 + x_2x_4 + x_3x_5 + x_4x_1 + x_5x_2\right).$$

In the special case considered here, $a = b = 0$, since we have

$$\sum_{j=1}^{5} x_j = \sum_{1 \le j < k \le 5} x_j x_k = 0$$

and $-\epsilon+\epsilon^2+\epsilon^3-\epsilon^4=-\sqrt{5}$, we obtain for the resolvent solution z the particularly simple representation[5]

$$z=25y^2=\frac{1}{5}(x_1x_2+x_2x_3+x_3x_4+x_4x_5+x_5x_1)^2.$$

Furthermore, with this representation it is clear that in the sense of Vandermonde, the existence of a rational solution of the bicubic resolvent can be interpreted as a relation between the solutions.

---

[5]A derivation of the bicubic resolvent based on Lagrange's universal approach (see Chapter 5) can be found in C. Runge, Über die auflösbaren Gleichungen der Form $x^5+ux+v=0$, *Acta Mathematica*, 7 (1885), pp. 173–186; see also Heinrich Weber, *Lehrbuch der Algebra*, volume I, Braunschweig, 1898, pp. 670–676: One first investigates the behavior of the slightly altered polynomial representation

$$y=\frac{\sqrt{5}}{50}(x_1x_2+x_2x_3+x_3x_4+x_4x_5+x_5x_1-x_1x_3-x_2x_4-x_3x_5-x_4x_1-x_5x_2)$$

under the 120 possible permutations of the five solutions $x_1,\ldots,x_5$. Ten of these permutations leave the polynomial unchanged. All of these are *even* permutations; that is, they belong to the collection of sixty permutations that leave unchanged the square root of the discriminant:

$$\sqrt{D}=\prod_{i<j}(x_i-x_j).$$

Furthermore, there are ten *odd* permutations whose effect on the polynomial y is to change its sign. Thus the sixty even permutations transform the polynomial y into six different polynomials $y_1=y,y_2,\ldots,y_6$, and the sixty odd permutations transform y into an additional six polynomials, namely $y_7=-y_1,\ldots,y_{12}=-y_6$. The first six polynomials are thus solutions of the sixth-degree equation

$$y^6+\lambda_5y^5+\cdots+\lambda_1y+\lambda_0=0,$$

whose coefficients $\lambda_0,\ldots,\lambda_5$ arise from the elementary symmetric polynomials in the polynomials $y_1,\ldots,y_6$. To obtain these coefficients in terms of c and d of the original equation $x^5+5cx+d=0$, the polynomials $y_1,\ldots,y_6$ are expressed in terms of the solutions $x_1,\ldots,x_5$. However, the resulting polynomials are only "almost" symmetric; namely, the polynomials of even degree (in the variables $y_1,\ldots,y_6$) are symmetric, while those of odd degree are altered by a sign change for odd permutations and are unchanged by even permutations. Using the fundamental theorem on symmetric functions and considering the degrees of $c,d,\sqrt{D},\lambda_0,\ldots,\lambda_5$ as polynomials in the variables $x_1,\ldots,x_5$ (namely 4, 5, 10, and $12-2j$ for λ_j), there must exist rational numbers μ_0,μ_1,μ_2,μ_4 satisfying

$$y^6+\mu_4cy^4+\mu_2c^2y^2+\mu_0c^3=\mu_1\sqrt{D}y.$$

After determining the constants, one finally obtains, after squaring the equation obtained, the form of the bicubic resolvent derived by a different route in the main text; here one determines $\sqrt{D}$ by observing that the discriminant D must be representable as a symmetric polynomial of degree 20 of the form $\alpha c^5+\beta d^4$ with two constants α and β, where the constants can be found using particular equations. One finally obtains $D=5^5\left(256c^5+d^4\right)$.

The Transformations of Tschirnhaus and of Bring and Jerrard

The first systematic attempt at a general solution method for equations of degree five was undertaken in 1683 by Ehrenfried Walther, Count of Tschirnhaus (1651–1708). Tschirnhaus's idea is based on the hope that one could generalize the well-known substitutions that cause the second-highest coefficient to disappear so that additional coefficients would disappear as well.

Instead of transforming a given equation

$$x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_1x + a_0 = 0$$

using the substitution

$$x = y - \frac{a_{n-1}}{n}$$

into an equation of reduced form

$$y^n + b_{n-2}y^{n-2} + \cdots + b_1y + b_0 = 0,$$

Tschirnhaus began his investigations with a substitution of the form

$$y = x^2 + px + q$$

with parameters p and q to be determined. The n solutions $x_1, \ldots, x_n$ of the original equation are transformed into the n solutions $y_1, \ldots, y_n$ with $y_j = x_j^2 + px_j + q$, where the coefficients of the powers of y^{n-1} and y^{n-2} are both zero precisely when the two conditions

$$\sum y_j = \sum y_j^2 = 0$$

are satisfied. If one starts with a reduced equation in which the coefficient of the second-highest power is already 0, then one obtains for the parameters p and q the following conditions that must be satisfied:

$$\begin{aligned} 0 = \sum y_j &= \sum \left(x_j^2 + px_j + q\right) = \sum x_j^2 + p\sum x_j + nq \\ &= \sum x_j^2 + nq, \\ 0 = \sum y_j^2 &= \sum \left(x_j^2 + px_j + q\right)^2 \\ &= \sum x_j^4 + 2p\sum x_j^3 + \left(p^2 + 2q\right)\sum x_j^2 + nq^2. \end{aligned}$$

The first of the two conditions immediately permits a unique determination of the parameter q. If one then substitutes the obtained value for q into the second condition, then one obtains for the parameter p a quadratic equation (except in the special case in which the coefficient of the third-highest power is already zero). Thus the so-called *Tschirnhaus*

transformation of a given nth-degree equation can always be parameterized such that the resulting equation has coefficients equal to zero for the powers y^{n-1} and y^{n-2}.

Tschirnhaus now believed that using transformations of higher degree, which of course contain more parameters to be chosen, would allow further simplification of the equations, so that every equation could be solvable in radicals. Although Tschirnhaus did not succeed in supporting his idea with concrete calculations, it is nevertheless possible to use a transformation of the form

$$y = x^4 + px^3 + qx^2 + rx + s$$

for his special case of a fifth-degree equation

$$x^5 + a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0 = 0,$$

resulting in an equation of the form

$$y^5 + b_1y + b_0 = 0.$$

The parameters can be determined by solving a cubic and a quardratic equation. This fact was first discovered in 1786 by the Swedish mathematician Erland Samuel Bring (1736–1798), though without the mathematical world taking proper note of his achievement. Only much later, in 1864, after George Birch Jerrard (1804–1863) had rediscovered the transformation, were Bring's investigations recalled. The transformation is today generally called the *Bring–Jerrard transformation*. However, its details are so complicated that the actual calculations are difficult to carry out.[6]

Literature on Equations of the Fifth Degree

R. Bruce King, *Behind the Quartic Equation*, Boston, 1996.

Samson Breuer,[7] *Über die irreduktiblen auflösbaren trinomischen Gleichungen fünften Grades*, Borna-Leipzig, 1918.

Sigeru Kabayashi, Hiroshi Nakagawa, Resolution of equation, *Math. Japonica*, 5 (1992), pp. 882–886.

[6] A description of the Bring–Jerrard transformation can be found in J. Pierpont, Zur Geschichte der Gleichung V. Grades (bis 1858), *Monatshefte für Mathematik und Physik*, 6 (1895), pp. 18–19.

[7] The sad fate of the victims of racial and political persecution demands that we recall here the 1933 expulsion of Samson Breuer (1891–1978). See Reinhard Siegmund Schultze, *Mathematiker auf der Flucht vor Hitler*, Braunschweig, 1998, pp. 109, 292.

Daniel Lazard, Solving quintics in radicals, in: Olav Arnfinn Laudal, Ragni Piene, *The Legacy of Niels Henrik Abel*, Berlin, 2004, pp. 207–225.

Blair K. Spearman, Kenneth S. Williams, Characterization of solvable quintics x^5+ax+b, *American Mathematical Monthly*, 101 (1994), pp. 986–992.

Blair K. Spearman, Kenneth S. Williams, On solvable quintics $X^5 + aX + b$ and $X^5 + aX^2 + b$, *Rocky Mountain Journal of Mathematics*, 26 (1996), pp. 753–772.

Exercises

(1) Solve the equation

$$x^5 + 15x + 12 = 0.$$

(2) Solve the equation

$$x^5 + 330x - 4170 = 0.$$

Chapter 9

The Galois Group of an Equation

How can one tell whether an equation of the fifth or higher degree is solvable in radicals?

9.1 The question thus formulated is a natural continuation of our previous results: if there is no solution to the general equation, what types of special equations are solvable in radicals? This question was answered by the twenty-year-old French mathematician Évariste Galois in 1832, shortly before his death in a duel.[1]

Galois, who grew up in the post-Napoleon restoration, appears to have studied the solution of equations in radicals entirely as a self-taught mathematician, although he did obtain a good education for the time, first at the Collège Louis-le-Grand, in Paris, and then at the École Préparatoire, later the École Normale. However, he twice failed the entrance examination for the École Polytechnique, and he was initially refused entrance to the École Préparatoire in early 1831 on account of his republican agitation. His membership in the Republican Guard later resulted in several months' imprisonment.

[1]The dramatic circumstances of Galois's discovery and the mysterious duel have frequently led to a romanticization of his life. An example of this is the novel by Tom Pertsinis, *The French Mathematician*, 1997. Those more interested in the cold facts should look at the article by A. Rothman: The short life of Évariste Galois, *Scientific American*, April 1982, pp. 112–120. See also T. Rothman: Genius and biographers: the fictionalization of Évariste Galois, *Amer. Math. Monthly* 89 (1982), 84–106, or the biography by Laura Toti Rigatelli, *Évariste Galois 1811–1832*, Birkhäuser, 1996 (Italian original, 1993).

Galois's attempts to publish his ideas failed due to the lack of understanding by the reviewers, resulting in part from the terse presentation, only much later recognized as correct. The first significant publication occurred only fourteen years after Galois's death, on the recommendation of Joseph Liouville (1809–1882).

Galois's thinking begins with the then current state of knowledge, which corresponds more or less with the content of the previous chapter. From that point, his interest may have been awakened to ask to what extent relations based on polynomial identities among the solutions arise that reduce the complexity of the equation in comparison to the normal case. Thus we have seen in Chapter 7 how Gauss and Vandermonde solved cyclotomic equations using such relationships. And Lagrange, too, whose work has been described in this book relatively briefly, investigated polynomial expressions in the solutions not only for the case of the general equation, but also for special equations.

Galois's central idea, for which there was no precedent at the time, but which later proved to be extremely fruitful in application to other problems in mathematics, consists in going beyond the current state of investigation by studying a characteristic object of greater simplicity.[2] Specifically, for each equation, Galois associated a mathematical object called a *group* in general, and for an equation the *Galois group*, in honor of Galois. The Galois group consists of a subset of the permutations of the solutions together with the operation of composition of permutations, as described in Chapter 5. The usefulness of this association is that it is possible to classify Galois groups in such a way that offers a classification of equations with respect to their solvability. In particular:

- All important properties of a given equation—irreducibility, solvability in radicals, and in the case of solvability the degree of the required root operations—can be determined without reference to the equation from properties of the Galois group.

---

[2] A telling example from a different area of mathematics is that of *knots*. For an elementary introduction, see Alexei Sossinsky, *Knots: Mathematics with a Twist*, Cambridge, Harvard University Press, 2004; Lee Neuwirth: The theory of knots, *Scientific American* 240 (June, 1979), pp. 110–124. Less-spectacular examples can be found in almost every mathematical area.

- Moreover, the number of different Galois groups is much smaller than the number of possible equations. Thus one can gain a complete understanding of all the Galois groups associated with equations of low degree.

9.2 As stated in the preface to this book, we are going to interest ourselves in the "modern point of view," that is, in Galois theory, as developed in the early twentieth century.[3] However, we first shall define the Galois group in an "elementary" way using the terminology that we have developed thus far. We shall follow Galois's path in broad outline, though without going into detail.[4] Furthermore, we are not going to offer complete proofs of anything, which would make little sense given the development thus far, which has been primarily of a motivational nature. Instead, we shall offer some concrete examples. In the next chapter we will return to fill in some of the gaps in this chapter's presentation.

In their analysis of the conditions for a general solution of equations of a particular degree, Abel, and before him Ruffini, had focused on the root operations, dealing apparently with the most striking places within a solution formula. Galois recognized how this procedure designed for the general equation could be applied to the solution of special equations and to the desired representation of their roots. To this end, he called a quantity *known* if it can be represented in terms of already known quantities using the four arithmetic operations. One begins by considering the coefficients of the equation as known, in analogy to the general equation, whose coefficients correspond to the elementary symmetric polynomials. One can enlarge the collection of known quantities by adding certain values, in particular, but not exclusively, the roots of already known quantities. Galois called such numbers added to the collection of known quantities *adjoined quantities*. The process itself he called *adjunction*. We

[3] How Galois theory has developed and changed over the years is well described in B. Melvin Kiernan, The development of Galois theory from Lagrange to Artin, *Archive for History of Exact Sciences*, 8 (1971/1972), pp. 40–154, as well as in an annotated revision of B. L. van der Waerden, Die Galois-Theorie von Heinrich Weber bis Emil Artin, *Archive for History of Exact Sciences*, 9 (1972), pp. 240–248.

[4] An extensively commented translation of Galois's original paper can be found in Harold M. Edwards, *Galois Theory*, New York, 1984. An overview of the history can be found in Erhard Scholz, Die Entstehung der Galois-Theorie, in: Erhard Scholz (ed.), *Geschichte der Algebra*, Mannheim, 1990, pp. 365–398.

note that Galois's concept of known quantities leads to sets of numbers that today are called *fields*, a concept about which we shall have more to say in the following chapter. However, since we would like to use this notion, we give the following definition.

Definition 9.1. A subset of the complex numbers is called a *field* if it is closed under the four arithmetic operations, that is, if the sum, difference, product, and quotient (aside from division by zero) of any two elements (not necessarily distinct) of the field is again in the field.[5]

The smallest collection of known quantities that results from an equation with rational coefficients is the field of rational numbers, denoted by $\mathbb{Q}$. Then building on this field, the solutions of an equation, such as, for example,

$$x^3 - 3x - 4 = 0,$$

represent a stepwise enlargement of the field of known quantities. Thus for the solution

$$x_1 = \sqrt[3]{2+\sqrt{3}} + \sqrt[3]{2-\sqrt{3}}$$

we first adjoin $\sqrt{3}$ to the rational numbers. We thereby obtain as our collection of known quantities resulting from the rational numbers with the adjunction of $\sqrt{3}$ the set

$$\mathbb{Q}(\sqrt{3}) = \left\{ a + b\sqrt{3} \mid a, b \in \mathbb{Q} \right\},$$

which it is easily shown to be closed under the four basic operations, so that $\mathbb{Q}\left(\sqrt{3}\right)$ is in fact a field. One speaks of this field as an *extension field* of the rational numbers obtained by adjoining to $\mathbb{Q}$ the number $\sqrt{3}$.

Now to obtain the solution x_1, one needs to adjoin only $\sqrt[3]{2+\sqrt{3}}$ as a second step, since then

$$\sqrt[3]{2-\sqrt{3}} = \frac{1}{\sqrt[3]{2+\sqrt{3}}}$$

---

[5]The notion of a field is usually defined in greater generality to include sets that are not subsets of the complex numbers. However, for our purposes, the definition given here will suffice.

is also a known quantity. The result is the extension field denoted by $\mathbb{Q}\left(\sqrt{3}, \sqrt[3]{2+\sqrt{3}}\right)$.[6]

9.3 Building on the terminology of known quantities, we may approach the definition of the central notion of Galois group by considering an equation, namely, the nth-degree equation

$$x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_1 x + a_0 = 0$$

with complex coefficients $a_{n-1}, \ldots, a_1, a_0$ and without multiple solutions. That is, we assume that all the n solutions are distinct.[7] The solution process of the equation, to the extent that it is possible, is now analyzed by associating with an arbitrary intermediate step a field K of "known quantities" that contains the coefficients $a_{n-1}, \ldots, a_1, a_0$.

We noted at the beginning of the chapter that an equation can be represented more simply than in the general case if there are polynomial relations among the solutions $x_1, \ldots, x_n$. Every such relation corresponds to a polynomial whose value is zero for the arguments $x_1, \ldots, x_n$. Thus, for example, the polynomial relation

$$x_1^2 = x_2 + 2$$

corresponds to the polynomial

$$h(X_1, \ldots, X_n) = X_1^2 - X_2 - 2,$$

---

[6] However, this extension field does not contain the other two solutions of the cubic equation. To get all three solutions into an extension field with the second adjunction, one could start with the field $\mathbb{Q}(\frac{1}{2} + \frac{1}{2}i\sqrt{3})$, which contains the cube roots of unity.

[7] For the case considered here of complex coefficients, this condition is not really a restriction, since multiple linear factors can be eliminated by factoring them out using only the four basic operations. This is done with the help of the Euclidean algorithm, described later, by finding the greatest common divisor of the given polynomial and its derivative. Since

$$\left((x-x_1)^j(x-x_2)^k\cdots\right)' = (x-x_1)^{j-1}(x-x_2)^{k-1}\times\cdots$$
$$\times\left(j(x-x_2)(x-x_3)\cdots + k(x-x_1)(x-x_3)\cdots + \cdots\right),$$

the greatest common divisor of the given polynomial and its derivative is equal to

$$(x-x_1)^{j-1}(x-x_2)^{k-1}\cdots.$$

Through dividing the original polynomial by this greatest common divisor, one thereby obtains a polynomial that has the same set of roots as the original polynomial, but with no multiple roots. The first person to employ such considerations in explicit transformations of polynomials was Jan Hudde (1628–1704), later the mayor of Amsterdam.

where here and in what follows we use lowercase letters for the solutions and uppercase letters for the variables of a polynomial. For each underlying field K, we denote by B_K the set of polynomials with coefficients in K that have the value 0 at the arguments $x_1, \ldots, x_n$.

Of course, the totality of all polynomials in B_K is much too great for a detailed listing. Even a complete description is not a trivial task. Galois himself took a route by which he used only a single polynomial created specifically for his purposes. He constructed this polynomial using what is now called the *Galois resolvent*, which is a special quantity in terms of which all the solutions $x_1, \ldots, x_n$ can be expressed using the four basic operations. We will go into this very explicit approach, though not in great detail, in the section on computing the Galois group (at the end of this chapter). Here it will suffice to note that the Galois resolvent can be sufficiently characterized without explicitly calculating the solutions.

Of course, there are always polynomials that obviously belong to the defined set B_K. Such examples can most easily be found among the symmetric polynomials. For the earlier example

$$x^3 - 3x - 4 = 0,$$

the three polynomials

$$X_1 + X_2 + X_3, \quad X_1X_2X_3 - 4, \quad X_1^2 + X_2^2 + X_3^2 - 6$$

belong to the set $B_{\mathbb{Q}}$. However, what is really of interest are the nonsymmetric polynomials, since only they reflect relations that allow for a reduction in the complexity of the equation. For Vandermonde's equation (see Chapter 7)

$$x^5 + x^4 - 4x^3 - 3x^2 + 3x + 1 = 0,$$

with solutions $x_{j+1} = 2\cos\left(\frac{2\pi 2^j}{11}\right)$ for $j = 0, 1, 2, 3, 4$, here are some polynomials that belong to $B_{\mathbb{Q}}$:

$$X_1^2 - X_2 - 2, \qquad X_2^2 - X_3 - 2, \qquad X_3^2 - X_4 - 2, \qquad \ldots,$$
$$X_1X_2 - X_1 - X_4, \quad X_2X_3 - X_2 - X_5, \quad \ldots.$$

9.4 While the set of polynomials B_K may seem rather abstract and not easily grasped, it must be made clear from the start that the set is even larger for less-complex equations, that is, those with particularly

many relations among the solutions. The set B_K is thus a sort of measure of the complexity of the underlying equation. A truly simple characterization of this complexity is obtained with the help of the Galois group, which by definition contains all permutations of the n variables $X_1, \ldots, X_n$ that transform a polynomial in B_K to another polynomial in the set. This leads to the following definition.

Definition 9.2. For a polynomial equation without multiple solutions whose coefficients lie in a field K, the *Galois group* (over the field K) is the set of all permutations σ in the symmetric group S_n that permute the indices $1, \ldots, n$ of the solutions $x_1, \ldots, x_n$ in such a way that for every polynomial $h(X_1, \ldots, X_n)$ with coefficients in K and $h(x_1, \ldots, x_n) = 0$, one has $h\big(x_{\sigma(1)}, \ldots, x_{\sigma(n)}\big) = 0$.

In the case that no nontrivial relations, that is, those not based on symmetric polynomials, exist, the Galois group consists of all $n!$ permutations, and indeed, every polynomial in the set B_K remains unchanged under all permutations. In contrast, the example given earlier of an equation of fifth degree first solved by Vandermonde leads to a drastically reduced set of only five permutations, whereby in this case the individual polynomials in the set $B_{\mathbb{Q}}$ are altered by the permutations, with only the value 0 resulting from evaluation at the solutions remaining unchanged. For example, the cyclic permutation $X_1 \to X_2 \to X_3 \to X_4 \to X_5 \to X_1$ transforms the polynomial $X_1^2 - X_2 - 2$ into the polynomial $X_2^2 - X_3 - 2$, which, however, again belongs to the set $B_{\mathbb{Q}}$. On the other hand, simply permuting indices 1 and 2 does not lead to an element of the Galois group, as one can see immediately by investigating the polynomial $X_1^2 - X_2 - 2$ and the result of the indicated permutation. The resulting polynomial, $X_2^2 - X_1 - 2$ does not belong to $B_{\mathbb{Q}}$, since $x_2^2 - x_1 - 2 = x_3 - x_1 \neq 0$.

Another example, one to which we shall return several times, is the biquadratic equation

$$x^4 - 4x^3 - 4x^2 + 8x - 2 = 0.$$

Without going into detail, we note first that the four solutions satisfy the identity $x_1x_3 + x_2x_4 = 0$, where the numbering of the solutions is explained in the section on computing the Galois group. In view of the solution procedure described in Chapter 3, one should keep

in mind that such an identity exists only when the cubic resolvent possesses a rational solution.

Since $x_1x_4 + x_2x_3 \neq 0$ and $x_1x_2 + x_3x_4 \neq 0$, only those permutations that leave the polynomial $X_1X_3 + X_2X_4$ can belong to the Galois group. The result is that sixteen of the twenty-four ($= 4!$) permutations of the numbers $1, 2, 3, 4$ are eliminated as candidates for membership in the Galois group. That the remaining eight permutations in fact "respect" *every* relationship among the solutions and therefore belong to the Galois group is demonstrated in the above-mentioned section, where it is shown how one can check whether a permutation belongs to the Galois group using a *single* polynomial in $B_{\mathbb{Q}}$, namely

$$(-X_2 + X_3 - 2X_4)^8 + 16(-X_2 + X_3 - 2X_4)^7 - \cdots \\ - 253184(-X_2 + X_3 - 2X_4) + 72256.$$

Here we shall content ourselves with an explicit enumeration of the permutations that belong to the Galois group. The following table shows how each of the eight permutations permutes the indices $1, 2, 3, 4$. The first permutation, here denoted by σ_0, is the identity, that is, the permutation that leaves every index unchanged:

	1	2	3	4
σ_0	1	2	3	4
σ_1	3	2	1	4
σ_2	1	4	3	2
σ_3	3	4	1	2
σ_4	2	1	4	3
σ_5	4	1	2	3
σ_6	2	3	4	1
σ_7	4	3	2	1

As already stated, it can be determined from the Galois group alone, without reference to the original equation, whether the equation is solvable and the degree of the roots that will appear in the solution. For such pronouncements it is not only the size of the Galois group that is at issue. The permutations themselves play a certain role, where what matters is the relations that exist among the permutations of the Galois group themselves, that is, relations in the sense

of composition of permutations, discussed in Chapter 5. In particular, if one performs Galois-group permutations σ and τ in succession, the result is another permutation. And this new permutation, denoted by $\tau \circ \sigma$, has, like σ and τ, the property of changing all the polynomials in the set B_K into polynomials in B_K and thus is itself a member of the Galois group.

9.5 A universally applicable procedure for documenting all the relations among the elements of the Galois group, though not particularly elegant due to its explicitness, is a table of the group operation. We shall see how this works in the context of the already mentioned biquadratic equation. As an example of composition of permutations we shall take the permutations σ_1 and σ_6 from the Galois group. The permutation σ_1 permutes the solution index 1 to the index $\sigma_1(1) = 3$. Since the index 3 is permuted by the second permutation σ_6 to $\sigma_6(3) = 4$, the net result is a shift of index 1 to $\sigma_6(\sigma_1(1)) = 4$. One does the analogous operations for the other three solution indices and obtains the following result:

	1	2	3	4
first σ_1 ...	3	2	1	4
... and then σ_6	4	3	2	1

A look at the table of the eight permutations of the Galois group shows that $\sigma_6 \circ \sigma_1 = \sigma_7$. The group table consists of all such combinations of two permutations, a sort of glorified multiplication table. Each entry in the table is the result of first applying the permutation from the top row and then the permutation in the left column:

	σ_0	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7
σ_0	σ_0	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7
σ_1	σ_1	σ_0	σ_3	σ_2	σ_6	σ_7	σ_4	σ_5
σ_2	σ_2	σ_3	σ_0	σ_1	σ_5	σ_4	σ_7	σ_6
σ_3	σ_3	σ_2	σ_1	σ_0	σ_7	σ_6	σ_5	σ_4
σ_4	σ_4	σ_5	σ_6	σ_7	σ_0	σ_1	σ_2	σ_3
σ_5	σ_5	σ_4	σ_7	σ_6	σ_2	σ_3	σ_0	σ_1
σ_6	σ_6	σ_7	σ_4	σ_5	σ_1	σ_0	σ_3	σ_2
σ_7	σ_7	σ_6	σ_5	σ_4	σ_3	σ_2	σ_1	σ_0

All relations among the permutations can be read off from the group table. One can no longer tell how the solutions and their indices are permuted by the permutations. However, one can tell—and we shall say it again because of its importance—from the group table alone whether the underlying equation is solvable in radicals and the degree of roots that are necessary for expressing the solutions.

