Graduate Texts in Mathematics

Thomas Becker
Volker Weispfenning
In Cooperation with Heinz Kredel

Gröbner Bases

A Computational Approach to Commutative Algebra

Springer

Graduate Texts in Mathematics **141**

Springer Science+Business Media, LLC

Graduate Texts in Mathematics

1 TAKEUTI/ZARING. Introduction to Axiomatic Set Theory. 2nd ed.

2 OXTOBY. Measure and Category. 2nd ed.

3 SCHAEFER. Topological Vector Spaces.

4 HILTON/STAMMBACH. A Course in Homological Algebra. 2nd ed.

5 MAC LANE. Categories for the Working Mathematician. 2nd ed.

6 HUGHES/PIPER. Projective Planes.

7 SERRE. A Course in Arithmetic.

8 TAKEUTI/ZARING. Axiomatic Set Theory.

9 HUMPHREYS. Introduction to Lie Algebras and Representation Theory.

10 COHEN. A Course in Simple Homotopy Theory.

11 CONWAY. Functions of One Complex Variable I. 2nd ed.

12 BEALS. Advanced Mathematical Analysis.

13 ANDERSON/FULLER. Rings and Categories of Modules. 2nd ed.

14 GOLUBITSKY/GUILLEMIN. Stable Mappings and Their Singularities.

15 BERBERIAN. Lectures in Functional Analysis and Operator Theory.

16 WINTER. The Structure of Fields.

17 ROSENBLATT. Random Processes. 2nd ed.

18 HALMOS. Measure Theory.

19 HALMOS. A Hilbert Space Problem Book. 2nd ed.

20 HUSEMOLLER. Fibre Bundles. 3rd ed.

21 HUMPHREYS. Linear Algebraic Groups.

22 BARNES/MACK. An Algebraic Introduction to Mathematical Logic.

23 GREUB. Linear Algebra. 4th ed.

24 HOLMES. Geometric Functional Analysis and Its Applications.

25 HEWITT/STROMBERG. Real and Abstract Analysis.

26 MANES. Algebraic Theories.

27 KELLEY. General Topology.

28 ZARISKI/SAMUEL. Commutative Algebra. Vol.I.

29 ZARISKI/SAMUEL. Commutative Algebra. Vol.II.

30 JACOBSON. Lectures in Abstract Algebra I. Basic Concepts.

31 JACOBSON. Lectures in Abstract Algebra II. Linear Algebra.

32 JACOBSON. Lectures in Abstract Algebra III. Theory of Fields and Galois Theory.

33 HIRSCH. Differential Topology.

34 SPITZER. Principles of Random Walk. 2nd ed.

35 ALEXANDER/WERMER. Several Complex Variables and Banach Algebras. 3rd ed.

36 KELLEY/NAMIOKA et al. Linear Topological Spaces.

37 MONK. Mathematical Logic.

38 GRAUERT/FRITZSCHE. Several Complex Variables.

39 ARVESON. An Invitation to C^*-Algebras.

40 KEMENY/SNELL/KNAPP. Denumerable Markov Chains. 2nd ed.

41 APOSTOL. Modular Functions and Dirichlet Series in Number Theory. 2nd ed.

42 SERRE. Linear Representations of Finite Groups.

43 GILLMAN/JERISON. Rings of Continuous Functions.

44 KENDIG. Elementary Algebraic Geometry.

45 LOÈVE. Probability Theory I. 4th ed.

46 LOÈVE. Probability Theory II. 4th ed.

47 MOISE. Geometric Topology in Dimensions 2 and 3.

48 SACHS/WU. General Relativity for Mathematicians.

49 GRUENBERG/WEIR. Linear Geometry. 2nd ed.

50 EDWARDS. Fermat's Last Theorem.

51 KLINGENBERG. A Course in Differential Geometry.

52 HARTSHORNE. Algebraic Geometry.

53 MANIN. A Course in Mathematical Logic.

54 GRAVER/WATKINS. Combinatorics with Emphasis on the Theory of Graphs.

55 BROWN/PEARCY. Introduction to Operator Theory I: Elements of Functional Analysis.

56 MASSEY. Algebraic Topology: An Introduction.

57 CROWELL/FOX. Introduction to Knot Theory.

58 KOBLITZ. p-adic Numbers, p-adic Analysis, and Zeta-Functions. 2nd ed.

59 LANG. Cyclotomic Fields.

60 ARNOLD. Mathematical Methods in Classical Mechanics. 2nd ed.

continued after index

Thomas Becker Volker Weispfenning
In Cooperation with Heinz Kredel

Gröbner Bases

A Computational Approach to Commutative Algebra

Springer

Thomas Becker
Fakultät für Mathematik
und Informatik
Universität Passau
Postfach 2540
8390 Passau
Germany

Volker Weispfenning
Fakultät für Mathematik
und Informatik
Universität Passau
Postfach 2540
8390 Passau
Germany

Heinz Kredel
Fakultät für Mathematik
und Informatik
Universität Passau
Postfach 2540
8390 Passau
Germany

Mathematics Subjects Classifications (1991): 13P10, 68Q40, 12Y05

Library of Congress Cataloging-in-Publication Data
Becker, Thomas.
Gröbner bases: a computational approach to commutative algebra/
Thomas Becker, Volker Weispfenning, in cooperation with Heinz
Kredel.
p. cm. – (Graduate texts in mathematics)
Includes bibliographical references and index.
ISBN 978-1-4612-6944-1 ISBN 978-1-4612-0913-3 (eBook)
DOI 10.1007/978-1-4612-0913-3
1. Gröbner bases. 2. Algebra – Data processing. I. Weispfenning,
Volker. II. Kredel, Heinz. III. Title. IV. Series.
QA251.3.B43 1993
512'.24 – dc20 92-37955
CIP

Printed on acid-free paper.

Production managed by Natalie Johnson; manufacturing supervised by Vincent Scelta.
Photocomposed using the authors' LaTeX Files.

9 8 7 6 5 4 3 2 (Corrected second printing, 1998)

ISBN 978-1-4612-6944-1

Preface

The origins of the mathematics in this book date back more than two thousand years, as can be seen from the fact that one of the most important algorithms presented here bears the name of the Greek mathematician Euclid. The word "algorithm" as well as the key word "algebra" in the title of this book come from the name and the work of the ninth-century scientist Mohammed ibn Mûsâ al-Khowârizmî, who was born in what is now Uzbekistan and worked in Baghdad at the court of Harun al-Rashid's son. The word "algorithm" is actually a westernization of al-Khowârizmî's name, while "algebra" derives from "al-jabr," a term that appears in the title of his book *Kitab al-jabr wa'l muqabala*, where he discusses symbolic methods for the solution of equations. This close connection between algebra and algorithms lasted roughly up to the beginning of this century; until then, the primary goal of algebra was the design of constructive methods for solving equations by means of symbolic transformations.

During the second half of the nineteenth century, a new line of thought began to enter algebra from the realm of geometry, where it had been successful since Euclid's time, namely, the *axiomatic method.* The starting point of the axiomatic approach to algebra is the question, What kind of object is a symbolic solution to an algebraic equation? To use a simple example, the question would be not only, What is a solution of $ax + b = 0$, but also, What are the properties of the objects a and b that allow us to form the object $-b/a$? The axiomatic point of view is that these are objects in a surrounding algebraic structure which determines their behavior. The algebraic structure in turn is described and determined by properties that are laid down in a set of axioms.

The foundations of this approach were laid by Richard Dedekind, Ernst Steinitz, David Hilbert, Emmy Noether, and many others. The axiomatic method favors abstract, non-constructive arguments over concrete algorithmic constructions. The former tend to be considerably shorter and more elegant than the latter. Before the arrival of computers, this advantage more or less settled the question of which one of the two approaches was to be preferred: the algorithmic results of mathematicians like Leopold Kronecker and Paul Gordan were way beyond the scope of what could be done with pencil and paper, and so they had little to offer except being more tedious than their non-constructive counterparts.

On the other hand, it would be a mistake to construe the axiomatic and the algorithmic method as being irreconcilably opposed to each other. As a matter of fact, significant algorithmical results in algebra were proved by the very proponents of axiomatic thinking such as David Hilbert and Emmy Noether. Moreover, mathematical logic—a field that centers around the axiomatic method—made fundamental contributions to algorithmic mathematics in the 1930s. Alan Turing and Alonzo Church for the first time made precise the concept of computability in what is known as Church's thesis, or also as the Church-Turing thesis. Kurt Gödel proved that certain problems inherently elude computability and decidability. This triggered a wave of new results by Alfred Tarski and other members of the Polish school of logicians on the algorithmic solvability or unsolvability of algebraic problems. Again, because of their enormous complexity, these algorithms were of no practical significance whatsoever. As a result, the beginning second half of this century saw an axiomatic and largely non-constructive approach to algebra firmly established in both research and teaching.

The arrival of computers and their breathtaking development in the last three decades then prompted a renewed interest in the problem of effective constructions in algebra. Many constructive results from the past were unearthed, often after having been rediscovered independently. Moreover, the development of new concepts and results in the area has now established *computer algebra* as an independent discipline that extends deeply into both mathematics and computer science.

There are many good reasons for viewing computer algebra as an independent field. However, the fact that the mathematical part of it is somewhat separated from the work of pure algebraists is, in our opinion, rather unfortunate and not at all justified. We feel that this situation must and will change in the near future. As a matter of fact, computational aspects are beginning to show up more and more in undergraduate-level textbooks on abstract algebra. There is, however, one particular contribution made by computational algebra that is in most dire need of being introduced in the mathematical mainstream, namely, the theory of *Gröbner bases.*

Gröbner bases were introduced by Bruno Buchberger in 1965. The terminology acknowledges the influence of Wolfgang Gröbner on Buchberger's work. To the reader who has any background in abstract algebra at all, the basic idea behind the theory is easily explained. Suppose you are given a finite set of polynomials in one variable over a field and you wish to decide membership in the ideal generated by these polynomials in the polynomial ring. What you must do is compute the greatest common divisor of the given polynomials by means of the Euclidean algorithm. Any given polynomial then lies in the ideal in question if and only if its remainder upon division by this gcd equals zero. Gröbner basis theory is the successful attempt to imitate this procedure for polynomials in several variables. Given a finite set of multivariate polynomials over a field, the *Buchberger algorithm* computes a new set of polynomials, called a Gröbner basis, which generates the same ideal as the original one and is an analogue to the gcd of the unvariate case in the following sense. A given polynomial lies in the ideal generated by the Gröbner basis if and only if a suitably defined normal form of the polynomial with respect to the Gröbner basis equals zero. The computation of this normal form is a rather straightforward generalization of long division of polynomials, except that we are looking at the division of one polynomial by a set of finitely many polynomials.

Considering both the outstanding importance of the Euclidean algorithm for the computation of gcd's of univariate polynomials and the scope of its implications in pure and computational algebra, it should come as no surprise that its multivariate analogue, the Buchberger algorithm for the computation of Gröbner bases, is of similar relevance. It leads to solutions to a large number of algorithmic problems that are related to polynomials in several variables. Most notably, algorithms that involve Gröbner basis computations allow exact conclusions on the solutions of systems of nonlinear equations, such as the (geometric) dimension of the solution set, the exact number of solutions in case there are finitely many, and their actual computation with arbitrary precision.

Most of the problems for which Gröbner bases provide algortihmic solutions were already known to be solvable in principle. Gröbner bases are a giant step forward insofar as actual implementations have become feasible and have actually provided answers to physicists and engineers. On the other hand, many problems of no more than moderate input size still defy computation. The mathematics behind the algorithms as well as the hardware that performs them have a long way to go before these problems can be considered solved to the satisfaction of the user.

The purpose of this book is to give a self-contained, mathematically sound introduction to the theory of Gröbner bases and to some of its applications, stressing both theoretical and computational aspects.

A book that would start out with Gröbner basis theory would have to direct its readers to a source for a large number of elementary results on commutative rings and, more specifically, on polynomials in several variables. These are of course all available somewhere, and certainly known to the mature mathematician. However, we found ourselves unable to name a reasonably small number of books that would enable the beginning graduate student or the non-mathematician with an interest in Gröbner bases to aquire this background within a reasonable amount of time. *We have therefore decided to write a book that requires no prerequisites other than the mathematical maturity of an advanced undergraduate student.* In particular, no prior knowledge of abstract algebra whatsoever is assumed. Under the European system, this means that the book can be used after the second semester of mathematics or computer science. People with different backgrounds will enter such a book at different points; for more details, we refer the reader to the comments on "How to Use This Book" on p. xi.

As for the overall concept, the book traverses three stages. Chapters 0–3 provide pre-Gröbner-bases results on commutative rings with an emphasis on polynomial rings, as well as the basics on vector spaces and modules. Chapters 4 and 5 then develop Gröbner basis theory. The definition of a Gröbner basis does not show up until Section 5.2, but the material of Chapter 4 and Section 5.1 is rather specific to Gröbner bases already. Chapters 6–10 cover a wide range of applications, intertwined with a development of post-Gröbner-bases algebra. Algorithms are presented using a semi-formalism that is self-explanatory even to those with no background in computer programming. Strong emphasis is placed on a mathematically sound verification of the algorithms. Each chapter closes with a "Notes" section that puts the material in a larger mathematical perspective by tracing its historical development and providing references to the literature.

Needless to say, the list of omissions is tremendous. If it is possible at all to write the definitive book on computational algebra, then this is not it.

More specifically, the choice of the material and the reasons for making it are as follows. The introductory chapters 0–3 are written mainly for the purpose of providing the necessary background for Gröbner bases and their applications. The solutions to algorithmic problems such as factorization of polynomials given there are strictly "in principle" solutions; implementations of any practical value involve considerably more mathematics. Our treatment is thus incomplete in a sense; on the other hand, we are laying firm mathematical foundations which can also be helpful for the reader who wishes to proceed to the advanced literature on topics in computational algebra other than Gröbner bases.

Chapters 4 and 5, the main chapters on Gröbner bases, are fairly complete both theoretically and algorithmically. The theory of orders and reduction relations of Chapter 4 is rather well-rounded. In Chapter 5, the theoretical aspects of Gröbner bases are explored extensively. The Buchberger algorithm for their computation is presented first in an "in principle" version and then in two real-life versions. The only major omission in these two chapters—and it is one that actually pervades the entire book—is the absence of any *complexity theory*, that is, the discussion of the time and space that an algorithm requires as a function of the size of its input. This omission is clearly a serious one. It was not made because we consider the issue to be of minor importance. On the contrary, we feel that complexitiy theory is too important an issue to be dealt with lightly. We hope that our effort will motivate others to treat these problems comprehensively in some kind of book format. A brief overview of complexity results for Gröbner basis constructions is given in the appendix "Outlook on Advanced and Related Topics" at the end of the book.

Once Gröbner bases have been introduced, there is an almost limitless choice of topics that one could cover. Our focus in Chapters 6–10 is on the theory of polynomial ideals. A large number of ready-to-use algorithms is presented. Furthermore, we demonstrate how Gröbner bases can often be used to give elegant an enlightening proofs of classical results, for example, in the area of algebraic field extensions. This shows that Gröbner bases are not only a powerful tool for actual computations, but also a cornerstone of commutative algebra.

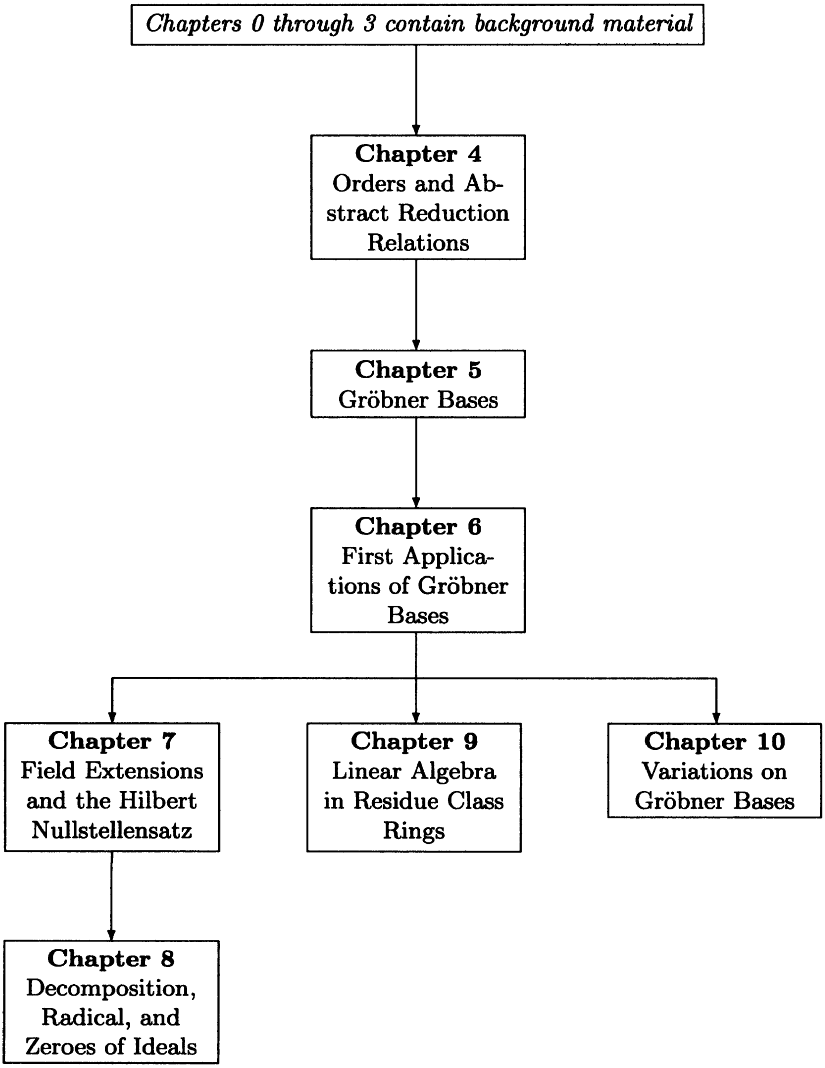
The book closes with an appendix that tries to at least partly make up for the incompleteness of this book. Here, we have given brief summaries of a number of recent results that surround or extend Gröbner basis theory. Each section explains a problem, outlines the solution, and provides a guide to the original literature.

The authors wish to thank Johannes Grabmeier, Alexander Knapp, Frank Lippold, Wolfgang Mark, Christian Münch, Michael Pesch, Gernot Schreib, and Thomas Sturm for reading parts of the manuscript. Gerlinde Kollmer kept us organized and did a lot of work in LaTeX along the way. The typesetting of the final manuscript was done by the first author in LaTeX—with the additional use of several AMSFonts—on an Atari Mega 2. Special thanks is due to Michael Pesch for his superb software consulting, and to Thomas Sturm for his competence and dedication.

Passau, Germany T.B., V.W., H.K.

How to Use This Book

Interdependence of Chapters


Chapters 0 through 3 contain background material
Chapter 4
Orders and Abstract Reduction Relations
Chapter 5
Gröbner Bases
Chapter 6
First Applications of Gröbner Bases
Chapter 7
Field Extensions and the Hilbert Nullstellensatz
Chapter 9
Linear Algebra in Residue Class Rings
Chapter 10
Variations on Gröbner Bases
Chapter 8
Decomposition, Radical, and Zeroes of Ideals

Sections 6.1, 6.4, and 7.5–7.7 are exempt from this flow diagram. They can be postponed or dropped altogether; details are to be found at the beginning of each of these sections.

Prerequisites

Chapters 0–3 of this book are written for the reader with very little or no background in abstract algebra. The prerequisite for this part is the mathematical maturity of an advanced undergraduate student. You may skip these chapters if you can answer the following questions.

> What is a commutative ring with unity, and when is it a field?
>
> What is an ideal, and what is a residue class ring?
>
> What does the Euclidean algorithm do with two univariate polynomials over a field, and how does it do it?
>
> What is a vector space?

If you failed the test, then you must read Chapters 0 and 1 and the first two sections of Chapter 2 to be able to understand the main part on Gröbner bases (Chapters 4 and 5). If you decide to continue on past Chapter 5 into the applications, you will soon feel the need to read the rest of Chapter 2 as well as Chapter 3.

If you passed the test or know you could, then for you, the book begins with Chapter 4. If you need to go back to one of the first four chapters for some specific definition or result that you have trouble with, then the index and the extensive cross-referencing of this book should make it easy for you to do so.

Exercises

There are two types of exercises: those printed in normal size, and those in small print. Normal size indicates that these exercises have the status of lemmas whose proof is left to the reader. Their statements will be used later on. None of them are hard; working them is also a good way of making sure that you are ready to grasp the material that is being presented next. Small print indicates exercises in the usual sense of application and extension of what has just been covered. The difficulty ranges from easy to moderate.

Use of Computer Algebra Systems

It is possible to view this as a mathematics textbook that can be read without the use of a computer. On the other hand, most of the mathematics presented here is application-oriented, and seeing things happen or making things happen on the screen will greatly enhance the experience of studying the material.

If a computer algebra system is at hand, then there are basically two things that you can do along with reading this book. Firstly, if an algorithm that you have just learned about is available on your system, you can simply run it on examples that you make up, get from the exercises, or find somewhere else. Although this is somewhat less than creative, you will be surprised how much it helps your understanding and motivation. The other thing is to implement algorithms from the book. Doing so from scratch will in general be a major endeavor. However, many algorithms in computational algebra are such that they allow a top-down approach, where good results can be obtained by tying together lower-level algorithms with relatively little effort. In order to do this, you need a system that provides a library of polynomial algorithms and the possibility to use them in your own programs. If you implement an algorithm that was already part of your system, then you have worked a useful exercise; if it was not, then you have extended the capabilities of your system.

Commercially available computer algebra systems that are suited to be used along with this book include Axiom, Macsyma, Maple, Mathematica, and Reduce. A system that the authors of this book recommend is MAS by Heinz Kredel. MAS makes available for interactive and programming use an extensive library of polynomial algorithms, including those that were developed for the system ALDES/SAC-2. In addition to such classics as greatest common divisors, factorization, and real root isolation, you will find the Buchberger algorithm for the computation of Gröbner bases as well as applications thereof such as ideal decomposition and real roots of polynomial systems. Of the more recent variants of the Buchberger algorithm, the non-commutative case (polynomial rings of solvable type), comprehensive Gröbner bases, and Gröbner bases over principal ideal domains and Euclidean domains are implemented. Programming in MAS is in a language that is based on MODULA-2. User-defined programs can be run interactively; if a MODULA-2 compiler is available, they can also be compiled, thus allowing a fair comparison between existing and user-defined versions of algorithms. MAS is available free of charge per anonymous ftp from alice.fmi.uni-passau.de and via World Wide Web from http://alice.fmi.uni-passau.de/mas.htm. Currently available is version 1.0 for UN*X workstations (e.g. IBM RS6000/AIX, HP 9000/HP-UX, NextStep, Sun Sparc with a Modula-2 to C translator) and PCs 386, 486, 586 (DOS, OS2 and Linux).

Use as a Textbook

It should be clear from the above discussion of prerequisites that this book allows a variety of uses as a textbook on the advanced undergraduate as well as the graduate level. There is at present no established way of including Gröbner bases in the mathematics/computer science curriculum. The fact that this book requires practically no specific prior knowledge should make it possible to experiment in this regard.

One conceivable situation that deserves perhaps some comment is the following. Suppose you are at a point where the basic theory of commutative rings and polynomial rings is available. Now you wish to cover Gröbner bases, but you do not have the time and/or the desire to get into the theory of orders and reduction relations to the extent that they are treated in Chapter 4. You may then essentially start with Section 4.5, which deals with reduction relations and Newman's lemma. This requires only a moderate amount of material from the earlier sections of Chapter 4, and you should have no trouble providing this material. You then jump ahead to Section 5.1. You will need some more material from Chapter 4, most of which is obvious and easily provided, such as the definition of a quasi-order. The only deeper results that you will need are Dickson's lemma, whose proof you lift from the proof of Proposition 4.49, the well-foundedness of term orders, which you prove using the comments in Exercise 4.63, and the properties of the induced quasi-order on the polynomial ring, which you transfer from Lemma 4.67 and Theorem 4.69.

Abbreviations

The following abbreviations will be used throughout this book.

cf., (Latin *confer*) compare
e.g., (Latin *exempli gratia*) for example
etc., (Latin *et cetera*) and so on
i.e., (Latin *id est*) that is
iff, if and only if
w.l.o.g., without loss of generality
w.r.t., with respect to

Moreover, a □ will indicate the end of a proof.

Numberings

Chapters and sections are numbered in the obvious way: Chapter 5, for example, consists of Sections 5.1–5.6. Definitions, lemmas, propositions, theorems, corollaries, and exercises are treated as one type of item and numbered consecutively within each chapter: Chapter 5 contains Exercise 5.1, Theorem 5.2, etc. Due to the fact that there is such an item on virtually every page, this should make it easy to locate referenced items. Algorithms are given in tables in order to prevent them from running across a page-break; these tables are also numbered within each chapter.

Contents

List of Algorithms

0

Basics

0.1 Natural Numbers and Integers

A mathematically rigorous definition of the number systems requires the use of axiomatic set theory. As with most of mathematics, however, the intuitive understanding of the natural numbers $\mathbb{N}$, the integers $\mathbb{Z}$, the rationals $\mathbb{Q}$, the reals $\mathbb{R}$, and the complex numbers $\mathbb{C}$ gained in elementary mathematics is sufficient for the beginning student of algebra. The occasional intrusion of set theory and foundational problems can be dealt with later. In this section, we discuss some properties of $\mathbb{N}$ and $\mathbb{Z}$ that are somewhat less than elementary. Throughout this book, we will use the convention that $0 \in \mathbb{N}$. The set $\mathbb{N} \setminus \{0\}$ of all positive natural numbers will be denoted by $\mathbb{N}^+$.

Theorem 0.1 (Induction Principle) *Let P be a property that a natural number may or may not have, and let us write "$P(n)$" for "the natural number n has property P." Assume that we can prove*

(i) $P(0)$, and

(ii) $P(n)$ implies $P(n+1)$ for all $n \in \mathbb{N}$.

Then $P(n)$ holds for all $n \in \mathbb{N}$.

The induction principle is too close to the axiomatic foundations of mathematics to be proved rigorously without the explicit knowledge and use of these foundations. We will therefore have to put it on the list of things that we accept by intuition. We can, however, make an intuitive argument for its plausibility. Assume that we can prove (i) and (ii) above. Now if someone hands us an arbitrary but fixed natural number m, then we can prove $P(m)$ as follows. We know that $P(0)$. Together with (ii) above, we conclude that $P(1)$. Using (ii) again, we can prove $P(2)$, and so on. Repeating the argument m times, we arrive at the conclusion $P(m)$. Being able to produce a proof of $P(m)$ for arbitrary $m \in \mathbb{N}$, we may with some plausibility claim to know that $P(m)$ holds for all $m \in \mathbb{N}$. Let us emphasize again that the above argument is not a mathematically satisfying proof since it makes a number of tacit assumptions on $\mathbb{N}$ (the existence of $\mathbb{N}$ being one of them) which would have to be postulated or derived from such postulates.

Recall that we have included 0 in $\mathbb{N}$. We mention that the induction principle may be applied with 0 replaced by any natural number k in (i) of Theorem 0.1. The conclusion of the theorem then states that $P(n)$ holds for all $n \in \mathbb{N}$ with $n \geq k$.

Exercise 0.2 Use the induction principle to prove that

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

for all $n \in \mathbb{N}$ with $n > 0$.

Once we have accepted the induction principle for one reason or another, we can prove two important corollaries, each of which is in fact equivalent to the induction principle.

Corollary 0.3 *Let P be as in Theorem 0.1, and assume that we can prove*

(i) $P(0)$, and

(ii) for all $n \in \mathbb{N}$: $P(m)$ for all $m \leq n$ implies $P(n+1)$.

Then $P(n)$ holds for all $n \in \mathbb{N}$.