Why the Galois group contains such information can be seen as plausible by breaking the solution process of the underlying equation into individual steps, each of which corresponds to the adjunction of a single quantity, and then investigating how this adjunction alters the Galois group. Namely, by extending the set of possible coefficients from a field K of "known quantities" to a larger field E, the set of "relational" polynomials B_K, which was used in the definition of the Galois group, is clearly enlarged to a set B_E. The more stringent requirement on the permutations linked with the extension of the field of coefficients may then lead to a restriction in the set of permutations belonging to the Galois group. It is thus seen that the possible reduction in size of the Galois group is intimately bound up with the properties of the adjoined values. To put it more concretely, under certain conditions, the adjunction of a value that is the mth root of an already known quantity has the effect of reducing the number of permutations in the Galois group by a factor of m.

9.6 We shall now take a detailed look at how the individual steps of solving a given equation are reflected in the corresponding reductions of the Galois group. We use our standard example of a biquadratic equation. Its solutions, since the cubic resolvent possesses a rational solution, can be expressed solely in terms of square roots:

$$x_{1,3} = 1 + \sqrt{2} \pm \sqrt{3 + \sqrt{2}},$$

$$x_{2,4} = 1 - \sqrt{2} \pm \sqrt{3 - \sqrt{2}}.$$

Beginning with the field of rational numbers as the set of a priori known quantities (given the rational coefficients of the equation), we add as our first additional known quantity the number $\sqrt{2}$. For the following two adjunctions in the solution process the numbers

$\sqrt{3+\sqrt{2}}$ and $\sqrt{3-\sqrt{2}}$ present themselves, in each case involving the square root of a known quantity.

Let us now observe how the adjunction of these three quantities reduces the Galois group. The initial adjunction of $\sqrt{2}$ to the base field $K = \mathbb{Q}$ leads to the result that among others, the polynomial

$$X_1 - X_2 + X_3 - X_4 - 4\sqrt{2}$$

belongs to the set $B_{\mathbb{Q}(\sqrt{2})}$, since $x_1 - x_2 + x_3 - x_4 = 4\sqrt{2}$. Since the four permutations $\sigma = \sigma_4, \sigma_5, \sigma_6, \sigma_7$ satisfy the condition $x_{\sigma(1)} - x_{\sigma(2)} + x_{\sigma(3)} - x_{\sigma(4)} = -4\sqrt{2}$, they can no longer belong to the Galois group after the extension of the field $K = \mathbb{Q}$ to $E = \mathbb{Q}\left(\sqrt{2}\right)$. Conversely, in analogy to the original field $K = \mathbb{Q}$, one can show using Galois's method that the permutations $\sigma_0, \sigma_1, \sigma_2, \sigma_3$ in fact belong to the Galois group. Thus the expansion of the set of "known quantities" to include the value $\sqrt{2}$ reduces the size of the Galois group by one-half.

With the subsequent adjunction of the second intermediate value, $\sqrt{3+\sqrt{2}}$, the set of polynomials that reflect the polynomial relations among the solutions is enlarged, for example, on account of the identity $x_1 - x_3 = 2\sqrt{3+\sqrt{2}}$, to include

$$X_1 - X_3 - 2\sqrt{3+\sqrt{2}}.$$

Since for the two permutations $\sigma = \sigma_1, \sigma_3$ the equation $x_{\sigma(1)} - x_{\sigma(3)} = -2\sqrt{3+\sqrt{2}}$ holds, these permutations are eliminated from the Galois group when $\sqrt{3+\sqrt{2}}$ is adjoined to the field $\mathbb{Q}\left(\sqrt{2}\right)$. Conversely, one can show that the two permutations σ_0, σ_2 in fact belong to the Galois group.

If finally the adjunction of $\sqrt{3-\sqrt{2}}$ is carried out, then all four solutions $x_1, \ldots, x_4$ can be expressed in terms of rational numbers and the adjoined numbers using the basic arithmetic operations. By the definition of the Galois group in terms of the obtained extension field, we need to consider the four polynomials $X_i - x_i$ for $i = 1, 2, 3, 4$. The result is that the Galois group contains only the identity permutation σ_0.

In Figure 9.1 the three adjunctions and their effect on the Galois group are illustrated. There the notation $K(a, b, \ldots)$ indicates the

extension field formed by the adjunction of the numbers $a, b, \ldots$ to a field K. That is, this field is defined as the totality of all numbers that can be obtained using the four arithmetic operations on the numbers $a, b, \ldots$ together with the numbers in the field K.

Steps in Solving the Equation	Fields of Current "Known Quantities"	Galois Group of the Equation
$\sqrt{3-\sqrt{2}}$	$\mathbb{Q}\left(\sqrt{2}, \sqrt{3+\sqrt{2}}, \sqrt{3-\sqrt{2}}\right)$	σ_0
↑ square root		
$\sqrt{3+\sqrt{2}}$	$\mathbb{Q}\left(\sqrt{2}, \sqrt{3+\sqrt{2}}\right)$	σ_0, σ_2
↑ square root		
$\sqrt{2}$	$\mathbb{Q}\left(\sqrt{2}\right)$	$\sigma_0, \sigma_1, \sigma_2, \sigma_3$
↑ square root		
coefficients of the equation	$\mathbb{Q}$	$\sigma_0, \sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6, \sigma_7$

Figure 9.1. Solving the equation $x^4 - 4x^3 - 4x^2 + 8x - 2 = 0$ by stepwise extension of the set of "known quantities" and how the associated extension fields reduce the Galois group.

It remains to note that the stepwise expansion of the set of known quantities by the square root of a previously known quantity does more than reduce the size of the Galois group by a factor of 2. Each of these adjunctions represents a decomposition of an appropriately arranged group table into four equal parts, each of which contains permutations only from one or the other half of the Galois group. For example, from the first adjunction one obtains the following decomposition, where the further decompositions can be seen in the upper-left-hand square:

	σ_0	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7
σ_0	σ_0	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7
σ_1	σ_1	σ_0	σ_3	σ_2	σ_6	σ_7	σ_4	σ_5
σ_2	σ_2	σ_3	σ_0	σ_1	σ_5	σ_4	σ_7	σ_6
σ_3	σ_3	σ_2	σ_1	σ_0	σ_7	σ_6	σ_5	σ_4
σ_4	σ_4	σ_5	σ_6	σ_7	σ_0	σ_1	σ_2	σ_3
σ_5	σ_5	σ_4	σ_7	σ_6	σ_2	σ_3	σ_0	σ_1
σ_6	σ_6	σ_7	σ_4	σ_5	σ_1	σ_0	σ_3	σ_2
σ_7	σ_7	σ_6	σ_5	σ_4	σ_3	σ_2	σ_1	σ_0

9.7 The correspondence that we have observed between adjunctions and decompositions of the group table holds generally for analogous decompositions into $m \times m$ squares, where m is a prime number,

under the assumption that the equation is irreducible and that the mth roots of unity have been adjoined in previous steps. Under these conditions, the following can be shown:

- The adjunction of an mth root, if it leads to an actual reduction in the Galois group, effects a decomposition of the group table as described into $m \times m$ squares.
- Conversely, for every such decomposition into $m \times m$ squares one can find an mth root whose adjunction reduces the group table to the upper-left-hand square.

Since mth roots of unity, as we described in Chapter 7, can always be expressed in radicals, this equivalence leads to the following result:

Theorem 9.3. *An irreducible equation is solvable in radicals precisely when the Galois group can be reduced step by step to a one-element Galois group, containing only the identity permutation, where each step corresponds to a decomposition of the (suitably ordered) group table into $m \times m$ squares each of which contains the mth part of the permutations.*

(In recognition of this equivalence, Galois groups allowing such a stepwise process are called *solvable*.)[8]

With this theorem, which to be sure we have not even begun to prove, it becomes clear why the Galois group is so valuable in analyzing the solvability of an equation: in principle, purely combinatorial considerations regarding the group table allow us to determine which root operations make progress in solving the original equation. Thus, for example, for the equation

$$x^5 - x - 1 = 0$$

it can be shown that the associated 120×120 group table permits a decomposition into squares only once: the resulting 60×60 square permits no further decomposition, and this is precisely the reason that the solutions of the given fifth-degree equation cannot be represented with nested root expressions with rational radicands.

[8] In the next chapter we shall learn a definition that will make dealing with the group table unnecessary.

Of course, the possibility of investigating the Galois group purely combinatorially via the group table is not the most elegant approach. How such an investigation can be simplified and why it works the way it does will be the topic of the next chapter.

9.8 We would like to use the remainder of this chapter to determine the Galois groups of some other equations, some of which have made an appearance in earlier chapters.[9] As with our standard example of the biquadratic equation, we will generally concentrate on that part of the demonstration in which it is shown that certain permutations cannot belong to the Galois group. The proof that the remaining permutations must belong to the Galois group can in principle always be carried out using the technique of the section below on computing the Galois group. In many cases, however, a simpler argument can be adduced. However, the theorems that are of use in this regard will be presented only in the following chapter.

In the definition of the Galois group we have generally excluded equations with multiple roots. In the remainder of the chapter we shall further reduce our considerations to irreducible equations. At the level of the Galois group this is equivalent, as will be shown in the next chapter, to the property that for every pair of solutions x_j and x_k there exists at least one permutation σ that permutes the solution x_j to x_k, that is, $\sigma(j) = k$. In such a case, one says that the Galois group *acts transitively* on the solutions of the equation.

9.9 Quadratic equations that are irreducible always possess a Galois group consisting of two permutations. In addition to the identity permutation σ_0, there is additionally a permutation σ_1 that permutes the two solutions. The group table has the following form:

	σ_0	σ_1
σ_0	σ_0	σ_1
σ_1	σ_1	σ_0

[9]The other examples have been taken, in part, from Leonhard Soicher, John McKay, Computing Galois groups over the rationals, *Journal of Number Theory*, 20 (1985), pp. 273–281. Their article also contains examples of equations of the sixth and higher degrees.

9.10 Regarding the Galois group of an irreducible cubic equation, there are two possibilities: Either the Galois group contains all six permutations of the three solutions or it contains three permutations that permute the solutions cyclically. An example of such an equation, derived from the cyclotomic equation of seventh degree, is

$$x^3 + x^2 - 2x - 1 = 0,$$

whose three solutions are $x_j = 2\cos\left(\frac{2\pi j}{7}\right)$, for $j = 1, 2, 3$. On account of the identities

$$x_2 = x_1^2 - 2, \quad x_3 = x_2^2 - 2, \quad x_1 = x_3^2 - 2,$$

the three polynomials

$$X_2 - X_1^2 + 2, \quad X_3 - X_2^2 + 2, \quad X_1 - X_3^2 + 2$$

belong to the set $B_{\mathbb{Q}}$ of polynomials to be considered in determining the Galois group. As a consequence, we have that a permutation σ belonging to the Galois group is already determined by its effect on a single index, for example on the index $\sigma(1)$. Therefore the Galois group consists of only the three permutations that permute the solutions cyclically:

	1	2	3
σ_0	1	2	3
σ_1	3	1	2
σ_2	2	3	1

The group table for the Galois group comprising these three permutations has the following form:

	σ_0	σ_1	σ_2
σ_0	σ_0	σ_1	σ_2
σ_1	σ_1	σ_2	σ_0
σ_2	σ_2	σ_0	σ_1

Of course, the Galois group of the previous equation can also be determined directly, that is, without knowledge of the solutions. Aside from Galois's general procedure, one may also calculate the difference product of the solutions, whose square is the discriminant.

This is most easily done with the general formula for cubic equations given in Chapter 5:

$$(x_1 - x_2)(x_2 - x_3)(x_1 - x_3) = \pm 6i\sqrt{3}\sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3},$$

where p and q are the coefficients of the reduced equation, for which in the case of the equation under investigation, $p = -\frac{7}{3}$ and $q = -\frac{7}{27}$. This leads to

$$(x_1 - x_2)(x_2 - x_3)(x_1 - x_3) = -7,$$

which shows at once that the odd permutations, that is, those that change the sign of the difference product, do not belong to the Galois group.

9.11 "Most" irreducible cubic equations, such as the equation

$$x^3 + x - 6 = 0$$

solved in Chapter 1 with the three solutions

$$x_{j+1} = \zeta^j \sqrt[3]{3 + \frac{2}{3}\sqrt{\frac{61}{3}}} + \zeta^{2j} \sqrt[3]{3 - \frac{2}{3}\sqrt{\frac{61}{3}}}$$

for $j = 0, 1, 2$, lead to a Galois group that contains all six permutations:

	1	2	3
σ_0	1	2	3
σ_1	3	1	2
σ_2	2	3	1
σ_3	1	3	2
σ_4	3	2	1
σ_5	2	1	3

The group table of the Galois group has the following form, where the decomposition into four 3×3 squares corresponding to the adjunction of the square root of the discriminant, which contains only three different permutations, is clear:

	σ_0	σ_1	σ_2	σ_3	σ_4	σ_5
σ_0	σ_0	σ_1	σ_2	σ_3	σ_4	σ_5
σ_1	σ_1	σ_2	σ_0	σ_5	σ_3	σ_4
σ_2	σ_2	σ_0	σ_1	σ_4	σ_5	σ_3
σ_3	σ_3	σ_4	σ_5	σ_0	σ_1	σ_2
σ_4	σ_4	σ_5	σ_3	σ_2	σ_0	σ_1
σ_5	σ_5	σ_3	σ_4	σ_1	σ_2	σ_0

The direct relationship between the Galois group and the solution of the equation is complicated in the case of the cubic equation in that the third root of unity $\zeta = -\frac{1}{2} + \frac{1}{2}i\sqrt{3}$ must be assumed as "known." When the equation has rational coefficients, this means that the direct correspondence between the solution steps of the equation and the steps in decomposition of the Galois group is ensured only when the field $\mathbb{Q}$ containing the coefficients is enlarged by the third root of unity to $\mathbb{Q}(\zeta)$.

To that extent an equation that looks simpler than the general one, such as the equation

$$x^3 - 3x^2 - 3x - 1 = 0,$$

which appears at the end of Chapter 1, with the three solutions

$$x_{j+1} = 1 + \zeta^j \sqrt[3]{2} + \zeta^{2j} \sqrt[3]{4},$$

$j = 0, 1, 2$, can have a Galois group comprising six permutations. For the difference product one obtains, since the coefficients of the reduced equation are $p = -6$ and $q = -6$, the value

$$(x_1 - x_2)(x_2 - x_3)(x_1 - x_3) = 6i\sqrt{3} \in \mathbb{Q}(\zeta),$$

with the result that the Galois group reduces to three permutations only when the third root of unity ζ is adjoined. In contrast, the Galois group over the field of rational numbers contains all six permutations.[10] As in the earlier investigated equation $x^3 + x^2 - 2x - 1 = 0$, even though a *casus irreducibilis* is present and therefore the three solutions are real, the three roots of unity must first be adjoined, with

[10] Moreover, since

$$x_j^2 - 3x_j - 2 = \zeta^{j-1} \sqrt[3]{2},$$

the *splitting field*, that is, the field arising from adjunction of all the solutions, is $\mathbb{Q}(\zeta, \sqrt[3]{2})$. The name "splitting field" comes from the fact that it is the smallest field in which the equation to be solved can be factored into linear factors.

the Galois group not being thereby reduced. Then the adjunction of a third root allows the solution of the equation.

9.12 Irreducible biquadratic equations can have Galois groups with 4, 8, 12, or 24 permutations. In some of these four cases there are different possibilities as to which permutations belong to the Galois group. However, from a qualitative point of view, in particular with regard to the solvability of the Galois group, two Galois groups will be considered the same if one can be transformed into the other by a possible rearrangement of the rows and columns of the group table. Such Galois groups are said to be *isomorphic.* With this notion of isomorphism of groups, there are then only five possibilities for the Galois group of an irreducible biquadratic equation: two with four permutations and one each with 8, 12, or 24 permutations. For each of these cases we will look at an equation.

We begin with the examples of biquadratic equations with the equation

$$x^4 + x^3 + x^2 + x + 1 = 0,$$

solved in Chapter 7, obtained from the fifth-degree cyclotomic equation, whose solutions are $x_{j+1} = \cos\left(\frac{2\pi 3^j}{5}\right) + i \sin\left(\frac{2\pi 3^j}{5}\right)$, for $j = 0, 1, 2, 3$, where the numbering is chosen as has already been done in the construction of the periods for the cyclotomic equations. In what follows, in the relations among the solutions the following symmetry will be apparent:

$$x_{j+1} = x_j^3.$$

Thus, as with every other equation for periods of a cyclotomic equation, a permutation belonging to the Galois group σ is determined by its action on a single index, for example by the index $\sigma(1)$. This leads to the result that only four permutations, those that permute the solutions cyclically, belong to the Galois group. The following tables show the permutations of the Galois group and the associated group table:

	1	2	3	4
σ_0	1	2	3	4
σ_1	2	3	4	1
σ_2	3	4	1	2
σ_3	4	1	2	3

	σ_0	σ_1	σ_2	σ_3
σ_0	σ_0	σ_1	σ_2	σ_3
σ_1	σ_1	σ_2	σ_3	σ_0
σ_2	σ_2	σ_3	σ_0	σ_1
σ_3	σ_3	σ_0	σ_1	σ_2

9.13 Similarly, the equation

$$x^4 + 1 = 0$$

leads to a Galois group with four permutations. This equation's solutions are four of the eighth roots of unity, namely

$$x_j = \cos\left(\frac{2\pi(2j-1)}{8}\right) + i\sin\left(\frac{2\pi(2j-1)}{8}\right), \quad \text{for } j = 1, 2, 3, 4.$$

First, in analogy to how we dealt with the previous equation, a permutation σ belonging to the Galois group is determined solely by $\sigma(1)$, that is, by its action on the first index, on account of the relation

$$x_j = x_1^{2j-1}.$$

Therefore, the permutations belonging to the Galois group are those listed in the left-hand table below. One then obtains the group table on the right, where the four identical permutations on the diagonal show that no renaming of the permutations can lead to the group table of the given equation:

	1	2	3	4
σ_0	1	2	3	4
σ_1	2	1	4	3
σ_2	3	4	1	2
σ_3	4	3	2	1

	σ_0	σ_1	σ_2	σ_3
σ_0	σ_0	σ_1	σ_2	σ_3
σ_1	σ_1	σ_0	σ_3	σ_2
σ_2	σ_2	σ_3	σ_0	σ_1
σ_3	σ_3	σ_2	σ_1	σ_0

9.14 An irreducible biquadratic equation with a Galois group consisting of eight permutations has already been investigated in Sections 9.4 and 9.5. A very simple equation that leads to an isomorphic Galois group is

$$x^4 - 2 = 0,$$

whose solutions are $x_j = i^{j-1}\sqrt[4]{2}$ for $j = 1, 2, 3, 4$. Since $x_1x_3 + x_2x_4 = 0$, we may proceed analogously to the previous equations investigated. We note only that the Galois group reduces to four

cyclic permutations if the field is extended to $\mathbb{Q}$ by the fourth root of unity i.

9.15 A Galois group comprising all twelve even permutations is obtained for the equation

$$x^4 + 8x + 12 = 0.$$

The reason is that using the formulas of Chapter 5 for the cubic resolvent

$$z^3 - 12z + 8 = 0,$$

one obtains for the difference product the value

$$\begin{aligned}\prod_{j>k}(x_j - x_k) &= 8\prod_{j>k}(z_j - z_k) = 48i\sqrt{3}\sqrt{\left(\frac{8}{2}\right)^2 + \left(\frac{-12}{3}^3\right)} \\ &= -576.\end{aligned}$$

9.16 An irreducible biquadratic equation with a Galois group of maximal size (24 permutations) is the following:

$$x^4 + x + 1 = 0.$$

9.17 In the case of irreducible equations of fifth degree, there are only five possibilities for the Galois group (up to isomorphism), being groups of 5, 10, 20, 60, 120 permutations. Equations whose Galois group falls into one of the first three cases are solvable in radicals, while those for the last two cases are not.

There are only five permutations in the Galois group of Vandermonde's equation $x^5 + x^4 - 4x^3 - 3x^2 + 3x + 1 = 0$, whose solutions are two-member periods of the cyclotomic equation of degree 11. As described in Sections 9.3 and 9.4, one can determine the permutations belonging to the Galois group with the help of polynomial relations such as $x_1^2 = x_2 + 2$ and $x_2^2 = x_3 + 2$.

The equation $x^5 - 5x + 12 = 0$ leads to a Galois group with ten permutations. Its solutions are

$$\begin{aligned} x_{j+1} = \epsilon^j \sqrt[5]{-1 + \frac{2}{5}\sqrt{5} - 3\sqrt{\frac{1}{5} - \frac{11}{125}\sqrt{5}}} \\ + \epsilon^{2j} \sqrt[5]{-1 - \frac{2}{5}\sqrt{5} + 3\sqrt{\frac{1}{5} + \frac{11}{125}\sqrt{5}}} \\ + \epsilon^{3j} \sqrt[5]{-1 - \frac{2}{5}\sqrt{5} - 3\sqrt{\frac{1}{5} + \frac{11}{125}\sqrt{5}}} \\ + \epsilon^{4j} \sqrt[5]{-1 + \frac{2}{5}\sqrt{5} + 3\sqrt{\frac{1}{5} - \frac{11}{125}\sqrt{5}}}, \end{aligned}$$

where $\epsilon = \cos\left(\frac{2\pi}{5}\right) + i \sin\left(\frac{2\pi}{5}\right)$ for $j = 0, 1, 2, 3, 4$. This can be seen from calculating the value of the bicubic resolvent $z = 5$ investigated in Chapter 8, which corresponds, up to a sign to be determined, to the identity

$$\begin{aligned} x_1x_2 + x_2x_3 + x_3x_4 + x_4x_5 + x_5x_1 \\ - x_1x_3 - x_2x_4 - x_3x_5 - x_4x_1 - x_5x_2 = -10. \end{aligned}$$

The Galois group of this equation appears as a group table in the preface to this book, in Figure 0.1.

The Galois group of the equation $x^5 - 2 = 0$, whose solutions are $x_j = \epsilon^{j-1} \sqrt[5]{2}$, for $j = 1, \ldots, 5$, contains twenty permutations. Since $x_j = x_1^{2-j} x_2^{j-1}$, each permutation of the Galois group is completely determined by its action on the two solutions x_1 and x_2. And in fact, all of the $5 \times 4 = 20$ ways of associating a pair of two different solutions with the two solutions x_1, x_2 result in a permutation belonging to the Galois group. Indeed, these twenty permutations are defined, for $p = 1, \ldots, 4$ and $q = 0, \ldots, 5$, by

$$\sigma_{p,q}\left(\epsilon^j \sqrt[5]{2}\right) = \epsilon^{pj+q} \sqrt[5]{2},$$

for $j = 0, 1, 2, 3, 4$.

The equation $x^5 + 20x + 16 = 0$ yields sixty permutations. These are the even permutations, those that leave unchanged the integer

value of the difference product. Using the formula given in footnote 5 in Chapter 8, the difference product is either +32000 or −32000.

An example of a fifth-degree equation with maximal Galois group of 120 permutations is $x^5 - x + 1 = 0$.

9.18 We close this chapter with a theorem of Galois containing his findings about the solvability of equations in the form of a "traditionally" formulated criterion. That equations satisfying this criterion are solvable was hypothesized before Galois by Abel, in 1828, in a letter to Crelle (1780–1855).[11]

Theorem 9.4. *An irreducible equation of prime degree is solvable in radicals if and only if all the solutions can be expressed as polynomials in two arbitrary solutions.*

In particular, an irreducible (over the rational numbers) fifth-degree equation with three real and two nonreal solutions cannot be solved in radicals. Thus, for example, the equation $x^5 - 17x - 17$ is immediately seen to be unsolvable, since by Eisenstein's irreducibility criterion the equation is irreducible, and furthermore, it has three real solutions, and there is no way that two of them can be used to express all the solutions in terms of polynomials (with rational coefficients).

An additional consequence of Galois's criterion is that the size of the Galois group of an irreducible solvable equation of prime degree n is always a divisor of $n(n-1)$ and a multiple of n.

Computing the Galois Group

As we have mentioned, the set B_K of polynomials used in the definition of the Galois group is far too large to list explicitly. Even a complete

[11] August Leopold Crelle is remembered primarily as the founder and editor of the first German mathematical journal. The *Journal für die Reine und Angewandte Mathematik* is even today frequently referred to as "Crelle's journal." Abel's proof of unsolvability was published in 1826 in volume 1 of Crelle's journal (Beweis der Unmöglichkeit, algebraische Gleichungen von höheren Graden, als dem vierten, allgemein aufzulösen, pp. 65–84). Our Theorem 9.4 is also formulated as Théorème IV (p. 143) in Abel's Mémoire sur une classe particulière d'équations résolubles algébriquement, *Crelle*, 4 (1829), pp. 131–156; see also Lars Gårding, Christian Skau, Niels Henrik Abel and solvable equations, *Archive for History of Exact Sciences*, 48 (1994), pp. 81–103.

description is not simple.[12] A way out of this dilemma, allowing for an explicit computation of the Galois group, is offered by the method used originally by Galois for defining the Galois group.

Galois constructed, for the hypothesized n distinct solutions $x_1, \ldots, x_n$ of a given nth-degree equation, the so-called *Galois resolvent*. He expressed this in the form

$$t = m_1 x_1 + m_2 x_2 + \cdots + m_n x_n$$

with suitably chosen numbers $m_1, \ldots, m_n$. Galois then observed that one can always find numbers $m_1, \ldots, m_n$ in the field K such that all $n!$ values

$$t_\sigma = m_1 x_{\sigma(1)} + m_2 x_{\sigma(2)} + \cdots + m_n x_{\sigma(n)}$$

resulting from permutations σ of the indices $1, \ldots, n$ are distinct.[13] A quantity t thus constructed now has, as Lagrange determined, the property that all solutions $x_1, \ldots, x_n$ can be represented by polynomials in t, in particular without using root operations: $x_1 = g_1(t), \ldots, x_n = g_n(t)$.[14] Every polynomial from the set B_K used to define the Galois group thus

---

[12] For readers who already know (almost) everything: The set B_K is an ideal in the polynomial ring $K[X_1, \ldots, X_n]$. From the Hilbert basis theorem, there exist finitely many (basis) polynomials $h_1, \ldots, h_m$ such that the set B_k comprises the polynomials of the form

$$f_1 h_1 + \cdots + f_m h_m$$

for some polynomials $f_1, \ldots, f_m$. If one could determine such basis polynomials $h_1, \ldots, h_m$, then one could compute the Galois group using the fact that each permutation can be individually checked with these polynomials for membership in the Galois group.