Proof For arbitrary $n \in \mathbb{N}$, let $Q(n)$ be the property "$P(m)$ for all $m \leq n$." Since obviously $Q(n)$ implies $P(n)$ for all $n \in \mathbb{N}$, it will suffice to prove $Q(n)$ for all $n \in \mathbb{N}$. We use the induction principle. $Q(0)$ is equivalent to $P(0)$ which we know to be true. Now assume that $n \in \mathbb{N}$ with $Q(n)$. This means that $P(m)$ for all $m \leq n$, and (ii) allows us to conclude $P(n+1)$. "$Q(n)$ and $P(n+1)$" is equivalent to $Q(n+1)$. We have thus verified (i) and (ii) of the induction principle for Q, so it follows that $Q(n)$ for all $n \in \mathbb{N}$ as desired. □

Corollary 0.4 *Let M be a non-empty subset of* $\mathbb{N}$. *Then M has a least element.*

Proof We will show that if M does not have a least element, then it is empty. Assume that M does not have a least element. To prove that M is empty, it suffices to show that every $n \in \mathbb{N}$ has the property "$n \notin M$," which will be achieved by means of the above version of the induction principle. If 0 were in M, then 0, being the least element of all of $\mathbb{N}$, would be a least element of M. We see that $0 \notin M$. Now assume that $n \in \mathbb{N}$, and $m \notin M$ for all $m \leq n$. If $n+1$ were in M, then it would be a least element of M since we have assumed that M does not contain any natural number that is less than $n+1$. We have proved $(n+1) \notin M$. □

Exercise 0.5 Show that Theorem 0.1, Corollary 0.3, and Corollary 0.4 are equivalent.

Corollary 0.4 is instrumental in the following rigorous proof of a widely known fact.

Proposition 0.6 *Let m, $n \in \mathbb{Z}$ with $n \neq 0$. Then there exist unique q, $r \in \mathbb{Z}$ such that $m = qn + r$ and $0 \leq r < |n|$.*

Proof We begin by proving uniqueness. Assume that we have q, r, q', $r' \in \mathbb{Z}$ with $m = qn + r = q'n + r'$ and $0 \leq r, r' < |n|$. Assume w.l.o.g. that $r \geq r'$. Then

$$(q' - q)n = r - r' \geq 0.$$

From $n \neq 0$ it follows that $r - r' = 0$ iff $q' - q = 0$. So if $r - r' = 0$, then we have $r = r'$ and $q = q'$ as desired. If $r - r' \neq 0$, then $0 < r - r' \leq r < |n|$, and this contradicts the inequality

$$r - r' = |r - r'| = |(q' - q)n| = |(q' - q)|\,|n| \geq |n|.$$

To prove existence, we distinguish between two cases.
Case 1: $0 < n$. We define a subset M of $\mathbb{N}$ by setting

$$M = \{\, m - sn \mid s \in \mathbb{Z} \,\} \cap \mathbb{N}.$$

Then M is not empty: if $0 \leq m$, then $m = m - 0 \cdot n \in M$, and if $m < 0$, then

$$m - mn = (-m)(-1 + n) = |m|(n - 1) \in M.$$

By Corollary 0.4, M has a least element r. Since $r \in M$, there must exist $q \in \mathbb{Z}$ with $r = m - qn$ which means $m = qn + r$. We claim that r satisfies $0 \leq r < |n| = n$. We must have $0 \leq r$ since $r \in \mathbb{N}$. Assume for a contradiction that $r \geq n$. Then $r > r - n \geq 0$ and thus

$$r - n = (m - qn) - n = m - (q + 1)n \in M,$$

contradicting the minimality of r.
Case 2: $n < 0$. By case 1 above, there exist q, $r \in \mathbb{Z}$ with $m = q(-n) + r$ and $0 \leq r < |-n| = |n|$. We see that $-q$ and r have the desired properties. □

The integers q and r of the proposition above are called, respectively, the **quotient** and **remainder** of m upon division by n. The proof that we have just given is a typical example of a non-constructive argument; it uses the existence of a least element in a set of natural numbers without giving a method to find it. The problem of effectively finding integer quotients and remainders will be taken up again in Section 0.3.

Proposition 0.6 has a number of very important consequences, some of which the reader is certainly familiar with. They will be proved and discussed thoroughly in Chapter 2. We list the more elementary ones here without proof for the sole purpose of providing a wider range of examples as we go along. None of the following results will be used in the development of the theory until it has been proved.

Let m, $n \in \mathbb{Z}$. We say that m **divides** n and write $m \mid n$ if there exists q in $\mathbb{Z}$ with $n = qm$. Here, q and m are called **divisors** of n. An integer d is called a **greatest common divisor**, or **gcd**, of m and n if d divides m and n and is divided by any common divisor of m and n. It is true that any two integers m and n have a gcd d in $\mathbb{Z}$. Moreover, there exist s, $t \in \mathbb{Z}$ with $d = sm + tn$, and d, s, and t can be computed effectively by the so-called **extended Euclidean algorithm** which is described in the proof of Theorem 2.32. Although it is embedded in a much more abstract context there, it is actually possible for the interested reader to look it up now. We mention that d is a gcd of m and n iff $-d$ is one, and that these are the only ones.

An integer $p \geq 2$ is called **prime**, or a **prime**, or a **prime number**, if it has the following property: whenever $p \mid mn$ with m, $n \in \mathbb{Z}$, then $p \mid m$ or $p \mid n$. It can be shown that this definition is equivalent to the condition that 1, -1, p, and $-p$ be the only divisors of p. One concludes immediately that a gcd of a prime with any other integer must be one of those four, and that a gcd of a prime p with any integer n satisfying $1 \leq n < p$ must be 1 or -1. Finally (still as a consequence of Proposition 0.6), it can be shown that every integer other than -1, 0, and 1 can be written as a product of primes and a possible factor of -1, and that this factorization is unique up to the order of the factors.

Once all this has been proved, one can easily show that there must be infinitely many primes, a fact that is frequently used in computer algebra. There are a number of important algorithms that require, in addition to the input they are supposed to manipulate, the input of finitely many primes. One knows in advance that the algorithm may, depending on the input, reject a finite number of finite sets of primes which one tries to input. To be able to assert that the algorithm can be made to run for any input, one therefore needs to know that there is an unlimited supply of primes.

Assume that there were only finitely many primes, say $p_1, \ldots, p_k$. Consider the integer $m = p_1 \cdot \cdots \cdot p_k + 1$. Then m can obviously be written as $m = q_i p_i + 1$ for each $1 \leq i \leq k$: take for q_i the product of the $k - 1$ primes different from p_i. If m were divisible by one of the k primes, say $m = qp_i$, then we would be contradicting the uniqueness of the quotient and remainder of Proposition 0.6. But m does have a factorization into prime numbers (possibly just one, namely, itself), so there must be more primes out there than $p_1, \ldots, p_k$.

0.2 Maps

Definition 0.7 Let A_1, ..., A_n be sets. Then the **Cartesian product**

$$A_1 \times \cdots \times A_n$$

of $A_1, \ldots, A_n$ is defined as the set of all ordered n-tuples $(a_1, \ldots, a_n)$ such that $a_i \in A_i$ for $1 \leq i \leq n$. This is sometimes also denoted by $\prod_{i=1}^n A_i$, and $\prod_{i=1}^n A$ is also written as A^n.

Definition 0.8 Let A and B be sets. A **map**, or **function**, with **domain** A and **range** B is a set $\varphi \subseteq A \times B$ such that for each $a \in A$, there exists exactly one $b \in B$ with $(a, b) \in \varphi$.

A map φ with domain A and range B is often given as a rule which assigns to each $a \in A$ exactly one $b \in B$, namely, the unique $b \in B$ with $(a, b) \in \varphi$. Adopting this point of view, we will speak of a map $\varphi : A \longrightarrow B$ from A to B and denote, for $a \in A$, the unique $b \in B$ with $(a, b) \in \varphi$ by $\varphi(a)$. The notation $a \longmapsto \varphi(a)$ is often used when a map is to be defined. For example,

$$\begin{array}{rccc} \varphi : & \mathbb{N} & \longrightarrow & \mathbb{N} \\ & n & \longmapsto & n+1 \end{array}$$

defines φ to be the map from $\mathbb{N}$ to $\mathbb{N}$ that satisfies $\varphi(n) = n + 1$ for all $n \in \mathbb{N}$. The set of all maps from A to B is denoted by B^A.

Definition 0.9 Let A and B be sets, $\varphi : A \longrightarrow B$ a map from A to B. If $a \in A$, then $\varphi(a)$ is called the **image of** a in B under φ. If $X \subset A$, then

$$\varphi(X) = \{ \varphi(a) \mid a \in X \}$$

is called the **image** of X in B under φ. If $b \in B$, then any $a \in A$ with $\varphi(a) = b$ (of which there may well be more than one, or none at all) is called a **preimage** of b in A under φ. If $Y \subseteq B$, then

$$\varphi^{-1}(Y) = \{ a \in A \mid \varphi(a) \in Y \}$$

is called the **inverse image** of Y in A under φ. If C is a subset of A, then the map $\psi : C \longrightarrow B$ defined by $\psi(c) = \varphi(c)$ for all $c \in C$ is called the **restriction** of φ to C. In this case, we write $\psi = \varphi \restriction C$.

By an abuse of notation which can become quite confusing, some people sometimes write $\varphi^{-1}(b)$ for $\varphi^{-1}(\{b\})$ when $b \in B$.

Lemma 0.10 Let $\varphi : A \longrightarrow B$ be a map.

(i) Let $Y_1, Y_2 \subseteq B$. Then

$$\varphi^{-1}(Y_1 \cup Y_2) = \varphi^{-1}(Y_1) \cup \varphi^{-1}(Y_2),$$

and

$$\varphi^{-1}(Y_1 \cap Y_2) = \varphi^{-1}(Y_1) \cap \varphi^{-1}(Y_2).$$

(ii) Let $X_1, X_2 \subseteq A$. Then

$$\varphi(X_1 \cup X_2) = \varphi(X_1) \cup \varphi(X_2),$$

and

$$\varphi(X_1 \cap X_2) \subseteq \varphi(X_1) \cap \varphi(X_2).$$

Proof (i) The first claim holds because for all $a \in A$, we have

$$\begin{aligned} a \in \varphi^{-1}(Y_1 \cup Y_2) &\iff \varphi(a) \in (Y_1 \cup Y_2) \\ &\iff \varphi(a) \in Y_1 \text{ or } \varphi(a) \in Y_2 \\ &\iff a \in \varphi^{-1}(Y_1) \text{ or } a \in \varphi^{-1}(Y_2) \\ &\iff a \in \left(\varphi^{-1}(Y_1) \cup \varphi^{-1}(Y_2)\right). \end{aligned}$$

Similarly,

$$\begin{aligned} a \in \varphi^{-1}(Y_1 \cap Y_2) &\iff \varphi(a) \in (Y_1 \cap Y_2) \\ &\iff \varphi(a) \in Y_1 \text{ and } \varphi(a) \in Y_2 \\ &\iff a \in \varphi^{-1}(Y_1) \text{ and } a \in \varphi^{-1}(Y_2) \\ &\iff a \in \left(\varphi^{-1}(Y_1) \cap \varphi^{-1}(Y_2)\right). \end{aligned}$$

(ii) The first claim follows from the following equivalence which holds for all $b \in B$.

$$\begin{aligned} b \in \varphi(X_1 \cup X_2) &\iff b = \varphi(a) \text{ for some } a \in (X_1 \cup X_2) \\ &\iff b = \varphi(a) \text{ for some } a \in X_1 \text{ or some } a \in X_2 \\ &\iff b \in \varphi(X_1) \text{ or } b \in \varphi(X_2) \\ &\iff b \in \left(\varphi(X_1) \cup \varphi(X_2)\right) \end{aligned}$$

Finally, let $b \in (\varphi(X_1 \cap X_2))$. Then there exists $a \in (X_1 \cap X_2)$ with $\varphi(a) = b$. Since $a \in X_1$ and $a \in X_2$, this shows that $b \in \varphi(X_1)$ and $b \in \varphi(X_2)$, i.e., $b \in (\varphi(X_1) \cap \varphi(X_2))$. □

The following example shows that the reverse inclusion for the intersection in (ii) above does not hold in general.

Example 0.11 Let $\varphi : \mathbb{Z} \longrightarrow \mathbb{N}$ be defined by $\varphi(n) = |n|$ for all $n \in \mathbb{Z}$, and let $X_1 = \{-1\}$, $X_2 = \{1\}$. Then $\varphi(X_1 \cap X_2) = \varphi(\emptyset) = \emptyset$, but $\varphi(X_1) \cap \varphi(X_2) = \{1\}$.

Definition 0.12 Let $\varphi : A \longrightarrow B$ be a map. Then φ is called

(i) **injective**, or **one-to-one**, if $\varphi(a_1) = \varphi(a_2)$ implies that $a_1 = a_2$ for all $a_1, a_2 \in A$, i.e., no two different elements of A ever have the same image in B under φ,

(ii) **surjective**, or **onto**, if $\varphi(A) = B$, i.e., for each $b \in B$, there exists an $a \in A$ with $\varphi(a) = b$, and

(iii) **bijective**, or a **bijection**, if it is both injective and surjective. A bijection from a finite set X to itself is also called a **permutation on** S.

Here is a simple reformulation of the above definition: φ is injective if for every $b \in B$, there is at most one preimage under φ, surjective if there is at least one, and bijective if there is exactly one. Verification of the following examples is left to the reader.

Examples 0.13 (i) The map $\varphi : \mathbb{N} \longrightarrow \mathbb{N}$ defined by $\varphi(n) = n + 1$ for all $n \in \mathbb{N}$ is injective but not surjective.

(ii) The map $\varphi : \mathbb{Z} \longrightarrow \mathbb{N}$ defined by $\varphi(m) = |m|$ for all $m \in \mathbb{Z}$ is surjective but not injective.

(iii) The map $\varphi : \mathbb{Z} \longrightarrow \mathbb{Z}$ defined by $\varphi(m) = -m$ for all $m \in \mathbb{Z}$ is bijective.

If A is a set, then the map $\varphi : A \longrightarrow A$ defined by $\varphi(a) = a$ for all $a \in A$ is called the **identity** on A and is denoted by id_A. If $X \subseteq A$, then the map $\iota : X \longrightarrow A$ defined by $\iota(a) = a$ for all $a \in X$ is called the **inclusion map** of X in A.

Exercise 0.14 Show that the identity map is always bijective, and inclusion maps are always injective.

Lemma 0.15 Let $\varphi : A \longrightarrow B$ be a map. Then the following hold:

(i) $X \subseteq \varphi^{-1}(\varphi(X))$ for all $X \subseteq A$. Moreover, $X = \varphi^{-1}(\varphi(X))$ for all $X \subseteq A$ iff φ is injective.

(ii) $\varphi(\varphi^{-1}(Y)) \subseteq Y$ for all $Y \subseteq B$. Moreover, $\varphi(\varphi^{-1}(Y)) = Y$ for all $Y \subseteq B$ iff φ is surjective.

Proof (i) Let $a \in X \subseteq A$. Using the definitions of image and inverse image, we see that $\varphi(a) \in \varphi(X)$ and thus $a \in \varphi^{-1}\varphi((X))$. Now assume that $X = \varphi^{-1}(\varphi(X))$ for all $X \subseteq A$, and let $a_1, a_2 \in A$ with $\varphi(a_1) = \varphi(a_2)$. Then

$$a_2 \in \varphi^{-1}\big(\varphi(\{a_1\})\big) = \{a_1\}$$

and thus $a_1 = a_2$. Conversely, assume that φ is injective, and let $X \subseteq A$ and $a \in \varphi^{-1}(\varphi(X))$. Then $\varphi(a) \in \varphi(X)$, and thus there exists $c \in X$ with $\varphi(c) = \varphi(a)$. Injectivity of φ implies $a = c$, and so $a \in X$.

(ii) Let $Y \subseteq B$ and $b \in \varphi(\varphi^{-1}(Y))$. Then there exists $a \in \varphi^{-1}(Y)$ with $\varphi(a) = b$, and we conclude that $b = \varphi(a) \in Y$. Now assume that $\varphi(\varphi^{-1}(Y)) = Y$ for all $Y \subseteq B$, and let $b \in B$. Then in particular,

$$\{b\} \subseteq \varphi\big(\varphi^{-1}(\{b\})\big), \quad \text{i.e.} \quad b \in \varphi\big(\varphi^{-1}(\{b\})\big),$$

and so there exists $a \in \varphi^{-1}(\{b\}) \subseteq A$ with $\varphi(a) = b$. Finally, assume that φ is surjective, and let $b \in Y \subseteq B$. Then there exists $a \in A$ with $\varphi(a) = b$. We see that $a \in \varphi^{-1}(Y)$ and thus $b \in \varphi(\varphi^{-1}(Y))$. □

Exercise 0.16 Use the examples of injective and surjective maps given earlier to understand that (with the notation of the last lemma), $\varphi^{-1}(\varphi(X)) \subseteq X$ does not hold in general for non-injective maps, and $Y \subseteq \varphi(\varphi^{-1}(Y))$ does not hold in general for non-surjective maps.

Definition 0.17 Let $\varphi : A \longrightarrow B$ and $\psi : B \longrightarrow C$ be maps. Then the **composition** $\psi \circ \varphi$ of φ and ψ is the map $\psi \circ \varphi : A \longrightarrow C$ defined by $(\psi \circ \varphi)(a) = \psi(\varphi(a))$ for all $a \in A$.

Exercise 0.18 Let $\varphi : A \longrightarrow B$, $\psi : B \longrightarrow C$, and $\chi : C \longrightarrow D$ be maps. Show that $(\chi \circ \psi) \circ \varphi = \chi \circ (\psi \circ \varphi)$.

Lemma 0.19 Let φ and ψ be as in the above definition. Then the following hold:

(i) If $\psi \circ \varphi$ is injective, then φ is injective.

(ii) If $\psi \circ \varphi$ is surjective, then ψ is surjective.

Proof (i) Let $a_1, a_2 \in A$ with $\varphi(a_1) = \varphi(a_2)$. Then $\psi(\varphi(a_1)) = \psi(\varphi(a_2))$, and so $a_1 = a_2$ since $\psi \circ \varphi$ is injective. To prove (ii), let $c \in C$. Since $\psi \circ \varphi$ is surjective, there exists $a \in A$ with $\psi(\varphi(a)) = c$, and we see that $\varphi(a)$ is the desired preimage of c under ψ in B. □

Exercise 0.20 Let φ and ψ be as above. Show that if both ψ and φ are injective (surjective), then the composition $\psi \circ \varphi$ is injective (surjective).

Lemma 0.21 Let $\varphi : A \longrightarrow B$ be a map. Then the following hold:

(i) If $A \neq \emptyset$, then φ is injective iff there exists a map $\psi : B \longrightarrow A$ with $\psi \circ \varphi = \mathrm{id}_A$.

(ii) φ is surjective iff there exists a map $\psi : B \longrightarrow A$ with $\varphi \circ \psi = \mathrm{id}_B$.

(iii) φ is bijective iff there is a map $\psi : B \longrightarrow A$ such that both $\psi \circ \varphi = \mathrm{id}_A$ and $\varphi \circ \psi = \mathrm{id}_B$ hold.

Proof The directions "$\Longleftarrow$" follow immediately from Lemma 0.19 together with the fact that identities are bijective. For "$\Longrightarrow$" of (i) we define $\psi(b)$, for $b \in B$, to be the unique preimage of b under φ if one exists at all, an arbitrary element of A otherwise. It is immediate from this definition that $\psi(\varphi(a)) = a$ for all $a \in A$. For "$\Longrightarrow$" of (ii) we define $\psi(b)$, for $b \in B$, to be any one of the preimages of b under f, knowing that there must be at least one. One verifies immediately that $\varphi(\psi(b)) = b$ for all $b \in B$. Finally, for "$\Longrightarrow$" of (iii), we set $\psi(b)$ equal to the unique preimage of b under φ, and it is now easy to check that $\psi \circ \varphi = \mathrm{id}_A$ and $\varphi \circ \psi = \mathrm{id}_B$. □

It is easy to see that for bijective φ, the map ψ of (iii) above is uniquely determined by φ. It is also called the **inverse** of φ and denoted by φ^{-1} (cf. the remark following Definition 0.9).

The proof of the existence of ψ in (ii) of the last lemma actually involves a set-theoretic subtlety. This will be discussed in Section 4.1; it need not bother us for the moment.

Exercise 0.22 Show that $\varphi : A \longrightarrow B$ is injective iff the reverse inclusion for the intersection in Lemma 0.10 (ii) holds for all X_1, $X_2 \subseteq A$.

If A is a set with finitely many elements, then we denote by $|A|$ the number of elements of A. ($|A|$ is also called the **cardinality** of A.)

Proposition 0.23 *Let A be a finite set, φ a map from A to itself. Then φ is injective iff it is surjective.*

Proof Assume that φ is surjective. Then each one of the $|A|$ elements of A has at least one preimage, and by the definition of maps, no two different elements can have a preimage in common. But there are only $|A|$ preimages available, so each $a \in A$ can have at most one of them. Conversely, assume that φ is injective. Then the images of the $|A|$ elements of A must be pairwise different, since φ never identifies two different ones. So there are $|A|$ many images, and thus they exhaust all of A. □

Functions from $\mathbb{N}$ to a set A are often defined recursively: one chooses a specific element of A as the image of 0 and then defines $f(n + 1)$, for all $n \in \mathbb{N}$, as some function of $f(n)$, or even of $\{f(1), \dots, f(n)\}$. Suppose, for example, that we wish to enumerate the prime numbers. This amounts to defining a function $f : \mathbb{N} \longrightarrow \mathbb{Z}$ such that $f(n)$ is the nth prime. This can be done by setting $f(0) = 2$ and $f(n+1) = F(f(n))$, where F is a function from $\mathbb{N}$ to $\mathbb{Z}$ which assigns to $m \in \mathbb{N}$ the least prime number greater than m. The *recursion principle* of set theory states that this is a legitimate way to define a function with domain $\mathbb{N}$. The recursion principle is closely related to the induction principle, and it can be similarly justified on an intuitive level.

Exercise 0.24 Give an *intuitive* argument to support the recursion principle on the basis of the assumption that a function f has been defined if we can tell what $f(a)$ is for each a in its domain.

0.3 Mathematical Algorithms

One of the main goals of this book is to demonstrate how one can often construct a mathematical object—whose existence is known—in finitely many steps from other objects that it depends upon. A method to achieve such a construction under a particular set of circumstances is called an **algorithm**. A prerequisite for the problem to make sense is of course that the object itself and the data that it depends on are of a finite nature, so that they may be represented by symbols on paper or by means of datatypes on a computer. Section 4.6 will discuss this in more detail; for the moment, let us note that we may certainly represent integers on a computer, that exact integer arithmetic can be implemented, and that there are not in principle any limits on the size of the integers we can compute with. We may now consider the non-constructive existence proof for the quotient and remainder upon division of one integer by another and ask whether, given m, $n \in \mathbb{Z}$ with $n \neq 0$, we can compute q, $r \in \mathbb{Z}$ with

$$m = qn + r \quad \text{and} \quad 0 \leq r < |n|.$$

Everybody knows of course that this is true; we use the problem solely to demonstrate how questions of computability will be formally handled in this book.

We will present algorithms in a semi-formal "programming language." It uses the arrow "←" to assign the value on the right-hand side of the arrow to the variable on the left; other than that, the "language" is modeled after Modula-2. However, all commands are self-explanatory even to those with no background at all in computer programming. It is clear that every algorithm requires verification: one must prove that it *terminates* after finitely many steps for every input as specified, and that it performs its task *correctly*, i.e., ouputs an object that has the desired properties.

An important tool for proving the correctness of an algorithm is the concept of the *loop invariant.* This is a mathematical statement or a mathematical object that remains unaffected by the execution of the loop in question. Typically, it will be an equation that holds before the loop is entered, while it involves one or more variables whose values are changed by the actions of the loop. A mathematical argument will then be required to prove that the equation continues to hold after each run through the loop.

An algorithm provides a recipe for performing the construction in question for every input that meets the given specification. We will also encounter the situation where we give an algorithm for the construction of some object without having an a priori abstract existence proof. It is then important to realize that the assignments of the algorithms in this book can also be interpreted as mathematical constructions. Mathematically speaking, an algorithm together with the proof of its correctness and termination is in fact nothing but a certain special type of mathematical existence proof. It differs from arbitrary existence proofs insofar as it does not allow non-constructive arguments of the type, "if object x did not exist, then there would follow a contradiction." For such an algorithmic existence proof to be valid we do not have to require that the objects involved are of a finite nature, i.e., can be represented on a computer; this latter condition is required only if we wish to assert that the algorithm can be physically performed. We wish to emphasize that we do not subscribe to constructivism as a philosophical tenet. We have no qualms, for example, about proving the termination of an algorithm by an argument of the type, "if the algorithm did not terminate, then there would result a contradiction."

For the more mathematically inclined reader, it is an interesting exercise to translate algorithms into existence proofs as they are normally given in mathematics. Arguments involving loop invariants will then turn into proofs by "finite induction," where a property P of a certain natural number n is proved by showing "$P(0)$" and "$P(m)$ implies $P(m+1)$ for all $m < n$."

To illustrate all this, we will now give and verify an algorithm for the computation of the quotient and remainder in $\mathbb{Z}$. The algorithm uses successive subtraction of the divisor from the dividend until the range for the remainder is reached. It thus also demonstrates that it is not in general a priority of this book to discuss efficiency of algorithms: the division method that is taught in elementary school is vastly superior to the algorithm DIVINT of the proposition below. We will, however, at least to some extent, avoid things that would make a programmer cringe. If, for example, an algorithm uses the absolute value of an integer repeatedly, then one should determine this absolute value just once and assign it to a variable rather than determine it over and over again.

The algorithm DIVINT below uses a function sgn on $\mathbb{Z}$ which is defined by

$$\operatorname{sgn}(n) = \begin{cases} -1 & \text{if} \quad n < 0 \\ 0 & \text{if} \quad n = 0 \\ 1 & \text{otherwise.} \end{cases}$$

We thus have $n = \operatorname{sgn}(n) \cdot |n|$ for all $n \in \mathbb{Z}$.

Proposition 0.25 *The algorithm* DIVINT *of Table* 0.1 *computes, for given m, $n \in \mathbb{Z}$ with $n \neq 0$, integers q and r such that $m = qn + r$ and $0 \leq r < |n|$.*

Proof *Termination*: If the algorithm did not terminate, then the set of values assigned to REM would be a set of natural numbers: each of them would have to be greater than or equal to $|n|$ due to the fact that the **while**-clause would never fail. Moreover, this set would not have a least element because, as one easily concludes from the fact that $n \neq 0$, the value of REM decreases strictly with each execution of the **while**-loop. We would thus obtain a contradiction to Corollary 0.4.

Correctness: The equation $|m| = \text{QUOT} \cdot |n| + \text{REM}$ is a loop invariant: it is trivially true after initalization, and during each execution of the **while**-loop, QUOT is increased by 1, while $|n|$ is subtracted from REM. From this together with the **while**-clause, we see that after the last execution of the **while**-loop, we have

$$|m| = \text{QUOT} \cdot |n| + \text{REM} \quad \text{and} \quad 0 \leq \text{REM} < |n|. \qquad (*)$$

If $m > 0$, then we have $|m| = m$ and $|n| = \operatorname{sgn}(n) \cdot n$, so $(*)$ implies that

$$m = \operatorname{sgn}(n) \cdot \text{QUOT} \cdot n + \text{REM},$$

TABLE 0.1. Algorithm DIVINT

```
Specification: $(q, r) \leftarrow \mathrm{DIVINT}(m, n)$
               Computation of quotient and remainder in $\mathbb{Z}$
Given: $m, n \in \mathbb{Z}$ with $n \neq 0$
Find: $q, r \in \mathbb{Z}$ with $m = qn + r$ and $0 \leq r < |n|$
begin
$M \leftarrow m$;  $N \leftarrow n$
REM $\leftarrow |M|$;  NABS $\leftarrow |N|$;  QUOT $\leftarrow 0$
while REM $\geq$ NABS do
    REM $\leftarrow$ REM $-$ NABS
    QUOT $\leftarrow$ QUOT $+ 1$
end
if $M \geq 0$ then QUOT $\leftarrow \mathrm{sgn}(N) \cdot$ QUOT
elsif REM $\neq 0$ then
    QUOT $\leftarrow -\mathrm{sgn}(N) \cdot$ (QUOT $+ 1$)
    REM $\leftarrow$ NABS $-$ REM
else QUOT $\leftarrow -\mathrm{sgn}(N) \cdot$ QUOT
end
return((QUOT, REM))
end DIVINT
```

and we see that the output has all the required properties. If $m < 0$ and REM $\neq 0$, then $m = -|m|$, and so $(*)$ tells us that

$$\begin{aligned} m &= -\mathrm{QUOT} \cdot |n| - \mathrm{REM} \\ &= -\mathrm{QUOT} \cdot |n| - |n| + (|n| - \mathrm{REM}) \\ &= -(\mathrm{QUOT} + 1) \cdot |n| + (|n| - \mathrm{REM}) \\ &= -\mathrm{sgn}(n) \cdot (\mathrm{QUOT} + 1) \cdot n + (|n| - \mathrm{REM}). \end{aligned}$$

Moreover, it follows easily from $0 < \mathrm{REM} < |n|$ that

$$0 < |n| - \mathrm{REM} < |n|,$$

and we see that again, we obtain the correct output. Finally, suppose $m < 0$ and REM $= 0$. Then it follows immediately from $(*)$ that

$$m = -\mathrm{sgn}(n) \cdot \mathrm{QUOT} \cdot n$$

as desired. □

Notes

For the longest time in the history of mathematics, the existence of the basic number systems $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$ was taken as self-evident, $\mathbb{R}$ being represented by a geometric line of points, $\mathbb{Z}$ and $\mathbb{Q}$ being obtained from the natural numbers $\mathbb{N}$ by means of inverting addition and multiplication. It is noteworthy that there seems to have been much more of a reluctance to admit negative numbers than there was to talk about fractions. The Greek mathematicians of antiquity were already aware of the difference between $\mathbb{Q}$ and $\mathbb{R}$. They made a sharp distinction between "geometric" entities such as $\sqrt{2}$ and "arithmetic" entities such as $2/3$. The ancient Greeks also had the concept of approximating real numbers by rational numbers, as exemplified by the approximations of areas and volumes given by Archimedes and Apollonios in the 3rd century B.C. on the basis of earlier work by Eudoxos. Their ideas were a remarkable anticipation of the rigorous definition of the real numbers as the result of infinite or limiting operations on rational numbers. This definition did not come about until the late nineteenth century, largely due to the German mathematicians Karl Weierstrass and Richard Dedekind.

The system $\mathbb{C}$ of complex numbers was for a long time considered to be of a rather dubious, "imaginary" rather than real nature. This vagueness was eventually removed by the geometric interpretation of $\mathbb{C}$ as the real plane. The idea of identifying complex numbers with points in the plane appears in the work of the English mathematician John Wallis in the 17th century. It was made precise around 1800 by the Norwegian surveyor Caspar Wessel and, independently, the Swiss bookkeeper Jean Robert Argand.

It was the eminent 17th-century French mathematician Pierre de Fermat who first recognized the induction principle as a rigorous method of proving theorems on natural numbers. He actually used the version of Corollary 0.4: he would prove that every $n \in \mathbb{N}$ has property P by showing that the set of natural numbers that do not have property P does not have a least element and is thus empty. The version of Theorem 0.1 was employed for the first time by the French philosopher and scientist Blaise Pascal in his 1665 treatise on the arithmetic triangle that later came to be known as the Pascal triangle. The induction principle as an axiom in a rigorous formal setting was formulated by Richard Dedekind, and also by the Italian mathematician Guiseppe Peano, both in the 1880s. Interestingly, mathematicians well into the 18th century were often content with verifying a conjecture concerning natural numbers on a finite number of examples, a method which was referred to as "proof by induction." This is why the induction principle is also known as the principle of *complete* induction.

A common foundation for the number systems—and in fact for almost all of modern mathematics—was only found in the late 19th century, when the Russian-German mathematician Georg Cantor introduced the theory of *sets.* Set theory also provided the first rigorous definition of a map as a set of ordered pairs, in contrast to the more vague concept of a "rule" assigning certain objects to given ones. For an account of set theory as a foundation of mathematics, we recommend Hrbacek and Jech (1984) or Kunen (1983). A good reference for the history of the number systems is Kline (1985).

1

Commutative Rings with Unity

1.1 Why Abstract Algebra?

The main objects of study in this book are polynomials. Only the most elementary mathematical skills are required to manipulate polynomials. However, in order to develop the theory of Gröbner bases it is necessary to work within the larger framework of abstract algebra. The concept of abstract algebra arises from the observation that certain operations such as addition and multiplication can be performed on a variety of objects, such as numbers, polynomials, functions, or matrices, to name just a few. Certain properties of these operations are often shared by different objects on which these operations are performed. As an example, consider addition, multiplication, and division with remainder of integers which we discussed in the first section. We will see that these operations can be performed not only with integers, but with polynomials in one variable over the rationals or reals as well (in fact, over any *field*), and that the same basic properties such as commutativity or the associative law hold. Now if two objects share structure and a property of that structure, then they also share all consequences of this property, and for economical reasons, one would want to derive these consequences simultaneously in a generalized setting. Indeed, all the results on integers that we mentioned before have precise counterparts for polynomials, and it would be a tremendous waste of time to prove them over again. This is exactly what abstract algebra is all about: investigating properties of operations and their consequences while neglecting the actual nature of the objects that these operations are performed on.

Definition 1.1 Let A be a set. A **binary operation** on A is a map from $A \times A$ to A.

A binary operation on A can thus be visualized as a rule that assigns to each pair (a, b) of elements of A a new element c of A.

Examples 1.2 (i) Addition and multiplication on integers, rationals, reals and complex numbers.

(ii) Let X be a set, $\mathbb{Z}^X$, $\mathbb{Q}^X$, $\mathbb{R}^X$, and $\mathbb{C}^X$ the set of all functions from X to $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$, respectively. Then pointwise addition and multiplication are binary operations on each of these. Here, $(f,g) \longmapsto f+g$ and $(f,g) \longmapsto fg$, where $(f+g)(x) = f(x)+g(x)$ and $(fg)(x) = f(x)g(x)$ for all $x \in X$.

(iii) If I is an interval on the real line, let us denote by $\mathrm{C}(I,\mathbb{R})$ the set of all continuous functions from I to the reals. Since sum and product of continuous functions are again continuous, pointwise addition and multiplication as defined in (ii) above are binary operations on $\mathrm{C}(I,\mathbb{R})$.

(iv) If $S(X)$ is the set of all maps from a set X to itself, then composition of maps, where $(\varphi,\psi) \longmapsto \psi \circ \varphi$, is a binary operation on $S(X)$.

(v) If X is a set and $\mathcal{P}(X)$ the power set of X, i.e., the collection of all subsets of X, then union and intersection, where $(U,V) \longmapsto U \cup V$ and $(U,V) \longmapsto U \cap V$, respectively, are binary operations on $\mathcal{P}(X)$.

The next step is to introduce sets with operations that satisfy certain axioms, and to investigate consequences of these axioms.

1.2 Groups

Definition 1.3 A **group** is a set G with a binary operation $(a,b) \longmapsto a \cdot b$ and a distinguished element $e \in G$ such that the following axioms hold:

(i) "$\cdot$" is associative, i.e., $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ for all a, b, $c \in G$.

(ii) $e \cdot a = a$ for all $a \in G$.

(iii) For all $a \in G$, there exists $b \in G$ with $b \cdot a = e$.

G is called an **Abelian group** if, in addition to (i)–(iii), "$\cdot$" is commutative, i.e., $a \cdot b = b \cdot a$ for all a, $b \in G$.

We will also write ab instead of $a \cdot b$, and, in view of (i) above, abc instead of $a(bc)$. Following the examples below, it will be shown that the distinguished element e of a group G is the only one satisfying (ii) above, and that it is also the only one satisfying $a \cdot e = a$ for all $a \in G$. Moreover, we will show that for each $a \in G$, the element $b \in G$ of (iii) above is uniquely determined by a, and that it is also the only one satisfying $a \cdot b = e$. In view of this, the distinguished element e is called the **neutral element** of G, and for $a \in G$, the element $b \in G$ of (iii) is called the **inverse** of a and is denoted by a^{-1}. The following examples are easy to verify.

Examples 1.4 (i) The integers $\mathbb{Z}$ with the operation $+$ form an Abelian group with neutral element 0 and $-a$ as the inverse of $a \in \mathbb{Z}$.

(ii) $\mathbb{Q}\setminus\{0\}$, i.e., the rationals without 0, is an Abelian group under multiplication with neutral element 1 and $1/a$ as the inverse of $a \in \mathbb{Q}$, $a \neq 0$.

(iii) If X is a set, then $\mathbb{Z}^X$, $\mathbb{Q}^X$, $\mathbb{R}^X$, $\mathbb{C}^X$, and $\mathrm{C}(\boldsymbol{I},\mathbb{R})$ with pointwise addition as defined in Example 1.2 (ii) and (iii) are Abelian groups. Here, the neutral element is the zero function $\mathbf{0}$ (where $\mathbf{0}(x) = 0$ for all x), and the inverse of a function f is its negative $-f$ (where $(-f)(x) = -f(x)$ for all x).

(iv) Let X be a non-empty set, $S(X)$ the set of all bijective maps from X to itself. Composition of maps is associative by Exercise 0.18, and id_X satisfies $\varphi \circ \mathrm{id}_X = \mathrm{id}_X \circ \varphi = \varphi$ for all $\varphi \in S(X)$. Moreover, for all $\varphi \in S(X)$, there exists an inverse map $\psi \in S(X)$ with $\psi \circ \varphi = \mathrm{id}_X$. Hence $S(X)$ with composition of maps and neutral element id_X is a group.

When dealing with Abelian groups, it is common to write the operation as $+$. It is then understood that the neutral element is denoted by 0 and the inverse of a by $-a$ in obvious reference to the standard example of the integers $\mathbb{Z}$.

Exercises 1.5 (i) Show that $S(X)$ as defined in (iv) above is not in general an Abelian group. (Hint: Take $X = \{1,2,3\}$, and find $\varphi, \psi \in S(X)$ with $\varphi \circ \psi \neq \psi \circ \varphi$.)

(ii) Let X be a set, G its power set $\mathcal{P}(X)$. For $A \in G$, denote by $\overline{A}$ the complement of A in X. Define a binary operation $\triangle$ on G by setting

$$A \triangle B = (A \cap \overline{B}) \cup (B \cap \overline{A})$$

for A, $B \in G$. ($A \triangle B$ is often called the **symmetric difference** of A and B, or, more obviously, the **union without the intersection** of A and B.) Show that G with the operation $\triangle$ is an Abelian group.

The next four lemmas provide the justification for the terminology "*the* neutral element" and "*the* inverse," and thus for the notation a^{-1} for the inverse of a. Throughout, let G be a group with distinguished element e.

Lemma 1.6 Let a, $b \in G$. If $ba = e$, then also $ab = e$.

Proof Let $c \in G$ with $cb = e$. From $ba = e$ we get

$$b = eb = (ba)b = b(ab).$$

Multiplying this equation by c from the left, we obtain $cb = (cb)(ab)$. If we replace (cb) by e, this becomes $e = e(ab) = ab$. □

Lemma 1.7 The element e also satisfies $ae = a$ for all $a \in G$.

Proof Let $a \in G$, and let $b \in G$ with $ba = e$. Then $ab = e$ by Lemma 1.6, and we get $a = ea = (ab)a = a(ba) = ae$. □

Lemma 1.8 For each $a \in G$, there is exactly one $b \in G$ with $ba = e$, and this is also the only element satisfying $ab = e$.

Proof Let a, $b \in G$ with $ba = e$. We already know that then $ab = e$ too. Now suppose that $c \in G$ with $ca = e$. Multiplying the equation $ba = ca$ by b from the right yields $be = ce$, hence $b = c$ by Lemma 1.7. If $ac = e$, then we can prove $b = c$ in a similar way by multiplying the equation $ab = ac$ by b from the left. □

Lemma 1.9 The distinguished element $e \in G$ is the only one satisfying $ea = a$ for all $a \in G$, and it is also the only one satisfying $ae = a$ for all $a \in G$.

Proof Suppose $e' \in G$ satisfies $e'a = a$ for all $a \in G$. Then in particular, $e'e = e$. Moreover, $e'e = e'$ by Lemma 1.7. We see that $e = e'$. If $e' \in G$ satisfies $ae' = a$ for all $a \in G$, then $ee' = e$, and this together with $ee' = e'$ implies $e = e'$. □

Exercise 1.10 Let G be a group with neutral element e. Show the following:

(i) $(a^{-1})^{-1} = a$ for all $a \in G$.

(ii) If a, b, $c \in G$ with $ab = ac$ or $ba = ca$, then $b = c$.

(iii) If $a \in G$ with $ab = b$ or $ba = b$ for some $b \in G$, then $a = e$.

One of the most important concepts in group theory is that of a subgroup.

Definition 1.11 Let G be a group and H a subset of G with $H \neq \emptyset$. H is called a **subgroup** of G if the following hold:

(i) a, $b \in H$ implies that $ab \in H$ for all a, $b \in G$.

(ii) For all $a \in G$, $a \in H$ implies that $a^{-1} \in H$.

The following exercise provides examples.

Exercises 1.12 (i) Let G be a group with neutral element e. Show that G and $\{e\}$ are subgroups of G.

(ii) Let $m \in \mathbb{Z}$, and denote by $m\mathbb{Z}$ the set $\{\, mk \mid k \in \mathbb{Z}\,\}$, i.e., $m\mathbb{Z}$ is the set of all integer multiples of m. Show that $m\mathbb{Z}$ is a subgroup of the additive group $\mathbb{Z}$.

Proposition 1.13 *Let G be a group with neutral element e, $\emptyset \neq H \subseteq G$. Then the following are equivalent:*

(i) *H is a subgroup of G.*

(ii) *H is closed under the group operation of G, $e \in H$ and H is again a group with neutral element e under this operation.*

(iii) *$a, b \in H$ implies $ab^{-1} \in H$ for all $a, b \in G$.*

Proof (i)$\Longrightarrow$(ii). The claim that H is closed under the group operation of G is simply a reformulation of condition 1.11 (i). It remains to verify Definition 1.3 (i)–(iii). The associative law 1.3 (i) holds for all $a, b, c \in G$, so in particular, it holds for all $a, b, c \in H$. For 1.3 (ii), pick any element $a \in H$. Then $a^{-1} \in H$ by 1.11 (ii), and thus $e = aa^{-1} \in H$ by 1.11 (i). Since e is a neutral element of G, it certainly satisfies $ea = a$ for all $a \in H$. Property 1.3 (iii) now follows immediately from 1.11 (ii).

(ii)$\Longrightarrow$(iii). We first show that the neutral element e' of the group H necessarily equals the neutral element e of the group G. Indeed, e' satisfies the equation $e' \cdot e' = e'$ in H, and viewing this as an equation in the group G, we may apply Exercise 1.10 (iii) to conclude that $e' = e$. Next, we claim that for each $a \in H$ the inverse of a in the group H equals its inverse in the group G. If we denote these inverses by b and c, respectively, then we have $e = ab = ac$ and hence $b = c$ by cancelation. Now let $a, b \in H$. Then $b^{-1} \in H$, and thus $ab^{-1} \in H$.

(iii)$\Longrightarrow$(i). Since H is not empty, we can pick $a \in H$ and conclude that $e = a \cdot a^{-1} \in H$. From this, we immediately get condition 1.11 (ii): if $a \in H$, then $a^{-1} = ea^{-1} \in H$. For 1.11 (i), let $a, b \in H$. Then $b^{-1} \in H$ and hence $ab = a(b^{-1})^{-1} \in H$. □

1.3 Rings

We have not even scratched the surface of the vast theory of groups with its countless applications, but our main interest in view of polynomials and Gröbner bases are rings.

Definition 1.14 A **ring** is a set R with two binary operations "$+$" and "$\cdot$," referred to as addition and multiplication, as well as a distinguished element 0 such that the following hold:

(i) R is an Abelian group w.r.t. addition with neutral element 0.

(ii) Multiplication is associative, i.e., $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all $a, b, c \in R$.

(iii) The distributive laws $a \cdot (b + c) = a \cdot b + a \cdot c$ and $(a + b) \cdot c = a \cdot c + b \cdot c$ hold for all $a, b, c \in R$.

If, in addition, multiplication is commutative too, i.e., $a \cdot b = b \cdot a$ for all $a, b \in R$, then R is called a **commutative ring.** R is called a **ring with 1,** or **ring with unity** if it contains a distinguished element 1 with $1 \neq 0$ and $1 \cdot a = a$ for all $a \in R$.

Again, we will write ab instead of $a \cdot b$ and abc instead of $a(bc)$. As usual, the inverse of $a \in R$ w.r.t. addition will be denoted by $-a$. Also, $a + (-b)$ will be written as $a - b$, and we will refer to this as subtracting b from a.

From now on, "ring" will mean "commutative ring with 1."

Verification of the following examples is left to the reader.

Examples 1.15 (i) The integers, rationals, reals, and complex numbers are rings with their natural addition and multiplication.

(ii) $\mathbb{Z}^X$, $\mathbb{Q}^X$, $\mathbb{R}^X$, and $\mathbb{C}^X$ with pointwise addition and multiplication as defined in Example 1.2 (ii) are rings with the constant functions $\mathbf{0}$ and $\mathbf{1}$ as zero element and unity, respectively.

(iii) $\mathrm{C}(\boldsymbol{I}, \mathbb{R})$ with pointwise addition and multiplication as defined in Example 1.2 (iii) is a ring.

(iv) Let p be a prime number, $\mathbb{Z}_p$ the set of all rational numbers whose denominator is not divisible by p after they have been reduced to lowest terms. It is not hard to see that $\mathbb{Z}_p$ is closed under addition and multiplication of rational numbers. Hence these are binary operations on $\mathbb{Z}_p$. Moreover, $0, 1 \in \mathbb{Z}_p$, and we see that $\mathbb{Z}_p$ is a ring.

Other important examples that will be introduced later are polynomial rings and residue class rings.

Definition 1.16 Let R be a ring, $a \in R$. Then a is called

(i) a **zero divisor** if $a \neq 0$ and there exists $0 \neq b \in R$ with $ab = 0$,

(ii) **invertible**, or a **unit**, if there exists $c \in R$ with $ac = 1$.

R is called an **integral domain**, or just **domain**, if it contains no zero divisors. R is called a **field** if every element of R other than 0 is invertible.

The following examples are easily verified.

Examples 1.17 (i) The integers $\mathbb{Z}$ form an integral domain whose only units are 1 and -1.

(ii) The rationals, reals, and complex numbers are fields.

(iii) For any interval $\boldsymbol{I}$ on the real line that does not consist of just one point, $\mathrm{C}(\boldsymbol{I}, \mathbb{R})$ is not an integral domain: one can easily construct continuous functions f and g on $\boldsymbol{I}$ such that f, $g \neq 0$, but for each $x \in \boldsymbol{I}$, either $f(x) = 0$ or $g(x) = 0$, so that $f \cdot g = \mathbf{0}$. A function $f \in \mathrm{C}(\boldsymbol{I}, \mathbb{R})$ is a unit iff $f(x) \neq 0$ for all $x \in \boldsymbol{I}$, since then $1/f$ is defined and continuous.

(iv) $\mathbb{Z}_p$ is an integral domain for all prime numbers p. A fraction $s/t \in \mathbb{Z}_p$ (reduced to lowest terms) is a unit iff p does not divide s, since then $t/s \in \mathbb{Z}_p$.

Note that in an integral domain, $ab = 0$ implies $a = 0$ or $b = 0$. We will now prove some elementary properties of ring elements, zero divisors, and units.

Lemma 1.18 Let R be a ring. Then the following hold:

(i) $a \cdot 0 = 0$ for all $a \in R$. In particular, 0 is not invertible.

(ii) If a, $b \in R$ with $ab \neq 0$, then ab is a zero divisor iff a or b is a zero divisor.

(iii) If a, $b \in R$, then ab is a unit iff both a and b are units.

(iv) Zero divisors are never invertible.

(v) The set U_R of all units of R is an Abelian group under ring multiplication. In particular, the multiplicative inverse of a unit $a \in R$ is uniquely determined (Lemma 1.8), and it will be denoted by a^{-1}.

Proof (i) Let $a \in R$. Then $a + a \cdot 0 = a \cdot 1 + a \cdot 0 = a(1 + 0) = a \cdot 1 = a$. Subtracting a on both sides of the equation yields $a \cdot 0 = 0$.

(ii) If ab is a zero divisor, then $abc = 0$ for some $0 \neq c \in R$. Furthermore, a, $b \neq 0$ by (i) since $ab \neq 0$. Now if $bc = 0$, then b is a zero divisor. If $bc \neq 0$, then $a(bc) = 0$ shows that a is a zero divisor. Conversely, assume that b is a zero divisor. Then $bc = 0$ for some $0 \neq c \in R$, so $(ab)c = a(bc) = a \cdot 0 = 0$ which shows that ab is a zero divisor. The case that a is a zero divisor can be handled in a similar way.

(iii) If ab is a unit, then $abc = 1$ for some $c \in R$. Now $a(bc) = b(ac) = 1$ shows that both a and b are units. Conversely, if a and b are units, then $ac = bd = 1$ for some c, $d \in R$, and hence $(ab)(cd) = (ac)(bd) = 1 \cdot 1 = 1$, which shows that ab is a unit.

(iv) Let $a \in R$ be a zero divisor, $0 \neq c \in R$ such that $ac = 0$. If a were a unit, there would have to exist $b \in R$ with $ab = 1$. We would obtain $0 = b \cdot 0 = b(ac) = (ab)c = 1 \cdot c = c$, a contradiction.

(v) If a, $b \in U_R$, then $ab \in U_R$ by (iii) above, hence ring multiplication is a binary operation on U_R. The associative and commutative law of multiplication hold in U_R because they hold in all of R. 1 is obviously in U_R and a neutral element w.r.t. multiplication. Finally, to prove the existence of inverses, let $a \in U_R$. Then there exists $b \in R$ with $ab = 1$. But this equation shows that b lies in U_R too. □

If K is a field, then the inverse of $0 \neq a \in K$ is often denoted by $1/a$, and $a(1/b)$ is written as a/b.

Lemma 1.19 (i) Let R be a domain. Then the following cancelation rule holds in R: if $ac = bc$ and $c \neq 0$, then $a = b$.

(ii) Every field is a domain.

(iii) Every finite domain is a field.

Proof (i) If $ac = bc$, then $(a - b)c = 0$. But $c \neq 0$ by assumption, and R has no zero divisors, so $a - b = 0$.

(ii) Let K be a field. Any zero divisor of K would by Lemma 1.18 (iv) be a non-zero, non-invertible element of K which is impossible in a field.

(iii) Let R be a finite domain, $0 \neq a \in R$. We have to find $b \in R$ with $ab = 1$. Consider the map $\varphi : R \longrightarrow R$ given by $\varphi(c) = ac$ for all $c \in R$. We claim that φ is injective: if $\varphi(c) = \varphi(c')$, then $ac = ac'$ and thus $c = c'$ by (i) above. Since an injective map from a finite set to itself is surjective by Proposition 0.23, we can find a preimage of 1 under φ, i.e., an element $b \in R$ with $ab = 1$. □

1.4 Subrings and Homomorphisms

As with most other algebraic theories, the concepts of *substructure* and *homomorphism* will play an important role in ring theory.

Definition 1.20 Let R be a ring, and $S \subseteq R$ such that

(i) $1 \in S$,

(ii) $a - b \in S$ for all a, $b \in S$, and

(iii) $ab \in S$ for all a, $b \in S$.

Then S is called a **subring** of R.

Following are some obvious examples.

Examples 1.21 (i) $\mathbb{Z}$ is a subring of $\mathbb{Z}_p$, and $\mathbb{Z}_p$ is a subring of $\mathbb{Q}$ for any prime p.

(ii) $C(\boldsymbol{I}, \mathbb{R})$ is a subring of $\mathbb{R}^{\boldsymbol{I}}$.

Lemma 1.22 Let R be a ring with unity 1, and let S be a subset of R with $1 \in S$. Then the following are equivalent:

(i) S is a subring of R.

(ii) The addition and multiplication of R, when restricted to elements of S, are binary operations on S, and S with these operations is a ring with the same zero element and unity as R.

Proof (i)$\Longrightarrow$(ii): Condition (ii) of the subring definition together with Proposition 1.13 implies that S is an additive subgroup of R with the same zero element as R. Condition (iii) of the definition tells us that multiplication too is a binary operation on R, and inspection of the remaining ring axioms shows that they trivially hold in S because they hold in all of R.

(ii)$\Longrightarrow$(i): Proposition 1.13 says that condition (ii) of the subring definition holds. Condition (iii) is simply a reformulation of the fact that multiplication is a binary operation on S, and condition (i) is immediate from the fact that S is a ring with the same unity as R. □

Note that by the above proposition, we have $0 \in S$, and $a \in S$ implies $-a \in S$ whenever S is a subring of R.

Exercise 1.23 Let R be a ring, S a subring of R. Give direct proofs of the facts that $0 \in S$, and $a \in S$ implies $-a \in S$.

If S is a subring of the ring R and S is actually a field, then it is called a **subfield** of R. Clearly, $\mathbb{Q}$ is a subfield of $\mathbb{R}$ which in turn is a subfield of $\mathbb{C}$. An example of a subfield of a ring where the latter is not a field is given by the set of all constant functions in $\mathrm{C}(\boldsymbol{I}, \mathbb{R})$.

Exercise 1.24 Let $D = \{\, a + bi\sqrt{5} \mid a, b \in \mathbb{Z}\,\}$, where $i^2 = -1$. Show that D is a subring of $\mathbb{C}$.

Exercise 1.25 Let R be a ring, $\{S_i\}_{i \in I}$ a family of subrings of R. Show that $\bigcap_{i \in I} S_i$ is again a subring of R.

Exercise 1.26 Let S be a subfield of R and $0 \neq a \in S$. Show that a is a unit of R whose inverse is the same as that in the field S.

Generally speaking, homomorphisms between algebraic structures are maps that preserve the operations, i.e., whenever an equation such as $ab = c$ holds in the domain of the map, then the images of a, b, and c must satisfy the same equation in the range.

Definition 1.27 Let R and S be rings and $\varphi : R \longrightarrow S$ a map. Then φ is called a **homomorphism of rings** if the following hold:

(i) $\varphi(a + b) = \varphi(a) + \varphi(b)$ for all a, $b \in R$.

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

(iii) $\varphi(1_R) = 1_S$.

Here, 1_R and 1_S denote the unities of R and S, respectively. We will often drop this distinction and just write "1" and "0" even when more than one ring is involved. A homomorphism φ is called an **embedding** if φ is injective, and an **isomorphism** if φ is bijective. A homomorphism from a ring R to itself is called an **endomorphism**, and an isomorphism from R to itself is called an **automorphism**.

The following easy exercises provide examples.

Exercises 1.28 Show the following:

(i) For any ring R, the identity map id_R is an automorphism of R.

(ii) If S is a subring of R, then the map $\varphi : S \longrightarrow R$ given by $\varphi(a) = a$ for all $a \in S$ is an embedding.

(iii) If $R = \mathrm{C}(\boldsymbol{I}, \mathbb{R})$ and $x_0 \in \boldsymbol{I}$, then $\varepsilon_{x_0} : R \longrightarrow \mathbb{R}$ given by $\varepsilon_{x_0}(f) = f(x_0)$ is a surjective homomorphism of rings. (Here, ε_{x_0} is the *evaluation* map: the image of f under ε_{x_0} is its value at x_0).

Lemma 1.29 Let $\varphi : R \longrightarrow S$ be a homomorphism of rings. Then $\varphi(0) = 0$ and $\varphi(-a) = -\varphi(a)$ for all $a \in R$.

Proof $\varphi(0) = \varphi(0 + 0) = \varphi(0) + \varphi(0)$. Subtracting $\varphi(0)$ on both sides yields $0 = \varphi(0)$. Furthermore,

$$0 = \varphi(0) = \varphi(a + (-a)) = \varphi(a) + \varphi(-a)$$

for all $a \in R$. Subtraction of $\varphi(a)$ on both sides yields $-\varphi(a) = \varphi(-a)$. □

The proof of the following simple facts is left to the reader.