[13] For the values $m_1, \ldots, m_n$ to be chosen, none of the equations

$$m_1 \left(x_{\sigma(1)} - x_{\tau(1)}\right) + \cdots + m_n \left(x_{\sigma(n)} - x_{\tau(n)}\right) = 0$$

may be satisfied for two distinct permutations σ and τ. Each of these $\frac{1}{2} n!(n! - 1)$ equations thus limits the possible selection of values $m_1, \ldots, m_n$ by a hyperplane in K^n: for $n = 2$ this is a straight line in K^2, for $n = 3$ a plane in K^3, and so on. Thus in any case, there remain infinitely many ways of choosing the values $m_1, \ldots, m_n$.

[14] The proof of this theorem—the "modern" variant $K(x_1, \ldots, x_n) = K(t)$ can be found in books on abstract algebra as a theorem on the existence of a primitive element—was only sketched by Galois. Corresponding to the constructed Galois resolvent t, Galois formed the polynomial

$$G(T, X_1, \ldots, X_n) = \prod_{\substack{\sigma \in S_n \\ \sigma(1)=1}} \left(T - (m_1 X_{\sigma(1)} + \cdots + m_n X_{\sigma(n)})\right)$$

$$= \sum_{k=0}^{(n-1)!} g_k(X_1, \ldots, X_n) T^k,$$

based on all permutations σ that fix the polynomial's $(n-1)!$ factors, where the coefficients $g_k(X_1, \ldots, X_n)$ appearing in the sum are polynomials in the variables $X_1, \ldots, X_n$ that are symmetric in the variables $X_2, \ldots, X_n$. If one now considers the polynomials $g_k(X_1, \ldots, X_n)$ as polynomials in the variable X_1, then one is dealing with symmetric polynomials in the variables $X_2, \ldots, X_n$, which therefore can be expressed as elementary symmetric polynomials in these variables. Since furthermore, each of these elementary symmetric polynomials can be expressed as a polynomial in the variable X_1 as well as the elementary symmetric polynomials in the variables

corresponds to a polynomial equation that is satisfied by the Galois resolvent t. And such polynomials in one variable, as explained in Point 3 in the upcoming section on computing with polynomials, are all multiples of an irreducible polynomial over the field K with t as a zero. Therefore every permutation can be examined with respect to this single equation to determine whether it belongs to the Galois group. And furthermore, this single irreducible polynomial with t as a zero can generally be found "easily" by employing the method of Lagrange and taking the product of all $n!$ linear factors $(T - t_\sigma)$ and then decomposing this degree-$n!$ equation into irreducible factors, whereby one then must seek the factor $\mathfrak{G}(T)$ that has t as a zero.

The investigation of a particular equation should help in clarifying these matters as well as in showing how in practice one can use the numerical solutions to compute the Galois group. In fact, due to the inevitable errors due to rounding, the numerical values are not suitable for proving an equality, though they can be used to prove an inequality, which often suffices. As an example, let us consider the equation

$$x^4 - 4x^3 - 4x^2 + 8x - 2 = 0,$$

$X_2, \ldots, X_n$ (for example, $X_2 + \cdots + X_n = (X_1 + \cdots + X_n) - X_1$), one obtains

$$g_k(X_1, \ldots, X_n) = \sum_{j=0,1,\ldots} h_{j,k}(X_1, \ldots, X_n) X_1^j,$$

where the polynomials $h_{j,k}(X_1, \ldots, X_n)$ are symmetric in the variables $X_1, \ldots, X_n$. One now defines the polynomial

$$F(X) = \sum_k \sum_j h_{j,k}(x_1, \ldots, x_n) t^k X^j.$$

The values $h_{j,k}(x_1, \ldots, x_n)$ can without exception be expressed as polynomials in the coefficients of the original equation, so that they must lie in the field K. As we shall show, the equation $F(X) = 0$ has only x_1 as a solution in common with the original equation, so that the linear factor $(X - x_1)$ can be computed using the Euclidean algorithm (see the following section) from the coefficients of the two equations, that is, from values in the field K together with t, using the four basic arithmetic operations. That one can in fact do without division can be shown by the methods of linear algebra (see Section 10.9).

We have still to investigate the zeros of the polynomial $F(X)$. Clearly, from the factor associated with the identity, that is, the permutation that leaves everything unchanged, we have

$$F(x_1) = \sum_k \sum_j h_{j,k}(x_1, \ldots, x_n) t^k x_1^j = \sum_k g_k(x_1, \ldots, x_n) t_k = G(t, x_1, \ldots, x_n) = 0.$$

Furthermore,

$$\begin{aligned} F(x_2) &= \sum_k \sum_j h_{j,k}(x_1, x_2, \ldots, x_n) t^k x_2^j = \sum_k \sum_j h_{j,k}(x_2, x_1, \ldots, x_n) t^k x_2^j \\ &= \sum_k g_k(x_2, x_1, \ldots, x_n) t^k = \prod_{\substack{\sigma \in S_n \\ \sigma(1)=2}} \left(t - (m_1 x_2 + m_2 x_{\sigma(2)} + \cdots + m_n x_{\sigma(n)})\right) \\ &\neq 0, \end{aligned}$$

with corresponding results for the other solutions $x_3, \ldots, x_n$.

already analyzed in Sections 9.4 and 9.5. It possesses four real solutions, whose numerical values can be found by any one of a number of approximation algorithms:

$$x_1 = 4.51521655\ldots, \qquad x_2 = 0.84506656\ldots,$$
$$x_3 = 0.31321057\ldots, \qquad x_4 = -1.67349368\ldots.$$

One now seeks a Galois resolvent by trial and error, where, for example, $t = -x_2 + x_3 - 2x_4$ possesses the required property: By numerical calculation one can show that the $4! = 24$ values $-x_{\sigma(2)} + x_{\sigma(3)} - 2x_{\sigma(4)}$ are distinct. Then, by multiplication by the 24 linear factors

$$\left(T - \left(-x_{\sigma(2)} + x_{\sigma(3)} - 2x_{\sigma(4)}\right)\right),$$

one finds for the Galois resolvent t a 24th-degree equation with integer coefficients, and therefore, with minimal rounding of the numerical results one can determine the nearest integer values exactly.

In the search for a factor $\mathfrak{G}(T)$ irreducible over K with rational coefficients that has the Galois resolvent t as a zero, knowledge about the numerical values can again be used to advantage. We need to check which of the twenty-four linear factors

$$\left(T - \left(-x_{\sigma(2)} + x_{\sigma(3)} - 2x_{\sigma(4)}\right)\right)$$

can be multiplied together to yield a polynomial with integer coefficients. Clearly, every combination of permutations can be rejected for which the numerically calculated product is not close to an integer polynomial. In the converse case, when the numerical result indeed corresponds approximately to an integer polynomial, this allegedly integer polynomial must be checked to see whether it is indeed a divisor of the degree-24 polynomial to be factored.

For our concrete example, one obtains for the Galois resolvent t an irreducible (over the rational numbers) eighth-degree polynomial $\mathfrak{G}(T)$ with $\mathfrak{G}(t) = 0$:

$$\mathfrak{G}(T) = T^8 + 16T^7 - 40T^6 - 1376T^5 - 928T^4 + 34048T^3 + 22208T^2 - 253184T + 72256.$$

The linear factors that constitute the polynomial $\mathfrak{G}(T)$ correspond to the following permutation of indices:

	1	2	3	4
σ_0	1	2	3	4
σ_1	3	2	1	4
σ_2	1	4	3	2
σ_3	3	4	1	2
σ_4	2	1	4	3
σ_5	4	1	2	3
σ_6	2	3	4	1
σ_7	4	3	2	1

The set of permutations that one obtains with this generally applicable method is the desired Galois group. Then, on the one hand, from the identity

$$(-x_2 + x_3 - 2x_4)^8 + 16(-x_2 + x_3 - 2x_4)^7 - \cdots$$
$$- 253184(-x_2 + x_3 - 2x_4) + 72256 = 0$$

one obtains a polynomial belonging to the set $B_{\mathbb{Q}}$. And this identity remains valid, on account of the underlying condition in the construction of the Galois resolvent, only for those permutations corresponding to one of the eight linear factors of the irreducible polynomial $\mathfrak{G}(T)$, which are precisely the permutations in the above table. On the other hand, one can prove conversely that every one of these permutations σ satisfies all other polynomial identities valid for the solutions $x_1, \ldots, x_n$. That is, one always has $h\left(x_{\sigma(1)}, \ldots, x_{\sigma(n)}\right) = 0$ whenever $h(x_1, \ldots, x_n) = 0$ holds.[15]

---

[15]Such a proof might begin with polynomial representations of the solutions in terms of the Galois resolvent t. That is,

$$x_1 = g_1(t), \quad \ldots, \quad x_n = g_n(t).$$

If one inserts these representations into the beginning equation $f(x) = 0$ to be solved, one obtains $f(g_j(t)) = 0$. From point 3 of the following section, the irreducible factor $\mathfrak{G}(T)$ belonging to the zero t of the constructed $n!$-degree polynomial must be a divisor of the polynomial $f(g_j(T))$. For each permutation σ sought on the basis of the Galois resolvent, the associated zero t_σ of $\mathfrak{G}(T)$ satisfies $f(g_j(t_\sigma)) = 0$ as well, so that every value $g_j(t_\sigma)$ must equal one of the solutions $x_1, \ldots, x_n$. If two such values $g_j(t_\sigma)$ and $g_k(t_\sigma)$ are equal for some permutation σ, then t_σ is a zero of the associated difference polynomial $g_j(T) - g_k(T)$, which therefore must be divisible by $\mathfrak{G}(T)$. It follows that $g_j(t) = g_k(t)$, that is, $x_j = x_k$. Altogether, this shows that the values $g_1(t_\sigma), \ldots, g_n(t_\sigma)$ correspond to a permutation of the solutions $x_1, \ldots, x_n$:

$$x_{\tau(1)} = g_1(t_\sigma), \quad \ldots, \quad x_{\tau(n)} = g_n(t_\sigma).$$

That the permutation τ is in fact the permutation σ can be seen from the fact that the polynomial

$$T - \left(m_1 g_1(T) + \cdots + m_n g_n(T)\right)$$

has $T = t$ as a zero, and therefore is divisible by the polynomial $\mathfrak{G}(T)$ and thus also has the value t_σ as a zero:

$$t_\sigma = m_1 g_1(t_\sigma) + \cdots + m_n g_n(t_\sigma).$$

For permutations τ corresponding to $g_1(t_\sigma), \ldots, g_n(t_\sigma)$ one therefore has $t_\sigma = t_\tau$. However, this equality can hold only for $\sigma = \tau$, since all $n!$ possible values of t_σ are

In summary, to compute the Galois group, first a Galois resolvent t is computed and the associated degree-$n!$ polynomial constructed. Among its irreducible factors, the factor $\mathfrak{G}(T)$ having t as a zero provides a complete description of the Galois group. In the process, every permutation of the solutions $x_1, \ldots, x_n$ is determined by changing a single value, namely, by passing from the Galois resolvent t to another zero t_σ of the irreducible factor $\mathfrak{G}(T)$. In this way, the permutations σ are implicitly determined by

$$t_\sigma = m_1 x_{\sigma(1)} + m_2 x_{\sigma(2)} + \cdots + m_n x_{\sigma(n)}.$$

Beginning with polynomial representations $x_1 = g_1(t), \ldots, x_n = g_n(t)$, one additionally has the formulas[16]

$$x_{\sigma(1)} = g_1(t_\sigma), \quad \ldots, \quad x_{\sigma(n)} = g_n(t_\sigma).$$

To finish, we note that an extension of the field K can lead in certain circumstances to the result that the polynomial $\mathfrak{G}(T)$, irreducible over K, can be factored into more than one factor. The factor that has the Galois resolvent t as a zero describes the Galois group defined in terms of the extension field. For this reason Galois himself studied the properties of such a decomposition—in particular, all the factors have the same degree—in order to discover the precise behavior of the Galois group in extension fields.

A Quick Course in Calculating with Polynomials

The determination of Galois groups, as described in the previous section, requires an extensive investigation of polynomials, where Galois's procedure allows us to restrict our attention to polynomials of a single variable, which are relatively easy to work with. For that reason, we have assembled here the most important properties of polynomials of one variable. It should be noted at the outset that with respect to divisibility and related properties, polynomials display striking analogies to the integers.

---

distinct. It follows that

$$x_{\sigma(1)} = g_1(t_\sigma), \quad \ldots, \quad x_{\sigma(n)} = g_n(t_\sigma).$$

If now some polynomial relation $h(x_1, \ldots, x_n) = 0$ is given, then one sees at once that the polynomial $h\big(g_1(T), \ldots, g_n(T)\big)$ is divisible by $\mathfrak{G}(T)$. This yields the desired result $0 = h\big(g_1(t_\sigma), \ldots, g_n(t_\sigma)\big) = h\left(x_{\sigma(1)}, \ldots, x_{\sigma(n)}\right)$.

It remains to note that here a single argument, namely the divisibility by $\mathfrak{G}$, has been used four times. This leads one to suspect that a universal principle is in play. That is indeed the case, as will be discussed in the following chapter.

[16] A proof of which appears in Point 3 of the upcoming section on calculating with polynomials.

Point 1. We begin with an analogue of integer division.

Theorem 9.5. *A polynomial $f(X)$ can be divided by a nonzero polynomial $g(X)$ with result a quotient $q(X)$ and remainder $r(X)$ such that*

$$f(X) = q(X)g(X) + r(X),$$

with the degree of the remainder polynomial $r(X)$ of lower degree than $g(X)$.

In practice, the computation of the two polynomials $q(X)$ and $r(X)$ is accomplished by means of a procedure analogous to familiar long division, which is no coincidence, since integers written in base 10 can be understood as the values of a polynomial evaluated at 10. However, in contrast to the long-division algorithm, there is no borrowing, so that the polynomial case is actually simpler. We will content ourselves with an example, since a general description of the algorithm would yield less clarity:

$$\begin{array}{l} (X^4 - 2X^3 + 3X^2 - X + 2) \div (X^2 - 2X - 1) = X^2 + 4 \\ \underline{(X^4 - 2X^3 - \ X^2)} \\ \qquad\qquad 4X^2 - \ X + 2 \\ \qquad\qquad \underline{4X^2 - 8X - 4} \\ \qquad\qquad\qquad 7X + 6 \end{array}$$

The result is therefore

$$\begin{aligned} X^4 - 2X^3 + 3X^2 - X + 2 \\ = (X^2 - 2X - 1)\,(X^2 + 4) + 7X + 6. \end{aligned}$$

Point 2. In analogy with the case of integers, the polynomial $g(X)$ is called a *divisor* of a polynomial $f(X)$ if $f(X)$ is divisible by $g(X)$ with zero remainder. A *greatest common divisor* of two polynomials $f(X)$ and $g(X)$, denoted by $\gcd\big(f(X), g(X)\big)$, is a polynomial of maximal degree that divides both $f(X)$ and $g(X)$. We note first that the greatest common divisor remains unchanged when $f(X)$ is replaced by $f(X) - h(X)g(X)$; that is,

$$\gcd\big(f(X), g(X)\big) = \gcd\big(f(X) - h(X)g(X), g(X)\big).$$

The reason for this is that every common divisor of $f(X)$ and $g(X)$ also divides $f(X) - h(X)g(X)$, and the converse holds because the given transformation is reversible.

Of special significance is the special case in which the polynomial $h(X)$ is equal to the quotient $q(X)$ resulting from division with remainder of $f(X)$ by $g(X)$: in this special transformation the two polynomials $f_1(X) = g(X)$ and $g_1(X) = f(X) - q(X)g(X)$ are obtained, whereby the degree of the second polynomial is smaller than that of the polynomial $g(X)$, since it is a remainder of a division.

As in the case of integers, by repeating such steps one can compute the greatest common divisor. The procedure, called the *Euclidean algorithm*, begins with the pair of polynomials $f_0(X) = f(X)$ and $g_0(X) = g(X)$ and reaches, at the jth step, the pair

$$f_j(X) = g_{j-1}(X), \quad g_j(X) = f_{j-1}(X) - q_{j-1}(X)g_{j-1}(X),$$

where the degree of the polynomial $g_j(X)$ is always less than that of the polynomial $g_{j-1}(X)$. Therefore, the algorithm must terminate after a finite number of steps with some $g_m(X) = 0$. One then has

$$\begin{aligned}\gcd\big(f(X), g(X)\big) &= \gcd\big(f_1(X), g_1(X)\big) = \cdots \\ &= \gcd\big(f_m(X), 0\big) = f_m(X).\end{aligned}$$

One now obtains, for every index j, for the greatest common divisor $f_m(X)$ thus obtained a representation $f_m(X) = u_j(X)f_j(X)+v_j(X)g_j(X)$, for suitable polynomials $u_j(X)$ and $v_j(X)$. This is clear at once inductively if working backward from the index $j = m$, for which such a representation is trivially satisfied, one considers the equation corresponding to the jth step:

$$\begin{aligned}f_m(X) &= u_j(X)f_j(X) + v_j(X)g_j(X) \\ &= u_j(X)g_{j-1}(X) + v_j(X)\big(f_{j-1}(X) - q_{j-1}(X)g_{j-1}(X)\big) \\ &= v_j(X)f_{j-1}(X) + \big(u_j(X) - v_j(X)q_{j-1}(X)\big)g_{j-1}(X).\end{aligned}$$

This equation, proved by induction, now reveals to us in the case $j = 0$ an important property of the greatest common divisor $f_m(X)$, as stated in the following theorem.

Theorem 9.6. *The greatest common divisor of two polynomials $f(X)$ and $g(X)$ can be expressed as $u(X)f(X) + v(X)g(X)$ for suitably chosen polynomials $u(X)$ and $v(X)$.*

Point 3. We now come to a result proved by Galois that played an important role in his investigations.

Theorem 9.7. *If the polynomial $f(X)$ is irreducible and possesses a common zero with the polynomial $g(X)$, then $g(X)$ is divisible by $f(X)$.*

We observe first that the formulation of the theorem is not quite exact, since irreducibility is always in reference to the set in which the polynomial's coefficients are presumed to belong. Here irreducibility is meant with respect to the field K.

Furthermore, the theorem may be interpreted to mean that the zeros of an irreducible polynomial are algebraically indistinguishable. That is, every property with respect to the four basic operations enjoyed by one zero holds as well for the other zeros of an irreducible polynomial.

For a proof, one first determines, using the Euclidean algorithm, the greatest common divisor of $f(X)$ and $g(X)$, which we shall denote by $d(X)$. Its coefficients must reside in the field K. Moreover, executing the Euclidean algorithm provides us with a representation $d(X) = u(X)f(X) + v(X)g(X)$ for suitably chosen polynomials $u(X)$ and $v(X)$. The common zero of the two polynomials $f(X)$ and $g(X)$ is therefore also a zero of $d(X)$, which implies that the degree of $d(X)$ is at least 1. This shows that the polynomial $d(X)$, as a divisor of the irreducible polynomial $f(X)$, must be equal to $f(X)$ up to some constant c in K. That is, $f(X) = c\,d(X)$ is a divisor of the polynomial $g(X)$.

Point 4. We also would like to make note of the following theorem.

Theorem 9.8. *The factorization of a polynomial into irreducible factors is unique up to the arrangement of the factors and multiplicative constants.*

For two factorizations $f_1(X)\cdots f_s(X) = g_1(X)\cdots g_r(X)$ into irreducible factors, the factor $f_1(X)$ must have a common zero with a factor $g_j(X)$. According to Theorem 9.7, therefore, $f_1(X)$ is a divisor of the polynomial $g_j(X)$ and conversely. Therefore, we have $f_1(X) = c\,g_j(X)$ for some number c in the field of coefficients K. If one now divides both sides of the original equation by $f_1(X)$, one can proceed, factor by factor, to complete the proof.

Literature on Galois Groups and Their History

H.-W. Alten (et al.), *4000 Jahre Algebra*, Berlin, 2003.

Edgar Dehn, *Algebraic Equations*, New York, 1960.

Harold M. Edwards, *Galois Theory*, New York, 1984.

Helmut Koch, *Einführung in die klassische Mathematik I*, Berlin, 1986.

Gerhard Kowol, *Gleichungen*, Stuttgart, 1990.

Ivo Radloff, Évariste Galois: principles and applications, *Historia Mathematica*, 29 (2002), pp. 114–137.

Jean-Pierre Tignol, *Galois' Theory of Algebraic Equations*, Singapore, 2001.

Exercises

(1) Determine using the Euclidean algorithm (in the "original" version for integers) the greatest common divisor of the integers 145673 and 2134197.

(2) Show that for an equation with rational coefficients and precisely two nonreal zeros, the Galois group over the rational numbers contains a permutation that exchanges the two nonreal solutions and leaves all other solutions unchanged.

(3) In what ways can the steps of the solution of the equation

$$x^4 - 2 = 0$$

be arranged? That is, over what fields between $\mathbb{Q}$ and $\mathbb{Q}\left(i, \sqrt[4]{2}\right)$ can the solution take place? Furthermore, provide for each chain of extension fields the associated decomposition of the Galois group.

Hint: Altogether, there are seven different possibilities for such chains of field extensions. One can establish this number for the associated decompositions of the Galois group with some effort. That this number carries over to the chains of field extensions will be shown in the next chapter.

(4) Determine the Galois group of the cyclotomic equation

$$x^{17} - 1 = 0$$

and show that there is precisely one stepwise decomposition process in the Galois group. Give as well the associated field extensions.

Chapter 10

Algebraic Structures and Galois Theory

The author of this book associates the following entry on Galois theory in the Brockhaus Encyclopedia, sixteenth edition, with his fruitless attempts as a fifteen-year-old student to understand why a general equation of the fifth degree should have no solution in radicals:

"According to Galois theory, solving an equation is equivalent to the construction of the field E over the field K of coefficients of the equation formed by adjoining the sought-for solutions. The set of permutations of the solutions of the equation induces a group of maps of E to itself (automorphisms) that leave unchanged the elements of K. By determining all the subgroups of this group of automorphisms, it is possible to construct the field E step by step using the subgroups that correspond to intermediate fields. The advantage of such a method is that the relations between fields, with their two operations, addition and multiplication, is replaced by relations among groups, which have a single operation."

What is the relation between this description of Galois theory and the material in the previous chapters?

10.1 This last chapter is meant to serve as a bridge between two points of view on Galois theory, namely, the "elementary" point of view of the previous chapters, focused on polynomials, and the "modern" point of view that became dominant at the beginning of the twentieth century. In the course of our discussion we shall show that this modern point of view, based as it is on abstract algebraic structures,

turns out to be easier to understand, despite, or indeed because of, its level of abstraction. To understand and appreciate this simple introduction to the modern theory, the reader should have had a semester course or the equivalent on abstract algebra and/or linear algebra and be familiar with such concepts as group, normal subgroup, quotient group, field, vector space, basis, dimension, homomorphism, and automorphism. Since this book is directed at readers who lack such preparation in whole or in part, the conceptual apparatus required for an understanding of the material will be introduced in bare outline as required in the course of the chapter.

As in the previous chapter, we begin with an nth-degree polynomial equation

$$x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_1x + a_0 = 0$$

with complex coefficients $a_{n-1}, \ldots, a_1, a_0$ and without multiple solutions. That is, the solutions $x_1, \ldots, x_n$ are assumed to be distinct. In contrast to the previous chapter, however, the field extension obtained by adjoining the solutions of the equations to the field of coefficients will be in the foreground: Again, we will begin with a field K that contains the coefficients of the equation. As we did in the previous chapter, we adjoin to this field K all the solutions $x_1, \ldots, x_n$, thereby obtaining the field $K(x_1, \ldots, x_n)$. This extension field was defined as the set of numbers that arise from applying the basic arithmetic operations to the elements of K and the solutions $x_1, \ldots, x_n$. As we shall see, the same set of numbers arises if we forgo division, so that every number of the field $K(x_1, \ldots, x_n)$ can be expressed as a polynomial in $x_1, \ldots, x_n$ with coefficients in K. We note finally that the field $K(x_1, \ldots, x_n)$ is called the *splitting field* of the given equation, since it is the smallest field in which the equation splits into linear factors.

10.2 The central notion of a field was defined in the previous chapter, where we limited ourselves, since it sufficed for our purposes, to subsets of the complex numbers closed under the four basic arithmetic operations. We would now like to go somewhat more deeply into the subject.[1] A brief look at the general definition of a field, one that

[1]From a historical point of view, fields were first defined by Richard Dedekind (1831–1916), in a paper that appeared in 1871. It was only twenty years later that the

goes beyond the complex numbers, is given in the section on groups and fields.

As we have seen in the previous chapter, the solution of a polynomial equation is intimately bound up with the fields that lie between K and $K(x_1, \ldots, x_n)$. Therefore, in what follows we shall attempt to classify systematically all such fields, and we shall do so in terms of the Galois group. But first, we need to come up with an alternative definition of the Galois group.

10.3 We have thus far used the word "group" as a component of the term "Galois group." Since in what follows we will need to go more deeply into the composition of permutations, we shall present a formal definition of a group, after stating the following theorem.

Theorem 10.1. *Every Galois group, when taken as a set of permutations together with the binary operation of composition of permutations, forms a group.*

That is, the Galois group satisfies the following definition.

Definition 10.2. A *group* consists of a set G on which is defined a binary operation $\circ$ (that is, for all $\sigma, \tau \in G$, it follows that $\sigma \circ \tau \in G$) such that the following conditions are satisfied:

- The binary operation is *associative*; that is, for $\sigma, \tau, \nu \in G$, one has $(\sigma \circ \tau) \circ \nu = \sigma \circ (\tau \circ \nu)$.
- The set G possesses an *identity element* ε such that $\varepsilon \circ \sigma = \sigma \circ \varepsilon = \sigma$ for all $\sigma \in G$.
- Each element σ of G possesses an *inverse*, denoted by σ^{-1}, under the operation $\circ$ such that $\sigma \circ \sigma^{-1} = \sigma^{-1} \circ \sigma = \varepsilon$.