Lemma 1.30 (i) If $\varphi : R \longrightarrow S$ and $\psi : S \longrightarrow T$ are homomorphisms (embeddings, isomorphisms) of rings, then $\psi \circ \varphi : R \longrightarrow T$ is a homomorphism (embedding, isomorphism) of rings.

(ii) If $\varphi : R \longrightarrow S$ is an isomorphism, then so is $\varphi^{-1} : S \longrightarrow R$.

(iii) If $\varphi : R \longrightarrow S$ is a homomorphism of rings, then the image $\varphi(R)$ of φ is a subring of S. □

Two rings R and S are called **isomorphic** if there exists an isomorphism from R to S, and this is denoted by $R \simeq S$. Statements (i) and (ii) of the lemma above say that $R \simeq S$ implies $S \simeq R$, and $R \simeq S$ together with $S \simeq T$ implies $R \simeq T$. Moreover, we saw in Exercise 1.28 (i) that $R \simeq R$ holds for any ring R. If $\varphi : R \longrightarrow S$ is a surjective homomorphism, then S is called a **homomorphic image** of R. Statement (iii) above thus tells us that $\varphi(R)$ is a homomorphic image of R for arbitrary homomorphism $\varphi : R \longrightarrow S$. The following definition and lemma provide a test for injectivity of a homomorphism.

Definition 1.31 Let $\varphi : R \longrightarrow S$ be a homomorphism of rings. We define the **kernel** of φ by setting

$$\ker(\varphi) = \{a \in R \mid \varphi(a) = 0\}.$$

Note that by Lemma 1.29, $0 \in \ker(\varphi)$ for every homomorphism φ.

Lemma 1.32 Let φ be as above. Then φ is injective iff $\ker(\varphi) = \{0\}$.

Proof Suppose φ is injective, and let $a \in \ker(\varphi)$. Then $0 = \varphi(a) = \varphi(0)$, hence $a = 0$ by injectivity. Conversely, if $\ker(\varphi) = \{0\}$ and a, $b \in R$ with $\varphi(a) = \varphi(b)$, then $0 = \varphi(a) - \varphi(b) = \varphi(a-b)$, which means $(a-b) \in \ker(\varphi)$. It follows that $a - b = 0$, and thus $a = b$. □

In a sense, the kernel of φ is a measure of how far φ is from being injective. The extreme case is φ injective with $\ker(\varphi) = \{0\}$. Then R and $\varphi(R)$ are isomorphic because φ, when viewed as a map from R to $\varphi(R)$, is injective and surjective, hence an isomorphism. We see that here, the structure of $\varphi(R)$ can be described completely just from knowing what $\ker(\varphi)$ was. We are now going to show that this is always the case: the kernel of φ determines the structure of $\varphi(R)$. We will have to build up quite a machinery until we can state the result in Corollary 1.56.

1.5 Ideals and Residue Class Rings

Definition 1.33 Let R be a ring and $\emptyset \neq I \subseteq R$. Then I is called an **ideal** of R if

(i) $a + b \in I$ for all a, $b \in I$, and

(ii) $ar \in I$ for all $a \in I$ and $r \in R$.

I is called **trivial** if $I = \{0\}$, **proper** if $I \neq R$.

Condition (ii) above is sometimes expressed by saying that I is closed under *inside-outside multiplication.* Examples of ideals are provided by the following lemma and exercise.

Lemma 1.34 Let $\varphi : R \longrightarrow S$ be a homomorphism of rings. Then $\ker(\varphi)$ is a proper ideal of R.

Proof $\ker(\varphi) \neq \emptyset$ since $0 \in \ker(\varphi)$. If a, $b \in \ker(\varphi)$, then $\varphi(a) = \varphi(b) = 0$, hence $\varphi(a + b) = \varphi(a) + \varphi(b) = 0 + 0 = 0$, and thus $a + b \in \ker(\varphi)$. If $a \in \ker(\varphi)$ and $r \in R$, then $\varphi(a) = 0$, hence

$$\varphi(ar) = \varphi(a) \cdot \varphi(r) = 0 \cdot \varphi(r) = 0$$

and thus $ar \in \ker(\varphi)$. The ideal $\ker(\varphi)$ is proper since $\varphi(1_R) = 1_S \neq 0$ and thus $1_R \notin \ker(\varphi)$. □

Exercises 1.35 Show the following:

(i) $\{0\}$ and R are ideals for any ring R.

(ii) Let R be a ring, $a \in R$, and denote by aR the set $\{ar \mid r \in R\}$ of all multiples of a. Then aR is an ideal of R. In particular, the set $m\mathbb{Z} = \{mk \mid k \in \mathbb{Z}\}$ of all integer multiples of $m \in \mathbb{Z}$ is an ideal of the ring $\mathbb{Z}$.

(iii) More generally, let R be a ring, $a_1, \ldots, a_n \in R$. Denote by $\sum_{i=1}^n a_iR$ the set

$$\left\{ \sum_{i=1}^n a_ir_i \;\middle|\; r_i \in R \text{ for } 1 \le i \le n \right\}$$

of all sums of multiples of the a_i. Then $\sum_{i=1}^n a_iR$ is an ideal of R.

(iv) Generalizing even further, let R be a ring, $A \subseteq R$. Then the set

$$\left\{ \sum_{i=1}^n a_ir_i \;\middle|\; 0 < n \in \mathbb{N},\; r_i \in R, \text{ and } a_i \in A \text{ for } 1 \le i \le n \right\}$$

of all "linear combinations" of elements of A is an ideal of R.

For examples (ii), (iii), and (iv) above, there is some standard terminology and notation:

Definition 1.36 Let R be a ring, $a \in R$. The ideal aR described in (ii) above is called the **principal ideal generated by** a, and it is also denoted by $\mathrm{Id}(a)$. If $a_1, \ldots, a_n \in R$, then the ideal $\sum_{i=1}^n a_iR$ described in (iii) above is called the **ideal generated by** $a_1, \ldots, a_n$. An ideal of this form is called **finitely generated**, and it will also be denoted by $\mathrm{Id}(a_1, \ldots, a_n)$. The ideal of (iv) above is called the **ideal generated by** A and will be denoted by $\mathrm{Id}(A)$. In this case, A is also called an **ideal basis** of $\mathrm{Id}(A)$. Here, we use the convention that the empty sum equals 0, so that the empty set generates the zero ideal.

We will also allow the "mixed notation" $\mathrm{Id}(A, a)$ instead of $\mathrm{Id}(A \cup \{a\})$.

Exercise 1.37 Let R be a ring, I an ideal of R. Show that if we regard R as just a group under addition, then I is a subgroup of R.

Note that you just showed that $0 \in I$ for any ideal I of R, and $a \in I$ implies $-a \in I$ for all $a \in R$. We will see that a proper ideal I of a ring R is never a subring of R since $1 \notin I$.

Exercise 1.38 Let R be a ring, I an ideal of R, $a \in R$. Show that $a \in I$ iff $aR \subseteq I$.

The following lemma is used frequently when working with ideals.

Lemma 1.39 Let R be a ring, I an ideal of R. Then the following are equivalent:

(i) I is proper.

(ii) $1 \notin I$.

(iii) $u \notin I$ for all units u of R.

Proof (i)$\Longrightarrow$(ii). Assume that I is proper. If 1 were in I, then $a = 1 \cdot a$ would have to be in I for all $a \in R$ by inside-outside multiplication, a contradiction.

(ii)$\Longrightarrow$(iii). Assume that $1 \notin I$. If u were in I for some unit u of R, then $1 = uu^{-1}$ would have to be in I too.

(iii)$\Longrightarrow$(i). If I contains no unit, then, in particular, $1 \notin I$, and thus $I \neq R$. □

Exercise 1.40 Show that a ring R is a field iff $\{0\}$ and R are the only ideals of R.

Exercise 1.41 Let I_1 and I_2 be ideals of the ring R. Show the following:

(i) The intersection $I_1 \cap I_2$ of I_1 and I_2 is again an ideal of R.

(ii) If we define the sum of I_1 and I_2 by setting

$$I_1 + I_2 = \{\, a_1 + a_2 \mid a_1 \in I_1,\ a_2 \in I_2 \,\},$$

then $I_1 + I_2$ is an ideal of R with $I_i \subseteq I_1 + I_2$ for $i = 1, 2$.

We saw in Lemma 1.34 above that kernels of homomorphisms are always proper ideals. But kernels are actually more than just another class of examples of ideals: we are going to show that conversely, every proper ideal of a ring R is in fact the kernel of some homomorphism φ from R to some ring S. Given I and R, we will now construct φ and S. The idea is the following: given a, $b \in R$, we must have $\varphi(a) = \varphi(b)$ iff $\varphi(a - b) = 0$ iff $(a - b) \in I$, since φ is to be a homomorphism with kernel I. We will achieve this by "lumping together" the elements of R in such a way that a, $b \in R$ belong to the same "lump" iff $(a - b) \in I$. The "lumps" will then be taken as the elements of S, and $\varphi(a)$ will be defined as the "lump" that a itself belongs to. That way, we will have $\varphi(a) = \varphi(b)$ iff $(a - b) \in I$ as desired. For the rest of this section, let R be a ring and I an ideal of R.

Definition 1.42 For each $a \in R$, we define the **residue class of a modulo I** to be the set $a + I = \{\, a + s \mid s \in I \,\}$. The set $\{\, a + I \mid a \in R \,\}$ of all residue classes will be denoted by R/I. Residue classes are sometimes also called **cosets**, and a is called a **representative** of $a + I$.

When it is obvious from the context what ideal I is being referred to, the residue class $a + I$ of a is often denoted by $\bar{a}$ or $[a]$. Note that I itself is a coset—namely, that of 0—and that $a \in a + I$ since $0 \in I$.

Example 1.43 Consider $6\mathbb{Z}$, the principal ideal generated by 6 in the ring of integers. $6\mathbb{Z}$ consists of all integer multiples of 6. Listing the integers in the following way will help to visualize the residue classes of $\mathbb{Z}$ modulo $6\mathbb{Z}$.

$$\begin{array}{rrrrrr}
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
-12 & -11 & -10 & -9 & -8 & -7 \\
-6 & -5 & -4 & -3 & -2 & -1 \\
0 & 1 & 2 & 3 & 4 & 5 \\
6 & 7 & 8 & 9 & 10 & 11 \\
12 & 13 & 14 & 15 & 16 & 17 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots
\end{array}$$

We see that the left-hand column equals $6\mathbb{Z}$. Moving over to the next column to the right amounts to adding 1 everywhere, so the second through sixth column are the residue classes $1+6\mathbb{Z}$ through $5+6\mathbb{Z}$. They are pairwise disjoint and exhaust all of $\mathbb{Z}$. It is customary to use the six representatives $0, 1, \ldots, 5$ when working in $\mathbb{Z}/6\mathbb{Z}$, but it is obvious that as a representative of a residue class, we could pick any one of its members (e.g., $5 + 6\mathbb{Z} = (-7) + 6\mathbb{Z}$). Residue classes $m_1 + 6\mathbb{Z}$ and $m_2 + 6\mathbb{Z}$ are actually one and the same iff $m_1 - m_2$ is a multiple of 6, i.e., is in $6\mathbb{Z}$. A similar picture can be drawn and the analogous statements hold if we replace 6 by any $m \in \mathbb{Z}$, $m > 0$. In each case, we would find that $\mathbb{Z}/m\mathbb{Z}$ consists of just m residue classes $0 + m\mathbb{Z}$, $1 + m\mathbb{Z}$, $\ldots$, $(m-1) + m\mathbb{Z}$, each of which has infinitely many members. Ideals $m\mathbb{Z}$ with $m < 0$ need not be considered since $m\mathbb{Z} = (-m)\mathbb{Z}$ for all $m \in \mathbb{Z}$. For the zero ideal $\{0\} = 0\mathbb{Z}$, each residue class $m + \{0\}$ consists of just m, and we see that here, we get infinitely many residue classes with one member each.

The statements of the above example will be proved rigorously and more generally in Lemma 1.44, Example 1.45, and Lemma 1.47. The reader is encouraged to go back to this example for an illustration of those proofs. The next lemma shows that residue classes are precisely the "lumps" of elements of R that we were looking for.

Lemma 1.44 For a, $b \in R$, we have $a + I = b + I$ iff $a - b \in I$.

Proof "$\Longrightarrow$": Assume that $a + I = b + I$. Since $a \in a + I$, it follows that $a \in b + I$. Hence there is $s \in I$ with $a = b + s$, which means that $a - b = s \in I$.

"$\Longleftarrow$": Assume that $a - b \in I$. We want to show that $a + I \subseteq b + I$ and $b + I \subseteq a + I$. For the first inclusion, let $c \in a + I$. Then $c = a + s$ for some $s \in I$, and we can write

$$c = b + (a - b + s). \tag{$*$}$$

Now $(a-b+s) \in I$ since $a-b \in I$ and $s \in I$. Hence the equation $(*)$ shows that $c \in b+I$. The reverse inclusion $b+I \subseteq a+I$ can be proved similarly. □

We can now say a little more about Example 1.43.

Example 1.45 Let $0 < m \in \mathbb{Z}$. Then $\mathbb{Z}/m\mathbb{Z}$ consists of the m elements $m\mathbb{Z}$, $1+m\mathbb{Z}$, ..., $(m-1)+m\mathbb{Z}$. An arbitrary $n \in \mathbb{Z}$ belongs to $r+m\mathbb{Z}$ where r is the remainder of n upon division by m. Both claims follow from Lemma 1.44: if $n \in \mathbb{Z}$ and $n = mq + r$, then $n - r = mq \in m\mathbb{Z}$, and hence $n \in n + \mathbb{Z} = r + \mathbb{Z}$. Since the remainder r can be found such that $0 \leq r \leq m-1$, we see that every residue class equals one of the sets $r+m\mathbb{Z}$, where $0 \leq r \leq m-1$.

Exercise 1.46 List the elements of $\mathbb{Z}/5\mathbb{Z}$ in a way similar to Example 1.43. Which of these residue classes do the following elements of $\mathbb{Z}$ belong to: -2, 0, 3, 9, 21, 329534, -329534?

In Lemma 1.44, we gave a criterion for two residue classes $a+I$ and $b+I$ to be equal. Now we ask a slightly different question: given $a+I$, for what b does $b+I$ equal $a+I$, i.e., what are the possible representatives of the coset $a+I$? The answer is that the *possible representatives of a residue class are precisely its own members.*

Lemma 1.47 Let a, $b \in R$. Then $b+I = a+I$ iff $b \in a+I$.

Proof By Lemma 1.44 it suffices to show that $b \in a+I$ iff $b-a \in I$. If $b \in a+I$, then $b = a+s$ for some $s \in I$, so $b-a = s \in I$. Conversely, if $b-a \in I$, then the equation $b = a+(b-a)$ shows that $b \in a+I$. □

The reader is advised to memorize Lemmas 1.44 and 1.47 carefully. These will be used constantly when working with residue classes.

We have already mentioned that each $a \in R$ lies in at least one residue class modulo I, namely, $a+I$. The next lemma shows that each $a \in R$ actually lies in *exactly* one residue class modulo I, so that R is the disjoint union of the different residue classes. This fact is sometimes expressed by saying that R/I is a *partition* of R.

Lemma 1.48 If a, $b \in R$, then either $a+I = b+I$ or $a+I \cap b+I = \emptyset$.

Proof We show that $a{+}I{\cap}b{+}I \neq \emptyset$ implies $a{+}I = b{+}I$. Let $c \in a{+}I{\cap}b{+}I$. Then $c = a+s = b+s'$ for some s, $s' \in I$. From the second equation we obtain $a-b = s'-s \in I$, and hence $a+I = b+I$ by Lemma 1.44. □

In order to complete the program that we outlined preceding Definition 1.42, it remains to turn R/I into a ring by defining an appropriate addition and multiplication, and to show that the map $\varphi : R \longrightarrow R/I$ with $\varphi(a) = a+I$ is a homomorphism. As we will see, this can be achieved in a very natural way. However, beginners often find it hard to believe that residue classes, which are by nature sets (or "lumps") of elements of R, can themselves be elements of a ring. But it is the very essence of abstract algebra that the nature of the elements of a group, ring, field, etc., remains unspecified by the definition. In a specific example, these elements can be numbers, functions, matrices, or, as in this case, sets of elements of a given ring.

The actual definition of "+" and "$\cdot$" on R/I refers to addition and multiplication on R in a natural way. We will not distinguish notationally between the operations on R and those on R/I.

Proposition 1.49 *Let I be a proper ideal of the ring R. For a, $b \in R$, set*

(i) $(a+I)+(b+I)=(a+b)+I$, *and*

(ii) $(a+I)(b+I)=ab+I$.

With these operations, R/I becomes a ring whose unity is the residue class $1+I$ and whose zero element is the residue class I.

Proof Verification of the ring axioms is actually going to be easy. However, there is a problem with the way the operations are defined that needs to be taken care of first. We have defined $(a+I)+(b+I)$ as $(a+b)+I$. Now someone else may form the same sum using different representatives, i.e., a', $b' \in R$ with $a+I=a'+I$ and $b+I=b'+I$. The result would be $(a'+b')+I$, and we must show that this is the same as our result $(a+b)+I$. This is also called showing that the operations are *well-defined*, or *independent of representatives*. So let a, b, a', $b' \in R$ with $a+I=a'+I$ and $b+I=b'+I$. Then $a-a' \in I$ and $b-b' \in I$, hence

$$(a+b)-(a'+b')=(a-a')+(b-b') \in I$$

and thus $(a+b)+I=(a'+b')+I$. To see that multiplication is well-defined too, let a, b, a', b' as before. We first note that by inside-outside multiplication, $(a-a')b=ab-a'b \in I$ and $a'(b-b')=a'b-a'b' \in I$. It follows that the sum

$$(ab-a'b)+(a'b-a'b')=ab-a'b'$$

is in I, too, and we see that $ab+I=a'b'+I$ as desired.

It is now easy to see that R/I with these operations satisfies the ring axioms: they are inherited from R. We verify distributivity as an example. For a, b, $c \in R$, we have

$$\begin{aligned}(a+I)\big((b+I)+(c+I)\big) &= (a+I)\big((b+c)+I\big)\\ &= \big(a(b+c)\big)+I\\ &= (ab+ac)+I\\ &= (ab+I)+(ac+I)\\ &= (a+I)(b+I)+(a+I)(c+I).\end{aligned}$$

The zero element of R/I is I since $(a+I)+(0+I)=(a+I)$ for all $a \in R$, and the unity of R/I is $1+I$ since $(a+I)(1+I)=(a+I)$ for all $a \in R$. $0+I \neq 1+I$ since otherwise 1 would have to be in I, which is not the case by Lemma 1.39 since I was assumed to be proper. □

Definition 1.50 R/I as described in the proposition above is called the **residue class ring of R modulo I**, or **mod I**.

We will use the notations 0, $0+I$, and I for the zero element of R/I, 1 and $1+I$ for its unity. Note that in R/I, we add and multiply residue classes by adding and multiplying their representatives.

Exercises 1.51 Let $\overline{n}$ stand for $n+m\mathbb{Z}$ $(m, n \in \mathbb{Z})$.

(i) What are the elements of $\mathbb{Z}/2\mathbb{Z}$? What is $\overline{1}+\overline{1}$ in the ring $\mathbb{Z}/2\mathbb{Z}$?

(ii) What is $\overline{9}(\overline{5}+\overline{3})$ in the ring $\mathbb{Z}/12\mathbb{Z}$?

(iii) Find all zero divisors and all units of the ring $\mathbb{Z}/12\mathbb{Z}$.

Exercise 1.52 Consider the map

$$\begin{array}{rccc} \varphi: & \mathbb{Z}/3\mathbb{Z} & \longrightarrow & \mathbb{Z}/6\mathbb{Z} \\ & m+3\mathbb{Z} & \longmapsto & 4m+6\mathbb{Z}. \end{array}$$

Show that φ satisfies properties (i) and (ii) of Definition 1.27, but not property (iii). Conclude that a ring may have a non-empty subset that satisfies properties (ii) and (iii) of Definition 1.20, but not property (i).

We are now finally in a position to define the homomorphism whose kernel is the given ideal I.

Proposition 1.53 *If I is a proper ideal of the ring R, then the map $\varphi : R \longrightarrow R/I$ defined by $\varphi(a) = a+I$ for all $a \in R$ is a surjective homomorphism of rings with* $\ker(\varphi) = I$.

Proof For all a, $b \in R$, we have

$$\varphi(a+b) = (a+b)+I = (a+I)+(b+I) = \varphi(a)+\varphi(b),$$

and similarly $\varphi(ab) = \varphi(a) \cdot \varphi(b)$. Moreover, $\varphi(1) = 1+I$, and the latter is the unity of R/I. We have proved that that φ is a homomorphism of rings. It is surjective because if $a+I \in R/I$, then $\varphi(a) = a+I$. Finally, $a \in \ker(\varphi)$ iff $\varphi(a) = 0$ iff $a+I = 0+I$ iff $a \in I$. □

Definition 1.54 The homomorphism φ described in the proposition above is called the **canonical homomorphism** from R to R/I.

We have now completed the program described preceding Definition 1.42. Let us emphasize again that the canonical homomorphism should be visualized as "lumping together" elements of R into residue classes modulo I, with $\varphi(a) = \varphi(b)$ iff $a-b \in I$ iff $b \in a+I$ iff $a \in b+I$. We mention at this point a different notation for the equivalent conditions above which is used primarily, but not exclusively, when working in the integers. One writes, instead of $(a+I)=(b+I)$,

$$a \equiv b \bmod I$$

and says "a is **congruent** b **modulo** I." Equivalence modulo I is thus equality in the residue class ring R/I. If I is a principal ideal, say $I = \mathrm{Id}(c)$, one also uses $a \equiv b \bmod c$. The congruence notation has the advantage that one can simply write a, b, ... instead of the longer $a+I$, $b+I$, ..., thus suppressing the fact that the elements of R/I are residue classes. It has the disadvantage that one might forget that fact.

1.6 The Homomorphism Theorem

Let us now once again consider a given homomorphism $\varphi : R \longrightarrow S$ of rings. For $a, b \in R$, we have $\varphi(a) = \varphi(b)$ iff $\varphi(a-b) = 0$ iff $(a-b) \in \ker(\varphi)$. So just like the canonical homomorphism, φ identifies those elements of R whose difference lies in the ideal $\ker(\varphi)$. Now if we are given an ideal $I \subseteq \ker(\varphi)$, then we can break up the action of φ into two steps: first we identify elements whose difference lies in I by passing to R/I. Because of $I \subseteq \ker(\varphi)$, these will have to be identified by φ anyway, so we can continue from R/I to S in such a way that the composition of the two steps yields φ. This is the content of the following theorem.

Theorem 1.55 (Homomorphism Theorem) *Let $\varphi : R \longrightarrow S$ be a homomorphism of rings, I an ideal of R with $I \subseteq \ker(\varphi)$. Denote the canonical homomorphism from R to R/I by χ. Then the map*

$$\begin{array}{cccc} \psi: & R/I & \longrightarrow & S \\ & (a+I) & \longmapsto & \varphi(a) \end{array}$$

is well-defined. ψ is a homomorphism of rings satisfying $\psi \circ \chi = \varphi$.

$$\begin{tikzcd} R \arrow[r, "\varphi"] \arrow[d, "\chi"'] & S \\ R/I \arrow[ur, "\psi"'] & \end{tikzcd}$$

The map ψ is surjective iff φ is surjective. It is injective iff $I = \ker(\varphi)$. □

Proof To say that ψ is well-defined is to say that the value $\varphi(a)$ of $(a+I)$ under ψ is independent of the representative of $a+I$ that one chooses. Let $a, a' \in R$ with $a+I = a'+I$. Then $a - a' \in I \subseteq \ker(\varphi)$, so

$$0 = \varphi(a-a') = \varphi(a) - \varphi(a')$$

and hence $\varphi(a) = \varphi(a')$. It is easy to see that ψ is a homomorphism: we have

$$\begin{aligned}\psi((a+I)+(b+I)) &= \psi((a+b)+I)\\ &= \varphi(a+b)\\ &= \varphi(a)+\varphi(b)\\ &= \psi(a+I)+\psi(b+I).\end{aligned}$$

Similarly, $\psi((a+I)(b+I)) = \psi(a+I)\psi(b+I)$. Also, $\psi(1+I) = \varphi(1_R) = 1_S$. To see that $\psi \circ \chi = \varphi$, let $a \in R$. Then $\psi(\chi(a)) = \psi(a+I) = \varphi(a)$.

If φ is surjective, then that means $\psi \circ \chi$ is surjective, and by Lemma 0.19 (ii), it follows that ψ is surjective. Conversely, if ψ is surjective, then φ, as the composition of two surjective maps, is surjective too. Now assume that ψ is injective. It suffices to show that $\ker(\varphi) \subseteq I$, the other inclusion being part of the assumption. Let $a \in \ker(\varphi)$. Then $\psi(a+I) = \varphi(a) = 0$. Since ψ is injective, it follows that $a+I$ equals zero in R/I. But that zero element is I, so we have $a+I = I$ and thus $a \in I$. Finally, assume that $I = \ker(\varphi)$. We want to show that ψ is injective. Using Lemma 1.32, we show that $\ker(\psi) = \{0\} = \{I\}$. Let $a+I \in \ker(\psi)$. Then $\varphi(a) = \psi(a+I) = 0$, which means that $a \in \ker(\varphi) = I$. By Lemma 1.47, it follows that $a+I = I$. □

Following Lemma 1.32, we announced that we were going to use the concepts of ideals and residue class rings in order to show how the kernel of a homomorphism φ determines the structure of $\varphi(R)$. We are now finally in a position to state this result. It is, in fact, an easy corollary to the homomorphism theorem.

Corollary 1.56 *Let $\varphi : R \longrightarrow S$ be a homomorphism of rings. Then*

$$R/\ker(\varphi) \simeq \varphi(R).$$

An isomorphism $\psi : R/\ker(\varphi) \longrightarrow \varphi(R)$ is given by $\psi(a+\ker(\varphi)) = \varphi(a)$.

Proof We apply Theorem 1.55 to φ and $I = \ker(\varphi)$. We obtain an injective homomorphism

$$\begin{array}{rccc}\psi : & R/\ker(\varphi) & \longrightarrow & S\\ & \psi(a+\ker(\varphi)) & \longmapsto & \varphi(a)\end{array}$$

It is easy to see that $\psi(R/\ker(\varphi)) = \varphi(R)$. So if we consider ψ as a map from $R/\ker(\varphi)$ to $\psi(R/\ker(\varphi))$ (and thus force it to be surjective too), then we are looking at an isomorphism between $R/\ker(\varphi)$ and $\varphi(R)$. □

Corollary 1.56 is often applied in the following way. An ideal I of a ring R is given, and one wants to find out more about R/I. Now if one can find a surjective homomorphism φ from R to some known ring S such that $\ker(\varphi) = I$, then one may conclude that $R/I = R/\ker(\varphi) \simeq \varphi(R) = S$. The results of the rest of this section illustrate this technique.

Example 1.57 Let $\boldsymbol{I}$ be an interval on the real line, $x_0 \in \boldsymbol{I}$, and J the set of all $f \in \mathrm{C}(\boldsymbol{I}, \mathbb{R})$ with $f(x_0) = 0$. It is easy to verify that J is an ideal of $\mathrm{C}(\boldsymbol{I}, \mathbb{R})$. We want to determine the structure of $\mathrm{C}(\boldsymbol{I}, \mathbb{R})/J$. All we have to do is note that $J = \ker(\varepsilon_{x_0})$, where ε_{x_0} is the evaluation map defined in Exercise 1.28 (iii). The image of ε_{x_0} was all of $\mathbb{R}$, so $\mathrm{C}(\boldsymbol{I}, \mathbb{R})/J$ is isomorphic to the reals. This fact plays an important role in the theory of rings of continuous functions.

Exercise 1.58 Let R be a ring. Show that $R/\{0\}$ is isomorphic to R.

The following easy exercise prepares the ground for the next theorem.

Exercises 1.59 Let R be a ring, S a subring of R, and I a proper ideal of R. Show the following:

(i) $S \cap I$ is a proper ideal of the ring S.

(ii) If we define $S + I$ to be the set $\{ s + a \mid s \in S,\ a \in I \}$, then $S + I$ is a subring of R that contains S.

(iii) I is a proper ideal of the ring $S + I$.

Theorem 1.60 (First Isomorphism Theorem) *Let R be a ring, S a subring of R, and I a proper ideal of R. Then*

$$S/(S \cap I) \simeq (S + I)/I.$$

Proof Consider the map $\varphi : S \longrightarrow (S + I)/I$ given by $\varphi(s) = s + I$. We see that $\varphi = \chi \circ \iota$, where $\iota : S \longrightarrow S + I$ is the natural embedding (where $\iota(s) = s$), and $\chi : (S + I) \longrightarrow (S + I)/I$ is the canonical homomorphism. It follows that φ, as the composition of two homomorphisms, is itself a homomorphism. By Corollary 1.56, we may conclude that

$$S/\ker(\varphi) \simeq \varphi(S).$$

The theorem can thus be proved by showing that $\ker(\varphi) = S \cap I$, and that $\varphi(S) = (S + I)/I$, i.e., that φ is surjective. Now if $s \in \ker(\varphi)$, then $\varphi(s) = s + I = 0 + I$, hence $s \in 0 + I = I$. But s was in S to begin with, so $s \in S \cap I$. Conversely, if $s \in S \cap I$, then $\varphi(s) = s + I = 0 + I$ and thus $s \in \ker(\varphi)$. To see that φ is surjective, let $b + I \in (S + I)/I$. Then $b = s + a$ for some $s \in S$ and $a \in I$, and the fact that $(s + a) - s = a \in I$ implies that $\varphi(s) = s + I = (s + a) + I = b + I$. □

In the following exercise, the reader will verify Theorem 1.60 explicitly for a specific example.

Exercise 1.61 Consider the ring $\mathbb{Z}_p$ of Example 1.15 (iv), where p is a fixed prime number. If we identify each integer m with the fraction $m/1$, then $\mathbb{Z}$ is obviously a subring of $\mathbb{Z}_p$. It is also obvious that $p\mathbb{Z}_p$, the principal ideal of $\mathbb{Z}_p$ generated by p, consists of all rational numbers which, after reduction to lowest terms, have a denominator that is not divisible by p and a numerator that is a multiple of p. Finally, one sees easily that $\mathbb{Z} \cap p\mathbb{Z}_p = p\mathbb{Z}$. Show the following:

(i) The subring $\mathbb{Z} + p\mathbb{Z}_p$ of $\mathbb{Z}_p$ actually equals all of $\mathbb{Z}_p$. (Hint: Use the fact that whenever p does not divide an integer m, then $\gcd(p, m) = 1$, and thus there exist integers s and t with $1 = sp + tm$. These facts have not been proved yet, but were mentioned in Section 0.1.)

(ii) Use (i) and the first isomorphism theorem to show that the residue class ring $\mathbb{Z}_p/p\mathbb{Z}_p$ is isomorphic to $\mathbb{Z}/p\mathbb{Z}$, and that $\mathbb{Z}_p/p\mathbb{Z}_p$ consisists of the p residue classes $p\mathbb{Z}_p$, $1 + p\mathbb{Z}_p$, ..., $(p-1) + p\mathbb{Z}_p$.

In order to state the next theorem, we need a lemma that relates the ideals of R to those of R/I. This will be a special case of the following more general lemma.

Lemma 1.62 Let $\varphi : R \longrightarrow S$ be a homomorphism of rings. Let us denote by $\mathcal{I}(R)$ and $\mathcal{I}(S)$ the set of all ideals of R and S, respectively, and by $\mathcal{I}_\varphi(R)$ the set of all those ideals I of R that satisfy $\ker(\varphi) \subseteq I$. Then the following hold:

(i) $\varphi^{-1}(J) \in \mathcal{I}_\varphi$ for all $J \in \mathcal{I}(S)$.

(ii) If φ is surjective, then $\varphi(I) \in \mathcal{I}(S)$ for all $I \in \mathcal{I}(R)$.

(iii) If φ is surjective, then the map

$$\begin{array}{rccc} \chi: & \mathcal{I}_\varphi(R) & \longrightarrow & \mathcal{I}(S) \\ & I & \longmapsto & \varphi(I) \end{array}$$

is bijective, and $\chi^{-1}(J) = \varphi^{-1}(J)$ for all $J \in \mathcal{I}(S)$.

Proof (i) Let J be an ideal of S and set $I = \varphi^{-1}(J)$. If a_1, $a_2 \in I$, then $\varphi(a_1)$, $\varphi(a_2) \in J$. It follows that

$$\varphi(a_1 + a_2) = \varphi(a_1) + \varphi(a_2) \in J$$

because J is an ideal, and we see that $a_1 + a_2 \in I$. If $a \in I$ and $r \in R$, then $\varphi(a) \in J$ and $\varphi(r) \in S$, and so $\varphi(ra) = \varphi(r) \cdot \varphi(a) \in J$, which means $ar \in I$. From $0 \in J$ it is clear that $\ker(\varphi) = \varphi^{-1}(\{0\}) \subseteq I$.

(ii) Assume that φ is surjective. Let I be an ideal of R, and set $J = \varphi(I)$. Then clearly $J \neq \emptyset$. If b_1, $b_2 \in J$, then $b_1 = \varphi(a_1)$ and $b_2 = \varphi(a_2)$ with a_1, $a_2 \in I$, and thus

$$b_1 + b_2 = \varphi(a_1) + \varphi(a_2) = \varphi(a_1 + a_2) \in J$$

because I is an ideal. If $b \in J$ and $s \in S$, then $b = \varphi(a)$ and $s = \varphi(r)$ with $a \in I$ and $r \in R$. We see that $sb = \varphi(r) \cdot \varphi(a) = \varphi(ra) \in J$.

(iii) In view of Lemma 0.21 (iii), it suffices to show that the map

$$\begin{array}{rccc} \kappa: & \mathcal{I}(S) & \longrightarrow & \mathcal{I}_\varphi(R) \\ & I & \longmapsto & \varphi^{-1}(I) \end{array}$$

satisfies $\kappa \circ \chi = \mathrm{id}_{\mathcal{I}_\varphi(R)}$ and $\chi \circ \kappa = \mathrm{id}_{\mathcal{I}(S)}$. The second equality states that $\varphi(\varphi^{-1}(J)) = J$ for all $J \in \mathcal{I}(S)$, which is true by Lemma 0.15 (ii). The first equality means $\varphi^{-1}(\varphi(I)) = I$ for all $I \in \mathcal{I}_\varphi(R)$. The inclusion "$\supseteq$" is true by Lemma 0.15 (i). For the reverse inclusion, let $I \in \mathcal{I}_\varphi(R)$ and $a \in \varphi^{-1}(\varphi(I))$ Then $\varphi(a) \in \varphi(I)$, and so there exists $a' \in I$ with $\varphi(a') = \varphi(a)$. It follows that $a - a' \in \ker(\varphi) \subseteq I$, and we may conclude that $a = a' + (a - a') \in I$. □

If I is a proper ideal of the ring R and J is an ideal of R with $I \subseteq J$, then we denote by J/I the subset

$$\{\, a + I \mid a \in J \,\}$$

of R/I, i.e., the image of J under the canonical homomorphism from R to R/I. The lemma above with the canonical homomorphism $\varphi : R \longrightarrow R/I$ taken for φ yields the following result.

Lemma 1.63 Let I be a proper ideal of the ring R. Then there is a one-to-one correspondence between those ideals of R that contain I and the ideals of R/I. A bijection from the former set of ideals to the latter is given by $J \longmapsto J/I$. The inverse of this map is the "lifting" of ideals, where an ideal J' of R/I is mapped to $\{\, a \in R \mid a + I \in J' \,\}$. □

Theorem 1.64 (Second Isomorphism Theorem) *Let R be a ring and I and J proper ideals of R with $I \subseteq J$. Then*

$$R/J \simeq (R/I)\,/\,(J/I).$$

Proof Consider the map

$$\begin{array}{rccl} \varphi : & R & \longrightarrow & (R/I)\,/\,(J/I) \\ & a & \longmapsto & (a + I) + J/I. \end{array}$$

We see that $\varphi = \chi_2 \circ \chi_1$, where

$$\chi_1 : R \longrightarrow R/I \quad \text{and} \quad \chi_2 : R/I \longrightarrow (R/I)\,/\,(J/I)$$

are the canonical homomorphisms. Being the composition of two surjective homomorphisms, φ is itself a surjective homomorphism, and by Corollary 1.56 we may conclude that

$$R/\ker(\varphi) \simeq \varphi(R) = (R/I)\,/\,(J/I).$$

It now suffices to show that $\ker(\varphi) = J$. If $a \in \ker(\varphi)$, then

$$\varphi(a) = (a + I) + J/I = J/I,$$

the latter being the zero element of the ring $(R/I)/(J/I)$. By Lemma 1.47, it follows that $a + I \in J/I$ and thus $a \in J$ by Lemma 1.63. Conversely, if

$a \in J$, then $a{+}I \in J/I$ by the definition of J/I, hence $\varphi(a) = (a{+}I){+}J/I = J/I$. We see that $a \in \ker(\varphi)$. □

The following example, though somewhat tedious, provides a good understanding of the second isomorphism theorem, and of the structure of residue class rings in general.

Example 1.65 Consider the ring $\mathbb{Z}$ and the two ideals $12\mathbb{Z}$ and $3\mathbb{Z}$. Then $12\mathbb{Z} \subset 3\mathbb{Z}$ since every multiple of 12 is a multiple of 3. We have

$$\begin{aligned}\mathbb{Z}/12\mathbb{Z} &= \{\, m + 12\mathbb{Z} \mid 0 \le m \le 11 \,\}, \quad \text{and}\\ \mathbb{Z}/3\mathbb{Z} &= \{\, m + 3\mathbb{Z} \mid 0 \le m \le 2 \,\}.\end{aligned}$$

The ideal $3\mathbb{Z}/12\mathbb{Z}$ of the ring $\mathbb{Z}/12\mathbb{Z}$ consists by definition of those residue classes of $\mathbb{Z}/12\mathbb{Z}$ whose representatives are in $3\mathbb{Z}$, i.e., multiples of 3. We see that

$$3\mathbb{Z}/12\mathbb{Z} = \{\, 0 + 12\mathbb{Z}, 3 + 12\mathbb{Z}, 6 + 12\mathbb{Z}, 9 + 12\mathbb{Z} \,\}.$$

Finally, we have

$$(\mathbb{Z}/12\mathbb{Z}) \,/\, (3\mathbb{Z}/12\mathbb{Z}) = \{\, (m + 12\mathbb{Z}) + 3\mathbb{Z}/12\mathbb{Z} \mid 0 \le m \le 11 \,\}.$$

We see that there are repetitions among these twelve elements: two residue classes $(m_1 + 12\mathbb{Z}) + 3\mathbb{Z}/12\mathbb{Z}$ and $(m_2 + 12\mathbb{Z}) + 3\mathbb{Z}/12\mathbb{Z}$ are equal iff

$$\begin{aligned}(m_1 + 12\mathbb{Z}) - (m_2 + 12\mathbb{Z}) \in 3\mathbb{Z}/12\mathbb{Z} \quad &\text{iff} \quad (m_1 - m_2) + 12\mathbb{Z} \in 3\mathbb{Z}/12\mathbb{Z}\\ &\text{iff} \quad m_1 - m_2 \text{ is divisible by 3.}\end{aligned}$$

So the ring $(\mathbb{Z}/12\mathbb{Z})/(3\mathbb{Z}/12\mathbb{Z})$ really consists of just the three residue classes

$$\{\, (m + 12\mathbb{Z}) + 3\mathbb{Z}/12\mathbb{Z} \mid 0 \le m \le 2 \,\},$$

and these behave like integers modulo $3\mathbb{Z}$. We have thus confirmed the statement of the second isomorphism theorem:

$$\mathbb{Z}/3\mathbb{Z} \simeq (\mathbb{Z}/12\mathbb{Z}) \,/\, (3\mathbb{Z}/12\mathbb{Z}).$$

Lemma 1.66 Let R and S be rings, $\varphi : R \longrightarrow S$ a surjective homomorphism of rings, I an ideal of R. The map

$$\begin{aligned}\psi : \quad R/I &\longrightarrow S/\varphi(I)\\ (a + I) &\longmapsto \varphi(a) + \varphi(I)\end{aligned}$$

is well-defined, and it is a surjective homomorphism of rings. If φ is an isomorphism, then so is ψ.

Proof We already know that $\varphi(I)$ is an ideal of S, and so the statement of the lemma is meaningful. Let $\chi : S \longrightarrow S/\varphi(I)$ be the canonical homomorphism, and consider the homomorphism

$$\chi \circ \varphi : R \longrightarrow S/\varphi(I).$$

Now $a \in I$ implies $\varphi(a) \in \varphi(I)$ and thus $\chi(\varphi(a)) = 0$. We see that $I \subseteq \ker(\chi \circ \varphi)$. We may thus apply Theorem 1.55 to conclude that the map

$$\psi : R/I \longrightarrow S/\varphi(I)$$

given by

$$\psi(a + I) = \chi\big(\varphi(a)\big) = \varphi(a) + \varphi(I)$$

is well-defined, and that it is a surjective homomorphism of rings. Now assume that φ is injective. We first note that $a \in \ker(\chi \circ \varphi)$ always implies $\varphi(a) \in \varphi(I)$ and thus $a \in \varphi^{-1}(\varphi(I))$. Since φ is assumed to be injective, the latter set equals I by Lemma 0.15 (i). So in this case, $\ker(\chi \circ \varphi) = I$, and hence ψ is injective by Theorem 1.55. □

1.7 Gcd's, Lcm's, and Principal Ideal Domains

In this section, we will explore some of the connections between ideal theory and the concept of *divisibility* in a ring.

Definition 1.67 Let R be a ring and $a, b \in R$. Then we say that a **divides** b and write $a \,|\, b$ if there exists $c \in R$ with $b = ac$. We call $a, b \in R$ **associated** if there exists a unit $u \in R$ with $a = bu$. If $a, b \in R$, then $d \in R$ is called a **greatest common divisor**, or **gcd**, of a and b if it has the following two properties:

(i) $d \,|\, a$ and $d \,|\, b$, i.e., d is a common divisor of a and b, and

(ii) whenever $d' \,|\, a$ and $d' \,|\, b$ for some $d' \in R$, then $d' \,|\, d$, i.e., any common divisor of a and b divides d.

Finally, a and b are called **relatively prime** if 1 is a gcd of a and b.

Note that if $a = bu$ for some unit u, then $b = au^{-1}$, and u^{-1} is again a unit. This justifies the symmetry in the definition of associated elements. Let us point out a subtlety of the terminology here: every $a \in R$ divides 0, since $0 = a \cdot 0$, but a is called a *zero divisor* of R only if there exists $0 \neq b \in R$ with $0 = ab$. Although all of the definitions in this section make sense for arbitrary rings, they turn out to be interesting mainly for integral domains.

The following exercise provides examples and some important elementary properties of divisibility, units, associated elements, and gcd's.

Exercises 1.68 Let R be a domain and $a, b, c, d, s, t, u \in R$. Show the following:

(i) If u is a unit, then $u \,|\, a$.

(ii) a and $-a$ are associated.

(iii) If $a \mid b$ and $b \mid c$, then $a \mid c$.

(iv) If $a \mid b$, then a is a gcd of a and b.

(v) a is a gcd of a and 0.

(vi) If $a \mid b$ and $a \mid c$, then $a \mid (sb + tc)$.

(vii) If $a \mid (b + c)$ and $a \mid b$, then $a \mid c$.

(viii) It is not true in general that $a \mid (b + c)$ implies $a \mid b$. (Make up a counterexample in $\mathbb{Z}$.)

(ix) It is not true in general that $a \mid bc$ implies $a \mid b$.

(x) If a and b are associated, then $a \mid c$ iff $b \mid c$, and $c \mid a$ iff $c \mid b$.

(xi) $a \mid b$ iff $bR \subseteq aR$, and $bR \subseteq aR$ iff $b \in aR$.

(xii) a and b are associated iff both $a \mid b$ and $b \mid a$ iff $aR = bR$.

(xiii) If a and b are associated and b and c are associated, then so are a and c.

(xiv) If $a \mid c$ and $b \mid d$, then $ab \mid cd$.

Lemma 1.69 Let R be a domain and a, $b \in R$. Then any two gcd's of a and b are associated. Conversely, if $d \in R$ is a gcd of a and b, then so is every $d' \in R$ that is associated to d.

Proof Let d and d' be gcd's of a and b. Then $d \mid d'$ since d is a common divisor and d' is a gcd of a and b. Similarly, $d' \mid d$. Hence d and d' are associated by Exercise 1.68 (xii). The second claim follows immediatetely from Exercise 1.68 (x): associated elements satisfy the exact same divisibility relations with any element of R. □

By a rather serious abuse of notation and terminology, it is common to speak of *the* gcd of a and b and to write $\gcd(a, b)$ in case one exists. It turns out that this causes no problems if the formula "$d = \gcd(a, b)$" is understood to mean "d is a gcd of a and b."

Greatest common divisors need not exist in general. Let, for example, D be the subring of $\mathbb{C}$ introduced in Exercise 1.24. Recall that for any complex number $z = a + ib$ $(a, b \in \mathbb{R})$, the norm $|z|$ of z is defined as $\sqrt{a^2 + b^2}$, and that $|z_1 z_2| = |z_1||z_2|$ for all z_1, $z_2 \in \mathbb{C}$. Note that the norm of an element of D is necessarily of the form $\sqrt{a^2 + 5b^2}$ with a, $b \in \mathbb{Z}$. Let us now consider

$$z_1 = 2 + 2i\sqrt{5} \in D \quad \text{and} \quad z_2 = 6 \in D.$$

Then $2 \mid z_1$ since $z_1 = 2(1 + i\sqrt{5})$, and $2 \mid z_2$ since $z_2 = 2 \cdot 3$. Furthermore, $1 + i\sqrt{5} \mid z_1$ again since $z_1 = 2(1 + i\sqrt{5})$, and $1 + i\sqrt{5} \mid z_2$ since

$$z_2 = (1 + i\sqrt{5})(1 - i\sqrt{5}).$$

Now if $d \in D$ were a gcd of z_1 and z_2 in D, we would have to have $d \mid z_1$, $d \mid z_2$, $2 \mid d$, and $1+i\sqrt{5} \mid d$. If we write out the meaning of these divisibilties, take norms everywhere and then square the equations, we see that there would have to be an integer m of the form $a^2 + 5b^2$ $(a, b \in \mathbb{Z})$ such that m divides 24 and 36 and is divided by 4 and 6 in $\mathbb{Z}$, and it is easy to see that this is impossible by checking the finitely many possibilities for m.

Gcd's may exist for a variety of reasons. A natural *sufficient* condition for the existence of a gcd of two elements a and b of a domain R is that the ideal

$$\mathrm{Id}(a, b) = aR + bR$$

(which consists of all "linear combinations" $sa + tb$) is principal: we will now show that every single generator of this ideal is then a gcd of a and b.

Lemma 1.70 Let R be a domain and a, b, $d \in R$. Then the following are equivalent:

(i) d is a common divisor of a and b, and there exist s, $t \in R$ with $d = sa + tb$.

(ii) d is a gcd of a and b, and there exist s, $t \in R$ with $d = sa + tb$.

(iii) $aR + bR = dR$.

Proof (i)$\Longrightarrow$(ii): We must show that $d' \mid a$ and $d' \mid b$ implies $d' \mid d$ for all $d' \in R$. In view of Execercise 1.68 (vi), this is immediate from the equation $d = sa + tb$.

(ii)$\Longrightarrow$(iii): From the fact that d is a common divisor of a and b, it follows with Exercise 1.68 (xi) that $aR \subseteq dR$ and $bR \subseteq dR$, and one easily concludes that $aR + bR \subseteq dR$. The equation $d = sa + tb$ states that $d \in aR + bR$, and it follows that $dR \subseteq aR + bR$.

(iii)$\Longrightarrow$(i): From $aR + bR \subseteq dR$ it follows that $aR \subseteq dR$ and $bR \subseteq dR$, and thus d is a common divisor of a and b. The existence of s and t with $d = sa + bt$ is an immediate consequence of

$$d \in dR \subseteq aR + bR. \ \square$$

An important consequence of the direction (i)$\Longrightarrow$(ii) above is this: if a and b are elements of a domain R and there exist s, $t \in R$ with $1 = sa+tb$, then 1 is a gcd of a and b in R.

In Section 2.3, we will encounter domains R in which any two elements a and b have a gcd despite the fact that $aR+bR$ is not in general a principal ideal. In that case, $\gcd(a, b)$ can not be written as a sum of multiples of a and b.

Exercise 1.71 Let R be a domain and a, b, $d \in R$ such that d is a gcd of a and b. Show that $aR + bR \subseteq dR$.

It is immediate from Lemma 1.70 that in a domain R where every ideal is principal, any two elements a and b will have a gcd, namely, any generator of the ideal $aR + bR$.

Definition 1.72 A **principal ideal ring** is a ring R with the property that every ideal of R is principal. A principal ideal ring which is also an integral domain is called a **principal ideal domain**, which is sometimes abbreviated to **PID**.

The following theorem provides an example. Its statement is one of the most important facts in algebra and number theory.

Theorem 1.73 *$\mathbb{Z}$ is a principal ideal domain.*

Proof Let I be an ideal of $\mathbb{Z}$. If $I = \{0\}$, then it is generated by 0. Otherwise, we consider the set

$$I^+ = \{ m \in I \mid m > 0 \}$$

of natural numbers. This set is not empty because $I \neq \{0\}$ and $k \in I$ implies $-k \in I$ for all $k \in \mathbb{Z}$. Since every non-empty set of natural numbers has a least element, we can find a least element n of I^+. We claim that $I = n\mathbb{Z}$. Since $n \in I$, we have $n\mathbb{Z} \subseteq I$ by Exercise 1.38. Conversely, let $m \in I$. Dividing m by n, we can find q, $r \in \mathbb{Z}$ with $m = nq + r$ and $0 \leq r \leq n - 1$. But $r = nq - m \in I$, so we must have $r = 0$ by the minimality of n. The equation $m = nq$ now shows that $m \in n\mathbb{Z}$. □

The proof of the following proposition is immediate from Lemma 1.70.

Proposition 1.74 *Let R be a PID and a, $b \in R$. Then a and b have a gcd d in R, and $aR + bR = dR$. In particular, there exist s, $t \in R$ with $d = sa + bt$.* □

We may now conclude that any two integers have a gcd in $\mathbb{Z}$. The proof that we have given is, however, highly non-constructive: the gcd is a single generator of a certain ideal, and such a generator was obtained as the least element of some infinite set of natural numbers. Gcd's in $\mathbb{Z}$—and in a variety of other domains—may actually be computed by means of the *Euclidean algorithm* which we will discuss in Section 2.2. For the moment, the reader who wants to see examples for the results in this section will have to rely on the elementary way of finding gcd's in $\mathbb{Z}$ by collecting common prime factors. Note that this method—as opposed to the extended Euclidean algorithm—does not provide s, $t \in \mathbb{Z}$ with $\gcd(a, b) = sa + tb$. Following are some more properties and examples of principal ideal rings.

Lemma 1.75 Every homomorphic image of a principal ideal ring is again a principal ideal ring.

Proof Let R be a principal ideal ring, S a ring, $\varphi : R \longrightarrow S$ a surjective homomorphism, and let J be any ideal of S. Then the ideal $I = \varphi^{-1}(J)$ of R (cf. Lemma 1.62) is principal, say $I = aR$. We claim that $J = \varphi(a)S$. Indeed, if $b \in J$, then $b = \varphi(c)$ for some $c \in I$. It follows that $c = ra$ for some $r \in R$ and thus

$$b = \varphi(c) = \varphi(ra) = \varphi(r)\varphi(a) \in \varphi(a)S. \ \square$$

The following lemma is now obvious from the fact that any residue class ring of a a ring R modulo an ideal is a homomorphic image of R under the canonical homomorphism.

Lemma 1.76 If R is a principal ideal ring, then so is R/I for every proper ideal I of R. In particular, $\mathbb{Z}/n\mathbb{Z}$ is a principal ideal ring for every $n \in \mathbb{Z} \setminus \{1, -1\}$. □

The example $\mathbb{Z}/n\mathbb{Z}$ shows that there are indeed principal ideal rings that are not domains: $2+6\mathbb{Z}$ is a zero divisor in $\mathbb{Z}/6\mathbb{Z}$. The next lemma provides more examples of PID's. It is important to note that in our development of the theory, this lemma has the status of just an example: its proof uses the unique prime factor decomposition of integers which we described but did not prove in Section 0.1. (The proof is to be found in Section 2.3, Theorem 2.51.)

Lemma 1.77 For any prime number p, the ring $\mathbb{Z}_p$ is a principal ideal domain whose non-trivial ideals are $\mathbb{Z}_p \supset p\mathbb{Z}_p \supset p^2\mathbb{Z}_p \supset p^3\mathbb{Z}_p \supset \cdots$.

Proof Using unique prime factor decomposition in the numerator, every non-zero element a of $\mathbb{Z}_p$ may be written in the form $a = p^n(r/s)$ with $n \in \mathbb{N}$, and r and s not divisible by p. The exponent n is then unique, and we will call it $h(a)$, the *height* of a. Now let I be a non-trivial ideal of $\mathbb{Z}_p$. The set

$$\{\, m \in \mathbb{N} \mid \text{there exists } a \in I \text{ with } h(a) = m \,\}$$

has a least element, say n, and we can find $a \in I$ with $h(a) = n$. We can then write $a = p^n(r/s)$ with r and s not divisible by p. It follows that $s/r \in \mathbb{Z}_p$ and hence

$$p^n = p^n(r/s)(s/r) \in I.$$

This shows that $p^n\mathbb{Z}_p \subseteq I$. Conversely, let $0 \neq b \in I$. Then $h(b) \geq n$, so $b = p^m(u/v)$ with $m \geq n$ and u, v not divisible by p. We can thus write

$$b = p^n p^{m-n}(u/v) = p^n c$$

with $c \in \mathbb{Z}_p$, which means that $b \in p^n\mathbb{Z}_p$. We have proved that every ideal I of $\mathbb{Z}_p$ is of the form $I = p^n\mathbb{Z}_p$ with $n \geq 0$, and it is now rather obvious that the ideals form the indicated chain under inclusion. □

Next, we show that our third standard example $\mathrm{C}(\boldsymbol{I}, \mathbb{R})$ is not in general a principal ideal ring.

Example 1.78 Let $\boldsymbol{I}$ be a proper interval on the real line, i.e., an interval that consists of more than just one point. Then $\mathrm{C}(\boldsymbol{I},\mathbb{R})$ is not a principal ideal ring. To see this, let x_0 be a point in the interior of $\boldsymbol{I}$, and let J be the ideal of all functions in $\mathrm{C}(\boldsymbol{I},\mathbb{R})$ that vanish at x_0. Assume for a contradiction that J were principal, generated by $f \in \mathrm{C}(\boldsymbol{I},\mathbb{R})$. Let g be any non-horizontal straight line with x-intercept x_0. Then $g \in J$ and hence $g = fh_1$ for some $h_1 \in \mathrm{C}(\boldsymbol{I},\mathbb{R})$. From $g(x) \neq 0$ for all $x \neq x_0$ we see that $f(x) \neq 0$ for all $x \in \boldsymbol{I} \setminus \{x_0\}$. Since f is a continuous real-valued function on $\boldsymbol{I}$, so is the function $f^{1/3}$. Moreover, $f^{1/3}$ is also in J, and thus there must exist $h_2 \in \mathrm{C}(\boldsymbol{I},\mathbb{R})$ with $f^{1/3} = fh_2$. Then $h_2(x) = (f(x))^{-2/3}$ for all $x \in \boldsymbol{I} \setminus \{x_0\}$. Since f is continuous with $f(x_0) = 0$, we must have

$$\lim_{x \longrightarrow x_0} f(x) = 0.$$

It follows that $\lim_{x \longrightarrow x_0} h_2(x)$ does not exist, contradicting the continuity of h_2 at x_0.