We have already asserted in the previous chapter that the composition $\sigma \circ \tau$ of two permutations σ and τ of the Galois group again belongs to the Galois group, and we have demonstrated this by means of several group tables. The reason for this is simply that the set of

---

notion of a field was extended to encompass other mathematical systems with analogous properties, even those that could not be considered as subsets of the complex numbers, for example by Heinrich Weber (1842–1913) in 1893. More details and references can be found in Erhard Scholz, Die Entstehung der Galois-Theorie, in: Erhard Scholz (ed.), *Geschichte der Algebra*, Mannheim, 1990, pp. 365–398.

polynomials B_K in the definition of the Galois group is carried to itself by the action of τ and again by the action of σ:

$$(\sigma \circ \tau)(B_K) = \sigma(\tau(B_K)) \subset \sigma(B_K) \subset B_K.$$

One may also verify associativity easily. Indeed, it holds for arbitrary functions and mappings. Thus for three permutations σ, τ, ν of the Galois group and for a given index j, one has

$$\begin{aligned}(\sigma \circ (\tau \circ \nu))(j) &= \sigma((\tau \circ \nu)(j)) = \sigma(\tau(\nu(j))) = (\sigma \circ \tau)(\nu(j)) \\ &= ((\sigma \circ \tau) \circ \nu)(j).\end{aligned}$$

We shall make use of the associative law in what follows mostly implicitly, namely by leaving off parentheses entirely. It is thanks to the associative law that we can indulge in such "imprecise" notation.

It is clear that the identity element of the Galois group is the identity permutation, that is, the permutation that leaves every index in its place. It is clear that this permutation belongs to every Galois group.

It is almost as obvious that for every permutation σ there is an inverse permutation τ with the desired properties. One simply defines $\tau(j) = k$ for each index j, where the index k is uniquely determined by $\sigma(k) = j$. Then we have the identity $\tau \circ \sigma = \sigma \circ \tau = \mathrm{id}$. On the other hand, it is much less clear whether for every permutation σ in the Galois group its inverse $\tau = \sigma^{-1}$ necessarily belongs to the Galois group as well. The simplest argument that this is so makes use of the fact that in any finite group, the sequence of powers $\sigma, \sigma^2 = \sigma \circ \sigma, \sigma^3 = \sigma \circ \sigma \circ \sigma, \ldots$ must contain two terms that are equal. From $\sigma^p = \sigma^q$, with $p > q$, it follows that $\sigma^{p-q} = \mathrm{id}$, so that σ^{p-q-1} is a representation of the inverse element $\tau = \sigma^{-1}$.

Thus we have shown that every Galois group is indeed a group. Further examples of groups appear in the section on groups and fields.

10.4 If a group G contains a subset U that is closed under the group operation and inverses, one calls U a *subgroup* of G. We will require the following theorem.

Theorem 10.3. *If a finite group G contains the subgroup U, then $|U|$, the number of elements of U, is a divisor of $|G|$, the number of elements in the group G.*

In the special case of a Galois group, this fact was first recognized by Galois, and implicitly by Lagrange previously.[2] To prove the theorem, one forms for every element σ of the group G the collection of *left cosets* of σ, defined by

$$\sigma U = \{\, \sigma \circ \tau \mid \tau \in U \,\}.$$

If two products $\sigma \circ \tau_1$ and $\sigma \circ \tau_2$ in σU are equal, then it follows that $\tau_1 = \sigma^{-1} \circ \sigma \circ \tau_1 = \sigma^{-1} \circ \sigma \circ \tau_2 = \tau_2$. This shows that every coset has precisely $|U|$ elements, and thus they are all of the same size. Furthermore, the cosets constitute a disjoint (that is, nonoverlapping) partition of the group. This is so because if two cosets had a nonempty intersection,

$$\sigma_1 U \cap \sigma_2 U \neq \emptyset,$$

there would exist two elements τ_1, τ_2 in the subgroup U such that $\sigma_1 \tau_1 = \sigma_2 \tau_2$, with the result that $\sigma_1 U = \sigma_2 \circ \tau_2 \circ \tau_1^{-1} U = \sigma_2 U$. That is, the two overlapping cosets are in fact one and the same. Since the entire group G can be partitioned into a set of, say, n cosets all of the same size $|U|$, it follows that $|G| = n|U|$, and so $|U|$ must be a divisor of $|G|$.

We can draw two direct consequences of the theorem that we have just proved.

Corollary 10.4. *For every element σ of a finite group G the smallest positive integer n such that $\sigma^n = \varepsilon$, called the* order *of σ, is a divisor of the number $|G|$ of elements of G.*

The truth of the corollary can be seen at once if for a given element σ we consider the subgroup $\{\, \varepsilon, \sigma, \sigma^2, \ldots \}$. The first duplication $\sigma^p = \sigma^q$ that appears in this listing, which must occur since the group is finite, must hold, as we showed in Section 10.3, for $p = 0$ and $q = n$, that is, for $\varepsilon = \sigma^n$. The subgroup then contains precisely n elements, and this number must be a divisor of $|G|$.

---

[2]The quotient $|G|/|U|$ is called the *index* of the subgroup U in G. However, in what follows we shall not use this terminology, in favor of a more direct characterization.

Corollary 10.5. *A group that contains a prime number of elements n can be listed completely as $\{\varepsilon, \sigma, \sigma^2, \ldots, \sigma^{n-1}\}$ for a suitably chosen element σ. Such a group is called a* cyclic group *of order n.*

The validity of this corollary is easily seen. Simply choose for σ any element other than the identity element. Then consider the subgroup $\{\varepsilon, \sigma, \sigma^2, \ldots, \sigma^{k-1}\}$, where k is the order of the element σ. Since $k > 1$ and k is a divisor of n, it follows that $k = n$, and the subgroup is in fact the entire group.

10.5 We would like now to use the knowledge that we have gained to expand our investigations of the previous chapter. Let us begin with the situation in which the field K that we have been considering is extended to a field L. The consequent expansion of the set of polynomials B_K to a set of polynomials B_L results in a shrinking of the Galois group, as we have seen in the previous chapter. However, we may now add that this extension must lead to a subgroup, and thus the number of elements of the Galois group must shrink to one of its divisors, as we have seen through numerous examples in the previous chapter.

Groups and Fields

The concept of a group extends far beyond sets of permutations with the operation of composition. Here are some examples of groups:

- The integers $\mathbb{Z}$ under addition.
- The rational numbers $\mathbb{Q}$ under addition, as well as the real numbers $\mathbb{R}$ and the complex numbers $\mathbb{C}$, also under addition.
- The nonzero rational numbers $\mathbb{Q}^\times$ under multiplication, and also the nonzero real numbers $\mathbb{R}^\times$ and the nonzero complex numbers $\mathbb{C}^\times$, also under multiplication.
- The n-dimensional real vector space $\mathbb{R}^n$ with coordinatewise addition.
- The set of $n \times m$ real matrices under addition.
- The set of real $n \times n$ matrices with nonzero determinant under matrix multiplication.
- The set of real 3×3 matrices with nonzero determinant whose associated mappings take the set of vertices of a Platonic solid with center

the origin into itself. The operation for these symmetry transformations is matrix multiplication, which is equivalent to composition of mappings.

- For a positive integer $n \geq 2$, the set of n remainders $\{0, 1, 2, \ldots, n-1\}$ together with addition "modulo n" forms a group; this is because, for example for $n = 3$, it makes no difference which numbers from the three cosets

$$\{0, 3, \ldots, -3, -6, \ldots\},$$
$$\{1, 4, \ldots, -2, -5, \ldots\},$$
$$\{2, 5, \ldots, -1, -4, \ldots\},$$

one chooses and then adds. The resulting coset always depends only on the two cosets chosen, not on the chosen numbers themselves. The resulting group is called the *cyclic group* or order n and denoted by $\mathbb{Z}/n\mathbb{Z}$. The elements of the group $\mathbb{Z}/n\mathbb{Z}$, that is, the cosets, are called *residue classes modulo n*. We have seen these groups in an equivalent (isomorphic) form, namely as the groups of nth roots of unity.

- For a prime number n, the residue classes obtained from division by n, excluding the zero class, form a group under multiplication.

With the exception of the multiplicative matrix groups, the symmetry groups for Platonic solids, and the permutation groups, all the groups mentioned here are *abelian*, that is, the group operation is *commutative*, i.e., for every pair of elements σ, τ in the group, $\sigma \circ \tau = \tau \circ \sigma$.

A *field* is defined as a set K together with two operations, denoted usually by $+$ (addition) and $\cdot$ (multiplication), such that the following conditions are satisfied:

- The set K is an abelian group under addition, with the identity element denoted by 0.
- The set $K - \{0\}$ is an abelian group under multiplication, with the identity element being denoted by 1.
- The *distributive law* holds; that is, for any three elements x, y, z in K, the identity $x \cdot (y + z) = x \cdot y + x \cdot z$ holds.

The most familiar examples of a field are the sets of rational numbers $\mathbb{Q}$, real numbers $\mathbb{R}$, and complex numbers $\mathbb{C}$ under the usual addition and multiplication. Another example is the field $\mathbb{Q}(a, b, c, \ldots)$ constructed by the adjunction of complex numbers $a, b, c, \ldots$ to the field of rational numbers.

Another example of a field is that of the *rational functions* in the variables $X_1, \ldots, X_n$. Such a function is a fraction whose numerator and denominator are polynomials in the variables $X_1, \ldots, X_n$ with coefficients in a field K. The denominator may not be the zero polynomial, of course.[3]

[3]For clarity, it should be pointed out that the nomenclature "rational function" is historical and not entirely correct, since a rational function is a formal quotient of

One obtains a subfield by allowing only those rational functions for which the variables, denoted by $A_0, \ldots, A_{n-1}$, are replaced by the elementary polynomials in the variables $X_1, \ldots, X_n$.[4] The extension of the field to the field of all rational functions forms the basis for studying the general polynomial equation in the context of fields (and their automorphisms).

Among the examples of groups given above one can find some *finite fields*. For a prime number n, $\mathbb{Z}/n\mathbb{Z}$ is a field. In contrast to the other fields we have looked at, namely subfields of the complex numbers, finite fields have the following property: The n-fold sum of the multiplicative identity 1 is equal to zero. One speaks in such a case of a *finite characteristic*, or of characteristic n, in contrast to *characteristic zero*, which is the designation for all subfields of the complex numbers. With some modifications one can develop a Galois theory for finite fields and other fields not contained in the complex numbers, whose results can be of help in computing Galois groups of extensions of the field of rational numbers.

10.6 The next concept that we wish to discuss is that of automorphism. We begin with some motivation. The concept of an automorphism will enable us to characterize the Galois group in terms of extensions from the field K to $K(x_1, \ldots, x_n)$ instead of as previously in terms of the equation; in particular, two equations with coefficients in K having identical splitting fields have identical Galois groups.

Up to now, the elements of the Galois group were viewed exclusively as permutations of the solutions. However, Galois, and before him Lagrange, had made intensive use of the fact that these permutations can be viewed as functions that map each value that can be represented by a polynomial in the solutions $x_1, \ldots, x_n$ to another such value. Thus if a polynomial $h(X_1, \ldots, X_n)$ is given with coefficients in the base field K, then one defines, for a permutation σ belonging to the Galois group,

$$\sigma\big(h(x_1, \ldots, x_n)\big) = h\left(x_{\sigma(1)}, \ldots, x_{\sigma(n)}\right).$$

For example, a number z represented as $z = x_2^2 - x_1 x_2 x_3$ is assigned the function value $\sigma(z) = x_{\sigma(2)}^2 - x_{\sigma(1)} x_{\sigma(2)} x_{\sigma(3)}$. In general, the given

---

polynomials and not a function in the usual sense of the term, although a function can be easily defined on the basis of a rational function.

[4]This field contains all rational functions that are symmetric: For every such rational function, whose numerator and denominator do not need a priori to be symmetric, the denominator can be made symmetric by extending the fraction. But then the numerator of the extended fraction must also be symmetric.

definition of $\sigma(z)$ makes sense only because it is independent of the polynomial representation $x = h(x_1, \ldots, x_n)$, which is obviously never unique. Namely, if one has two polynomials h_1 and h_2 with identical values $h_1(x_1, \ldots, x_n) = h_2(x_1, \ldots, x_n)$, then the difference of these two polynomials belongs to the set B_K that was used in the previous chapter in defining the Galois group. And indeed, the Galois group was defined in such a way that the difference belonging to B_K results in the equality $h_1\left(x_{\sigma(1)}, \ldots, x_{\sigma(n)}\right) = h_2\left(x_{\sigma(1)}, \ldots, x_{\sigma(n)}\right)$, with the result that it makes no difference whether the value of $\sigma(z)$ is defined in terms of h_1 or h_2.

With the given definition we have thus succeeded in extending the definition of permutations of the set $\{ x_1, \ldots, x_n \}$ belonging to the Galois group to the set of values representable by polynomials in the solutions $x_1, \ldots, x_n$. As we have already mentioned, though without proof, this set is the splitting field $K(x_1, \ldots, x_n)$. But even without reference to this unproved fact, for the permutations belonging to the Galois group one can obtain an extension of the realm of definition to the entire splitting field $K(x_1, \ldots, x_n)$ if one allows rational functions in the polynomials in the solutions $x_1, \ldots, x_n$. No problems with vanishing denominators arise, since as we shall see, for $y \neq 0$ it is always the case that $\sigma(y) \neq 0$.

These extended permutations, in reference to their domain of definition, now satisfy the properties that we wish to invoke in the notion of a field automorphism. First, we see at once from the definition of the extended permutations two identities that hold for two arbitrary values y and z representable polynomially in the solutions $x_1, \ldots, x_n$:

$$\sigma(y + z) = \sigma(y) + \sigma(z),$$
$$\sigma(yz) = \sigma(y)\sigma(z).$$

More difficult is the proof that $\sigma(y) = 0$ can hold only for $y = 0$ and that the mapping σ is invertible.[5]

---

[5] Given a value y with $\sigma(y) = 0$ and a polynomial representation $y = g(t)$ associated with a Galois resolvent t, one has $0 = \sigma(y) = g(t_\sigma)$, that is, t_σ is a zero of the polynomial $g(T)$. In analogy to argumentation that we have used previously, it follows that the polynomial $g(t)$ is divisible by the irreducible polynomial $\mathfrak{G}(T)$ constructed by Galois. And therefore $g(T)$ also has t as a zero, that is, $0 = g(t) = y$.

We shall soon see that the mappings that we have constructed here are linear operators onto the finite-dimensional K-vector space $K(X_1, \ldots, X_n)$. Therefore, since $\sigma(y) \neq 0$ for $y \neq 0$, every one of the mappings σ is invertible.

Finally, it follows that $\sigma(y) = y$ for all $y \in K$. To see this, for determining the value $\sigma(y)$ of a permutation σ, one simply chooses the constant polynomial $h(X_1, \ldots, X_n) = y$.

It is important that the four given properties also serve conversely to determine a permutation in the Galois group. Given a function (more usual in this context would be the term *mapping*), denoted by σ, defined on the field $K(x_1, \ldots, x_n)$, one can use σ on both sides of the original equation. For $j = 1, \ldots, n$ one thereby obtains

$$\sigma(x_j)^n + a_{n-1}\sigma(x_j)^{n-1} + \cdots + a_1\sigma(x_j) + a_0 = 0,$$

that is, $\sigma(x_j)$ is also a solution of the equation. Since for $j \neq k$ one has

$$\sigma(x_j) - \sigma(x_k) = \sigma(x_j) + \sigma(-1)\sigma(x_k) = \sigma(x_j - x_k) \neq 0,$$

the values $\sigma(x_1), \ldots, \sigma(x_n)$ are distinct, so that one is acutally dealing with a permutation of the solutions. For a polynomial $h(X_1, \ldots, X_n)$ with coefficients in the field K and $h(x_1, \ldots, x_n) = 0$, one also has $h\big(\sigma(x_1), \ldots, \sigma(x_n)\big) = \sigma\big(h(x_1, \ldots, x_n)\big) = 0$. Thus the permutation defined by the function σ in fact belongs to the Galois group.

With the proof of this equivalence complete, we have obtained a third characterization of the Galois group. In addition to the original definition in terms of the polynomial set B_K and the characterization in terms of the Galois resolvent, as described in the section on computing the Galois group in Chapter 9, we have the following theorem.

Theorem 10.6. *A Galois group of a given equation over the field K can be obtained as the set* $\mathrm{Aut}\big(K(x_1, \ldots, x_n) \,|\, K\big)$*, the set of all automorphisms of the splitting field $K(x_1, \ldots, x_n)$ that leave unchanged all elements of the field K.*

The term *automorphism* is defined as follows.

Definition 10.7. An *automorphism* of a field L is an invertible mapping σ that assigns to each value $y \in L$ a value $\sigma(y) \in L$ such that the following conditions are satisfied:

$$\sigma(y + z) = \sigma(y) + \sigma(z),$$
$$\sigma(yz) = \sigma(y)\sigma(z).$$

In this third characterization of the Galois group the original equation appears only implicitly, namely, in the form of the field extension of K to $K(x_1, \ldots, x_n)$ in terms of the solutions $x_1, \ldots, x_n$. As we have stated, two equations with coefficients in K and identical splitting fields automatically have the same Galois group. Another advantage of this third characterization, which nowadays is generally taken as *the* definition of the Galois group, is its universality. Automorphisms can be studied for field extensions that do not correspond to a splitting field, though in such a case, the properties of interest to us here are to be found only in part.

10.7 We now would like to investigate the properties of the Galois group, where from now on we shall adopt the definition of the Galois group as a group of automorphisms $G = \mathrm{Aut}\left(K(x_1, \ldots, x_n) \mid K\right)$, and therefore we no longer need to assume that the solutions of the equation are all distinct.

The first of three important theorems that we are going to derive were known in essence by Galois.

Theorem 10.8. *If one adjoins to a field K all the roots $x_1, \ldots, x_n$ of an equation with coefficients in K, then the set of values in the splitting field $K(x_1, \ldots, x_n)$ that remain unchanged by every automorphism in the Galois group is the field K.*

We thus need to show that a value z for which $\sigma(z) = z$ for all automorphisms σ in the Galois group belongs to the field K. We have seen the argumentation that we need for the proof aleady in the special situation of solving the cyclotomic equation.[6]

[6] In terms of our previous discussions, a proof can be most simply given using the Galois resolvent t, for which we have proved the properties

$$x_{\sigma(1)} = g_1(t_\sigma), \quad \ldots, \quad x_{\sigma(n)} = g_n(t_\sigma).$$

If we consider σ as an automorphism, then these identities can be reformulated as

$$\sigma(x_1) = g_1(t_\sigma), \quad , \ldots, \quad \sigma(x_n) = g_n(t_\sigma).$$

For a value $z = h(x_1, \ldots, x_n)$ that remains unchanged under all automorphisms σ of the Galois group, one sums the different representations of the value z and obtains in this way, where $|G|$ is the number of elements in the Galois group,

$$|G|\, z = \sum_\sigma \sigma(z) = \sum_\sigma h\big(\sigma(x_1), \ldots, \sigma(x_n)\big) = \sum_\sigma h\big(g_1(t_\sigma), \ldots, g_n(t_\sigma)\big);$$

10.8 It is high time that we solidified the theoretical discussion with some concrete examples. For the quadratic equation

$$x^2 - 6x + 1 = 0,$$

the two solutions are used to extend the field of the rational numbers $\mathbb{Q}$ to the field

$$\mathbb{Q}(\sqrt{2}) = \left\{ a + b\sqrt{2} \mid a, b \in \mathbb{Q} \right\}.$$

The Galois group consists of two permutations, where the non-identity permutation σ_1 exchanges the two solutions $3 \pm 2\sqrt{2}$. It can be extended to the mapping $(a, b \in \mathbb{Q})$

$$\sigma_1 \left(a + b\sqrt{2} \right) = a - b\sqrt{2}.$$

We wish to use this example to offer another interpretation of automorphisms. That is, the automorphisms of the Galois groups are *linear transformations*, where the extension field is viewed as a *vector space* defined over the base field K. Those readers familiar with these concepts will certainly understand what we mean. For the remaining readers with limited knowledge of analytic geometry and coordinatewise calculations with *vectors*, and perhaps even *matrices*, the vector representation

$$\sigma_1 \left[\begin{pmatrix} a \\ b \end{pmatrix} \right] = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a \\ -b \end{pmatrix},$$

related to the *basis* $\{ 1, \sqrt{2} \}$, is more suggestive. However, we must stress that such coordinatewise representations are of use only for a bit of emphasis. The advantage in using constructs from linear algebra consists in allowing one to avoid such explicit representations, which always depend on the particular choice of basis. It is important only that such representations exist and that the number of elements in a basis is independent of the particular choice. This invariant of a vector space, known as the *dimension*, is of course associated with

---

because of the symmetry in the values t_σ, the sum on the right-hand side can be expressed in terms of the polynomial that has these values t_σ as zeros. This polynomial is the irreducible factor $\mathfrak{G}(T)$ divisible by $T - t$, as was constructed by Galois from the nth-degree resolvent equation (see the section in Chapter 9 on computing the Galois group). Since the coefficients of this polynomial $\mathfrak{G}(T)$ are in the field K, the value z also belongs to K.

every other field extension, where one speaks of the *degree* of the extension. We have then the following definition.

Definition 10.9. The *degree* of an extension field E of a base field K is equal to the natural number m that is the size of a basis of E as a vector space over K; that is, one can find exactly m values $e_1, \ldots, e_m$ in E such that every value in the field E can be expressed uniquely in the form

$$k_1 e_1 + \cdots + k_m e_m$$

with values (coordinates) $k_1, \ldots, k_m$ from the field K.[7]

For example, the degree of the field extension $\mathbb{Q}(\sqrt{2})$ of $\mathbb{Q}$ is equal to 2. Another example is the adjunction of the fifth root of unity $\zeta = \cos\left(\frac{2\pi}{5}\right) + i\sin\left(\frac{2\pi}{5}\right)$ to obtain the field $\mathbb{Q}(\zeta)$. To show that the degree of the extension from $\mathbb{Q}$ to $\mathbb{Q}(\zeta)$ is 4, we consider the equation, already investigated in Chapters 7 and 9,

$$x^4 + x^3 + x^2 + x + 1 = 0,$$

whose four solutions are the complex fifth roots of unity $\zeta, \zeta^2, \zeta^3, \zeta^4$. A basis of the vector space, which also allows a simple multiplication and division in the field $\mathbb{Q}(\zeta)$, is the four values $1, \zeta, \zeta^2, \zeta^3$. The proof that these four values actually form a basis is most instructive, since the arguments used can easily be generalized. First, we observe that every number that can be expressed as a polynomial in the rational numbers and the root of unity ζ is of the form $k_0 1 + k_1\zeta + \cdots + k_s\zeta^s$ with rational coefficients $k_0, k_1, \ldots$. However, since $\zeta^4 = -1 - \zeta - \zeta^2 - \zeta^3$, one can always find a representation with $k_4 = k_5 = \cdots = 0$. In this case, the coefficients k_0, k_1, k_2, k_3 are uniquely determined. For otherwise, that is, if there were two different coordinate representations for one and the same value, one would have for the root of unity ζ an equation of at most the third degree with rational coefficients. Since the above

---

[7] Those who wish to "experiment" with this definition may wish to prove the following degree formula for towers of extension fields:

Theorem 10.10. *For a nested tower of fields $K \subset L \subset E$, the degree of the total extension of E over K is equal to the product of the degree of E over L and that of L over K.*

To prove the theorem, one selects two bases for the two intermediate extensions, forms all products of one element from each basis, and convinces oneself that the result is a basis of the total extension.

This degree formula is used in the proofs of the impossibility of the classical construction problems; see the section on this topic at the end of the chapter.

fourth-degree equation is irreducible, none of its solutions can be a solution of an equation of lower degree (see Theorem 9.7).

It remains only to show that the set

$$\left\{ k_0 + k_1\zeta + k_2\zeta^2 + k_3\zeta^3 \mid k_0, k_1, k_2, k_3 \in \mathbb{Q} \right\}$$

is closed under division, so that this set is in fact the field $\mathbb{Q}(\zeta)$. Indeed, for two polynomials $f(X)$ and $g(X)$ with rational coefficients, if $g(\zeta) \neq 0$, we have

$$\frac{f(\zeta)}{g(\zeta)} = \frac{f(\zeta)g\left(\zeta^2\right)g\left(\zeta^3\right)g\left(\zeta^4\right)}{g(\zeta)g\left(\zeta^2\right)g\left(\zeta^3\right)g\left(\zeta^4\right)}.$$

The fraction on the right-hand side of this identity has the desired representation of the form $k_0 1 + k_1\zeta + k_2\zeta^2 + k_3\zeta^3$, since the denominator is rational. To see this, one could multiply out the denominator for a number of the form $g(\zeta) = k_0 1 + k_1\zeta + k_2\zeta^2 + k_3\zeta^3$. Fortunately, one can avoid such drudgery if one uses the generally valid argument that will be presented in the following section.

10.9 To generalize the result that we have established for the field $\mathbb{Q}(\zeta)$, we would like to prove the following theorem.

Theorem 10.11. *If one adjoins to a field K all the solutions of an equation with coefficients in K, that is, $x_1, \ldots, x_n$, then the degree of this field extension is equal to $|G|$, the number of elements in the Galois group $G = \mathrm{Aut}\left(K(x_1, \ldots, x_n) \mid K\right)$.*

To prove this theorem, we shall use an argument using the Galois resolvent t that is completely analogous to that used for the field $\mathbb{Q}(\zeta)$, whereby every value in the field $K(x_1, \ldots, x_n)$ can be represented by a quotient in which numerator and denominator take the form $k_0 + k_1 t + \cdots + k_m t^m$ and all coefficients k_j belong to the field K. Moreover, the highest power m can be restricted to the value $|G| - 1$, since $|G|$ is the degree of the polynomial $\mathfrak{G}(T)$, which has t as a zero. It remains to show that every such quotient can be expressed as a polynomial in t. To that end, we consider in the $|G|$-dimensional K-vector space $K[t] = \left\{ k_0 + \cdots + k_{|G|-1}t^{|G|-1} \mid k_0, \ldots, k_{|G|-1} \in K \right\}$ the associated mapping

$$h(t) \in K[t] \mapsto g(t)h(t)$$

defined by multiplication by an element $g(t) \in K$. This mapping is linear. Furthermore, the interpretation of the mapping as multiplication shows that in the case $g(t) \neq 0$, no nonzero element is mapped to zero. Results from linear algebra on systems of linear equations tell us that every element of $K[t]$ possesses a preimage, where in particular, the preimage of the number 1 is the number $1/g(t)$. This preimage allows us to represent expressions with $g(t)$ in the denominator as polynomials in t, so that every value in the splitting field $K(x_1, \ldots, k_n)$ can be expressed polynomially in terms of the Galois resolvent t. The theorem is proved.

Moreover, in considering $t = m_1x_1 + \cdots + m_nx_n$, one sees that every number in the splitting field $K(x_1, \ldots, x_n)$ can be expressed as a polynomial in the solutions $x_1, \ldots, x_n$. Thus in adjoining the solutions, one can do without the operation of division. This fact, which was previously mentioned but not proved, has now been established.

10.10 The close relation that we have observed between the Galois group and its underlying field extension will now be extended to the following theorem.

Theorem 10.12. *If one adjoins the solutions $x_1, \ldots, x_n$ of an equation with coefficients in a field K, then in the resulting splitting field $K(x_1, \ldots, x_n)$, the set of values that remain unchanged under the action of every automorphism of a subgroup U of the Galois group $G = \mathrm{Aut}\left(K(x_1, \ldots, x_n) \mid K\right)$ is equal to K only when U is the full Galois group.*

That all elements of the field K remain unchanged under all automorphisms of the subgroup U is clear. What is important, however, is the converse: If aside from the field K there are no elements that are unchanged by every automorphism of the subgroup U, then this subgroup must be the full Galois group. In other words, every proper subgroup of the Galois group leaves unchanged some elements not in K.

To prove this theorem, one begins with a subgroup U of the Galois group in which the elements of the field K are the only numbers in the extension field $K(x_1, \ldots, x_n)$ that remain fixed by all the automorphisms of the Galois group. With the Galois resolvent t one

forms the polynomial

$$\prod_{\sigma \in U} (X - \sigma(t)).$$

If one applies an automorphism τ of the subgroup U to the coefficients of this polynomial, this corresponds to a permutation of the linear factors, so that the coefficients remain unchanged.[8] According to our assumption, the polynomial's coefficients must lie in the field K, since they are invariant under all automorphisms of U. If U were a proper subgroup, then the Galois resolvent t would be a solution of an equation with coefficients in K whose degree was less than the number of elements $|G|$ of the Galois group, which contradicts the fact that the Galois resolvent t is a zero of $\mathfrak{G}(T)$ according to Galois's construction, that is, of an irreducible polynomial of degree $|G|$ over K.

A quite similar line of reasoning can be used to prove the fact noted in Section 9.8 that the Galois group G of an irreducible equation always acts transitively on the solutions, that is, that for every pair of solutions x_j, x_k there is an automorphism σ of the Galois group for which $\sigma(x_j) = x_k$.[9]

10.11 The three theorems of Sections 10.7, 10.9, and 10.10 now allow us to prove in a few quick strokes the fundamental theorem of Galois theory. Given our journey thus far, it may come as a surprise that this theorem has no immediate relation to the question whether an equation is solvable in radicals. Rather, it establishes a one-to-one correspondence between the subgroups of the Galois group $\mathrm{Aut}\left(K(x_1, \ldots, x_n) \mid K\right)$ and the fields lying between K and $K(x_1, \ldots, x_n)$. The fundamental theorem thereby makes a coherent theory out of all of our investigations and examples of the previous

[8] For two distinct automorphisms σ_1 and σ_2 of U, it follows that $\tau \circ \sigma_1$ and $\tau \circ \sigma_2$ are also distinct. Furthermore, every automorphism ν of U is obtained in this way, namely, in the form $\tau \circ \left(\tau^{-1} \circ \nu\right) = \nu$.

[9] If one applies an automorphism τ of the Galois group to the coefficients of the polynomial

$$\prod_{\sigma \in G} (X - \sigma(x_j)),$$

the result is a permutation of the linear factors, so that the coefficients remain unchanged. By Theorem 10.8, the coefficients must lie in K. Because of the irreducibility of the original polynomial and Theorem 9.7, all the solutions, and in particular the solution x_k, must appear in the product.

chapter. See, for example, Figure 9.1 in Chapter 9. Since subgroups of a finite group are relatively easy to find—indeed, in the worst case one can try out a finite number of possibilities—one obtains from the fundamental theorem a complete classification of intermediate fields. It is then possible in special cases to obtain immediate information on which intermediate fields, if any, arise by the adjunction of roots.

We now state the fundamental theorem of Galois theory.

Theorem 10.13. *If one adjoins to a subfield K of the complex numbers all the solutions $x_1, \ldots, x_n$ of an equation with coefficients in K, then the Galois group $G = \mathrm{Aut}\left(K(x_1, \ldots, x_n) \mid K\right)$ of the field extension thereby obtained, that is, the group of all automorphisms of the field $K(x_1, \ldots, x_n)$ that leave the base field K fixed, possesses the properties enumerated below. In detail, these properties are with regard to the intermediate fields L, that is, fields such that $K \subset L \subset K(x_1, \ldots, x_n)$, and to each such field the associated subgroup $\mathrm{Aut}\left(K(x_1, \ldots, x_n) \mid L\right)$, which comprises all automorphisms of the Galois group that leave every element of L fixed. (See also Figure 10.1.)*

(1) *The mapping that associates the subgroup $\mathrm{Aut}\left(K(x_1, \ldots, x_n) \mid L\right)$ with the intermediate field L establishes a one-to-one correspondence (that is, a* bijection*) between the intermediate fields and the subgroups of the Galois group G.*

(2) *The degree of the field extension from L to $K(x_1, \ldots, x_n)$ is equal to the number of elements in the associated subgroup*

$$\mathrm{Aut}\left(K(x_1, \ldots, x_n) \mid L\right)$$

of the Galois group. This is the number of automorphisms that fix every element of L.

(3) *If an intermediate field $L = K(y_1, \ldots, y_m)$ is obtained by adjoining to K the roots $y_1, \ldots, y_m$ of an equation with coefficients in K, all lying in the field $K(x_1, \ldots, x_n)$, then the Galois group $\mathrm{Aut}(L \mid K)$ contains $|G| / \left|\mathrm{Aut}\left(K(x_1, \ldots, x_n) \mid L\right)\right|$ automorphisms. One can thus obtain all automorphisms of this Galois group $\mathrm{Aut}(L \mid K)$ by restricting the domain of definition $K(x_1, \ldots, x_n)$ of the automorphisms belonging to G to the intermediate field $L = K(y_1, \ldots, y_m)$.*

$$
\begin{array}{ccc}
K(x_1, \ldots, x_n) & & \{\mathrm{id}\} \\
\cup & & \cap \\
\vdots & & \vdots \\
\cup & & \cap \\
L & \overset{1:1}{\longleftrightarrow} & U = \mathrm{Aut}\left(K(x_1, \ldots, x_n) \mid L\right) \\
\cup & & \cap \\
\vdots & & \vdots \\
\cup & & \cap \\
K & & G = \mathrm{Aut}\left(K(x_1, \ldots, x_n) \mid K\right)
\end{array}
$$

Figure 10.1. The fundamental theorem of Galois theory: The intermediate fields L, that is, the fields L such that $K \subset L \subset K(x_1, \ldots, x_n)$, are in one-to-one correspondence with the subgroups U of the Galois group $G = \mathrm{Aut}\left(K(x_1, \ldots, x_n) \,|\, K\right)$.

A proof of the fundamental theorem consists basically in referring to the theorems already proven in Sections 10.7, 10.9, and 10.10. We investigate the extension of an intermediate field L to the field $L(x_1, \ldots, x_n) = K(x_1, \ldots, x_n)$ using those three theorems. It remains to note that the fields $L(x_1, \ldots, x_n)$ and $K(x_1, \ldots, x_n)$ are identical, since the field L is simultaneously an extension field of K and a subfield of $K(x_1, \ldots, x_n)$.

With these observations out of the way, one part of the assertion, namely, that an intermediate field L is uniquely determined by the associated subgroup $\mathrm{Aut}\left(K(x_1, \ldots, x_n) \mid L\right)$ (that is, that the mapping is *injective*), is clear, since it is a consequence of Theorem 10.8, according to which L can be characterized as the set of all elements of the field $K(x_1, \ldots, x_n)$ that are fixed by all automorphisms of $\mathrm{Aut}\left(K(x_1, \ldots, x_n) \mid L\right)$. To show that the mapping is in fact a bijection, we need to show that for every subgroup U of the Galois group there is always a corresponding intermediate field L (thus that the mapping is *surjective*). We can construct such a field L with the desired property $U = \mathrm{Aut}\left(K(x_1, \ldots, x_n) \mid L\right)$ by taking all elements of $K(x_1, \ldots, x_n)$ that are fixed by every automorphism in U. That is,

$$L = \{\, z \in K(x_1, \ldots, x_n) \mid \sigma(z) = z \text{ for all } \sigma \in U \,\}.$$

That the set L thus defined is in fact a field—on account of construction it is referred to as a *fixed field*—can easily be shown by checking that for every pair of elements of L, their sum, product,

quotient, and difference are again in L. It is clearer that the fixed field L contains the field K and is a subfield of $K(x_1, \ldots, x_n)$. And now comes the main point: The group $U = \mathrm{Aut}\left(K(x_1, \ldots, x_n) \mid L\right)$ associated with the intermediate field L thus constructed contains the group U, since every automorphism in U fixes every element of L. Moreover, this subgroup has the property, given the construction of the field L, that only elements of L are fixed by all the automorphisms of the subgroup. According to Theorem 10.12, this has the consequence that the subgroup U must equal the entire Galois group $\mathrm{Aut}\left(K(x_1, \ldots, x_n) \mid L\right)$.

The second part of the fundamental theorem follows at once from Theorem 10.11 applied to the extension of L to $K(x_1, \ldots, x_n)$.

The third part of the fundamental theorem refers to the situation that we have examined by means of a number of examples in the previous chapter when *all* the solutions $y_1, \ldots, y_m$ of a resolvent equation are adjoined to the field K. The elements of such intermediate fields L are mapped again to L because of the special conditions on all automorphisms σ of the Galois group G. To see this, one need only apply such an automorphism σ of the Galois group $G = \mathrm{Aut}\left(K(x_1, \ldots, x_n) \mid K\right)$ to both sides of the underlying equation of the intermediate field L. This shows that each of the solutions y_j is mapped by σ to another solution $\sigma(y_j) = y_k$.

Since the field L is mapped to itself by all the automorphisms of the Galois group G, for each automorphism in G one can restrict its definition from the field $K(x_1, \ldots, x_n)$ to the intermediate field $L = K(y_1, \ldots, y_m)$, thereby obtaining an automorphism of the group $\mathrm{Aut}(L \mid K)$. Here two automorphisms $\sigma, \tau \in G$ yield the same restriction precisely when $\sigma^{-1} \circ \tau$ is the identity on L, that is, when $\sigma^{-1} \circ \tau$ belongs to the subgroup $\mathrm{Aut}\left(K(x_1, \ldots, x_n) \mid L\right)$ associated with the intermediate field L. That one obtains all automorphisms in this way follows most simply from the degree formula for nested field extensions. The fundamental theorem of Galois theory is now completely proven.

The Fundamental Theorem of Galois Theory: An Example

As an example of computing all the intermediate fields between the base field and the splitting field, we return to the biquadratic equation that we made use of in Chapter 9, namely,

$$x^4 - 4x^3 - 4x^2 + 8x - 2 = 0.$$

The four solutions of this equation are, as previously stated,

$$x_{1,3} = 1 + \sqrt{2} \pm \sqrt{3 + \sqrt{2}},$$
$$x_{2,4} = 1 - \sqrt{2} \pm \sqrt{3 - \sqrt{2}}.$$

The Galois group was determined in Chapter 9 to be a group with eight elements. If the elements are considered permutations, they act on the solutions as follows:

	1	2	3	4
σ_0	1	2	3	4
σ_1	3	2	1	4
σ_2	1	4	3	2
σ_3	3	4	1	2
σ_4	2	1	4	3
σ_5	4	1	2	3
σ_6	2	3	4	1
σ_7	4	3	2	1

The possible subgroups can be determined by trial and error. In addition to the whole group and the group consisting of the identity alone, any subgroup must have either two or four elements. The two-element subgroups arise from each of the five elements of order two:

$$\{\sigma_0, \sigma_1\}, \ \{\sigma_0, \sigma_2\}, \ \{\sigma_0, \sigma_3\}, \ \{\sigma_0, \sigma_4\}, \ \{\sigma_0, \sigma_7\}.$$

In addition, one can easily find, using the group table computed in the previous chapter, three subgroups with four elements:

$$\{\sigma_0, \sigma_1, \sigma_2, \sigma_3\}, \ \{\sigma_0, \sigma_3, \sigma_4, \sigma_7\}, \ \{\sigma_0, \sigma_3, \sigma_5, \sigma_6\}.$$

It remains to observe that the last of these groups is cyclic of order four, while the other two four-element subgroups are isomorphic to the group $(\mathbb{Z}/2\mathbb{Z})^2$.

According to the fundamental theorem of Galois theory, there is a one-to-one correspondence between this collection of subgroups and the fields

that lie between the two fields $\mathbb{Q}$ and

$$\mathbb{Q}(x_1, x_2, x_3, x_4) = \mathbb{Q}\left(\sqrt{3+\sqrt{2}}, \sqrt{3-\sqrt{2}}\right).$$

Moreover, these fields can be determined from the subgroups by determining the fixed fields. To do so, we shall use the identities

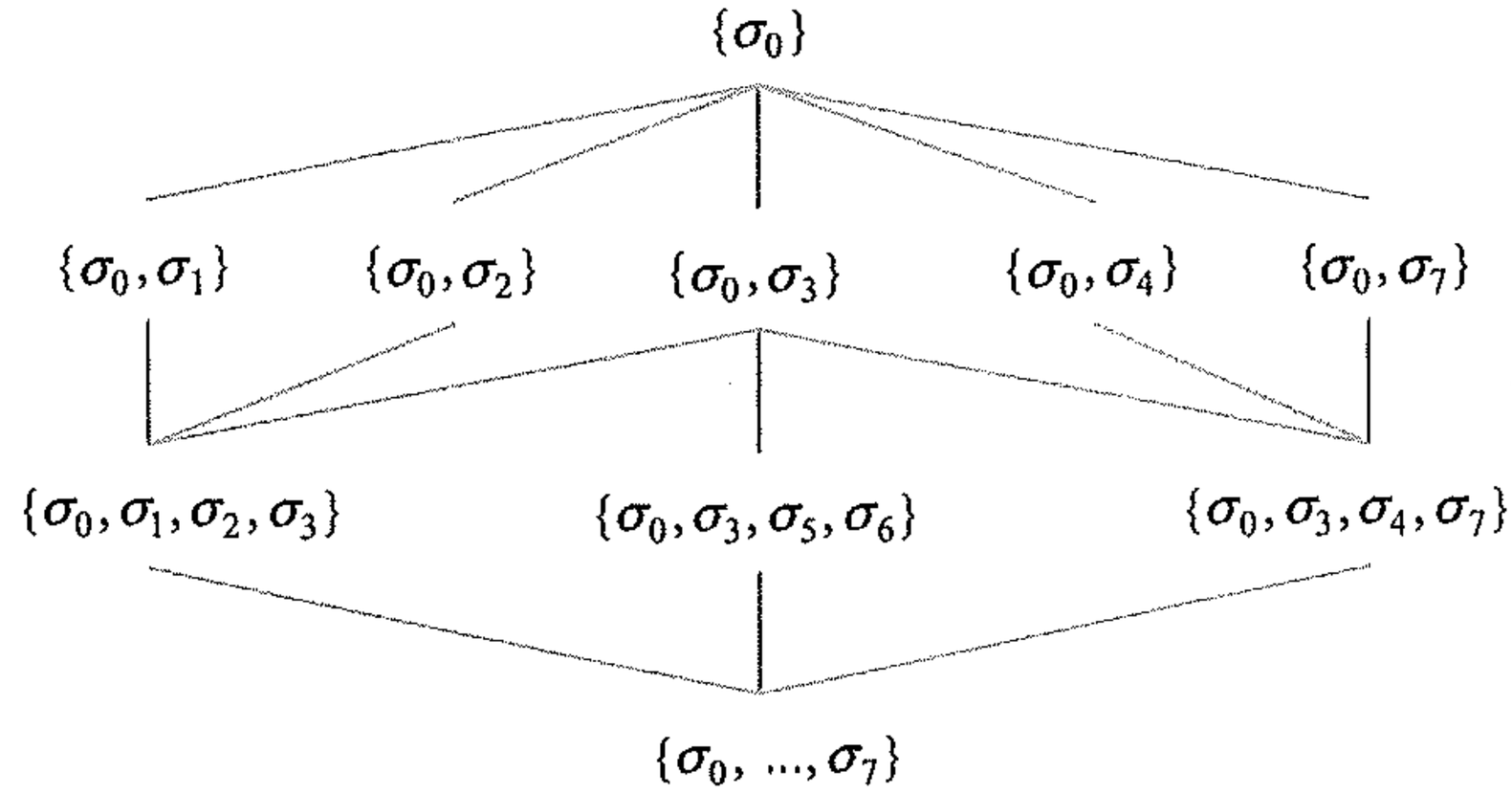
$$\begin{aligned}
\sqrt{2} &= \frac{1}{4}(x_1 - x_2 + x_3 - x_4),\\
\sqrt{3+\sqrt{2}} &= \frac{1}{2}(x_1 - x_3),\\
\sqrt{3-\sqrt{2}} &= \frac{1}{2}(x_2 - x_4),\\
\sqrt{7} &= \frac{1}{4}(x_1 - x_3)(x_2 - x_4).
\end{aligned}$$

With these identities we can determine the images of these numbers under the automorphisms directly from the tabulated permutations. As a direct consequence, the fixed fields can be determined. In Figure 10.2, the subgroups are represented together with the corresponding intermediate fields showing the inclusion relations among the various objects.

10.12 Now that we have proved the fundamental theorem of Galois theory, we would like to say a bit more about its significance. As we have seen, for every intermediate field L there corresponds precisely one subgroup G of the Galois group $\mathrm{Aut}\left(K(x_1, \ldots, x_n) \mid L\right)$ and conversely.

A particular intermediate field L will not necessarily be mapped to itself by an automorphism σ of the Galois group G. But since the image $\sigma(L)$, as one can readily check, is necessarily a field, and thus another intermediate field, it must have a corresponding subgroup. And this subgroup can in fact be determined directly from the subgroup U associated with L. Namely, the group must be a *conjugate* subgroup of the form $\sigma U \sigma^{-1}$. Its automorphisms leave an element $\sigma(z)$ unchanged precisely when the automorphisms of the subgroup U leave the value z unchanged. This is illustrated in Figure 10.3.

If for a subgroup U all the conjugate subgroups are simply the subgroup U, then U is called a *normal subgroup*. In the situation here, in reference to the third item in the fundamental theorem of Galois theory, this implies that every automorphism of $K(x_1, \ldots, x_n)$



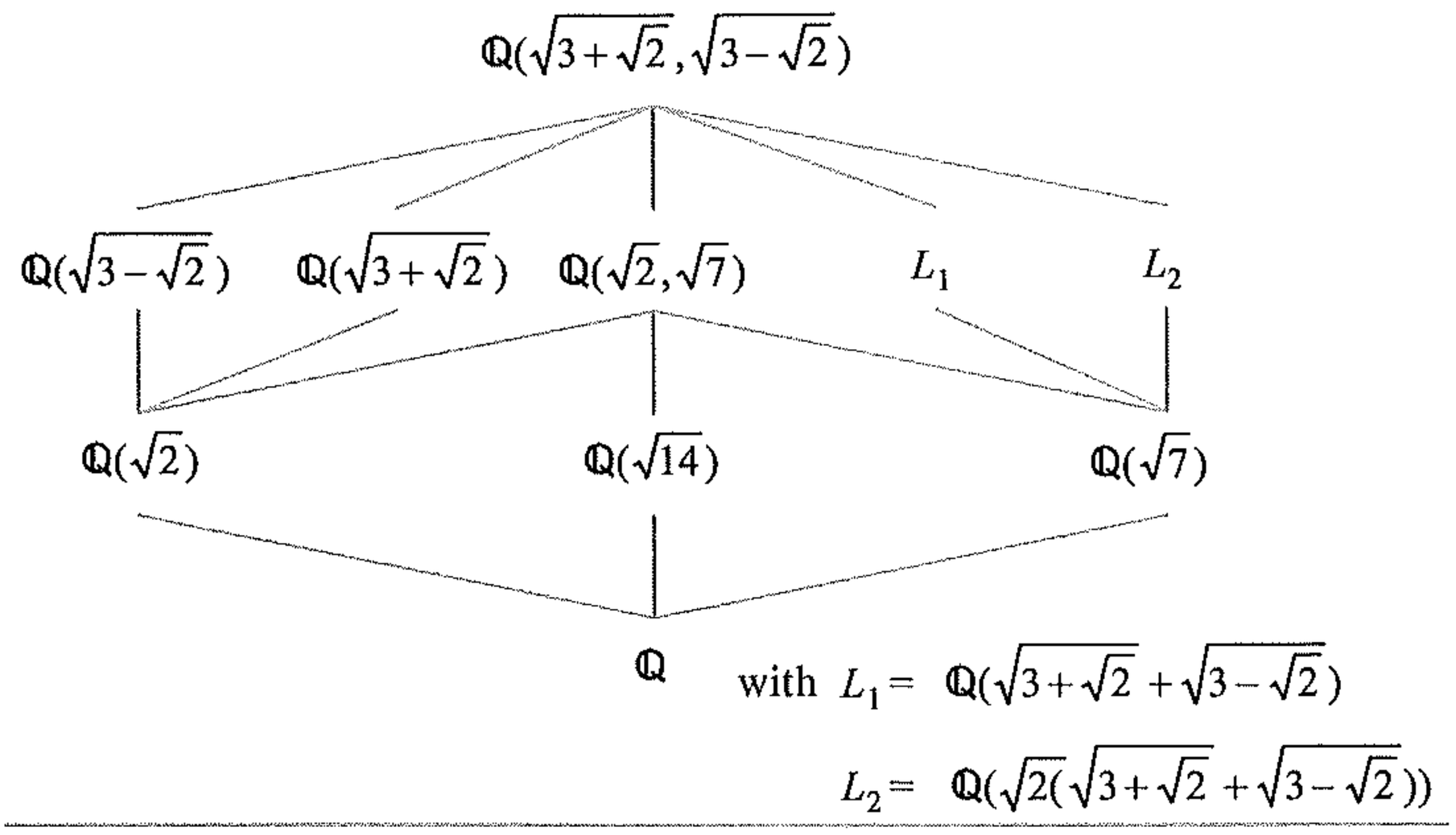


Figure 10.2. The subgroups of the Galois group are in one-to-one correspondence with the fields between the base field and the splitting field.

restricted to L yields an automorphism of L.[10] We have already mentioned that two automorphisms σ and τ in $\mathrm{Aut}\left(K(x_1,\ldots,x_n) \mid K\right)$ are the same when restricted to L precisely when $\sigma^{-1}\circ\tau$ is the identity mapping on L and therefore belongs to the subgroup $U = \mathrm{Aut}\left(K(x_1,\ldots,x_n) \mid L\right)$. Then the Galois group $\mathrm{Aut}(L \mid K)$ is at once recognizable as the set G/U of cosets of the subgroup U, where

[10] An example of a subgroup that is not equal to a conjugate subgroup, thus leading to two distinct intermediate fields, can be found in a previously examined example (see Figure 10.2), namely $\mathbb{Q}\left(\sqrt{3+\sqrt{2}}\right)$ and $\mathbb{Q}\left(\sqrt{3-\sqrt{2}}\right)$.

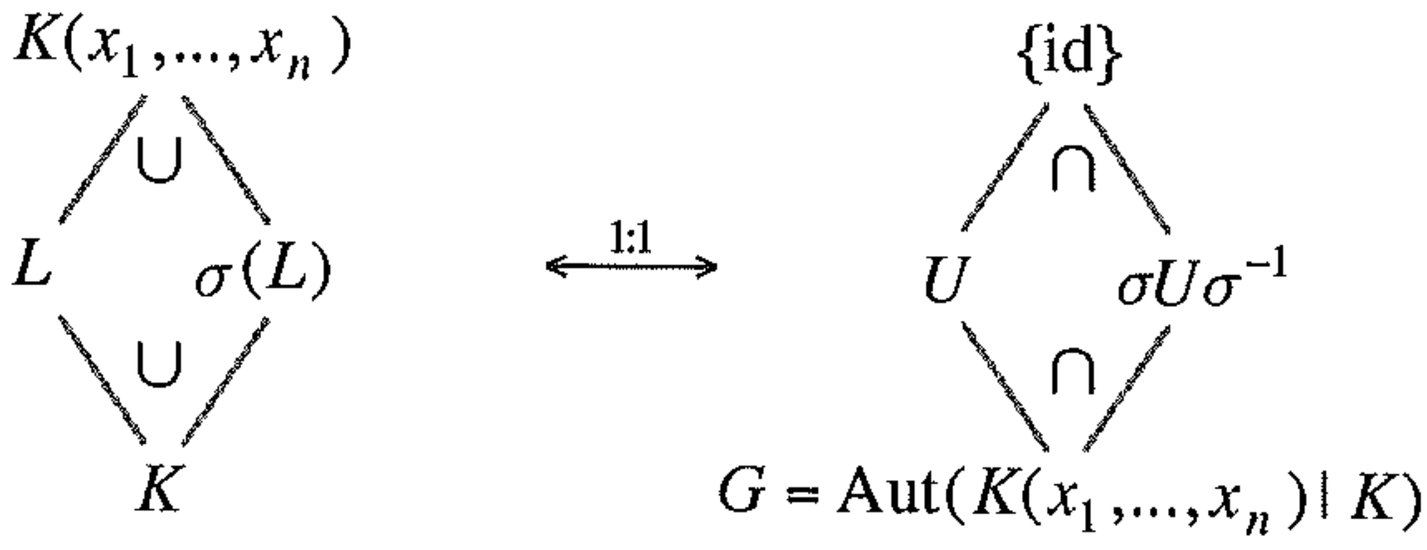


Figure 10.3. The fundamental theorem of Galois theory, showing the relationships among intermediate fields L, their images $\sigma(L)$ under automorphisms σ of the Galois group, and their associated subgroups. Due to the one-to-one correspondence between intermediate fields and subgroups, $L = \sigma(L)$ holds precisely when $U = \sigma U \sigma^{-1}$.

of course the composition of automorphisms within the group $\mathrm{Aut}(L \mid K)$ can be derived from the composition of the group G.