We now resume our discussion of gcd's in arbitrary domains. The concept of gcd's can easily be generalized to more than two elements. Let R be a domain, $2 \leq m \in \mathbb{N}$, and $a_1, \ldots, a_m \in R$. Then $d \in R$ is called a **greatest common divisor**, or **gcd**, of $a_1, \ldots, a_m$ if

(i) $d \,|\, a_i$ for $1 \leq i \leq m$, and

(ii) $d' \,|\, d$ whenever $d' \in R$ with $d' \,|\, a_i$ for $1 \leq i \leq m$.

Furthermore, $a_1, \ldots, a_m$ are called **relatively prime** if they have 1 as a gcd. It is easy to see that the obvious generalization of Lemma 1.69 holds, and we will again frequently allow ourselves to speak of *the* gcd of ring elements. The next lemma shows that gcd's of more than two elements hardly pose any new problems.

Lemma 1.79 Let R be a domain, $2 \leq m \in \mathbb{N}$, and $a_1, \ldots, a_m \in R$. Then the following hold:

(i) If $d_{m-1} \in R$ is a gcd of $a_1, \ldots, a_{m-1}$ and $d_m \in R$ is a gcd of d_{m-1} and a_m, then d_m is a gcd of $a_1, \ldots, a_m$.

(ii) If R contains a gcd for any two elements, then it has one for every finite set of elements.

(iii) $d \in R$ is a gcd of $a_1, \ldots, a_m$ iff it is a gcd of $a_1, \ldots, a_m$, 0.

(iv) If R contains a gcd for any two elements and $0 \neq d \in R$, then d is a gcd of $da_1, da_2, \ldots, da_m$ iff $a_1, \ldots, a_m$ are relatively prime.

Proof (i) d_m divides a_m and the gcd of $a_1, \ldots, a_{m-1}$ and hence it divides each of $a_1, \ldots, a_m$. If $d' \in R$ has this latter property too, then it must divide d_{m-1} and hence the gcd d_m of d_{m-1} and a_m.

Statement (ii) follows easily from (i) together with an induction on the number of elements whose gcd we are considering, and (iii) is trivial.

(iv) Suppose d is a gcd of $da_1, \ldots, da_m$, and let $d' = \gcd(a_1, \ldots, a_m)$. Then dd' is a conmmon divisor of $da_1, \ldots, da_m$ by Exercise 1.68 (xiv), so $dd' \mid d$, say $d = udd'$, and cancelation of d shows that d' is a unit. Conversely, assume that $a_1, \ldots, a_m$ are relatively prime, and this time, let

$$d' = \gcd(da_1, \ldots, da_m).$$

Now d is a common divisor of $da_1, \ldots, da_m$, and so $d \mid d'$, say $d' = dc$. It follows that $dc \mid da_i$ for $1 \leq i \leq m$, and it is now easy to see that c is a common divisor of $a_1, \ldots, a_m$ and thus a unit. □

From the lemma above together with Proposition 1.74 we conclude that in a PID, every finite set of elements has a gcd. Although we are not yet concerned with actual computations of gcd's here, we note that by (i) above, gcd's of finitely many elements can be computed as soon as the gcd of any two elements can be computed.

Exercise 1.80 Let R be a domain, $2 \leq m \in \mathbb{N}$, and $d, a_1, \ldots, a_m \in R$. Show that the following are equivalent:

(i) d is a common divisor of $a_1, \ldots, a_m$, and there exist $s_1, \ldots, s_m \in R$ with $d = s_1a_1 + \cdots + s_ma_m$.

(ii) d is a gcd of $a_1, \ldots, a_m$, and there exist $s_1, \ldots, s_m \in R$ with $d = s_1a_1 + \cdots + s_ma_m$.

(iii) $dR = a_1R + \cdots + a_mR$.

We conclude this section with a discussion of least common multiples.

Definition 1.81 Let R be a domain and $a, b \in R$. A **least common multiple**, or **lcm**, of a and b is an element $c \in R$ that is a common multiple of a and b (i.e., $a \mid c$ and $b \mid c$) and divides any other common multiple of a and b.

Lemma 1.82 Let R be a domain and $a, b \in R$. Then $c \in R$ is an lcm of a and b iff $cR = aR \cap bR$.

Proof "$\Longrightarrow$": Let c be any lcm of a and b. Being a common multiple of a and b, c is in $aR \cap bR$, which implies

$$cR \subseteq aR \cap bR.$$

For the reverse inclusion, we note that every element of $aR \cap bR$, being a common multiple of a and b, is divided by c, which means that it is in cR.

"$\Longleftarrow$": Let $c \in R$ such that $cR = aR \cap bR$. Then $c \in aR$ and $c \in bR$, which means that c is a common multiple of a and b. For arbitrary $c' \in R$ to be a common multiple of a and b means that

$$c' \in aR \cap bR = cR,$$

and thus $c \mid c'$. This shows that c is actually the least common multiple of a and b. □

The following lemma is an easy consequence of the above one together with Exercise 1.68 (xii).

Lemma 1.83 Let R be a domain and a, b, $c \in R$ such that c is an lcm of a and b. Then the set of all lcm's of a and b consists precisely of those $c' \in R$ that are associated to c. □

As with gcd's, it is common to speak of *the* lcm of two ring elements a and b and to write "$c = \mathrm{lcm}(a, b)$" for "c is an lcm of a and b."

From Lemma 1.82, we may conclude that lcm's always exist in PID's. The next proposition, however, is not only more general, but also shows an important connection between gcd's and lcm's. From a computational point of view, it guarantees that lcm's can be computed as soon as gcd's can be found and divisions can be performed, i.e., when $a \mid b$, then one can find c with $ac = b$.

Proposition 1.84 *Let R be a domain in which any two elements have a gcd. Then any two elements have an lcm. Moreover, whenever a, $b \in R$, then* $\mathrm{lcm}(a, b) = a'b$, *where* $a' \cdot \mathrm{gcd}(a, b) = a$.

Proof Let a, $b \in R$, and let $d = \mathrm{gcd}(a, b)$. Furthermore, let a', $b' \in R$ such that $a = da'$ and $b = db'$. Then $a'b = ab'$, and we see that $a'b$ is a common multiple of a and b. Now let $c \in R$ be any common multiple of a and b, with $c = ra$. Lemma 1.79 (iv) tells us that a' and b' are relatively prime, and that

$$rd = \mathrm{gcd}(rda', rdb') = \mathrm{gcd}(ra, rb) = \mathrm{gcd}(c, rb).$$

Now b divides c and rb, and so it must divide their gcd rd, say $rd = sb$. We thus obtain

$$c = ra = rda' = sba',$$

and we see that indeed $a'b \mid c$. □

Corollary 1.85 *Let R be a domain in which any two elements have a gcd, and let a, $b \in R$ be relatively prime. Then*

$$ab = \mathrm{lcm}(a, b) \quad \text{and} \quad abR = aR \cap bR. \quad □$$

Lcm's of finitely many ring elements $a_1, \ldots, a_m$ with $m \geq 2$ are defined in the obvious manner as common multiples of $a_1, \ldots, a_m$ that divide any other common multiple of these.

Exercise 1.86 Let R be a domain and c, $a_1, \ldots, a_m \in R$. Show that the following are equivalent:

(i) c is an lcm of $a_1, \ldots, a_m$.

(ii) $cR = \bigcap_{i=1}^{m} a_i R$.

Exercise 1.87 Let R be a domain and $a_1, \ldots, a_m \in R$. Show the following:

(i) If $c_{m-1} \in R$ is an lcm of $a_1, \ldots, a_{m-1}$ and $c_m \in R$ is an lcm of c_{m-1} and a_m, then c_m is an lcm of $a_1, \ldots, a_m$.

(ii) If R contains an lcm for any two elements, then it contains one for any finite set of elements.

The exercise above reduces the computation of lcm's to the computation of lcm's of pairs of ring elements and thus, via Proposition 1.84, to the computation of gcd's of pairs of ring elements. It is worth noting, however, that the plain analogue of Proposition 1.84 fails for more than two elements: the lcm of more than two elements is *not* in general obtained by dividing their gcd out of their product. (Make up a counterexample in $\mathbb{Z}$.) The next proposition, which is preceded by a lemma, shows how Corollary 1.85 continues to hold for more than two elements if R is a PID.

Lemma 1.88 Let R be a PID, $2 \leq m \in \mathbb{N}$, and suppose $a_1, \ldots, a_m \in R$ such that a_1 and a_i are relatively prime for $2 \leq i \leq m$. Then

$$\gcd(a_1, a_2 \cdot \cdots \cdot a_m) = 1.$$

Proof We proceed by induction on m. If $m = 2$, then there is nothing to prove. Now let $m > 2$. Then

$$1 = \gcd(a_1, a_2) = \gcd(a_1, a_3 \cdot \cdots \cdot a_m)$$

by induction hypothesis, and so there exist s, t, u, $v \in R$ with

$$1 = sa_1 + ta_2 \quad \text{and} \quad 1 = ua_1 + va_3 \cdot \cdots \cdot a_m.$$

Multiplying these two equations with each other, we see that

$$1 = (sua_1 + sva_3 \cdot \cdots \cdot a_m + tua_2)a_1 + tva_2 \cdot \cdots \cdot a_m. \ \square$$

Proposition 1.89 *Let R be a PID, $2 \leq m \in \mathbb{N}$, and suppose $a_1, \ldots, a_m \in R$ are pairwise relatively prime. Then*

$$a_1 \cdot \cdots \cdot a_m = \operatorname{lcm}(a_1, \ldots, a_m) \quad \text{and} \quad a_1 \cdot \cdots \cdot a_m R = \bigcap_{i=1}^{m} a_i R.$$

Proof We proceed by induction on m. For $m = 2$, the claim is identical with Corollary 1.85. Now let $m > 2$. Using Exercise 1.87, the previous lemma and the trivial fact that $a_2, \ldots, a_m$ are again pairwise relatively prime, we see that

$$\begin{aligned} \operatorname{lcm}(a_1, \ldots, a_m) &= \operatorname{lcm}(a_1, \operatorname{lcm}(a_2, \ldots, a_m)) \\ &= \operatorname{lcm}(a_1, a_2 \cdot \cdots \cdot a_m) \\ &= a_1 \cdot \cdots \cdot a_m. \end{aligned}$$

The second claim is now immediate from Exercise 1.86. $\square$

1.8 Maximal and Prime Ideals

We will now discuss how special properties of an ideal I of a ring R correspond to properties of the residue class ring R/I.

Definition 1.90 An ideal I of the ring R is called **prime** if $I \neq R$ and $ab \in I$ implies $a \in I$ or $b \in I$ for all $a, b \in R$. I is called **maximal** if $I \neq R$ and for all ideals J of R, $I \subseteq J$ implies that $I = J$ or $J = R$ (i.e., I is proper but not properly contained in any proper ideal of R).

The following lemma and exercise provide examples. Recall that a prime number is an integer $p \geq 2$ with the property that if p divides a product of two integers, then it divides at least one of the factors.

Lemma 1.91 Let $2 \leq p \in \mathbb{Z}$. Then p is a prime number iff the ideal $p\mathbb{Z}$ of $\mathbb{Z}$ is prime.

Proof The definition of primeness of $p\mathbb{Z}$ is hardly more than a reformulation of primeness of p:

$$\begin{aligned} p \text{ prime} &\iff p \,|\, mn \text{ implies } p \,|\, m \text{ or } p \,|\, n \text{ for all } m, n \in \mathbb{Z} \\ &\iff mn \in p\mathbb{Z} \text{ implies } m \in p\mathbb{Z} \text{ or } n \in p\mathbb{Z} \text{ for all } m, n \in \mathbb{Z} \\ &\iff p\mathbb{Z} \text{ prime. } \square \end{aligned}$$

Exercise 1.92 Let I be an interval on the real line, $x_0 \in I$. Show that the ideal

$$\{ f \in \mathrm{C}(I, \mathbb{R}) \mid f(x_0) = 0 \}$$

of the ring $\mathrm{C}(I, \mathbb{R})$ is prime.

It will be shown below that in both cases above, the ideals are actually maximal. We first prove a lemma that is often used to verify that a given ideal is maximal.

Lemma 1.93 Let R be a ring, I a proper ideal of R. Then I is maximal iff for all $a \in R \setminus I$, there exists $r \in R$ and $b \in I$ such that $b + ar = 1$.

Proof "$\Longrightarrow$": Let $a \in R \setminus I$. Then the ideal $J = I + aR$ as defined in Exercise 1.41 (ii) satisfies $I \subseteq J$. We have $I \neq J$ since $a \in J \setminus I$, so it follows that $J = R$ by the maximality of I. In particular, $1 \in J$, so 1 can be written in the form $b + ar$ with $b \in I$ and $r \in R$.

"$\Longleftarrow$": Let J be an ideal of R with $I \subseteq J$. We must show that $I \neq J$ implies $J = R$. Now if $I \neq J$, then there is $a \in J \setminus I$. By assumption, we can find $b \in I$ and $r \in R$ with $1 = b + ar \in J$, hence $J = R$ by Lemma 1.39. □

Proposition 1.94 *Let I be a proper ideal of R. Then the following hold:*

(i) I is prime iff R/I is a domain.

(ii) I is maximal iff R/I is a field.

Proof We first note that for any proper ideal I of R and a, $b \in R$, we have $(a+I)(b+I) = 0$ iff $ab+I = 0$ iff $ab \in I$. Now R/I is an integral domain iff $(a+I)(b+I) = 0$ implies $a+I = 0$ or $b+I = 0$ for all a, $b \in R$. With the above observation, this translates into $ab \in I$ implies $a \in I$ or $b \in I$. I is maximal iff I and R are the only ideals that contain I iff $\{0+I\}$ and R/I are the only ideals of R/I (see Lemma 1.63) iff R/I is a field (see Exercise 1.40). □

Exercise 1.95 Show that the ideal of Exercise 1.92 is actually maximal. (Hint: Use Example 1.57.)

Corollary 1.96 *Every maximal ideal I of a ring R is prime.*

Proof If I is maximal, then R/I is a field and hence an integral domain by Lemma 1.19 (ii), and so I is prime. □

We will later see that the converse of the corollary above is not true in general. It does hold, however, in PID's.

Proposition 1.97 *Let R be a principal ideal domain. Then every non-trivial prime ideal is maximal.*

Proof Let aR be a prime ideal of R with $a \neq 0$, and let bR be an ideal with $aR \subseteq bR$. We want to show that $bR = aR$ or $bR = R$. From $aR \subseteq bR$ we conclude that $a \in bR$, hence $a = bc$ for some $c \in R$. Since aR is prime and $bc = a \in aR$, we must have $b \in aR$ or $c \in aR$. In the former case, $bR \subseteq aR$ and thus $bR = aR$. In the latter case, $c = ad$ for some $d \in R$, hence

$$a = bc = bad = abd$$

and thus $bd = 1$ by Lemma 1.19 (i). It follows that $1 \in bR$ and so $bR = R$ by Lemma 1.39. □

We can now combine Theorem 1.73, Lemma 1.91, Corollary 1.96, and Proposition 1.97 to obtain the following complete description of the ideals of $\mathbb{Z}$.

Proposition 1.98 *Every ideal of $\mathbb{Z}$ is principal. A non-trivial ideal I of $\mathbb{Z}$ is prime iff it is maximal iff it is of the form $p\mathbb{Z}$ for some prime number $p \in \mathbb{Z}$.* □

It is clear from the description of the ideals of $\mathbb{Z}_p$ given in Lemma 1.77 that $p\mathbb{Z}_p$ is the only maximal ideal of $\mathbb{Z}_p$. It is also the only non-trivial prime ideal, because $p^n\mathbb{Z}_p$ contains the product pp^{n-1} but none of the factors whenever $n > 1$.

Proposition 1.97 does not apply to $C(\boldsymbol{I}, \mathbb{R})$, because $C(\boldsymbol{I}, \mathbb{R})$ is not a PID. As a matter of fact, one can construct non-maximal prime ideals in $C(\boldsymbol{I}, \mathbb{R})$, but one needs set-theoretical techniques that are not available to us at this point. Examples of non-maximal prime ideals will occur naturally in the next chapter.

1.9 Prime Rings and Characteristic

We will now use the homomorphism theorem to obtain some structural results on rings.

Definition 1.99 Let R be a ring. For $a \in R$ and any natural number n with $n > 0$ we define

$$n \cdot a = \underbrace{a + a + \cdots + a}_{n \text{ times}},$$

and we set $0_{\mathbb{Z}} \cdot a = 0_R$. (The distinction between the two zero elements $0_{\mathbb{Z}}$ and 0_R will be suppressed in the sequel.) If there exists a natural number $n > 0$ with $n \cdot 1 = 0$, then we call the least such n the **characteristic** of R. If no such n exists, we say that the characteristic of R is 0. We will write $\text{char}(R)$ for the characteristic of R.

Obviously, the characteristics of $\mathbb{Z}$, $\mathbb{Z}_p$, and $\text{C}(\boldsymbol{I}, \mathbb{R})$ are 0. The characteristic of a ring cannot be 1 because we have required that $1 \neq 0$.

Exercise 1.100 Let $n \in \mathbb{Z}$ with $n > 1$. Show that $\text{char}(\mathbb{Z}/n\mathbb{Z}) = n$.

Lemma 1.101 Let R be a ring with $\text{char}(R) = n$, and let $m \in \mathbb{N}$. Then the folllowing hold:

(i) $m \in n\mathbb{Z}$ iff $m \cdot 1 = 0$.

(ii) If $a \in R$, then $m \in n\mathbb{Z}$ implies $m \cdot a = 0$.

(iii) If in addition, R is a domain and $0 \neq a \in R$, then $m \in n\mathbb{Z}$ iff $m \cdot a = 0$.

Proof (i) If $m \in n\mathbb{Z}$, then $m = qn$ for some integer q, and thus

$$m \cdot 1 = qn \cdot 1 = q \cdot (n \cdot 1) = q \cdot 0 = 0.$$

Now assume that $m \cdot 1 = 0$. If the characteristic n of R equals 0, then it follows immediately that $m = 0$. If $n > 0$, then there exist q, $r \in \mathbb{Z}$ with $m = qn + r$ and $0 \leq r < n$. We may conclude that

$$0 = (qn + r) \cdot 1 = (qn) \cdot 1 + r \cdot 1 = r \cdot 1,$$

and so $r = 0$ by the minimality of n.

(ii) If $m \in n\mathbb{Z}$ and $a \in R$, then $m \cdot a = a(m \cdot 1) = 0$ by (i) above.

(iii) In view of (ii), it remains to prove the implication from right to left. Assume that $m \cdot a = 0$. If the characteristic n of R equals 0, then we may argue that

$$0 = m \cdot a = a(m \cdot 1)$$

and so $m \cdot 1 = 0$ since $a \neq 0$ and R is a domain, and we see that $m = 0$. If $n > 0$, then we can once again find q, $r \in \mathbb{Z}$ with $m = qn + r$ and $0 \leq r < n$. We now write

$$0 = (qn + r) \cdot a = (qn) \cdot a + r \cdot a = r \cdot a = a(r \cdot 1)$$

and conclude from $a \neq 0$ that $r \cdot 1 = 0$. From the minimality of n, it now follows that $r = 0$. □

Exercise 1.102 Let R and S be rings with $\mathrm{char}(R) \neq 0$, and let $\varphi : R \longrightarrow S$ be a homomorphism of rings. Show that $\mathrm{char}(S)$ divides $\mathrm{char}(R)$.

The main result of this section states that every ring of characteristic 0 contains an isomorphic copy of $\mathbb{Z}$ as a subring, whereas a ring of characteristic n contains an isomorphic copy of $\mathbb{Z}/n\mathbb{Z}$. We extend the definition of $n \cdot a$ to all of $\mathbb{Z}$ in the obvious way by setting $n \cdot a = -((-n) \cdot a)$ for $n < 0$.

Proposition 1.103 *Let R be a ring. If* $\mathrm{char}(R) = 0$*, then the map $\varphi : \mathbb{Z} \longrightarrow R$ given by $\varphi(n) = n \cdot 1$ is an embedding of rings. If* $\mathrm{char}(R) = m$*, then the map $\psi : \mathbb{Z}/m\mathbb{Z} \longrightarrow R$ given by $\psi(n + m\mathbb{Z}) = n \cdot 1$ is an embedding of rings.*

Proof It is easy to see from the definition of $n \cdot 1$ that the map $\varphi : \mathbb{Z} \longrightarrow R$ given by $\varphi(n) = n \cdot 1$ is always a homomorphism. If $\mathrm{char}(R) = 0$, then $\varphi(n) \neq 0$ for all $n \neq 0$, hence $\ker(\varphi) = \{0\}$ and φ is an embedding. If $\mathrm{char}(R) = m$, then $\ker(\varphi) = m\mathbb{Z}$ by Lemma 1.101 (i). By Corollary 1.56, the map $\psi : \mathbb{Z}/m\mathbb{Z} \longrightarrow \varphi(\mathbb{Z})$ given by $\psi(n + m\mathbb{Z}) = \varphi(n) = n \cdot 1$ is an isomorphism, so if we regard it as a map from $\mathbb{Z}/m\mathbb{Z}$ to all of R, it becomes an embedding. □

It is an immediate consequence of the proposition above that a ring of characteristic zero must have infinitely many elements. The image of $\mathbb{Z}$ or $\mathbb{Z}/m\mathbb{Z}$, respectively, in R as described in the above proposition obviously consists of all sums $n \cdot 1$ in R, where $n \in \mathbb{Z}$. It is also called the **prime ring** of R. R itself is called a prime ring if it equals its own prime ring. We see that $\mathbb{Z}$ and its residue class rings are prime rings. If p is a prime number, then the field $\mathbb{Z}/p\mathbb{Z}$ is a also called the **prime field** of characteristic p.

Exercise 1.104 Let R be a ring. Show that the prime ring of R equals the intersection of all subrings of R.

Recall from Section 0.1 that a prime number is an integer p with the property that if p divides a product of two integers, then it divides at least one of the factors.

Proposition 1.105 *Let R be an integral domain. Then* $\mathrm{char}(R)$ *equals either 0 or some prime number p.*

Proof Assume for a contradiction that $\mathrm{char}(R) = n$, where $n \neq 0$ and n is not prime. Then $\mathbb{Z}/n\mathbb{Z}$ is not an integral domain by Lemma 1.91 and Proposition 1.94 and thus contains non-zero elements a and b with $ab = 0$. Now if $\varphi : \mathbb{Z}/n\mathbb{Z} \longrightarrow R$ is the embedding of the last proposition, then $\varphi(a) \cdot \varphi(b) = 0$ with $\varphi(a), \varphi(b) \neq 0$, a contradiction. □

If a is an element of any ring and $n \in \mathbb{N}$, then the notation a^n is rather self-evident: $a^0 = 1$, $a^1 = a$, and for $2 \leq n$, a^n is a product of n factors each of which equals a. The following lemma shows that the equation $(a+b)^2 = a^2 + b^2$ is not as ludicrous after all as we have been taught.

Lemma 1.106 Let R be a domain with $\operatorname{char}(R) = p \neq 0$, and let $a, b \in R$. Then $(a+b)^p = a^p + b^p$.

Proof By the binomial theorem, we have

$$(a+b)^p = a^p + \binom{p}{1} \cdot a^{p-1}b + \cdots + \binom{p}{p-1} \cdot ab^{p-1} + b^p.$$

(We have not proved the binomial theorem for rings in general, but any one of the proofs given in elementary mathematics for the reals carries over verbatim.) The binomial coefficients are defined as

$$\binom{p}{i} = \frac{(p-i+1)(p-i+2)\cdot \cdots \cdot p}{1 \cdot 2 \cdot \cdots \cdot i}$$

for $0 < i < p$. The numerator contains the factor p which does not divide any one of the factors in the denominator since $i < p$. But p is a prime number and thus does not divide the entire denominator. We see that p cannot be canceled, so all binomial coefficients above are multiples of the characteristic p. By Lemma 1.101 (ii), it follows that all summands in the expansion of $(a+b)^p$ vanish except for a^p and b^p. □

Recall that a finite domain is automatically a field. It is clear that a finite field cannot have characteristic zero, since in the latter case the elements $n \cdot 1$ with $n \in \mathbb{N}$ are all different.

Lemma 1.107 Let K be a finite field with $\operatorname{char}(K) = p$. Then every element of K has a pth root, i.e., for all $a \in K$, there exists $b \in K$ with $b^p = a$.

Proof We claim that the map $\varphi : K \longrightarrow K$ defined by $\varphi(a) = a^p$ is injective. Indeed, $a^p = b^p$ implies $a^p - b^p = (a-b)^p = 0$ by Lemma 1.106, and so $a - b = 0$ since K, being a field, has no zero divisors. K was assumed to be finite, so by Proposition 0.23, φ is surjective, which is what we have claimed. □

Exercise 1.108 (Fermat's Theorem) Show that if p is a prime number, then $a^p \equiv a \bmod p$ for all $a \in \mathbb{Z}$. (Hint: Argue that it suffices to prove the claim for $a \in \mathbb{N}$, then use induction on a. See the end of Section 1.5 for an explanation of the congruence notation.)

1.10 Adjunction, Products, and Quotient Rings

In this section, we discuss three constructions of rings from given ones. We begin with *ring adjunction.*

Definition 1.109 Let S be a ring, R a subring of S, and $M \subseteq S$. Then we define $R[M]$, the ring obtained by **adjunction** of M to R, as the intersection of all subrings of S that contain both R and M. If $M = \{m_1, \ldots, m_n\}$, then we write $R[m_1, \ldots, m_n]$ instead of $R[\{m_1, \ldots, m_n\}]$. In this case, $R[M]$ is also called a **finitely generated** extension ring of R.

Note that in the above definition, the set of all subrings of S that contain both R and M is not empty since S itself is such a subring, and that $R[M]$ is a subring of S by Exercise 1.25. We obviously have R, $M \subseteq R[M]$. It is important to note that although S is not part of the notation, the ring $R[M]$ depends not just on the elements of M, but on the structure of S as well, i.e., on the definition of the operations in S. It is clear that $R[\emptyset] = R$, so that the concept of adjunction is interesting only for non-empty M. $R[M]$ can be described explicitly as follows.

Lemma 1.110 Let S, R, $M = \{m_1, \ldots, m_n\}$ be as in the definition above, and set

$$T = \{\, m_1^{\nu_1} \cdot \cdots \cdot m_n^{\nu_n} \mid \nu_1, \ldots\ , \nu_n \in \mathbb{N} \,\}.$$

Then $R[M]$ consists of all sums of the form $\sum_{i=1}^{k} r_i t_i$, where $r_i \in R$ and $t_i \in T$ for $1 \leq i \leq k$.

Proof Let B be the set of all such sums. Since subrings are closed under addition and multiplication, every subring of S that contains R and M must contain B, hence B is contained in their intersection $R[M]$. To prove the reverse inclusion, we note that B itself is such a subring. □

Examples 1.111 (i) Let $S = \mathbb{C}$, $R = \mathbb{Z}$, and $M = \{i\sqrt{5}\}$. It is then easy to see that every element of $R[i\sqrt{5}]$ can be simplified to the form $a + ib\sqrt{5}$ with a, $b \in \mathbb{Z}$. We see that what we obtain is the subring D of $\mathbb{C}$ introduced in 1.24.

(ii) If $S = \mathbb{Q}$, $R = \mathbb{Z}$, p is a fixed prime number, and

$$M = \{\, 1/n \mid n \in \mathbb{N},\ p \text{ does not divide } n \,\},$$

then it is not hard to see that $R[M] = \mathbb{Z}_p$.

The following lemma is often useful.

Lemma 1.112 Let R and S be rings, M_1, $M_2 \subseteq S$. Then the following hold:

(i) $R[M_1][M_2] = R[M_1 \cup M_2]$.

(ii) If $M_1 \subseteq M_2$, then $R[M_1] \subseteq R[M_2]$.

Proof (i) It suffices to show that the set of all subrings S' of S that contain both $R[M_1]$ and M_2 equals the set of those subrings that contain R and $M_1 \cup M_2$. We know that R, $M_1 \subseteq R[M_1]$, so if S' contains $R[M_1]$ and M_2, then it contains R, M_1, and M_2, and thus R and $M_1 \cup M_2$. Conversely, if S' contains R and $M_1 \cup M_2$, then it contains R and M_1 and thus $R[M_1]$, and it also contains M_2.