10.13 There is a construction principle at the heart of what we have just described that can be greatly generalized, even beyond the bounds of Galois theory, by defining a composition for the cosets formed from a normal subgroup U of a group G:

$$(\sigma_1 U) \circ (\sigma_2 U) = (\sigma_1 \circ \sigma_2) U.$$

The fact that U is a normal subgroup is what guarantees that this definition is well defined. That is, the definition does not depend on the choice of the elements σ_1 and σ_2: One must check that the result of a composition remains unchanged when σ_1 or σ_2 is replaced by $\sigma_1 \circ \tau$ or $\sigma_2 \circ \tau$ for some automorphism τ in U. For σ_2, this is clear. However, for σ_1, it holds only on account of $((\sigma_1 \circ \tau) \circ \sigma_2)U = \left(\sigma_1 \circ \sigma_2 \circ \left(\sigma_2^{-1} \circ \tau \circ \sigma_2\right)\right) U$, with $\sigma_2^{-1} \circ \tau \circ \sigma_2 \in U$ for $\tau \in U$.

Much less surprising than such a definition is the fact that the composition thus defined leads to the set of cosets G/U being a group in its own right. In the section on groups and fields we have already made use of such notation with the example $\mathbb{Z}/n\mathbb{Z}$, where the subgroup $n\mathbb{Z}$ is obviously a normal subgroup because the group $\mathbb{Z}$ is abelian (commutative).

We have already encountered such *quotient groups* G/U in the previous chapter, in the form of decompositions of the group table

(see also Figure 10.4). Moreover, in the proof of the fundamental theorem of Galois theory we have been involved with such mechanisms, when we restricted automorphisms of the Galois group of the field $K(x_1, \ldots, x_n)$ to the intermediate field $L = K(y_1, \ldots, x_m)$: Whenever the subgroup U was a normal subgroup, the intermediate field L was mapped to itself by all the automorphisms of G, so that the restriction of the definition to L yields an automorphism group $\mathrm{Aut}(L \mid K)$ with $|G|/|U|$ elements.

G:

σ \ τ	σ_0	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7
σ_0	σ_0	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7
σ_1	σ_1	σ_0	σ_3	σ_2	σ_6	σ_7	σ_4	σ_5
σ_2	σ_2	σ_3	σ_0	σ_1	σ_5	σ_4	σ_7	σ_6
σ_3	σ_3	σ_2	σ_1	σ_0	σ_7	σ_6	σ_5	σ_4
σ_4	σ_4	σ_5	σ_6	σ_7	σ_0	σ_1	σ_2	σ_3
σ_5	σ_5	σ_4	σ_7	σ_6	σ_2	σ_3	σ_0	σ_1
σ_6	σ_6	σ_7	σ_4	σ_5	σ_1	σ_0	σ_3	σ_2
σ_7	σ_7	σ_6	σ_5	σ_4	σ_3	σ_2	σ_1	σ_0

G/U:

σ \ τ	σ_0	σ_1	σ_2	σ_3
σ_0	σ_0	σ_1	σ_2	σ_3
σ_1	σ_1	σ_0	σ_3	σ_2
σ_2	σ_2	σ_3	σ_0	σ_1
σ_3	σ_3	σ_2	σ_1	σ_0

U:

Figure 10.4. The decomposition of the group table into four 4×4 squares, as results from the example shown in Figure 9.1 of Chapter 9, indeed for the first adjunction shown in that figure. Above is shown the quotient group G/U constructed from the cosets.

The great significance of Point 3 of the fundamental theorem of Galois theory (Theorem 10.13) is above all that complicated field extensions can be broken down into a succession of suitable subextensions, and indeed, in such a way that with the first extension a situation is achieved that satisfies the initial assumptions. Here the

requirement for an intermediate field resulting from the adjunction of all the solutions of an equation with coefficients in K is equivalent to the normal group property of associated subgroups.[11] For a clearer overview, Figure 10.5 shows the objects involved in the fundamental theorem of Galois theory together with some of its assertions.

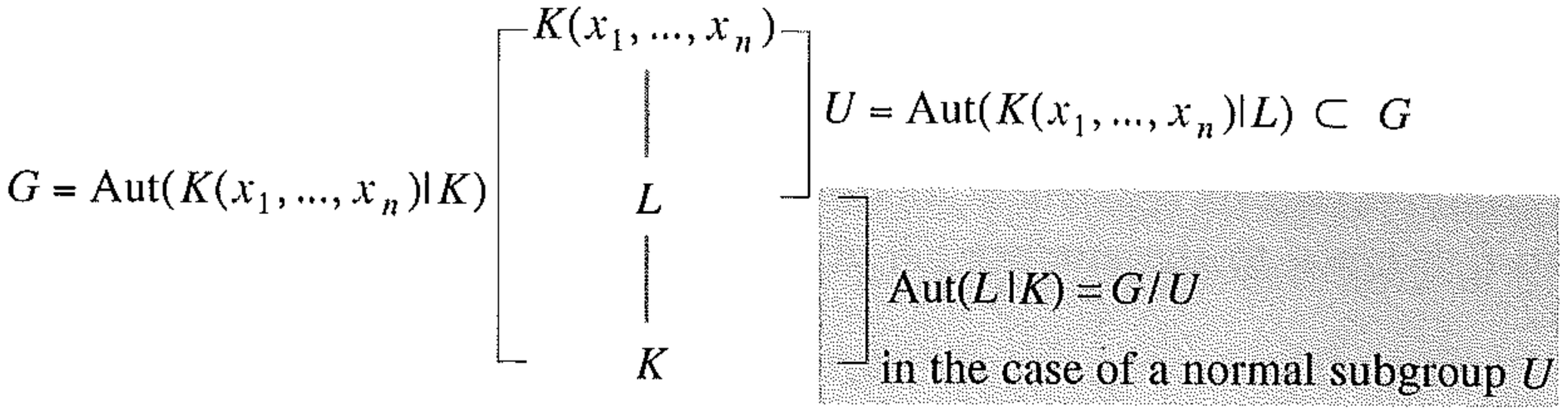


Figure 10.5. The fundamental theorem of Galois theory: an intermediate field L and the associated subgroup U. In the case that U is a normal subgroup (or equivalently, if L is a splitting field), the Galois group $\mathrm{Aut}(L \mid K)$ can also be determined, namely, as the quotient group G/U. We have yet to determine which properties of a normal subgroup U characterize the situation in which the extension of K to L can be achieved through the adjunction of a root $\sqrt[n]{a}$.

Artin's Version of the Fundamental Theorem of Galois Theory

In his 1942 book *Galois theory*, Emil Artin presented a proof of the fundamental theorem of Galois theory quite different from the one presented here.[12] In contrast to earlier proofs, Artin's proof is accomplished without use of the Galois resolvent (or without use of the corresponding theorem of the primitive element). Instead, Artin built up Galois theory in a way that

[11] For an intermediate field $L = K(y_1, \ldots, y_m)$ whose associated subgroup U is a normal subgroup of the Galois group $\mathrm{Aut}(L \mid K)$, one may form, for each index $j = 1, \ldots, m$, the equation

$$\prod_{\sigma \in G/U} (X - \sigma(y_j)) = 0,$$

whose coefficients are in the field K and all of whose solutions belong to the field L. Then in addition to $y_1, \ldots, y_m$, the solutions $\sigma(y_j)$ can be adjoined, without the field $L = K(y_1, \ldots, y_m)$ having to be enlarged.

[12] For a discussion, see B. L. van der Waerden, Die Galois-Theorie von Heinrich Weber bis Emil Artin, *Archive for History of Exact Sciences*, 9 (1972), pp. 240–248.

makes it possible to prove the fundamental theorem, reformulated only in some of its assumption, using exclusively concepts from linear algebra. In particular, he investigates the set of solutions of systems of linear equations. An argumentation involving polynomials and their solutions is then unnecessary for a proof of the fundamental theorem in this variant (though it is, of course, for its subsequent interpretation).

The situation investigated by Artin is the following: Given a field[13] E and a finite group G of automorphisms of the field, a field K is constructed consisting of those elements of E that are fixed by all automorphisms of G.

For this situation, Artin first proves, by analyzing various systems of equations that he constructs, that the degree of the field extension from K to E is equal to the number $|G|$ of automorphisms constituting the group G.

With this theorem in hand, Artin's version of the fundamental theorem of Galois theory is relatively easy to prove. In a version simplified as to terminology, it goes like this: In the situation of a field E, a finite group G of automorphisms of this field, and a fixed field K for the group G, the following assertions hold:

- The mapping that associates each intermediate field L with the subgroup U of automorphisms in G that are the identity on L is a bijection, that is, a one-to-one mapping, between the intermediate fields and the subgroups of G.
- The degree of the field extension of L to E is equal to the number of elements $|U|$ of the subgroup U.
- The subgroup U is a normal subgroup of G precisely when the field K can also be represented as a fixed field of a group of automorphisms of the intermediate field L.

10.14 In the remaining sections of this chapter we would like to fill some of the holes left in the exposition of the previous chapter. To this end we will consider, in light of the fundamental theorem of Galois theory, what structure a Galois group of an equation should have in order that the equation be solvable in radicals. In particular, we must specialize the situation in Point 3 of the theorem to the case of a successive adjunction of roots, where each radicand belongs to a field that arises from one of the previous adjunction steps. Since root expressions can always be reformulated so that only roots with

[13] The theorem is valid for all fields, not only those that are subfields of the complex numbers.

prime degree are involved, it suffices for us to restrict our attention to prime degrees.

We begin with the case of a one-step adjunction, in which an nth root $\sqrt[n]{a}$ is adjoined to a field K, with the radicand a in K. For Point 3 of the fundamental theorem to be applicable, its conditions must be satisfied. To ensure that this is the case, we assume that the nth roots of unity $\zeta, \zeta^2, \ldots, \zeta^{n-1}$ are already in the field K, so that all solutions of the equation $x^n - a = 0$ lie in the field $K(\sqrt[n]{a})$.

If one now extends the field K, which for some prime n contains all nth roots of unity, to a field L by adjoining all the solutions of an equation with coefficients in K, then using the Galois group one can determine whether this extension can be obtained through adjunction of an nth root $\sqrt[n]{a}$ of some number a in K. Namely, we have the following theorem.

Theorem 10.14. *Given a field K that contains all nth roots of unity $\zeta, \zeta^1, \ldots, \zeta^{n-1}$ for a prime number n and an extension field L that arises from the adjunction of all solutions of an equation with coefficients in K, there exists in L an nth root $\sqrt[n]{a}$ such that $a \in K$ but $\sqrt[n]{a} \notin K$ and such that $L = K(\sqrt[n]{a})$ if the Galois group $\mathrm{Aut}(L \mid K)$ is cyclic of order n, that is, if $\mathrm{Aut}(L \mid K) = \{\mathrm{id}, \sigma, \sigma^2, \ldots, \sigma^{n-1}\}$ for a suitably chosen automorphism σ.*

To prove this theorem, we begin with the situation $L = K(\sqrt[n]{a})$. In this case, every automorphism σ of the Galois group

$$\mathrm{Aut}\left(K\left(\sqrt[n]{a}\right) \mid K\right)$$

is uniquely determined by its effect on the element $\sqrt[n]{a}$. Note that on account of $\left(\sigma\left(\sqrt[n]{a}\right)\right)^n = \sigma(a) = a$, the element $\sqrt[n]{a}$ must be mapped by σ to $\zeta^k \sqrt[n]{a}$ for some exponent k. Moreover, exponents add under compositions of automorphisms: If $\sigma\left(\sqrt[n]{a}\right) = \zeta^k \sqrt[n]{a}$ and $\tau\left(\sqrt[n]{a}\right) = \zeta^j \sqrt[n]{a}$, then $(\sigma \circ \tau)\left(\sqrt[n]{a}\right) = (\tau \circ \sigma)\left(\sqrt[n]{a}\right) = \zeta^{k+j} \sqrt[n]{a}$. The Galois group therefore "corresponds to," that is, is isomorphic to, a subgroup of the cyclic group $\mathbb{Z}/n\mathbb{Z}$. Since n is a prime number, the Galois group must consist of one automorphism or of all n of them. Since $\sqrt[n]{a} \notin K$, the first possibility is excluded, so that the Galois group can be given as

$$\mathrm{Aut}\left(K\left(\sqrt[n]{a}\right) \mid K\right) = \left\{\mathrm{id}, \sigma, \sigma^2, \ldots, \sigma^{n-1}\right\}$$

with $\sigma\left(\sqrt[n]{a}\right) = \zeta\sqrt[n]{a}$.

We now begin, conversely, with an extension of K to a field L whose Galois group is equal to $\mathrm{Aut}(L \mid K) = \left\{\mathrm{id}, \sigma, \sigma^2, \ldots, \sigma^{n-1}\right\}$ for a suitably chosen automorphism σ. For every element b of the field L one can form the Lagrange resolvent, which we met in Chapters 5 and 7:

$$(\zeta, b) = b + \zeta\sigma(b) + \zeta^2\sigma^2(b) + \cdots + \zeta^{n-1}\sigma^{n-1}(b).$$

From the definition, one immediately obtains the identity

$$\sigma\big((\zeta, b)\big) = \sigma(b) + \zeta\sigma^2(b) + \zeta^2\sigma^3(b) + \cdots + \zeta^{n-1}\sigma^n(b) = \zeta^{-1} \cdot (\zeta, b),$$

and thus $\sigma\left((\zeta, b)^n\right) = (\zeta, b)^n$, so that $(\zeta, b)^n$ must be in the field K. If one finds an element b in the field L whose Lagrange resolvent (ζ, b) is not equal to zero, then since $\sigma^j\big((\zeta, b)\big) = \zeta^{-j}(\zeta, b)$, none of the automorphisms other than the identity leaves all the elements of the field $K\big((\zeta, b)\big)$ unchanged. From the fundamental theorem of Galois theory, the field $K\big((\zeta, b)\big)$ can therefore not be a proper subfield of L; that is, $K\big((\zeta, b)\big) = L$. Therefore, the field L arises from the field K by adjunction of an nth root of an element of the field K, namely, $a = (\zeta, b)^n$.

We have still to prove that one can always find in L an element b whose Lagrange resolvent (ζ, b) does not vanish. We can extend the selection to Lagrange resolvents $\left(\zeta^k, b\right)$, with $k = 1, 2, \ldots, n-1$, for an arbitrary nth root of unity different from 1, since such Lagrange resolvents, if they are nonzero, can be used in accord with the previous construction. If one now forms the sum of the Lagrange resolvents

$$\left(\zeta^k, b\right) = b + \zeta^k\sigma(b) + \zeta^{2k}\sigma^2(b) + \cdots + \zeta^{(n-1)k}\sigma^{n-1}(b),$$

for the exponents $k = 0, 1, \ldots, n-1$, one obtains, since $1 + \zeta^j + \zeta^{2j} + \cdots + \zeta^{(n-1)j} = 0$ (for $j = 1, \ldots, n-1$),

$$\sum_{k=1}^{n-1} \left(\zeta^k, b\right) = nb.$$

Were all the Lagrange resolvents $\left(\zeta^k, b\right)$ for the exponents $k = 1, 2, \ldots, n-1$ to vanish, then one would have the equality $(1, b) = nb$, since only the first summand in the above sum would remain. Since the value $(1, b)$ remains unchanged under all automorphisms, the

number $(1, b)/n$ must lie in the field K. Thus every choice of element b not belonging to K leads to at least one nonvanishing Lagrange resolvent (ζ^k, b).

10.15 From this theorem we can prove an immediate corollary by making use again of Point 3 of the fundamental theorem of Galois theory.[14] This corollary answers the question as to the circumstances in which for the solution of an equation, an intermediate field can be generated by adjunction of a root. Such a field is called a *radical extension.*

Corollary 10.15. *If a field K containing all nth roots of unity for a prime number n is extended to a field L by adjoining all solutions of an equation with coefficients in K, then there is a normal subgroup U of the Galois group $G = \mathrm{Aut}(L \mid K)$ with $|G| = n|U|$ precisely when in the extension field L there is an element b that is not itself in K but such that its nth power b^n is in K (from which it follows that the field $K(b)$ is a radical extension of K lying within L).*

We have only to recall that the criterion just formulated has already been mentioned in the terminology of group tables (see Section 9.7). Whether the extension of the first radical extension can be correspondingly decomposed can then be read off analogously from the subgroup $U = \mathrm{Aut}(L \mid K(b))$. Because of the necessary assumptions about the roots of unity, the criterion must, however, be seen as unsatisfactory if we are seeking a direct answer as to whether the solutions of an equation can be expressed in terms of radicals. We need particularly to explain what change in the Galois group is caused when the required roots of unity are adjoined.

10.16 To be able to use the previous theorem, the roots of unity of the relevant degree must already belong to the base field K. To be sure, in the examples of the previous chapter we often had the rational numbers $\mathbb{Q}$ as the base field, on the basis of which we initiated our investigations into determination of the Galois group. To use the previous theorem, it is necessary first to adjoin appropriate elements—elements that do not necessarily belong to the splitting

[14] Including the extension discussed in Section 10.13.

field $K(x_1, \ldots, x_n)$ of the equation under investigation. Then instead of the field K, there appears an extension field K', and analogously, instead of the splitting field $K(x_1, \ldots, x_n)$ of the current equation, the field $K'(x_1, \ldots, x_n)$. We shall now see how the change in the Galois group takes place, as illustrated in Figure 10.6.

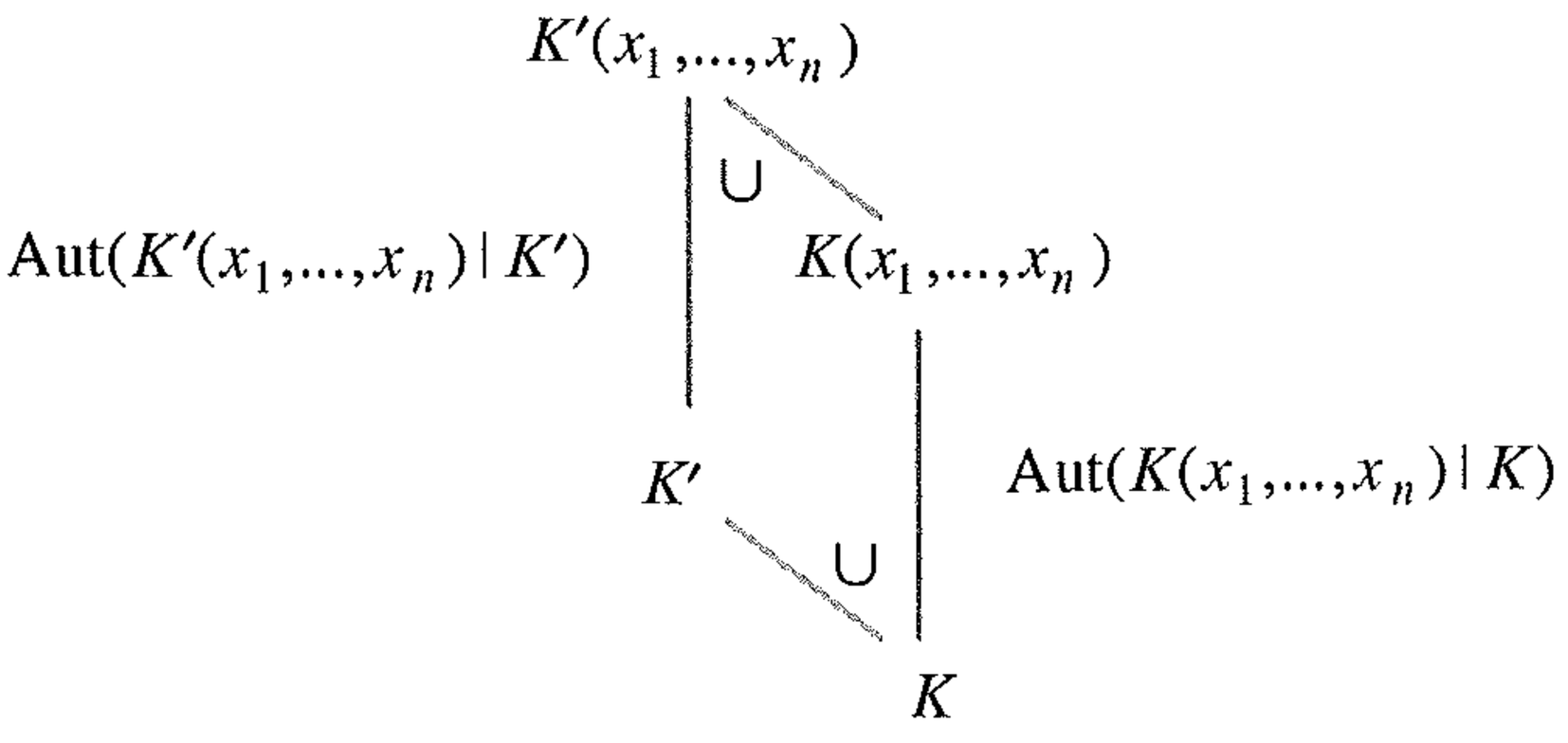


Figure 10.6. Extension of the base field K to a field K' and the resulting change in the Galois group.

An automorphism of the "new" Galois group

$$\mathrm{Aut}\left(K'(x_1, \ldots, x_n) \mid K\right)$$

is determined by its images of the solutions $x_1, \ldots, x_n$. Thus it is uniquely determined by the possible restriction of the domain of definition to the field $K(x_1, \ldots, x_n)$. Then the Galois group

$$\mathrm{Aut}\left(K'(x_1, \ldots, x_n) \mid K'\right)$$

is a subgroup of the original Galois group $\mathrm{Aut}\left(K(x_1, \ldots, x_n) \mid K\right)$.[15]

10.17 The cyclotomic equation $x^n - 1 = 0$ with prime degree n represents a very instructive and at the same time important application of these considerations. Indeed, we have already mentioned, in Chapter 7, the stepwise path to solving such cyclotomic equations, but it is appropriate now to do this again in a more systematic fashion. In particular, we wish to use the results of the previous section to prove the following theorem.

[15] It remains to observe that this subgroup relation, which results from an extension of the field K to the field K', is just the relation that Galois established in his original investigations.

Theorem 10.16. *The cyclotomic equation $x^n - 1 = 0$ with prime degree n is solvable in radicals; that is, each of its solutions can be represented by an expression involving nested roots and rational numbers.*

If ζ is an nth root of unity different from 1, then the field extension from $\mathbb{Q}$ to $\mathbb{Q}(\zeta)$ is of degree $n-1$, as we have seen already in Section 10.8, where $\zeta, \zeta^2, \ldots, \zeta^{n-1}$ is a vector-space basis. Each of the automorphisms σ of the Galois group is uniquely determined by the value of $\sigma(\zeta)$, which is always a power of ζ. Therefore, the Galois group $\mathrm{Aut}(\mathbb{Q}(\zeta) \mid \mathbb{Q})$ consists precisely of the automorphisms determined by $\sigma_k(\zeta) = \zeta^k$ for $k = 1, 2, \ldots, n-1$.

To obtain expressions in radicals for the solutions of the cyclotomic equation, we seek intermediate fields of the extension from $\mathbb{Q}$ to $\mathbb{Q}(\zeta)$ that correspond to the steps of a solution in radicals. Based on the fundamental theorem of Galois theory, there exist corresponding subgroups of the Galois group $\mathrm{Aut}(\mathbb{Q}(\zeta) \mid \mathbb{Q})$. However, to determine these is by no means simple, unless the powers ζ^k are represented again as in Chapter 7 in the form ζ^{g^j}, where g is a primitive root modulo n.[16] The Galois group can then be written in the form

$$\left\{ \mathrm{id}, \sigma_g, \sigma_g^2, \ldots, \sigma_g^{n-1} \right\},$$

so that for each divisor f of $n-1$, setting $e = \frac{n-1}{f}$, we obtain one (and only one) subgroup with f elements, namely,

$$U_e = \left\{ \mathrm{id}, \sigma_g^e, \sigma_g^{2e}, \ldots, \sigma_g^{e(f-1)} \right\}.$$

One obtains the associated subfield by taking a generic element z from the field $\mathbb{Z}(\zeta)$, representing it in coordinate form

$$z = m_1\zeta + m_2\zeta^2 + \cdots + m_{n-1}\zeta^{n-1}$$

with rational coordinates $m_1, \ldots, m_{n-1}$ and then checking under what circumstances this element z remains unchanged under the automorphism σ_g^e. This is the case precisely for

$$m_{g^0} = m_{g^e} = m_{g^{2e}} = \cdots, \quad m_{g^1} = m_{g^{e+1}} = m_{g^{2e+1}} = \cdots, \quad \ldots,$$

[16] From a purely group-theoretic point of view, a primitive root modulo n yields an isomorphism between the cyclic group $\mathbb{Z}/(n-1)\mathbb{Z}$ and the multiplicative group $\mathbb{Z}/n\mathbb{Z}-\{0\}$. This isomorphism makes it much easier to find the subgroups of $\mathbb{Z}/n\mathbb{Z}-\{0\}$.

so that the element z can be expressed in terms of the f-member period as described in Chapter 7:

$$z = m_{g^0}\eta_0 + m_{g^1}\eta_1 + \cdots + m_{g^{e-1}}\eta_{e-1}.$$

The f-member periods

$$\eta_0 = P_f(\zeta), \quad \eta_1 = P_f\left(\zeta^g\right), \quad \ldots, \quad \eta_{e-1} = P_f\left(\zeta^{g^{e-1}}\right),$$

now possess the property that they are fixed by all the automorphisms of the subgroup U_e on the one hand, and on the other are changed by every automorphism not in U_e. Therefore, the fields $\mathbb{Q}(\eta_0), \mathbb{Q}(\eta_1), \ldots, \mathbb{Q}(\eta_{e-1})$ must all be the same according to the fundamental theorem of Galois theory. Now for a solution of the cyclotomic equation in radicals, it suffices to find, for every possible period length f, an expression in radicals for a single f-member period, say η_0, since the other f-member periods can be calculated from η_0 using the four basic arithmetic operations.

The steps that make possible the solution of the nth-degree cyclotomic equation in radicals can now be planned on the basis of a decomposition of the integer $n-1$ into (not necessarily distinct) prime factors: $n-1 = p_1 p_2 \cdots p_s$. One may assume inductively that all the cyclotomic equations of degree p_j have been solved in radicals. We shall denote by K' the field that arises from the field $\mathbb{Q}$ by adjoining these roots of unity. Beginning with the increasing chain of fields

$$\begin{aligned}\mathbb{Q} &\subset \mathbb{Q}\left(P_{(n-1)/p_1}(\zeta)\right) \subset \mathbb{Q}\left(P_{(n-1)/(p_1p_2)}(\zeta)\right) \subset \cdots \\ &\subset \mathbb{Q}\left(P_{p_s}(\zeta)\right) \subset \mathbb{Q}(\zeta),\end{aligned}$$

where the Galois group of each extension step is cyclic of degree p_j, one then considers the chain of extension fields

$$\begin{aligned}K' &\subset K'\left(P_{(n-1)/p_1}(\zeta)\right) \subset K'\left(P_{(n-1)/(p_1p_2)}(\zeta)\right) \subset \cdots \\ &\subset K'\left(P_{p_s}(\zeta)\right) \subset K'(\zeta).\end{aligned}$$

According to the results of Section 10.6, each extension step has a Galois group that is a subgroup of the corresponding group for the original chain of fields. However, groups with a prime number of elements have only themselves and the one-element group as subgroups. Therefore, for each *actual* extension step of the second chain, the Galois group is the same as that of the cyclic Galois group of the

corresponding step in the first chain. Using Lagrange resolvents, the extension step can therefore be generated by adjunction of a single root.