(ii) Here, it suffices to note that every subring of S that contains R and M_2 contains R and M_1, so the set of subrings of S whose intersection is $R[M_2]$ is a subset of the set of those whose intersection is $R[M_1]$. □

The second construction that we discuss is that of *direct products.* The proof of the following proposition is left to the reader as an easy though slightly tedious exercise.

Proposition 1.113 *Let $R_1, \ldots, R_n$ be rings, and let R be the set of all n-tuples $(a_1, \ldots, a_n)$ such that $a_i \in R_i$ for $1 \le i \le n$ with the following operations:*

(i) $(a_1, \ldots, a_n) + (b_1, \ldots, b_n) = (a_1 + b_1, \ldots, a_n + b_n)$, *and*

(ii) $(a_1, \ldots, a_n) \cdot (b_1, \ldots, b_n) = (a_1 b_1, \ldots, a_n b_n)$.

Then R is a ring whose zero is $(0, \ldots, 0)$ and whose unity is $(1, \ldots, 1)$. □

R as defined above is called the **(finite) direct product** of $R_1, \ldots, R_n$ and is denoted by $\prod_{i=1}^n R_i$, or by $R_1 \times \cdots \times R_n$. An n-fold direct product $\prod_{i=1}^n R$ of a ring with itself is also denoted by R^n.

Exercises 1.114 (i) What is the negative of $(a_1, \ldots, a_n)$ in a direct product $R = \prod_{i=1}^n R_i$ of rings? When is $(a_1, \ldots, a_n)$ a unit in R?

(ii) Show that the direct product of two integral domains is not an integral domain.

Products of rings have the following universal property.

Lemma 1.115 Let $R_1, \ldots, R_n$ be rings. Then the following hold:

(i) For each $1 \le i \le n$, the map

$$\begin{array}{rccc} \pi_i : & R_1 \times \cdots \times R_n & \longrightarrow & R_i \\ & (a_1, \ldots, a_n) & \longmapsto & a_i \end{array}$$

is a surjective homomorphism of rings. π_i is called the ith **projection** of the product $R_1 \times \cdots \times R_n$.

(ii) Whenever R is a ring and $\varphi_i : R \longrightarrow R_i$ is a homomorphism of rings for $1 \le i \le n$, then the map

$$\begin{array}{rccc} \varphi : & R & \longrightarrow & R_1 \times \cdots \times R_n \\ & a & \longmapsto & (\varphi_1(a), \ldots, \varphi_n(a)) \end{array}$$

is a homomorphism of rings that satisfies $\pi_i \circ \varphi = \varphi_i$ for $1 \le i \le n$. Moreover, the kernel of φ equals the intersection of the kernels of the homomorphisms φ_i.

Proof The proof of (i) is a straightforward verification of the homomorphism properties. It is equally easy to see that φ of (ii) is a homomorphism of rings, and the equation $\pi_i \circ \varphi = \varphi_i$ can be read off directly from the definitions. Finally, we have, for all $a \in R$,

$$\begin{aligned} a \in \ker(\varphi) &\iff \varphi(a) = (0, \ldots, 0) \\ &\iff \varphi_i(a) = 0 \text{ for } 1 \le i \le n \\ &\iff a \in \ker(\varphi_i) \text{ for } 1 \le i \le n \\ &\iff a \in \bigcap_{i=1}^{n} \ker(\varphi_i). \ \square \end{aligned}$$

The following lemma is obtained by specializing (ii) of the above lemma to the case where each φ_i is a canonical homomorphism from R to a residue class ring of R modulo a proper ideal of R.

Lemma 1.116 Let R be a ring and $I_1, \ldots, I_n$ proper ideals of R. Then

$$\varphi : R \longrightarrow \prod_{i=1}^{n} R/I_i$$

defined by $\varphi(a) = (a + I_1, \ldots, a + I_n)$ is a homomorphism of rings whose kernel equals $\bigcap_{i=1}^{n} I_i$. □

Finally, we discuss the construction of *quotient rings*. This is easy to understand if one recalls the construction of the rationals from the integers, which can be described informally as follows. One considers pairs (s, t) of integers with $t \neq 0$. Two such pairs (s, t), (q, r) are then considered to be equal iff $sr = qt$. (Formally, this identification is being made by means of an *equivalence relation*, but it turns out to be more convenient for practical purposes to just *consider* such pairs as equal.) With the notation s/t for (s, t) one then defines

$$(s/t)(q/r) = (st/qr) \quad \text{and} \quad (s/t) + (q/r) = (sr + tq)/tr$$

and shows that these definitions are consistent with the above identification of certain pairs. Under these operations, the set of all such fractions becomes a field $\mathbb{Q}$ whose zero element is $0/1$ and whose 1 is $1/1$. The inverse of an element $s/t \neq 0$ of $\mathbb{Q}$ is t/s. The ring $\mathbb{Z}$ can be embedded into $\mathbb{Q}$ by mapping $n \in \mathbb{Z}$ to the fraction $n/1$, i.e., $\mathbb{Z}$ is a subring of $\mathbb{Q}$ if one writes s for $s/1$. A formal proof of these facts is tedious but straightforward. Looking at such a proof, the reader will find that the exact same procedure goes through if one starts with any integral domain R instead of $\mathbb{Z}$. Moreover, all arguments continue to hold if one does not consider all fractions s/t with $t \neq 0$, but only those with $t \in M$ where M is a given subset of R that is closed under multiplication and satisfies $1 \in M$ and $0 \notin M$. The original proof is then just the special case $M = R \setminus \{0\}$. (This is a multiplicatively closed set since R is an integral domain.) The only difference is that for general M, the result of the construction is only an integral domain and not necessarily a field. This will be illustrated by an example below. Other than that, we will forego the proof entirely because the reader may in good conscience rely on the intuitive understanding of rational numbers when working with such general rings of quotients.

Theorem 1.117 *Let R be an integral domain, M a subset of R that is closed under multiplication, with $1 \in M$ and $0 \notin M$. Let R_M be the set of all formal fractions s/t where $s \in R$ and $t \in M$, with s/t, q/r considered equal iff $sr = qt$. Then the operations $s/t \cdot q/r = sq/tr$ and $s/t + q/r = (sr + qt)/tr$ are well-defined, and R_M becomes an integral domain under these operations whose zero is $0/1$ and whose unity is $1/1$. R can be embedded into R_M by mapping $a \in R$ to $a/1$, i.e., R_M contains R if we write a for $a/1$. If we choose $M = R \setminus \{0\}$, then R_M becomes a field. In this case, $(s/t)^{-1} = t/s$ whenever s, $t \neq 0$.*

Definition 1.118 R_M as described above is called the **ring of quotients** of R w.r.t. M. If $M = R \setminus \{0\}$, then R_M is called the **field of quotients**, or **quotient field**, or **field of fractions**, of R and is denoted by Q_R.

It is easy to see that in R_M, $t/t = 1/1 = 1$ for all $t \in M$, and that for all $s \in R$ and $t \in M$, we have $s/t = 0$ iff $s = 0$.

Examples 1.119 (i) If $R = \mathbb{Z}$ and $M = \mathbb{Z} \setminus \{0\}$, then $R_M = \mathbb{Q}$, i.e., we have $Q_{\mathbb{Z}} = \mathbb{Q}$.

(ii) If $R = \mathbb{Z}$, p is a fixed prime number and $M = \{m \in \mathbb{Z} \mid p \text{ does not divide } m\}$, then $R_M = \mathbb{Z}_p$. Here, R_M is not a field since $p/1$ is not invertible.

(iii) More generally, let R be any integral domain, I a prime ideal of R. Then a, $b \notin I$ implies $ab \notin I$ by primeness of I. Hence the set $M = R \setminus I$ is multiplicatively closed. Moreover, $0 \notin M$ since $0 \in I$, and $1 \in M$ since I is proper. Hence we may form R_M, which consists of all fractions s/t with s, $t \in R$ and $t \notin I$. In this case, R_M is called the **localization** of R at I. By an abuse of notation, this is sometimes also denoted by R_I, or even by R_p if $I = (p)$. We see that for any prime number p, $\mathbb{Z}_p$ is the localization of $\mathbb{Z}$ at $p\mathbb{Z}$.

Exercises 1.120 (i) Let R be a ring, M a multiplicatively closed subset of R with $1 \in M$ and $0 \notin M$, and let s, $t \in M$. Show that $s/t \in R_M$ is a unit of R_M. Give a counterexample for the converse, i.e., an example where $s \notin M$ and yet s/t is invertible in R_M. (Hint: Take $R = \mathbb{Z}$ and $M = \mathbb{Z}^+$.)

(ii) Let R be a ring and I a prime ideal of R. Show that the set of all non-units forms an ideal J of R_I, and that J is a maximal ideal which contains every proper ideal of R_I.

The above exercise explains the term *localization*: a ring is called **local** if it has exactly one maximal ideal, and we just saw that for any ring R and a prime ideal I of R, R_I is a local ring.

Rings of quotients have the following universal embedding property.

Lemma 1.121 Let R and S be integral domains, M a multiplicatively closed subset of R with $1 \in M$ and $0 \notin M$. Let $\varphi : R \longrightarrow S$ be an embedding of rings such that every element of $\varphi(M)$ is a unit of S. Then there exists a unique embedding $\overline{\varphi} : R_M \longrightarrow S$ with $\overline{\varphi} \upharpoonright R = \varphi$.

$$\begin{tikzcd}
R \arrow[r, "\varphi"] \arrow[d, phantom, "\rotatebox[origin=c]{-90}{$\subseteq$}"] & S \\
R_M \arrow[ur, "\overline{\varphi}"'] &
\end{tikzcd}$$

In particular, every embedding of an integral domain into a field extends uniquely to an embedding of Q_R into that field.

Proof We define

$$\overline{\varphi}(s/t) = \varphi(s) \cdot \left(\varphi(t)\right)^{-1} \qquad (s \in R,\ t \in M).$$

We claim that then $\overline{\varphi}$ is well defined. Indeed, let $s/t = q/r$ in R_M. Then $sr = qt$, hence $\varphi(s) \cdot \varphi(r) = \varphi(q) \cdot \varphi(t)$, and so

$$\varphi(s) \cdot (\varphi(t))^{-1} = \varphi(q) \cdot (\varphi(r))^{-1}.$$

It is now a straightforward exercise to prove that $\overline{\varphi}$ is a homomorphism of rings. It is clear that $\overline{\varphi}$ extends φ:

$$\overline{\varphi}(a) = \overline{\varphi}(a/1) = \varphi(a) \cdot (\varphi(1))^{-1} = \varphi(a).$$

To see that $\overline{\varphi}$ is an embedding, assume that $\overline{\varphi}(s/t) = 0$. Then $\varphi(s) = 0$, hence $s = 0$ since φ was an embedding, and thus $s/t = 0$. Finally, let $\psi : R_M \longrightarrow S$ be another embedding of rings with $\psi \upharpoonright R = \varphi$. Then we must have, for all $t \in M$,

$$1_S = \psi(1_R) = \psi(t \cdot (1/t)) = \psi(t) \cdot \psi(1/t) = \varphi(t) \cdot \psi(1/t),$$

and thus $\psi(1/t) = (\varphi(t))^{-1}$. So whenever $s/t \in R_M$, we have

$$\psi(s/t) = \psi(s)\psi(1/t) = \varphi(s) \cdot (\varphi(t))^{-1} = \overline{\varphi}(s/t). \quad \square$$

For the rest of this section, let R be a ring, M a multiplicatively closed subset of R with $1 \in M$ and $0 \notin M$. We want to investigate the behavior of ideals under the passage from R to R_M and vice versa. If I is an ideal of R, then the **extension** I^{e} of I to R_M is the ideal generated by the set I in the ring R_M. If J is an ideal of R_M, then the **contraction** J^{c} of J to R is defined as $J \cap R$. For the rest of this section, all extensions will be to R_M, and all contractions will be to R. The following trivial observation will be used repeatedly: if $a \in R_M$, then there exists $s \in M$ with $sa \in R$.

Lemma 1.122 Let I be an ideal of R. Then the following hold:

(i) $I^{\mathrm{e}} = \{ a/s \mid a \in I,\ s \in M \}$.

(ii) I^{e} is proper iff $I \cap M = \emptyset$.

(iii) $I \subseteq I^{\mathrm{ec}}$.

(iv) If I is prime and $I \cap M = \emptyset$, then I^{e} is prime and $I = I^{\mathrm{ec}}$.

Proof (i) The inclusion "$\supseteq$" is trivial. Now let $b \in I^{\mathrm{e}}$. Then there exist $r_1, \ldots, r_n \in R_M$ and $a_1, \ldots, a_n \in I$ such that $b = \sum_{i=1}^{n} r_i a_i$. Let $s_1, \ldots, s_n \in M$ with $s_i r_i \in R$, and set $s = s_1 \cdot \cdots \cdot s_n$. Then

$$b = \frac{1}{s} \sum_{i=1}^{n} s r_i a_i ,$$

and this is of the desired form.

(ii) If $s \in I \cap M$, then $1 = s/s \in I^{\mathrm{e}}$ and so $I^{\mathrm{e}} = R_M$. Conversely, if $I^{\mathrm{e}} = R_M$, then $1 \in I^{\mathrm{e}}$ and so there exist $s \in M$ and $a \in I$ with $1 = a/s$. We see that $a = s \in I \cap M$.

(iii) This is immediate from the definitions of extension and contraction.

(iv) Assume that I is a prime ideal of R. We already know that $I \subseteq I^{\mathrm{ec}}$. Now let $b \in I^{\mathrm{ec}}$. By (i) above, there exists $s \in M$ with $sb \in I$. Since $s \notin I$ and I is prime, we must have $b \in I$. To show that I^{e} is prime, we let a, $b \in R_M$ with $ab \in I^{\mathrm{e}}$. There are r, $s \in M$ with ra, $sb \in R$ and thus

$$(ra)(sb) = (rs)(ab) \in I^{\mathrm{ec}} = I.$$

It follows that $ra \in I$ or $sb \in I$, and so $a = (ra)/r \in I^{\mathrm{e}}$ or $b = (sb)/s \in I^{\mathrm{e}}$. □

Lemma 1.123 Let J be an ideal of R_M. Then the following hold:

(i) J^{c} is an ideal of R.

(ii) If J is proper, then $J^{\mathrm{c}} \cap M = \emptyset$.

(iii) If J is prime, then so is J^{c}.

(iv) $J = J^{\mathrm{ce}}$.

Proof (i) Both J and R are closed under addition and under multiplication with elements of R, so the same is true for their intersection J^{c}. Moreover, J^{c} is not empty because $0 \in J \cap R$.

(ii) If J is a proper ideal of R_M, then it does not contain any units of R_M; in particular, it does not contain any elements of M.

(iii) Assume that J is a prime ideal of R_M. Then J^{c} is proper by (ii) above. To see that J^{c} is a prime ideal of R, we let $a, b \in R$ with $ab \in J^{\mathrm{c}}$. Since $J^{\mathrm{c}} \subseteq J$ and J is prime, we must have $a \in J$ or $b \in J$ and hence $a \in J^{\mathrm{c}}$ or $b \in J^{\mathrm{c}}$.

(iv) Let $b \in J$ and $s \in M$ with $sb \in R$. Then $sb \in J \cap R = J^{\mathrm{c}}$ and thus $b = (sb)/s \in J^{\mathrm{ce}}$. Conversely, let $b \in J^{\mathrm{ce}}$. Then $b = a/s$ with $a \in J^{\mathrm{c}} \subseteq J$ and $s \in M$, and so

$$b = a/s = (1/s)a \in J. \quad \square$$

Lemma 1.124 Denote by $\mathcal{I}(R)$ and $\mathcal{I}(R_M)$ the set of all ideals of R and R_M, respectively. Then the following hold:

(i) The map $\varphi : \mathcal{I}(R) \longrightarrow \mathcal{I}(R_M)$ defined by $\varphi(I) = I^{\mathrm{e}}$ for all $I \in \mathcal{I}(R)$ is surjective.

(ii) The map $\psi : \mathcal{I}(R_M) \longrightarrow \mathcal{I}(R)$ defined by $\varphi(J) = J^{\mathrm{c}}$ for all $J \in \mathcal{I}(R_M)$ is injective.

Proof To prove (i), let $J \in \mathcal{I}(R_M)$. Then by (iv) of the previous lemma, $J = \varphi(I)$ for $I = J^{\mathrm{c}}$. For the proof of (ii), suppose $J_1, J_2 \in \mathcal{I}(R_M)$ with $\psi(J_1) = \psi(J_2)$. Then again by (iv) of the previous lemma,

$$J_1 = J_1^{\mathrm{ce}} = \left(\psi(J_1)\right)^{\mathrm{e}} = \left(\psi(J_2)\right)^{\mathrm{e}} = J_2^{\mathrm{ce}} = J_2. \quad \square$$

Exercise 1.125 Show that the map $J \longmapsto J^{\mathrm{c}}$ is a bijection between the set of all prime ideals of R_M and the set of all those prime ideals of R that do not intersect M. What is the inverse of this map?

Notes

Well into the 19th century, the subject matter of algebra was the search for algorithmic solutions of algebraic equations in number systems such as $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, or $\mathbb{C}$. Within a period of about 50 years around the turn of the 19th century, algebra gradually changed its appearance and turned into a theory of algebraic structures such as groups, rings, and fields. This transition occurred for a variety of reasons. One of them was certainly the elegance of this axiomatic approach, which clarified the foundations such as the nature of an algebraic object and distilled from the traditional proofs the essence of the arguments. This helped to economize algebraic research by avoiding repetition of similar arguments in different contexts.

The concept of an abstract group was developed to various degrees of generality in papers by Cayley, Dedekind, and Kronecker, starting around 1850. Abstract fields first appear in the work of Dedekind and Weber who named them "Körper" (German for "body"), and of Kronecker who used the term "Rationalitätsbereich" (German for "rational domain"). A landmark in the study of the structure of fields is Steinitz's *Algebraische Theorie der Körper* (1910). The structure of finite fields had already been investigated in 1830 by Evariste Galois.

The concept of an ideal was introduced by Dedekind as a set theoretic version of Kummer's "ideal number," which was invented in order to circumvent the failure of unique factorization in certain natural extensions of the domain $\mathbb{Z}$ (The domain D of Exercise 1.24 is a case in point.) The relevance of ideals in the theory of polynomial rings was highlighted by the Hilbert basis theorem (cf. the discussion in the Notes to Chapter 4 on p. 183). The systematic development of ideal theory in more general rings is largely due to E. Noether. In the older literature the term "module" is sometimes used for "ideal" (cf. Macaulay, 1916). The term "ring" seems to be due to D. Hilbert; Kronecker used the term "order" for ring.

2

Polynomial Rings

In this chapter, we will define and investigate polynomials, the main object of study in this book. We will occasionally find it convenient to work on the higher level of abstraction of general ring theory, but the focus remains on polynomial rings. Only Sections 1 and 2 of this chapter are directly relevant for the theory of Gröbner bases. We will also discuss a number of algorithms, such as greatest common divisor or factorization, which are often used in connection with Gröbner basis techniques. When talking about these algorithms, we will *not* attempt to give fast, up-to-date versions. We will be content to give a method for solving the respective problem in finitely many steps. The reader who has a further interest in these algorithms is thus provided with the theoretical background to proceed to the advanced literature.

2.1 Definitions

To arrive at a rigorous and sufficiently general definition of polynomials we first remind the reader of a familiar special case: polynomials in one variable with real coefficients. Such a polynomial is usually written in the form

$$f = \sum_{i=0}^{m} a_i X^i$$

with $a_i \in \mathbb{R}$ for $0 \leq i \leq m$. Clearly f is uniquely determined by the a_i. We can thus think of f as being given by a sequence $\{a_i\}_{i \in \mathbb{N}}$ of real numbers where $a_i = 0$ for all but finitely many $i \in \mathbb{N}$. But such a sequence is nothing but a function $F : \mathbb{N} \longrightarrow \mathbb{R}$, with a_i being the function value $F(i)$. We will now generalize this in two directions. Firstly, the reals will be replaced by an arbitrary ring R, which does not cause any problems in the definition. Secondly, we wish to allow several variables, i.e., we need coefficients not just for powers X^i of X, but for power products of variables, such as $X_1^{\nu_1} \cdot \cdots \cdot X_n^{\nu_n}$. The function $F : \mathbb{N} \longrightarrow \mathbb{R}$ will thus have to be replaced by a function $\mathbb{N}^n \longrightarrow R$ that assigns a coefficient in the ring R to each n-tuple $(\nu_1, \ldots, \nu_n)$, where the latter represents in a mathematically sound way the power product $X_1^{\nu_1} \cdot \cdots \cdot X_n^{\nu_n}$.

Exercise 2.1 Imagine the mathematical definition of polynomials in the variables $X_1, \ldots, X_n$ over $\mathbb{Q}$ as "coefficient functions" from $\mathbb{N}^n$ to $\mathbb{Q}$ has been achieved. Which function do you think will correspond to the constant polynomial c (with $c \in \mathbb{Q}$), and which one to a plain variable X_i? Which one will bear the name of the monomial $cX_1^{\nu_1} \cdot \cdots \cdot X_n^{\nu_n}$?

The structure on the set $\mathbb{N}^n$ that is relevant here is that of a *monoid.*

Definition 2.2 A **monoid** is a set M together with a binary operation "$\cdot$" and a distinguished element $1 \in M$ such that the following hold:

(i) "$\cdot$" is associative, i.e., $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all a, b, $c \in M$.

(ii) $1 \cdot a = a \cdot 1 = a$ for all $a \in M$.

The distinguished element $1 \in M$ is also referred to as the **neutral element** of M. M is called **Abelian,** or **commutative,** if in addition, "$\cdot$" is commutative, i.e., $a \cdot b = b \cdot a$ for all a, $b \in M$. We will also write ab instead of $a \cdot b$.

Examples 2.3 (i) Every group is a monoid, every Abelian group is an Abelian monoid. In particular, every ring is an Abelian monoid under addition.

(ii) Every ring is an Abelian monoid under multiplication. (Recall that by "ring" we always mean "commutative ring with 1.")

(iii) Let $1 \leq n \in \mathbb{N}$. Then it is easy to verify that the set $\mathbb{N}^n$ of all n-tuples of natural numbers with the operation of componentwise addition, where

$$(\nu_1, \ldots, \nu_n) + (\mu_1, \ldots, \mu_n) = (\nu_1 + \mu_1, \ldots, \nu_n + \mu_n)$$

is an Abelian monoid with neutral element $(0) = (0, \ldots, 0)$. We will call this monoid the **additive monoid** $\mathbb{N}^n$.

We mention that $\mathbb{N}^n$ can also be turned into a monoid by taking componentwise multiplication as the operation and $(1, \ldots, 1)$ as the distinguished element, but this structure is of no relevance to us. To avoid confusion in this respect, we will often write $(M, 1, \cdot)$ for the monoid M with operation "$\cdot$" and neutral element 1. The additive monoid $\mathbb{N}^n$ thus becomes $(\mathbb{N}^n, (0), +)$.

A **homomorphism** from a monoid M to a monoid N is a map $\varphi : M \longrightarrow N$ with the following two properties:

(i) $\varphi(a)\varphi(b) = \varphi(ab)$ for all a, $b \in M$, and

(ii) $\varphi(1_M) = 1_N$.

Condition (ii) above is not redundant: the map φ of Exercise 1.52 provides a counterexample when the rings involved are viewed as just multiplicative monoids. A homomorphism φ of monoids is called an **embedding** (of monoids) if it is injective, an **isomorphism** (of monoids) if it is bijective. The following lemma, whose proof is straightforward from the definitions, provides an example that will be useful later on.

Lemma 2.4 Let S be a ring and $c_1, \ldots, c_n \in S$, where $1 \leq n \in \mathbb{N}$. Then the map

$$\begin{array}{cccc} \sigma : & \mathbb{N}^n & \longrightarrow & S \\ & (\nu_1, \ldots, \nu_n) & \longmapsto & c_1^{\nu_1} \cdot \cdots \cdot c_n^{\nu_n} \end{array}$$

is a homomorphism from $(\mathbb{N}^n, (0), +)$ to $(S, 1, \cdot)$. □

From now on, let R be a ring and M an Abelian monoid. If f is a function from M to R, then the **support** of f is defined as

$$\operatorname{supp}(f) = \{\, u \in M \mid f(u) \neq 0 \,\}.$$

Denote by RM the set of all functions $f : M \longrightarrow R$ with finite support, i.e., $f(u) \neq 0$ for only finitely many $u \in M$. Define an addition and a multiplication on RM by setting, for f, $g \in M$ and all $u \in M$,

$$\begin{aligned} (f+g)(u) &= f(u) + g(u), \quad \text{and} \\ (fg)(u) &= \sum_{\substack{v,w \in M \\ vw=u}} f(v)g(w). \end{aligned}$$

Note that there are only finitely many non-zero summands in the second sum since there are only finitely many v, $w \in M$ such that $f(v)g(w) \neq 0$. What we really mean here is the sum over all non-zero summands; it will often be convenient to work with sums that may be formally infinite. Both $f + g$ and fg are again in RM, i.e., take non-zero value for only finitely many $u \in M$:

$$\begin{aligned} &\operatorname{supp}(f+g) \subseteq \operatorname{supp}(f) \cup \operatorname{supp}(g), \quad \text{and} \\ &\quad \operatorname{supp}(fg) \subseteq \{\, u \in M \mid u = vw \text{ with } v \in \operatorname{supp}(f),\ w \in \operatorname{supp}(g) \,\}, \end{aligned}$$

and both sets on the right-hand side are finite since f, $g \in RM$.

Proposition 2.5 *RM as defined above is a ring whose 0 is the function f that satisfies $f(u) = 0$ for all $u \in M$, and whose 1 is the function defined by*

$$1_{RM}(u) = \begin{cases} 1 & \text{if} \quad u = 1_M \\ 0 & \text{otherwise.} \end{cases}$$

Proof We show associativity of "$+$" and the distributive law and leave verification of the remaining axioms as exercises. Let f, g, $h \in RM$. Then, for all $u \in M$,

$$\begin{aligned}\big((f+g)+h\big)(u) &= (f+g)(u)+h(u)\\ &= \big(f(u)+g(u)\big)+h(u)\\ &= f(u)+\big(g(u)+h(u)\big)\\ &= f(u)+(g+h)(u)\\ &= \big(f+(g+h)\big)(u),\end{aligned}$$

and thus $(f+g)+h = f+(g+h)$. Again, for all $u \in M$,

$$\begin{aligned}\big(f(g+h)\big)(u) &= \sum_{vw=u} f(v)(g+h)(w)\\ &= \sum_{vw=u} f(v)\big(g(w)+h(w)\big)\\ &= \sum_{vw=u} f(v)g(w)+f(v)h(w)\\ &= \sum_{vw=u} f(v)g(w)+\sum_{vw=u} f(v)h(w)\\ &= (fg)(u)+(fh)(u)\\ &= (fg+fh)(u). \ \square\end{aligned}$$

Exercise 2.6 Complete the proof of the above proposition.

Definition 2.7 The ring RM defined above is called the **monoid ring** over R and M. If M is the additive monoid $\mathbb{N}^n$, then it is called the **polynomial ring** in n variables over R. The elements of the polynomial ring are called **polynomials**. The polynomial ring and its elements are called **univariate** if $n = 1$, **multivariate** otherwise.

A function $f \in RM$ is called a **monomial** if it has only one non-zero value, i.e., $\operatorname{supp}(f)$ is a singleton $\{u\}$ with $u \in M$.

Lemma 2.8 Let $f \in RM$. Then f has a unique representation as a sum of monomials with pairwise different support which is given by

$$f = \sum_{u \in \operatorname{supp}(f)} f_u, \qquad \text{where} \quad f_u(v) = \begin{cases} f(u) & \text{if} \quad v = u\\ 0 & \text{otherwise,}\end{cases}$$

with the understanding that the empty sum is the zero element of RM.