Altogether, we have shown that cyclotomic equations can be solved in radicals. It remains to observe that in comparison to the considerations of Chapter 7, here complex calculations have been completely avoided. However, the price paid is a higher degree of abstraction.

10.18 We now come finally to the criterion (and its proof) that allows one to determine whether an equation is solvable in radicals. The basis of the theorem is the notion of solvability of a group, which is defined as follows.

Definition 10.17. A finite group G is called *solvable* if there is a chain of groups

$$\{\mathrm{id}\} = G_0 \subset G_1 \subset G_2 \subset \cdots \subset G_{k-1} \subset G_k = G$$

for which the subgroup G_j is a normal subgroup of the next group in the chain G_{j+1}, such that the quotient group G_{j+1}/G_j is cyclic of prime order.[17]

Calling a group thus defined solvable makes sense only because the solvability of an equation and the solvability of a group are so closely related. We have the following theorem, whose statement in terms of group tables was given in the previous chapter.

Theorem 10.18. *An equation is solvable in radicals, that is, all of its solutions can be expressed in terms of nested roots whose radicands can be expressed in terms of the coefficients using the four basic operations, if and only if its Galois group is solvable.*

To prove the theorem, we begin with a solvable equation. The field containing the coefficients can be extended stepwise through adjunction of roots of prime degree to a field that contains the solutions $x_1, \ldots, x_n$.[18] We shall look at only a single step, in which the field K

[17] Apparently weaker, but in fact equivalent, is a modification in the definition whereby the quotient group has only to be abelian.

[18] It is by no means excluded that the field extensions lead outside the field $K(x_1, \ldots, x_n)$.

arising from the previous radical extensions is extended by adjunction of pth roots, p a prime, of an element of K to a field L. We assume such a sequence of adjunction steps in which the necessary radical extensions for the solution of the pth cyclotomic equation have already been carried out, so that K contains the pth roots of unity. Now one sees that the four fields K, L, $K(x_1, \ldots, x_n)$, and $L(x_1, \ldots, x_n)$ are in relation to one another as shown in Figure 10.7. In particular, the degree of the extension of the field $K(x_1, \ldots, x_n)$ to $L(x_1, \ldots, x_n)$ is either 1 or p, depending on whether the adjoined pth root of unity is or is not an element of $K(x_1, \ldots, x_n)$.

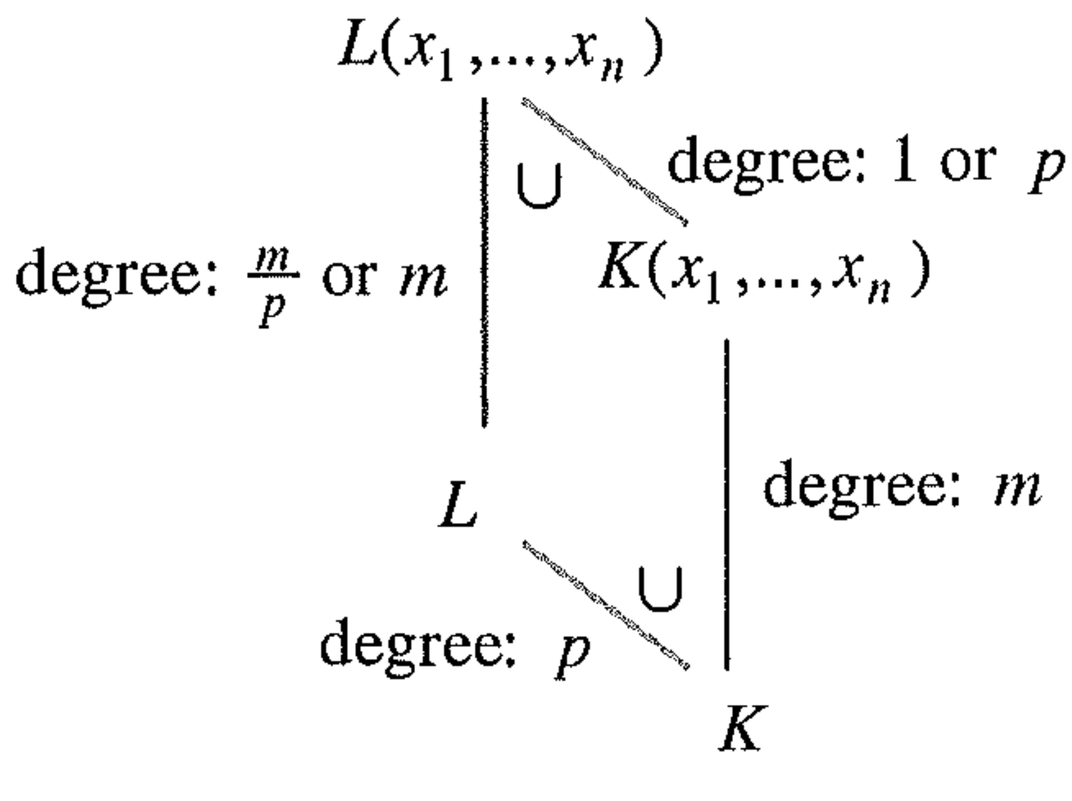


Figure 10.7. Extension of the field K to a field L by the adjunction of pth roots.

These two cases appear as follows:

- In the case

$$K(x_1, \ldots, x_n) = L(x_1, \ldots, x_n),$$

the field L is a subfield of the field $K(x_1, \ldots, x_n)$. The Galois group $\mathrm{Aut}\left(K(x_1, \ldots, x_n) \mid L\right)$ is then, by Point 3 in the fundamental theorem, a normal subgroup of the Galois group $\mathrm{Aut}\left(K(x_1, \ldots, x_n) \mid K\right)$, where the associated quotient group is cyclic with order the prime p.

- In the second case, in which the field extension of $K(x_1, \ldots, x_n)$ to $L(x_1, \ldots, x_n)$ has degree p, we have that the Galois group $\mathrm{Aut}\left(L(x_1, \ldots, x_n) \mid L\right)$ is "equal" to the Galois group

$$\mathrm{Aut}\left(K(x_1, \ldots, x_n) \mid K\right);$$

that is, all automorphisms of the latter group result from those of the former when its domain of definition is restricted.

Step by step, that is, with additional adjunctions of roots to the field L, one obtains the desired chain

$$\cdots \subset \mathrm{Aut}\left(L(x_1,\ldots,x_n) \mid L\right) \subset \mathrm{Aut}\left(K(x_1,\ldots,x_n) \mid K\right) \subset \cdots$$

of subgroups of the Galois group.

It remains to prove the converse. That is, we now begin with the assumption that the Galois group $G = \mathrm{Aut}\left(K(x_1,\ldots,x_n) \mid K\right)$ is solvable, that is, that one has the requisite chain of subgroups $\{\mathrm{id}\} = G_0 \subset G_1 \subset G_2 \subset \cdots \subset G_{k-1} \subset G_k = G$. We now begin to create a field K' from the field K by adjunction of suitable roots of unity. In particular, we adjoin the roots of unity of every prime degree less than or equal to the greatest proper divisor of the number $|G|$ of elements of G. The Galois group $H = \mathrm{Aut}\left(K'(x_1,\ldots,x_n) \mid K'\right)$ is a subgroup of the solvable Galois group G and is itself solvable. This can be seen by modifying the above chain of subgroups as follows:

$$\{\mathrm{id}\} \subset G_1 \cap H \subset G_2 \cap H \subset \cdots \subset G_{k-1} \cap H \subset G_k \cap H = H.$$

Each of these groups is a normal subgroup of the next group in the chain. Moreover, we may view each of the associated quotient groups $(G_{j+1} \cap H)/(G_j \cap H)$ as a subgroup of the quotient group G_{j+1}/G_j, which comprises precisely those cosets that contain at least one element of H. Therefore, either $G_{j+1} \cap H = G_j \cap H$ or the quotient group $(G_{j+1} \cap H)/(G_j \cap H)$ is cyclic of prime order. This shows that the subgroup H itself is solvable. Thus for the rest of the proof we assume without loss of generality that the chain of subgroups has been shortened such that all the subgroups

$$\{\mathrm{id}\} = H_0 \subset H_1 \subset \cdots \subset H_{k-1} \subset H_k = H$$

that appear are distinct.

The penultimate group, H_{k-1}, is a normal subgroup of the group H. Furthermore, the associated quotient group is cyclic, of order a prime number p. According to Section 10.15, there is thus a pth root of an element a in the field K' whose adjunction, as shown in Figure 10.8, yields the first intermediate field. Then in an analogous fashion the remaining fields can be constructed stepwise in relation

to the subgroups $H_{k-2}, \ldots, H_1$ as continuing extensions up through the field $K'(x_1, \ldots, x_n)$, whose Galois group is H_{k-1}. All in all, this shows that the solutions $x_1, \ldots, x_n$ can be expressed in terms of nested radicals with the radicands in the field K.

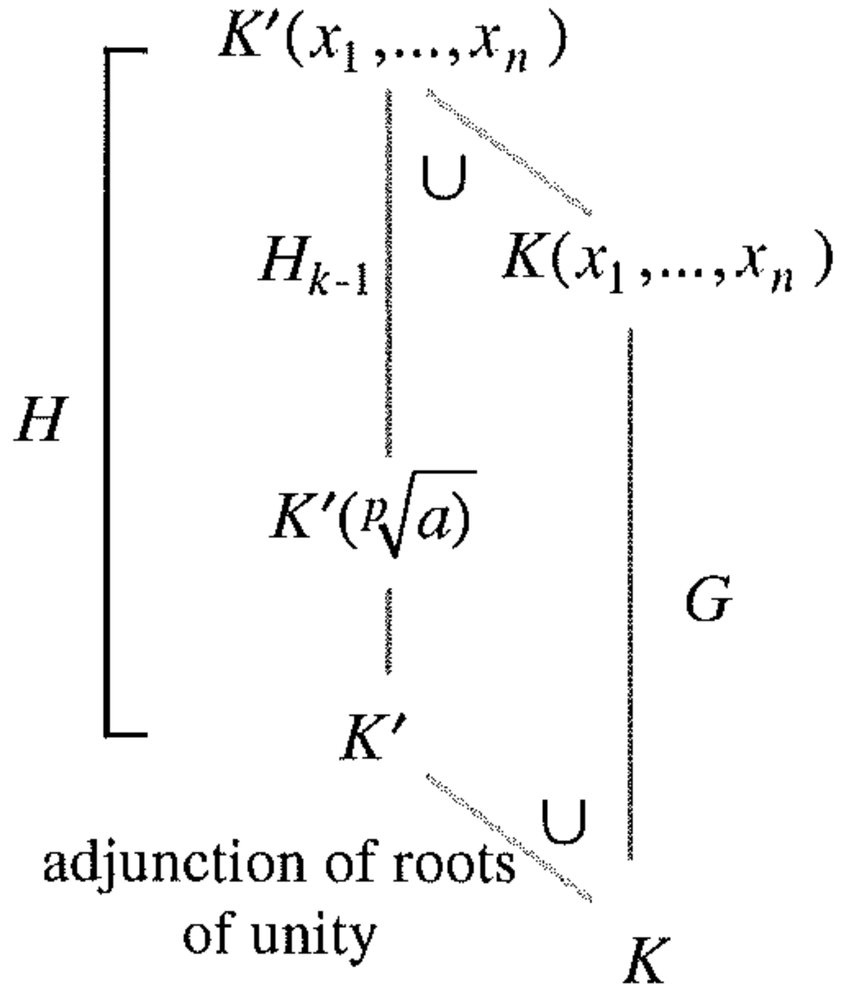


Figure 10.8. How to find a solution of the underlying equation for solvable groups.

10.19 With the completion of the proof above we have accomplished the goal of this book. As announced in the introduction, using Galois theory alone, one can determine whether a given equation is solvable in radicals. Properties of an equation and of the field generated by the solutions can be obtained through the investigation of much simpler objects, namely groups. Of course, in a specific case, it is not recommended to test the solvability of a group by a step-by-step decomposition of the group table, as we have done in the previous chapter. Much simpler is to use methods obtained from the general theory of finite groups. However, we have deliberately refrained from explaining such methods in this book in order to limit the scope of this introductory excursion into the theory of polynomial equations. We note here only that the symmetric group S_n, that is, the group of all permutations on n elements, is not solvable for every $n > 5$. A proof of this fact is given in the epilogue. Equations of the fifth degree with Galois group S_5 are therefore not solvable in radicals. An

example of such an equation—and indeed, such equations are the rule rather than the exception—was given in Section 9.17.

The Unsolvability of the Classical Construction Problems

We stated in Chapter 7 that the classical problems of squaring the circle, doubling the cube, and trisecting an angle have no solution. The reason that the circle cannot be doubled lies not in Galois theory, but in the transcendence of the number π.

The proofs of the impossibility of the other two constructions also do not rest on the deep results of Galois theory. In general, it suffices, as we shall now see, to consider the formula for the degrees of towers of field extensions.

To complete the proofs, then, we shall reformulate the results of Chapter 7 in terms of fields: If a construction can be carried out with straightedge and compass that leads to the construction of a point z in the complex plane, this is algebraically equivalent to the point z lying in a field obtainable from the field $\mathbb{Q}$ of rational numbers by stepwise extensions of degree 2. According to the degree formula for nested field extensions, the number z must therefore belong to a field whose degree over the field $\mathbb{Q}$ of rational numbers is a power of 2.

Since the field $\mathbb{Q}\left(\sqrt[3]{2}\right)$ has degree three over the rational numbers, $\sqrt[3]{2}$ cannot, based on the degree formula, lie in a field whose degree is a power of 2. Therefore, a segment of length $\sqrt[3]{2}$ cannot be constructed with straightedge and compass. Thus the problem of doubling the cube is unsolvable.[19]

In the problem of trisecting an angle, again the key is in the construction of a cubic extension field. The problem is equivalent to constructing a number z, given a number a on the unit circle, satisfying $z^3 = a$. Beginning with the field $\mathbb{Q}(\zeta, a)$, where ζ is a cube root of unity not equal to 1, the adjunction of the number z yields one of the following two scenarios:

(1) The number z lies in the field $\mathbb{Q}(\zeta, a)$, and no genuine field extension arises.
(2) The number z does not lie in the field $\mathbb{Q}(\zeta, a)$, and the result is a field extension of degree 3.

[19] A way of solving this problem, as well as the other classical construction problems, in an elementary fashion, without using the degree formula or other result of Galois theory, is described by Detlef Laugwitz in Eine elementare Methode für die Unmöglichkeit bei Konstruktionen mit Zirkel und Lineal, *Elemente der Mathematik*, 17 (1962), pp. 54–58. In addition there is a discussion of how few square roots are necessary to represent the roots of cubic equations.

In case (2), which occurs, for example, in the problem of trisecting an angle of 120 degrees, the construction of the point z with straightedge and compass is impossible.

We note finally that one can prove the constructibility or nonconstructibility of regular polygons using the degree formula for nested field extensions.

Literature on Galois Theory and Algebraic Structures

Emil Artin, *Galois Theory*, Notre Dame, 1942.

Siegfried Bosch, *Algebra*, Berlin, 2001.

Jean Pierre Escofier, *Galois Theory*, New York, 2001.

Serge Lang, *Algebra*, Springer, 2002.

Bartel Leendert van der Waerden, *Algebra I*, Berlin, 1971.

Jean-Pierre Tignol, *Galois' Theory of Algebraic Equations*, Singapore, 2001.

Exercises

(1) Prove the degree formula for nested field extensions $K \subset L \subset E$, that is, that the degree of the extension E over K is equal to the products of the degrees of E over L and L over K.

(2) Show that the number of elements of a finite field must be a power of a prime.

(3) If for a prime number n, a permutation group $G \subset S_n$ operates transitively on the set $\{1, 2, \ldots, n\}$, then any normal subgroup $H \subset G$ with $H \neq \{\mathrm{id}\}$ also operates transitively on $\{1, 2, \ldots, n\}$. Hint: Decompose the set $\{1, 2, \ldots, n\}$ into so-called *orbits*, each of which is the set of all elements that can be translated one into the other by a suitable permutation in H. Why must these orbits all be of the same size?

Conclude, moreover, that if the Galois group of an irreducible equation of prime degree n is solvable, then the penultimate group in the solution is cyclic of order n.

(4) Show that for a prime number n, the set of linear transformations, that is, functions $f_{a,b} : \mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ such that $f_{a,b}(x) = ax + b$ for given residue classes $a, b \in \mathbb{Z}/n\mathbb{Z}$, form a group.

In addition, prove the following:

- No linear transformation has two or more fixed points, that is, for each transformation $f_{a,b}$ there is at most one residue class x with $f_{a,b}(x) = x$.
- Every element of order n has the form $f_{1,b}$ with $b \neq 0$.
- Every *linear group*, that is, a group of linear transformations including the transformation $f_{1,1}$, is solvable. Hint for the last part: The subgroup generated by $f_{1,1}$ is a component of the solution.

(5) Among the examples of Galois groups of equations of fifth degree in Section 9.17, it was stated that the equation $x^5 - 2 = 0$ over the field of rational numbers leads to a Galois group with twenty elements, while the equation $x^5 - 5x + 12 = 0$, despite the more complicated root structure, results in a Galois group of only ten elements. How is this phenomenon to be explained in the light of our considerations in this chapter on radical extensions of the underlying field and subgroups of the Galois group?

Epilogue

In the end was the beginning. Both historically and in relation to the thematic framework of this introduction, the end result creates a new beginning: Although the problem of solving polynomial equations in radicals posed by Cardano and Ferrari was able to be answered, the objects involved in the solution, groups and fields, raise many new questions about their general properties, and not only in the sense of "art for art's sake." The knowledge that these objects, and the associated applications and techniques, are applicable in many fields of inquiry has allowed algebra, that is, the subfield of mathematics that deals with basic arithmetic operations, to establish itself as a major mathematical discipline. In the field of abstract algebra, the objects of consideration are defined and "classified" in the broadest possible generality and categorized according to their basic structure. To do this with maximum efficiency, general classifications are refined as needed, for example, groups and fields with their subcategories abelian groups and finite fields; and such classifications are also generalized, for example, with the definition of a commutative ring, which satisfies all the requirements of a field except for the invertibility of multiplication.[1]

[1]The best-known examples of rings that are not fields are the integers, the set of polynomials in one or several variables, and the set of residue classes $\mathbb{Z}/n\mathbb{Z}$ for n not a prime.

There are several advantages to developing mathematical objects by such an axiomatic method:

- Mathematics becomes more transparent. In particular, one can recognize fundamental properties in a collection of various mathematical objects that exhibit a number of properties in common.
- Mathematics becomes liberated from fundamental "truths" taken for granted once it has been freed from particular interpretations and applications. Thus, for example, it was with the generalization of the parallel postulate to non-Euclidean geometries that it became possible to establish the unprovability of the parallel axiom, a problem that had been festering since antiquity.
- Such an approach is more economical, at least with respect to mathematics as a whole, since important facts do not have to be proved over and over in different situations. Moreover, these general principles, which in fact are of central interest in mathematics, can often be derived as special cases from more generally valid theorems.

Although such an axiomatically constituted mathematics diverges from the descriptive natural sciences in being only indirectly connected with our physical perception of the world, one should note that classification plays an important role in those sciences as well, from the Linnaean taxonomic system of biological classification to the periodic table of the elements to the classification of symmetries of fundamental particles.

If this book employed such a structural approach only in the last chapter, and perhaps half-heartedly but pragmatically in the chapter before that, the reason was to minimize the difficulties for the interested nonmathematician. The multiplicity of definitions and concepts that seem opaque on first contact presents an almost insuperable barrier to the nonmathematician. Perhaps some readers of the last chapter will have received such an impression, despite the contrary intention of the author.

To avoid unnecessary complications, some things were deliberately excluded, some of which are related to polynomials. We tacitly accepted, without a formal definition, a polynomial as a formal sum

of products of one or more variables $X, Y, \ldots$, and coefficients taking values in some fixed set. Generally, this set was a particular field, but one could also have taken the ring of integers or indeed the set of all polynomials in additional variables.

Such formal polynomials are to be distinguished from the functions that such polynomials define when the variables are replaced by concrete values $a, b, \ldots$ from some set of numbers. Now one can calculate both with polynomials themselves, taking their sums and products, and with their functional values. It is clear that the two forms of calculation are compatible, for example that one has $(f \cdot g)(a) = f(a) \cdot g(a)$. However, one should prove that this is the case.

The simplification of our presentation also serves the purpose of specializing the discussion to subfields of the complex numbers. It was clear, on account of the fundamental theorem of algebra, that a splitting field exists for every polynomial with complex coefficients. Despite the practicality of such an approach, and despite the importance of the fundamental theorem of algebra, the form of argumentation has little to do with algebra. It is not only that the fundamental theorem is proved using mathematical analysis (calculus), an argument involving estimates of distance and intermediate values, which renders the theorem's appellation a historical artifact. It is also that a generalization to other cases, for example that of finite fields, cannot be carried out by such methods.

For these reasons it is understandable that in algebra a completely different tack is generally taken for constructing the splitting fields that are crucial to Galois theory. Beginning with a field K and a polynomial irreducible over K, a field extension containing the elements that solve the corresponding equation

$$x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_1 x + a_0 = 0$$

is constructed in a completely formal way. One does this by the adjunction of a formal value α, where in calculating with expressions of the form

$$k_0 + k_1\alpha + k_2\alpha^2 + \cdots + k_m\alpha^m,$$

with $k_0, \ldots, k_m \in K$, we employ the simplification

$$\alpha^n = -a_{n-1}\alpha^{n-1} - a_{n-2}\alpha^{n-2} - \cdots - -a_1\alpha - \alpha_0,$$

so that it always can be achieved that $m \leq n-1$. One can then show that the set

$$K[\alpha] = \{\, k_0 + k_1\alpha + k_2\alpha^2 + \cdots + k_{n-1}\alpha^{n-1} \mid k_j \in K \,\}$$

forms a field that clearly contains a solution, namely α, of the given equation.[2] What is tricky here is the proof that the set $K[\alpha]$ is closed under division.[3]

If the polynomial is then factored over the field $K[\alpha]$ into irreducible factors, one can proceed with additional adjunction steps. In this way, one finally obtains a completely algebraically constructed splitting field.[4] It is uniquely determined, as can be shown, in that every other splitting field is isomorphic to this one, that is, that the elements are related by a one-to-one correspondence that is compatible with the basic arithmetic operations.[5]

With the formulation thus described, the general equation can now be made amenable to treatment by Galois theory in terms of purely algebraic methods. We have seen the general equation in Chapter 5 as the equation in which formal variables $x_1, \ldots, x_n$ in

---

[2]From a formal point of view, this approach is similar to that of a quotient group from a normal subgroup. It is an example of a ring of residue classes, which can be constructed from a ring and a subset of a ring called an *ideal*. It is such methods of constructing new objects that requires the axiomatic definition of such objects as groups and fields, not just as subgroups of the symmetric group, as we were always able to do in the case of finite groups, and subfields of the complex numbers.

[3]Essentially, the arguments from Section 10.9 can be easily extended. That is, one investigates the linear system of equations that corresponds to multiplication by the inverse of an element of $K[\alpha]$. However, it is also necessary for the considerations of Section 10.9 to prove that the product of two nonzero elements is again nonzero.

[4]This purely algebraic construction can in fact be used to prove the fundamental theorem of algebra using complete induction. (The induction is over the highest power of 2 that divides the degree of the equation.) Analytic arguments enter the picture only in the form of that fact, provable by the intermediate value theorem, that every odd-degree polynomial with real coefficients has a real zero. See Jean-Pierre Tignol, *Galois' theory of Algebraic Equations*, Singapore, 2001, pp. 119, 121–122, and Exercise 5 at the end of this chapter.

[5]A field automorphism is simply an isomorphism of a field with itself.

the associated elementary symmetric polynomials

$$\begin{aligned} s_1(x_1,\ldots,x_n) &= x_1 + x_2 + \cdots + x_n, \\ s_2(x_1,\ldots,x_n) &= x_1x_2 + x_1x_3 + \cdots + x_{n-1}x_n, \\ &\cdots \\ s_n(x_1,\ldots,x_n) &= x_1x_2\cdots x_n, \end{aligned}$$

are to be determined.

In the language of field extensions, this corresponds to the situation in which beginning with a field K of polynomial coefficients, one is to investigate the extension of the field $K(s_1,\ldots,s_n)$ to the field $K(x_1,\ldots,x_n)$. Due to the uniqueness theorem for symmetric polynomials (see the section on this topic in Chapter 5), one may treat the symmetric polynomials in $K(s_1,\ldots,s_n)$ as though they were formal variables with no formal polynomial relations among them (one speaks of *algebraically independent* quantities). One thereby obtains an additional, fully equivalent, interpretation of the general equation, in which now the equation's coefficients $a_0, a_1, \ldots, a_{n-1}$ are variables for which, as described, a splitting field can be constructed. Since the solutions have no relations among themselves—in the first place by definition and in the second place by the equivalence[6]—the Galois group of the general equation is the full symmetric group.

Theorem E.1. *The Galois group of the general nth-degree equation is the symmetric group S_n; that is, it contains all permutations of the n solutions $x_1,\ldots,x_n$.*

As a consequence, the results for the general equation, as first discovered by Lagrange, appear as a special case of Galois theory. Here every intermediate field is generated by polynomials in the variables

[6]Of course, a direct proof is also possible: Beginning with a given polynomial $h(X_1,\ldots,X_n)$ with $h(x_1,\ldots,x_n) = 0$, one forms the product

$$g(X_1,\ldots,X_n) = \prod_{\sigma\in S_n} h\left(X_{\sigma(1)},\ldots,X_{\sigma(n)}\right).$$

Since the polynomial g is symmetric in the variables $X_1,\ldots,X_n$, it can be expressed as a polynomial in the elementary symmetric polynomials in these variables. There is thus a polynomial $u(Y_1,\ldots,Y_n)$ such that the polynomial $g(X_1,\ldots,X_n)$ can be expressed in the form $g(X_1,\ldots,X_n) = u\left(s_1(X_1,\ldots,X_n),\ldots,s_n(X_1,\ldots,X_n)\right)$. If one substitutes the solutions $x_1,\ldots,x_n$ into this identity, then one obtains $0 = g(x_1,\ldots,x_n) = u(a_{n-1},\ldots,a_0)$. This shows immediately that $u = 0$. The previous polynomial identities finally show that $g = 0$ and $h = 0$.

$x_1, \ldots, x_n$ that remain unchanged under the automorphisms of the associated group of permutations. Furthermore, it naturally follows that the solvability of the general equation of a particular degree n is equivalent to the solvability of the symmetric group S_n. Abel's impossibility theorem thus corresponds to the following group-theoretic theorem.

Theorem E.2. *The symmetric group S_n is unsolvable for $n \geq 5$.*

In textbooks, this proof of a group-theoretic theorem usually is used to derive Abel's theorem. A proof is possible with arguments similar to those used by Ruffini (see the section on this topic at the end of Chapter 5). To this end, one first proves the following theorem.

Theorem E.3. *If G is a subgroup of the symmetric group S_n for $n \geq 5$ containing all three-cycles, that is, all cyclic permutations of the form $a \to b \to c \to a$ of three distinct elements a, b, c, and if N is a normal subgroup of G with commutative quotient group G/N, then this normal subgroup also contains all the three-cycles.*

To prove this preparatory theorem one represents an arbitrary three-cycle $a \to b \to c \to a$ as the product

$$(d \to b \to a \to d)^{-1} \circ (a \to e \to c \to a)^{-1} \circ (d \to b \to a \to d) \circ (a \to e \to c \to a),$$

where d and e are arbitrary distinct elements that are also distinct from a, b, c. Since the quotient group is commutative, the product must lie in the coset that represents the identity, that is, in N. As asserted, then, every three-cycle belongs to the normal subgroup N.

On the basis of the theorem just proved, it can now be deduced step by step that every group in an ascending chain corresponding to a solution of the symmetric group S_n must contain all three-cycles. The chain can therefore not end in the trivial group containing a single element, and so the symmetric group cannot be solvable.

Moreover, the same argument can be applied to the alternating group A_n, defined as the group of all even permutations. With reference to the alternating group A_n, we note that it is a normal subgroup of the symmetric group S_n, since the quotient group is a commutative

two-element group. In the case of the general equation, the alternating group corresponds to the intermediate field that arises through adjunction of the square root of the discriminant.

To the extent that the base field K for the general equation is a subgroup of the complex numbers, the implicitly assumed possibility of extending Galois theory and its applications to radical extensions is unproblematic. In an extension of Galois theory to arbitrary fields, however, two additional complicating factors need to be considered:

- The generalization works only if every irreducible polynomial possesses distinct zeros. Otherwise, not every automorphism of the splitting field is associated uniquely with a permutation of the zeros, and moreover, the construction of Galois resolvents can be problematic. Nevertheless, fields of characteristic zero and finite fields cause no problems in this respect.
- The characterization of radical extensions in terms of Lagrange resolvents assumes that one can divide by the degree of the field extension (see the end of the proof in Section 10.14). In fields with finite characteristic this is not necessarily possible.[7]

Another hole in the preceding chapter relates to finite fields, which we have used only indirectly, other than giving some examples in Chapter 10, namely, in the form of fields of residue classes modulo a prime. In particular, we made use of the existence of a primitive root modulo n, so that the cyclotomic equation could be solved using suitable sums of roots of unity, that is, the periods. Thus for prime numbers n we assumed the existence of an integer g such that the numbers $g^1, g^2, \ldots, g^{n-1}$ represent distinct nonzero residue classes $1, 2, \ldots, n-1$.

Using algebraic structures, this fact can be formulated in slightly greater generality.

Theorem E.4. *Every finite subgroup of the multiplicative group of a field is cyclic.*

The application of interest here, relating to subgroups of a finite field $\mathbb{Z}/p\mathbb{Z}$, was first proved, in a formulation as a statement about

[7] Indeed, the general quadratic equation over the two-element field $\mathbb{Z}/2\mathbb{Z}$, for example, is not solvable in radicals. See B. L. van der Waerden, *Algebra I*, Section 62.

residue classes, by Legendre (1752–1833). Earlier proofs by Euler must be considered incomplete. Although a proof can be given based on extensive computations in residue classes,[8] we would like to offer a proof of the generalized theorem, which is shorter and more easily understood.

We begin with an investigation of Euler's phi function, which associates with a natural number d the number of integers in the set $\{1, 2, \ldots, d\}$ relatively prime to d. For example, $\varphi(6) = 2$, since 1 and 5 are the only integers between 1 and 6 relatively prime to 6; and $\varphi(8) = 4$, with $1, 3, 5, 7$ relatively prime to 8. Euler's phi function satisfies the relation

$$\sum_{d|n} \varphi(d) = n.$$

We will first justify this formula, where the sum is over all divisors d of n. To this end, consider for each residue class j modulo n, represented for example by the n integers $0, 1, \ldots, n-1$, its order d as an element of the group $\mathbb{Z}/n\mathbb{Z}$. Each such order d must be a divisor of n, and the residue class j will have order d precisely when it is represented by an integer $m \cdot \frac{n}{d}$, so that j must lie in the subgroup generated by the residue class associated with $\frac{n}{d}$. This subgroup is cyclic of order d, and thus isomorphic to $\mathbb{Z}/d\mathbb{Z}$, and therefore contains precisely $\varphi(d)$ elements of order d. The partition of the n-element group $\mathbb{Z}/n\mathbb{Z}$ thus obtained corresponds precisely to the summation formula.

After these preparations we can address the actual content of the theorem, namely, a finite subgroup U of the multiplicative group of a field. If d is a natural number such that there is an element x in U for which the group generated by x is the group $\{1, x, x^2, \ldots, x^{d-1}\}$ of d elements, then according to Section 10.4, d is a divisor of $|U|$, the number of elements in U. Since $x^d = 1$, every element of this group is a zero of the polynomial $X^d - 1$. Since we know from Section 4.2 that for each zero of a polynomial we may split off a linear factor, and thus this polynomial can have at most d zeros, there cannot exist an element of U outside of the subgroup $\{1, x, x^2, \ldots, x^{d-1}\}$ that generates a d-element subgroup. Therefore, in the group U there is either no element that generates a d-element subgroup or there are

[8] See, for example, Jay R. Goldman, *The Queen of Mathematics*, Wellesley, 1998, Chapter 10.

$\varphi(d)$ of them. If one again decomposes the group U as we earlier decomposed the group $\mathbb{Z}/n\mathbb{Z}$ according to the size of the subgroup that each element generates, then for $n = |U|$ we obtain the summation formula

$$n = \sum_{d|n} \varphi(d) \cdot \delta_d,$$

where each of the numbers δ_d is either 0 or 1. One then sees immediately a similarity to the previously derived summation formula, that for divisors d of n we must always have $\delta_d = 1$. In particular, there are $\varphi(n)$ elements of U that generate an n-element subgroup, that is, the entire group U. The group U is therefore cyclic.

Exercises

(1) Prove *Fermat's little theorem*: For a prime number n and a positive integer a relatively prime to n, the number $a^{n-1} - 1$ is divisible by n.

(2) Prove *Wilson's theorem*: For a prime number n, the number $(n-1)! + 1$ is divisible by n. Then conclude from this that a natural number $n \geq 2$ is prime if and only if $(n-1)! + 1$ is divisible by n.

(3) Prove the generalization of Fermat's little theorem that for a natural number n and a natural number a relatively prime to n, the number $a^{\varphi(n)} - 1$ is divisible by n. Hint: First show that the residue classes in $\mathbb{Z}/n\mathbb{Z}$ represented by integers relatively prime to n form a group under multiplication.

(4) Prove that if $n = pq$ is a product of two distinct prime numbers p and q and if u and v are two natural numbers such that $uv - 1$ is divisible by $(p-1)(q-1)$, then for every natural number a, the number $a^{uv} - a$ is divisible by n.[9] In such a case, the pairs (u, n) and (v, n) can serve as cryptographic keys, where one is used for encryption, the other for decryption. One speaks of an *asymmetric cryptographic algorithm*. In contrast to symmetric

[9] The significance of this exercise is that the two residue class mappings $x \mapsto x^u$ and $x \mapsto x^v$ of $\mathbb{Z}/n\mathbb{Z}$ into itself are inverses of each other. Such a construction is used in cryptography, in the *RSA encryption procedure*. Here very large primes, that is, of several hundred digits each, are used, so that determining two such prime numbers p, q given their product $n = pq$ would be impossible even after millions of years of computation on today's fastest computers.

algorithms, in which encryption and decryption use a single key, with the RSA algorithm one of the keys, that for encoding, say, can be published without fear that unauthorized persons will be able to decode encrypted messages. One therefore refers to the RSA algorithm as *public key encryption*. Hint: The assertion can be reduced to that of Exercise 3. To include the case in which a is divisible by p or q, one might demonstrate the divisibility of $a^{uv} - a$ by p and by q separately.

(5) Prove the fundamental theorem of algebra in an algebraic way by proving that for a nonconstant polynomial $f(X)$ with complex coefficients, if one factors $f(x)$ into linear factors as

$$f(X) = (X - x_1) \cdots (X - x_n)$$

in some algebraic extension field of $\mathbb{C}$ (which is always possible via an algebraic construction), one in fact has that $x_1, \ldots, x_n \in \mathbb{C}$. First show the following:

- The theorem holds for quadratic polynomials $f(X)$ (see also Exercise 1 in Chapter 2).
- It suffices to prove the existence of a single complex solution x_j. Moreover, one can restrict attention to polynomials with real coefficients.

Now the proof can be carried out using mathematical induction on the highest power of 2 that divides the degree of the polynomial. For the base step of the induction, one uses the version of the theorem for real polynomials of odd degree, which is proved using mathematical analysis (calculus). For the induction step, one investigates polynomials of the form

$$g_c(X) = \prod_{i<j} (X - (x_i + x_j + c x_i x_j))$$

for a suitably chosen parameter c.

(6) For a prime number m and a natural number a relatively prime to m, the *Legendre symbol* is defined as follows:

$$\left(\frac{a}{m}\right) = \begin{cases} +1 & \text{if } a = s^2 + km \text{ for suitable integers } s \text{ and } k, \\ -1 & \text{if } a \text{ does not have such representation.} \end{cases}$$

The Legendre symbol therefore tells whether the residue class represented by a is a square in the multiplicative group of residue classes $\mathbb{Z}/m\mathbb{Z} - \{0\}$. Even though the value of the Legendre symbol can be determined by finite trial and error, one is naturally interested in a direct calculation. Therefore, show first that

$$a^{(m-1)/2} - \left(\frac{a}{m}\right)$$

is divisible by m.

Other properties of the Legendre symbol can be obtained with the use of roots of unity. For a second prime number $n \geq 3$ we again let ζ denote the nth root of unity $\zeta = \cos\left(\frac{2\pi}{n}\right) + i\sin\left(\frac{2\pi}{n}\right)$, while the periods of length $(n-1)/2$ (see Section 7.2) are denoted by $\eta_0 = P_{(n-1)/2}(\zeta)$ and $\eta_1 = P_{(n-1)/2}(\zeta^g)$, where the integer g again represents a primitive root modulo n. Show that

$$(\eta_0 - \eta_1)^m - \left(\frac{m}{n}\right)(\eta_0 - \eta_1) = m\left(a_0 + a_1\zeta + \cdots + a_{n-2}\zeta^{n-2}\right)$$

with integers $a_0, a_1, \ldots, a_{n-2}$. Show also (if you have not already done so in Exercise 2 of Chapter 7) that

$$(\eta_0 - \eta_1)^2 = (-1)^{(n-1)/2} n.$$

Finally, show how the resulting identity

$$\left[(-1)^{\frac{n-1}{2}\frac{m-1}{2}}\left(\frac{n}{m}\right) - \left(\frac{m}{n}\right)\right](\eta_0 - \eta_1)^2 = m\left(b_0 + b_1\zeta + \cdots + b_{n-2}\zeta^{n-2}\right),$$

with integers $b_0, b_1, \ldots, b_{n-2}$, yields the *law of quadratic reciprocity*[10]

$$\left(\frac{m}{n}\right) = (-1)^{\frac{n-1}{2}\frac{m-1}{2}}\left(\frac{n}{m}\right).$$

(7) For a group G whose number of elements $|G|$ is divisible by a prime number p, one defines the mapping

$$\varphi(g_1, g_2, \ldots, g_p) = (g_2, \ldots, g_p, g_1)$$

---

[10]The law of quadratic reciprocity was first proved by Carl Friedrich Gauss on April 8, 1796, as documented by an entry in his diary. It is a fundamental result of number theory with many ramifications. Furthermore, using the law of quadratic reciprocity together with some other elementary properties of integers, one can compute the values of arbitrary Legendre symbols rather quickly.

for $g_1, g_2, \ldots, g_p \in G$, as well as the set

$$X = \{\, (g_1, g_2, \ldots, g_p) \in G^p \mid g_1 g_2 \cdots g_p = \epsilon \,\},$$

where ϵ again denotes the group identity.

Prove the following:

- $|X| = |G|^{p-1}$.
- The mapping φ maps the set X into itself.
- If the identity $\varphi^k(x) = x$ holds for an element $x \in G^p$ and an integer k not divisible by p, then all the coordinates of x are equal.
- Every *orbit* $\{\, x, \varphi(x), \varphi^2(x), \ldots \}$, where $x \in G^p$, consists of either one element or p elements.
- The number of one-element orbits in X is divisible by p.

Assuming that there exists an element $x \in X$ with a one-element orbit, conclude that there exists at least one other element with a one-element orbit, and thereby prove the existence of an element of G of order p (Cauchy's theorem).[11]

[11] Cauchy's theorem is usually formulated in a more general form, which is named after Ludwig Sylow (1832–1918). The Sylow theorems make assertions about subgroups of a group of order the power of a prime.

Index

Abel, Niels Henrik, xii, xix, 49, 50, 53, 55, 95, 114
absolute value, 14, 32, 34
Acampora, Renato, 1, 4, 9
adjunction, 95–97, 102–105, 109
algebraic structure, 171
algebraically independent, 169
al-Khwarizmi, 2, 3, 5
alternating group, 170
analytic geometry, 136
Archimedes, 4
Argand, Jean Robert, 34
Ars Magna, x, xi, 5, 6, 8–10, 18, 23, 24, 29, 30
Artin, Emil, xix, 95, 149, 150, 162
associative law, 128
automorphism, 126, 132–135, 139–143, 145–147, 151, 152, 154–156
automorphism group, 125, 135, 148
axiomatic foundations of mathematics, 166, 168
Ayoub, Raymond G., 51

Bachmann, Paul, 79
basis, 6, 115, 126, 136, 137, 155
Bézout, Etienne, 46, 83
bicubic resolvent, 85, 86, 88, 113
bijection, 141, 142, 150
biquadratic equation, xi, xii, xviii, 24–28, 37, 38, 40–42, 44, 53, 76, 99, 101, 102, 106, 110–112
Bombelli, Rafael, 11, 18
Bos, Henk, 28, 70
Bosch, Siegfried, 162
Breuer, Samson, 90
Bring, Erland Samuel, 83, 89, 90
Bring–Jerrard transformation, 83, 90
Buchmann, Johannes, xvi

Cantor, Moritz, 11
Cardano
  formula, 6, 7, 11
  suspension, 4
  wave, 4
Cardano, Geronimo, xi, xvi, 23, 24
Cartesian coordinates, 69, 70
casus irreducibilis, 10, 16, 19, 20, 75
Cauchy's theorem, 176
Cauchy, Augustin-Louis, 34
characteristic of a field, 132, 171
closed, xiv, xix, 96, 126, 128, 138, 168
complex
  analysis, 14
  number, 10–13, 31, 32, 34, 64, 74, 76, 96
  plane, 14, 15, 64, 161
composition of permutations, 43, 94, 101, 127
computer algebra system, 58
conjugate subgroup, 145, 146
continuity of a function, 14, 31–33
coset, 129, 131, 146, 147, 159, 170
Crelle, August Leopold, 114
cryptography, xvi, 173
cubic equation, x, xi, 1, 4–10, 17–20, 24, 25, 38, 40, 41, 53, 55, 75, 97, 107–109, 161
cubic resolvent, 100, 102, 112
cycle, 54
cyclotomic equation, xviii, 17, 39, 59, 60, 65, 67, 70, 72, 73, 75, 77, 79, 80, 94, 107, 110, 112, 123, 135, 154–158, 171

Davies, Philip J., xi
Dedekind, Richard, 126

degree formula for a tower of fields, 137, 143, 161
Dehn, Edgar, 122
Delahaye, Jean-Paul, 75
Descartes, René, 28–30, 69, 70
difference product, 108, 109, 112, 114
dimension, 126, 136
discriminant, 40, 41, 46, 88, 107, 108, 171
disjoint partition, 129
distributive law, 13, 131
Dörrie, Heinrich, 26
doubling the cube, 161

Edwards, Harold M., 95, 122
Eisenstein irreducibility criterion, 59, 60
Eisenstein, Ferdinand Gotthold Max, 59
ElGamal encryption procedure, xvi
elliptic curve, xvi
encryption
  asymmetric, xvi
  symmetric, 173
Escofier, Jean Pierre, 162
Euclidean algorithm, 73, 97, 116, 122, 123
Euler's phi function, 172
Euler, Leonhard, 16, 46, 81–83, 172
extension field, 96, 97, 104, 119, 126, 136, 137, 139, 142, 151, 153, 154, 161, 174

factorial function, 43
Fermat prime, 72, 73
Fermat's little theorem, 173
Ferrari, Ludovico, 1, 23, 165
Ferro, Scipione del, 4
field, xiv, xvi, xix, 13, 16, 131, 165
  finite, xvi, 132, 162, 165, 167, 171
field extension, 135, 137, 139, 141, 148, 150, 158, 161, 162, 169
  degree of, 137, 150
fifth-degree equation, *see* quintic equation
Fior, Antonio, 1, 3, 4
fixed field, 142, 143, 145, 150
Fontana, Niccolò, *see* Tartaglia
Führer, Lutz, 21
function theory, 14
fundamental particle, 166
fundamental theorem of algebra, 30–32, 34, 37, 167, 168, 174
fundamental theorem of Galois theory, 140–145, 147–150, 152, 153, 155
fundamental theorem of symmetric polynomials, 45, 46

Galois group, xiii–xv, xix, 93–95, 97–116, 118, 119, 122, 123, 127–130, 132–136, 138–151, 153–160, 162, 163, 169
  solvable, 105
Galois resolvent, 98, 115, 117–119, 133–135, 139, 140, 149, 171
Galois theory, xii, xiv, xv, xviii, xix, 53, 75, 80, 82, 95, 125, 132, 147, 149, 160–162, 167–169, 171
Galois, Évariste, xii, xv, 93, 122
Gårding, Lars, 114
Gauss, Carl Friedrich, 12, 31, 56, 63–65, 67, 69, 70, 75, 77, 94, 175
Gaussian plane, 65
general equation, 38, 39, 42, 45, 49, 50, 52, 53, 55, 77, 80, 93–95, 168–171
Girard, Albert, 30
Goldman, Jay R., 172
Gottlieb, Christian, 73
greatest common divisor, 120
group
  abelian, 131, 165
  alternating, 170
  cyclic, 130, 131, 151, 155
  of integers mod n, 131, 132, 151, 163, 165, 172, 173
  linear, 163
  solvable, 160
  symmetric, 43, 99, 160, 169, 170
group table, xiii, xiv, 101, 102, 104–108, 110, 111, 113, 127, 144, 147, 153, 160

Henn, Hans-Wolfgang, 73
heptadecagon, xi, xviii, 63, 68, 70, 72
Hermes, Johann Gustav, 73
Hilbert basis theorem, 115
homomorphism, 126
Hudde, Jan, 97

ideal, 115, 168
identity element, 43, 127, 128, 130, 131
imaginary part, 14, 19, 21
imaginary unit, 14
injective, 142
integers, 14, 65, 73, 130, 165
intermediate value theorem, 31, 168
inverse element, 13, 128
irreducible, 59–61, 76, 82, 105–108, 110–112, 114, 116–119, 121, 122, 133, 136, 138, 140, 162, 167, 168, 171
isomorphic groups, 110, 111, 144, 151, 172
isomorphism, 110, 155, 168

Jerrard, George Birch, 90
Jörgensen, Dieter, 4

Kabayashi, Sigeru, 90
Katscher, Friedrich, 1
Kiernan, B. Melvin, 95
King, R. Bruce, 90
Klein, Felix, 73
known quantity, 97, 102–104
Koch, Helmut, 122
Kowol, Gerhard, 122

Lagrange interpolation, 35
Lagrange resolvent, 46, 78, 152, 157
Lagrange, Joseph Louis, 43–47, 49, 52, 77, 82, 83, 94, 95, 115, 116, 129, 132, 169
Lang, Serge, 162
Laugwitz, Detlef, 161
Lazard, Daniel, 91
Legendre symbol, 174, 175
Legendre, Adrien-Marie, 172
Leibniz, Gottfried Wilhelm, 12
lexicographic order, 47, 48
Lindemann, Carl Louis Ferdinand von, 75
linear algebra, 116, 150
linear equation, 139, 150
linear factor, 29, 30, 38, 45, 54–56, 58–60, 76, 83, 97, 109, 116–118, 126, 140, 172, 174
linear transformation, 136, 163
Liouville, Joseph, 94

Malfatti, Giovanni Francesco, 82, 83, 86, 87
mapping, 134
matrix, 130, 131
Matthiessen, Ludwig, 26, 37
McKay, John, 106
modular arithmetic, 66, 71, 77, 80, 131, 155
de Moivre, Abraham, 16, 76
de Moivre's formula, 16, 32, 34
monomial, 47, 48
multiple root, 39, 40, 97, 106

Nahin, Paul J., 21
Nakagawa, Hiroshi, 90
Neuwirth, Lee, 94
Newton, Isaac, 45
non-Euclidean geometry, 166
nonsymmetric polynomial, 98
normal subgroup, 126, 145, 147–150, 157–159, 168, 170

orbit, 176
order of a group element, 130

parallel postulate, 166
pentagon, 63, 73
period, 3, 37, 66–68, 70–74, 77–80, 110, 112, 156, 171, 175
periodic table, 166
permutation
  cyclic, 43, 170
  even, 88, 113, 170
  identity, 43, 103, 105, 106
  odd, 88, 108
Pertsinis, Tom, 93
Pesic, Peter, 53
Pieper, Herbert, 12
Pierpont, James, 82, 90
Platonic solid, 130, 131
polar coordinates, 19
polynomial
  elementary symmetric, 38, 42, 44, 45, 47, 48, 50, 88, 115, 169
  monic, 56–59, 61, 86
polynomial ring, 115
primitive element, 115, 149
primitive root, 65, 66, 71, 73, 77, 80, 155, 171, 175
public key encryption, xvi, 174

quadratic equation, ix, 1–3, 5, 27, 38, 40, 55, 67, 68, 71–74, 89, 171
quadratic reciprocity, 175
quartic equation, *see* biquadratic equation
quintic equation, xii, 37, 49, 81, 82, 86, 99, 163
quotient group, 126, 147, 149, 157, 159, 168, 170

radical extension, 153, 158, 163, 171
Radloff, Ivo, 51, 122
Ramanujan, Srinivasa, xi
rational function, 131–133
rational numbers, 31, 59–61, 69, 74, 75, 82, 96, 102, 109, 114, 117, 123, 130–132, 136, 153, 155, 161, 163
real numbers, 13, 14, 16, 31, 32, 130
regular polygon, 63, 64, 70, 162
Reich, Karin, 28, 70
relational polynomials, 102
residue class, xvi, 131, 163, 165, 168, 171–173, 175
resultant, 54
Ribenboim, Paulo, 72
Richelot, F. J., 73
Rigatelli, Laura Toti, 93
root of unity, 46, 68, 70, 73, 78, 80, 109, 137, 152, 155, 158, 161, 175
Rothman, A., 93
RSA encryption, xvi, 173, 174
Ruffini, Paolo, 49–53, 82, 95, 170
Runge, C., 88
Scholz, Erhard, 28, 70, 95, 127
Schultz, Phillip, 4

Schultze, Reinhard Siegmund, 90
seventeen-gon, *see* heptadecagon
Skau, Christian, 51, 114
Soicher, Leonhard, 106
solution formula, xii, xviii, xix, 2, 4, 37, 42, 44, 46, 50, 52, 55, 76, 77, 95
solvability in radicals, xii–xiv, xix, 50, 81, 94, 157
solvability of a group, 157, 160
Sossinsky, Alexei, 94
Spearman, Blair K., 91
splitting field, 109, 132–135, 139, 144, 146, 149, 153, 154, 167
squaring the circle, 161
Stewart, Ian, 69
Stillwell, John, 53
straightedge and compass, xi, 64, 69, 70, 74, 75, 161, 162
Stubhaug, Arild, 50
subgroup, xv, 125, 128–130, 139–147, 151, 153–157, 159, 160, 163, 168, 170–173, 176
  normal, 153
subset, 14, 94, 96, 126–128, 168
surjectivity, 142
Sylow, Ludwig, 176
symmetric polynomial, 42, 44, 45, 47, 48, 50, 52, 88, 95, 98, 99, 115
Tartaglia, 1, 4, 9
three-cycle, 52, 170
Tietze, Heinrich, 69
Tignol, Jean-Pierre, 122, 162, 168
transcendence, 161
transitive operation, 106, 140, 162
transposition, 54
triangle inequality, 32
trisecting an angle, 161
Tschirnhaus transformation, 90
Tschirnhaus, Ehrenfried Walther Graf von, 83, 89, 90

uniqueness theorem for symmetric polynomials, 48, 169

van der Waerden, Bartel Leendert, xix, 19, 95, 149, 162, 171
Vandermonde, Alexandre Théophile, 47, 77, 78, 82, 83, 88, 94, 99
vector space, 126, 130, 133, 136–138
Viète, François, 27
Viète's root theorem, 28, 29, 38, 40, 67, 86

Weber, Heinrich, 88, 95, 127, 149
Wessel, Caspar, 12
Williams, Kenneth, S., 91
Wilson's theorem, 173

zero of a function, 118, 122, 133, 140, 172