Proof We proceed by induction on the number $k = |\operatorname{supp}(f)|$ of elements of $\operatorname{supp}(f)$. If $k = 0$, then $f = 0$ and the indicated sum is empty. If $0 < k$, then we may choose $v \in \operatorname{supp}(f)$ and consider $g = f - f_v$. It is easy to see from the definition of addition in RM that $\operatorname{supp}(g) = \operatorname{supp}(f) \setminus \{v\}$ and $f(u) = g(u)$ for all $u \in \operatorname{supp}(g)$. We may thus apply the induction hypothesis to g to obtain

$$f = f_v + g = f_v + \sum_{u \in \operatorname{supp}(g)} g_u = \sum_{u \in \operatorname{supp}(f)} f_u.$$

To prove uniqueness, let N be a finite subset of M and $g_u \in RM$ monomials, one for each $u \in N$, such that

$$f = \sum_{u \in N} g_u.$$

It is now an immediate consequence of the definition of addition in RM that $N = \operatorname{supp}(f)$, and $g_u = f_u$ for all $u \in N$. □

If one insists that a summation symbol always implies a certain ordering of the summands, then the representation of the lemma above is unique only up to the order of the summands. In the presence of commutativity, however, there is no harm in allowing the "unordered sum" of a finite set of ring elements. As an unordered sum, the representation of the lemma is uniquely determined by f.

Next, we show that both R and M can be embedded into RM in a natural way. A ring element a will be sent to the function that says a at 1_M and 0 otherwise, while an element u of M will be sent to the "characterisic function of $\{u\}$," which says 1_R at u and zero otherwise.

Lemma 2.9 Define a map $\iota : R \longrightarrow RM$ by setting, for all $a \in R$ and $u \in M$,

$$\iota(a)(u) = \begin{cases} a & \text{if} \quad u = 1_M \\ 0 & \text{otherwise.} \end{cases}$$

Then ι is an embedding of rings.

Proof Let $a, b \in R$. Then we have, for all $u \in M$

$$\begin{aligned} \big(\iota(a) + \iota(b)\big)(u) &= \iota(a)(u) + \iota(b)(u) \\ &= \begin{cases} a+b & \text{if} \quad u = 1_M \\ 0 & \text{otherwise} \end{cases} \\ &= \iota(a+b)(u), \end{aligned}$$

and

$$\begin{aligned} \big(\iota(a) \cdot \iota(b)\big)(u) &= \sum_{\substack{v,w \in M \\ vw = u}} \iota(a)(v) \cdot \iota(b)(w) \\ &= \begin{cases} ab & \text{if} \quad u = 1_M \\ 0 & \text{otherwise} \end{cases} \\ &= \iota(ab)(u), \end{aligned}$$

the second equality being true because the only way to obtain a non-trivial summand is to have $u = v = w = 1_M$. Finally, it is obvious that $\iota(1_R) = 1_{RM}$, and that ι is injective. □

Lemma 2.10 Define a map $\eta : M \longrightarrow RM$ by setting, for all $u, w \in M$,

$$\eta(u)(w) = \begin{cases} 1_R & \text{if} \quad w = u \\ 0 & \text{otherwise.} \end{cases}$$

Then η is an embedding of $(M, 1_M, \cdot)$ into $(RM, 1_{RM}, \cdot)$.

Proof Let u, $v \in M$. Then we have

$$\begin{aligned} (\eta(u) \cdot \eta(v))(w) &= \sum_{\substack{w_1, w_2 \in M \\ w_1 w_2 = w}} \eta(u)(w_1) \cdot \eta(v)(w_2) \\ &= \begin{cases} 1_R & \text{if} \quad w = uv \\ 0 & \text{otherwise} \end{cases} \\ &= \eta(uv)(w), \end{aligned}$$

the second equality being true because the only way to get a non-trivial summand is to have $w_1 = u$ and $w_2 = v$. It is obvious that $\eta(1_M) = 1_{RM}$. The map η is injective because $\operatorname{supp}(\eta(u)) = \{u\}$ and functions with different support are clearly different. □

The images in RM of elements of M under η and of ring elements under ι are obviously monomials. We will now show that every monomial in RM is of the form $\iota(a) \cdot \eta(u)$.

Lemma 2.11 Let $f \in RM$ be a monomial. Then there exists $a \in R$ and $u \in M$ with $f = \iota(a) \cdot \eta(u)$. Moreover, a and u are uniquely determined by f: u is the element of $\operatorname{supp}(f)$ and $a = f(u)$.

Proof Let u be the element of $\operatorname{supp}(f)$ and $a = f(u)$. Then we have, for all $w \in M$,

$$\begin{aligned} (\iota(a) \cdot \eta(u))(w) &= \sum_{\substack{w_1, w_2 \in M \\ w_1 w_2 = w}} \iota(a)(w_1) \cdot \eta(u)(w_2) \\ &= \begin{cases} a & \text{if} \quad w = u \\ 0 & \text{otherwise} \end{cases} \\ &= f(w), \end{aligned}$$

the second equality being true because the only way to see a non-trivial summand is to take $w_1 = 1_M$ and $w_2 = u$. The equality of the first and third expressions above also shows that a and u are uniquely determined by f: taking different a or u will result in a different value or support and thus in a different function. □

The following proposition is a simple combination of Lemma 2.8 with the one above. It explains the notation RM for the monoid ring: if we identify

R and M with their respective images under the natural embeddings in RM, then RM consists of all sums of products of the form ru with $r \in R$ and $u \in M$.

Proposition 2.12 *Every $f \in RM$ has a unique representation as a sum of monomials with pairwise different support which is given by*

$$f = \sum_{u \in \operatorname{supp}(f)} \iota(f(u)) \cdot \eta(u). \ \square$$

The remarks concerning uniqueness that were made following Lemma 2.8 apply verbatim to the uniqueness of the representation of the proposition above.

As happens so often when a new algebraic structure is constructed from one or several given ones, the monoid ring has a certain universal property. If S is a ring, then by the *multiplicative monoid of S* we understand the monoid which is obtained from S by disregarding addition.

Proposition 2.13 (Universal Property of Monoid Rings) *Let R and S be rings, M an Abelian monoid. Suppose $\varphi : R \longrightarrow S$ is a ring homomorphism and $\sigma : M \longrightarrow S$ is a homomorphism from M to the multiplicative monoid of S. Then there exists a unique homomorphism $\overline{\varphi} : RM \longrightarrow S$ with $\overline{\varphi} \circ \iota = \varphi$ and $\overline{\varphi} \circ \eta = \sigma$.*

$$\begin{tikzcd}
R \arrow[r, "\iota"] \arrow[dr, "\varphi"'] & RM \arrow[d, "\overline{\varphi}"'] & M \arrow[l, "\eta"'] \arrow[dl, "\sigma"] \\
& S &
\end{tikzcd}$$

Proof *Uniqueness*: By Proposition 2.12, every element of RM is a sum of products of elements of the form $\iota(a)$ and $\eta(u)$ with $a \in R$ and $u \in M$. A homomorphism from RM to any ring S is therefore uniquely determined by its values on elements of this form. For $\overline{\varphi}$, these values are prescibed to be $\varphi(a)$ and $\sigma(u)$.

Existence: We define $\overline{\varphi}$ by setting, for $f \in RM$,

$$\overline{\varphi}(f) = \sum_{u \in \operatorname{supp}(f)} \varphi(f(u)) \cdot \sigma(u) = \sum_{u \in M} \varphi(f(u)) \cdot \sigma(u).$$

(Note that we are once again working with a formally infinite sum here.)

We have, for all f, $g \in RM$,

$$\begin{aligned}
\overline{\varphi}(f) + \overline{\varphi}(g) &= \sum_{u \in M} \varphi\big(f(u)\big) \cdot \sigma(u) + \sum_{u \in M} \varphi\big(g(u)\big) \cdot \sigma(u) && \text{(Def. } \overline{\varphi}\text{)} \\
&= \sum_{u \in M} \Big(\varphi\big(f(u)\big) + \varphi\big(g(u)\big)\Big) \cdot \sigma(u) && \text{(Distrib. in } S\text{)} \\
&= \sum_{u \in M} \varphi\big(g(u) + f(u)\big) \cdot \sigma(u) && (\varphi \text{ hom.}) \\
&= \sum_{u \in M} \varphi\big((f+g)(u)\big) \cdot \sigma(u) && \text{(Def. + in } RM\text{)} \\
&= \overline{\varphi}(f+g), && \text{(Def. } \overline{\varphi}\text{)}
\end{aligned}$$

and

$$\begin{aligned}
\overline{\varphi}(f) \cdot \overline{\varphi}(g) &= \Big(\sum_{v \in M} \varphi\big(f(v)\big) \cdot \sigma(v)\Big) \cdot \Big(\sum_{w \in M} \varphi\big(g(w)\big) \cdot \sigma(w)\Big) && \text{(Def. } \overline{\varphi}\text{)} \\
&= \sum_{v,w \in M} \varphi\big(f(v)\big) \cdot \varphi\big(g(w)\big) \cdot \sigma(v) \cdot \sigma(w) && \text{(Distrib. and comm. in } RM\text{)} \\
&= \sum_{v,w \in M} \varphi\big(f(v) \cdot g(w)\big) \cdot \sigma(vw) && (\varphi \text{ and } \sigma \text{ hom.}) \\
&= \sum_{u \in M} \sum_{\substack{v,w \in M \\ vw=u}} \varphi\big(f(v) \cdot g(w)\big) \cdot \sigma(u) && \text{(Grouping summands by } vw\text{)} \\
&= \sum_{u \in M} \varphi\Bigg(\sum_{\substack{v,w \in M \\ vw=u}} f(v) \cdot g(w)\Bigg) \cdot \sigma(u) && (\varphi \text{ hom.}) \\
&= \sum_{u \in M} \varphi\big(fg(u)\big) \cdot \sigma(u) && \text{(Def. } \cdot \text{ in } RM\text{)} \\
&= \overline{\varphi}(fg). && \text{(Def. } \overline{\varphi}\text{)}
\end{aligned}$$

Moreover, we have

$$\begin{aligned}
\overline{\varphi}(1_{RM}) &= \sum_{u \in \operatorname{supp}(1_{RM})} \varphi\big(1_{RM}(u)\big) \cdot \sigma(u) \\
&= \varphi\big(1_{RM}(1_M)\big) \cdot \sigma(1_M) \\
&= \varphi(1_R) \cdot \sigma(1_M) \;=\; 1_S \cdot 1_S \;=\; 1_S.
\end{aligned}$$

To show that $\overline{\varphi} \circ \iota = \varphi$, let $a \in R$. Then

$$\begin{aligned}
\overline{\varphi}\big(\iota(a)\big) &= \sum_{u \in \operatorname{supp}(\iota(a))} \varphi\big(\iota(a)(u)\big) \cdot \sigma(u) \\
&= \varphi\big(\iota(a)(1_M)\big) \cdot \sigma(1_M) \\
&= \varphi(a) \cdot 1_S \;=\; \varphi(a).
\end{aligned}$$

To see that $\overline{\varphi} \circ \eta = \sigma$, let $w \in M$. Then

$$\begin{aligned}\overline{\varphi}(\eta(w)) &= \sum_{u \in \operatorname{supp}(\eta(w))} \varphi(\eta(w)(u)) \cdot \sigma(u) \\ &= \varphi(\eta(w)(w)) \cdot \sigma(w) = \varphi(1_R) \cdot \sigma(w) \\ &= 1_S \cdot \sigma(w) = \sigma(w). \quad \square\end{aligned}$$

We will now introduce the notation and terminology that will make polynomials look and behave as we know them from elementary algebra. (Cf. the remarks at the beginning of this section, where we explained how to get to the abstract point of view from the elementary one.) Let $1 \le n \in \mathbb{N}$ and $M = \mathbb{N}^n$, so that RM becomes the polynomial ring in n variables over R. We will use the notation $(\nu) = (\nu_1, \ldots, \nu_n)$ for elements of $M = \mathbb{N}^n$, and (0) for its neutral element $(0, \ldots, 0)$ Note that the monoid operation on M is now denoted by $+$.

First of all, we will identify each $a \in R$ with its image $\iota(a)$ in RM, so that a now stands for both the ring element a and the function that says a at (0) and 0 otherwise. This does not cause any trouble because ι is an embedding of rings and thus R is isomorphic to $\iota(R)$. Under this point of view, we may treat R as a subring of RM. In particular, we now have $1_R = 1_{RM}$. A polynomial of the form a with $a \in R$ (it is in fact a *monomial*) is called a **constant**, or a **constant polynomial**.

It would be entirely possible to do the same thing about η, i.e., to identify $\eta((\nu))$ with (ν) for all $(\nu) \in M = \mathbb{N}^n$, but we can do better than that. For $1 \le i \le n$, we set

$$(\varepsilon_i) = (0, \ldots, 0, \underset{\substack{\uparrow \\ i\text{th place}}}{1}, 0, \ldots, 0) \in M,$$

and we let $X_i \in RM$ be the monomial $\eta((\varepsilon_i))$, i.e.,

$$X_i((\nu)) = \begin{cases} 1 & \text{if } (\nu) = (\varepsilon_i) \\ 0 & \text{otherwise.} \end{cases}$$

X_i is called the ith **variable**, or **indeterminate**. (It is clear that any other letter is as good as X to denote variables, and this is frequently done.) It is quite obvious that each $(\nu) \in M$ has a unique representation of the form

$$(\nu_1, \ldots, \nu_n) = \sum_{\nu_1 \text{ summands}} (\varepsilon_1) + \cdots + \sum_{\nu_n \text{ summands}} (\varepsilon_n).$$

(A rigorous proof is by induction on $\nu_1 + \cdots + \nu_n$.) Applying η to the equation, we see that every monomial in RM of the form $\eta((\nu))$ with $(\nu) \in M$ can be written as

$$\begin{aligned}\eta((\nu)) &= \prod_{\nu_1 \text{ factors}} X_1 \cdot \cdots \cdot \prod_{\nu_n \text{ factors}} X_n \\ &= X_1^{\nu_1} \cdot \cdots \cdot X_n^{\nu_n},\end{aligned}$$

and different exponent tuples give rise to different monomials because η is an embedding. A monomial of the form $\eta((\nu)) = X_1^{\nu_1} \cdot \cdots \cdot X_n^{\nu_n}$ with $(\nu) \in M$ is called a **term**, and the set of all terms is denoted by $T(X_1, \ldots, X_n)$, or simply by T when it is clear from the context what the number of variables is. Being the image of $M = \mathbb{N}^n$ under the embedding η, T forms an Abelian monoid under multiplication in the ring RM, and the **exponent map**

$$\begin{array}{rccc} \eta: & (\mathbb{N}^n, (0), +) & \longrightarrow & (T, 1_R, \cdot) \\ & (\nu_1, \ldots, \nu_n) & \longmapsto & X_1^{\nu_1} \cdot \cdots \cdot X_n^{\nu_n} \end{array}$$

is an isomorphism of monoids. We see that the structure of $(T, 1, \cdot)$ is independent of the ring R, and we may thus talk about the **monoid of terms** in the variables $X_1, \ldots, X_n$ without specifying a ring R. If $t = X_1^{\nu_1} \cdot \cdots \cdot X_n^{\nu_n} \in T$, then the **total degree** of t is defined as

$$\deg(t) = \sum_{i=1}^{n} \nu_i.$$

Lemma 2.11 translated into our new notation now tells us that every monomial m of the polynomial ring has a representation of the form

$$m = aX_1^{\nu_1} \cdot \cdots \cdot X_n^{\nu_n} \qquad (a \in R,\ (\nu) \in \mathbb{N}^n),$$

and m uniquely determines a and the exponent tuple (ν). Here, a is called the **coefficient** of m, whereas $X_1^{\nu_1} \cdot \cdots \cdot X_n^{\nu_n}$, rather obviously, is called its *term*.

If we now apply Proposition 2.12 to the polynomial ring and translate its statement into our new notation, then we see that every polynomial $f \in RM$ has a unique representation as a sum of monomials with pairwise different terms which is given by

$$f = \sum_{(\nu) \in \mathrm{supp}(f)} f((\nu)) \cdot X_1^{\nu_1} \cdot \cdots \cdot X_n^{\nu_n}. \qquad (*)$$

In other words, for each $f \in RM$, there exists a unique finite subset N of $M = \mathbb{N}^n$ and a unique set $\{ a_{(\nu)} \mid (\nu) \in N \}$ of non-zero elements of R such that

$$f = \sum_{(\nu) \in N} a_{(\nu)} X_1^{\nu_1} \cdot \cdots \cdot X_n^{\nu_n}.$$

Note that this representation of f really displays f as a function $\mathbb{N}^n \longrightarrow R$: the set N of exponent tuples is the support of f, and $a_{(\nu)}$ is the value of f at (ν). It is now easy to see from the definition of addition and multiplication in the monoid ring that if we represent polynomials in this way, then they behave the way they do in elementary algebra: we add them by combining like terms, and we multiply them by first multiplying out and then combining like terms. It is now even more obvious than before that the product of two monomials is again a monomial, and thus the set of monomials, like the set of terms, forms an Abelian monoid under ring multiplication. This monoid, however, will be of much less interest to us than the monoid of terms.

In the univariate case $n = 1$, it is customary to write X instead of X_1, and to add zero summands to the representation $(*)$ to obtain

$$f = \sum_{i=0}^{m} a_i X^i \qquad (m \in \mathbb{N},\ a_i \in R).$$

This is then a representation that is unique up to zero summands at the top. It becomes unique if one requires that in addition, a_m is not zero whenever $f \neq 0$. In that case, a_m is called the **head coefficient** of f, and m is the **degree** of f. Here, f is called **monic** if $f \neq 0$ and its head coefficient equals 1.

Now consider the ring $R[X_1, \dots, X_n]$ obtained by adjoining the variables $X_1, \dots, X_n$ to R within RM in the sense of Definition 1.109. By Lemma 1.110, the result is all of RM. It is customary to use this alternate notation: whenever R is a ring and $M = \mathbb{N}^n$ with $1 \leq n \in \mathbb{N}$, then then RM is denoted by

$$R[X_1, \dots, X_n],$$

or $R[\underline{X}]$ for short. Now let $f \in R[\underline{X}]$. Then we define the set $M(f)$ of **monomials of** f as the set of summands occuring in the unique representation $(*)$ above, i.e.,

$$M(f) = \left\{ f((\nu)) \cdot X_1^{\nu_1} \cdot \cdots \cdot X_n^{\nu_n} \;\middle|\; (\nu) \in \operatorname{supp}(f) \right\}.$$

The set $T(f)$ of **terms of** f is defined as the set of terms of elements of $M(f)$, i.e.,

$$T(f) = \{ X_1^{\nu_1} \cdot \cdots \cdot X_n^{\nu_n} \mid (\nu) \in \operatorname{supp}(f) \}.$$

The **total degree** of a non-zero polynomial f is defined as

$$\deg(f) = \max\{ \deg(t) \mid t \in T(f) \}.$$

For a univariate polynomial, the total degree thus coincides with the degree. Finally, we define the set $C(f)$ of **coefficients of** f as the set of all coefficients of elements of $M(f)$, i.e.,

$$C(f) = \left\{ f((\nu)) \;\middle|\; (\nu) \in \operatorname{supp}(f) \right\}.$$

If $at \in M(f)$, then the coefficient a of this monomial is, rather obviously, also referred to as the *coefficient of* t *in* f. With this notation, the unique representation $(*)$ of $f \in R[\underline{X}]$ may be rewritten in two possible ways:

$$f = \sum_{t \in T(f)} a_t t = \sum_{m \in M(f)} m.$$

The following lemma states one of the most frequently used properties of polynomials.

Lemma 2.14 Let $a_1, \dots, a_m \in R$, and suppose $t_1, \dots t_m \in T$ are pairwise different. Then, in $R[\underline{X}]$,

$$\sum_{j=1}^{m} a_j t_j = 0 \quad \text{iff} \quad a_1 = \cdots = a_m = 0.$$

Proof The direction "$\Longleftarrow$" is trivial since $R[\underline{X}]$ is a ring. For "$\Longrightarrow$," assume for a contradiction that not all a_j equal zero. Dropping those a_j that are zero and renumbering, we may assume that $a_j \neq 0$ for $1 \leq j \leq m$. Now the sum on the left-hand side is the unique representation of a polynomial with support $\{t_1, \dots, t_m\}$, while the support of 0 in $R[\underline{X}]$ is the empty set, a contradiction. □

Lemma 2.15 (Universal Property of Polynomial Rings) Let R and S be rings, $\varphi : R \longrightarrow S$ a homomorphism of rings, $c_1, \dots, c_n \in S$. Then there is a unique homomorphism $\overline{\varphi} : R[X_1, \dots, X_n] \longrightarrow S$ of rings which extends φ, i.e., $\overline{\varphi} \restriction R = \varphi$, and satisfies $\overline{\varphi}(X_i) = c_i$ for $1 \leq i \leq n$. Here,

$$\overline{\varphi}\Bigl(\sum_{j=1}^{m} a_j X_1^{\nu_{j1}} \cdot \cdots \cdot X_n^{\nu_{jn}}\Bigr) = \sum_{j=1}^{m} \varphi(a_j) c_1^{\nu_{j1}} \cdot \cdots \cdot c_n^{\nu_{jn}}. \qquad (*)$$

Proof Let

$$\begin{array}{cccc} \sigma : & \mathbb{N}^n & \longrightarrow & S \\ & (\nu_1, \dots, \nu_n) & \longmapsto & c_1^{\nu_1} \cdot \cdots \cdot c_n^{\nu_n} \end{array}$$

be the homomorphism of Lemma 2.4. Then Proposition 2.13 provides a unique homomorphism $\overline{\varphi} : R[\underline{X}] \longrightarrow S$ with the properties $\overline{\varphi} \circ \iota = \varphi$ and $\overline{\varphi} \circ \eta = \sigma$. Now our notation is such that ι is the identity map on R, so the first property becomes $\overline{\varphi} \restriction R = \varphi$. The second property states that for all $(\nu) \in \mathbb{N}^n$,

$$\overline{\varphi}\Bigl(\eta((\nu))\Bigr) = \sigma((\nu)).$$

With our notation for the term $\eta((\nu))$ and by our choice of σ, this turns into

$$\overline{\varphi}(X_1^{\nu_1} \cdot \cdots \cdot X_n^{\nu_n}) = c_1^{\nu_1} \cdot \cdots \cdot c_n^{\nu_n}. \qquad (**)$$

In particular, we have $\overline{\varphi}(X_i) = c_i$ for all $1 \leq i \leq n$. Moreover, every homomorphism that satisfies this latter condition must also satisfy $(**)$, and thus $\overline{\varphi}$ must be the unique homomorphism provided by Proposition 2.13. The last statement of the proposition can now easily be verified by looking up the definition of $\overline{\varphi}$ in the proof of Proposition 2.13. □

From the equation $(*)$ in the proposition above, we can easily determine under what condition the homomorphism $\overline{\varphi}$ will be an embedding: $\overline{\varphi}$ will be injective iff φ and $c_1, \dots, c_n \in S$ satisfy the following conditions:

(i) φ is injective, and

(ii) whenever $b_1, \ldots, b_m \in \varphi(R)$ and $(\nu_{j1}, \ldots, \nu_{jn}) \in \mathbb{N}^n$ are pairwise distinct for $1 \leq j \leq m$, then

$$\sum_{j=1}^{m} b_j c_1^{\nu_{j1}} \cdot \cdots \cdot c_n^{\nu_{jn}} = 0 \quad \text{implies} \quad b_1 = \cdots = b_m = 0.$$

Intuitively speaking, (ii) means that the c_i "behave like indeterminates over $\varphi(R)$." Again from the equation $(*)$ of the last proposition, together with Lemma 1.110, we see that $\overline{\varphi}$ will be surjective iff $S = \varphi(R)[c_1, \ldots, c_n]$. We can thus formulate the following lemma.

Lemma 2.16 Let R and S be rings, $\varphi : R \longrightarrow S$ an embedding, and suppose $c_1, \ldots, c_n \in S$ are such that $S = \varphi(R)[c_1, \ldots, c_n]$ and condition (ii) above is satisfied. Then $S \simeq R[\underline{X}]$, and an isomorphism from $R[\underline{X}]$ to S is given by $\overline{\varphi}$ as described in $(*)$ of the last proposition. □

Now let $2 \leq n$ and $1 \leq i \leq n$. We may then form the monoid ring RM with $M = \mathbb{N}^n$. We know that it equals the result of adjoining to R the variables $X_1, \ldots X_n$, which fact is expressed in our notation $R[X_1, \ldots, X_n]$ for RM. We may also adjoin $X_1, \ldots, X_i$ to R in RM, and Lemma 1.112 (ii) tells us that

$$R[X_1, \ldots, X_i] \subseteq R[X_1, \ldots, X_n].$$

Note that in a strict formal sense, the ring on the left is different from $R[X_1, \ldots, X_i]$ when formed as the monoid ring RM with $M = \mathbb{N}^i$: in the latter ring, X_1 is a function from $\mathbb{N}^i$ to R, whereas in the former, it is a function from $\mathbb{N}^n$ to R. Our notational convention is such that it covers this distinction up, and this is in fact quite desirable: it is an easy though notationally tedious exercise to prove that the two rings that $R[X_1, \ldots, X_i]$ stands for are naturally isomorphic by Lemma 2.16. A similar qualification applies to the following equation which is immediate from Lemma 1.112 (i):

$$R[X_1, \ldots, X_i][X_{i+1}, \ldots, X_n] = R[X_1, \ldots, X_n].$$

Let us describe this equality a little more explicitly. An element of

$$R[X_1, \ldots, X_i][X_{i+1}, \ldots, X_n]$$

is of the form $\sum f_t t$ with

$$f_t \in R[X_1, \ldots, X_i] \subset R[X_1, \ldots, X_n]$$

and

$$t \in T(X_{i+1}, \ldots, X_n) \subset R[X_1, \ldots, X_n].$$

It may hence be viewed as a sum of products of elements of $R[X_1, \ldots, X_n]$, i.e., an element of $R[X_1, \ldots, X_n]$. Conversely, elements of $R[X_1, \ldots, X_n]$ are of the form $f = \sum a_t t$ with

$$a_t \in R \quad \text{and} \quad t \in T(X_1, \ldots, X_n).$$

Every $t \in T(X_1, \dots, X_n)$ can be uniqely written in the form uv with $u \in T(X_1, \dots, X_i)$ and $v \in T(X_{i+1}, \dots, X_n)$. Hence each monomial $a_t t$ of f can be viewed as a monomial $(a_{uv}u)v$ of the ring $R[X_1, \dots, X_i][X_{i+1}, \dots X_n]$. Then

$$f = \sum (a_{uv}u)v,$$

and we see that f can also be viewed as an element of

$$R[X_1, \dots, X_i][X_{i+1}, \dots, X_n].$$

This identification of $R[X_1, \dots, X_n]$ and $R[X_1, \dots, X_i][X_{i+1}, \dots, X_n]$ will turn out to be a simple but powerful tool in the theory of polynomial rings. If we take, for example,

$$f = X^3Y^3Z^3 + 3X^3Y^3Z - X^2Y^3Z^3 + 2XY^2Z^3 - 1,$$

in $\mathbb{Z}[X, Y, Z]$ then, as an element of $\mathbb{Z}[X, Y][Z]$,

$$f = (X^3Y^3 - X^2Y^3 + 2XY^2)Z^3 + (3X^3Y^3)Z - 1,$$

whereas, as an element of $\mathbb{Z}[X][Y, Z]$,

$$f = (X^3 - X^2)Y^3Z^3 + (3X^3)Y^3Z + (2X)Y^2Z^3 - 1.$$

Since $R[X_1, \dots, X_n]$ is a commutative ring which is obtained from R by adjoining $X_1, \dots, X_n$, it is true that

$$R[X_1, \dots, X_n] = R[X_{\pi(1)}, \dots, X_{\pi(n)}]$$

for every permutation π of the indices $\{1, \dots, n\}$. If $0 \neq f \in R[X_1, \dots, X_n]$ and $1 \leq i \leq n$, then the **degree of f in X_i**, denoted by $\deg_{X_i}(f)$, is defined as the degree of f when viewed as a univariate polynomial in X_i, i.e., as an element of

$$R[X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n][X_i].$$

The next lemma lists two important special cases of Proposition 2.15.

Lemma 2.17 (i) If R is a subring of S and $c_1, \dots, c_n \in S$, then the map

$$\begin{array}{ccc} R[X_1, \dots, X_n] & \longrightarrow & S \\ \displaystyle\sum_{j=1}^{m} a_j X_1^{\nu_{j1}} \cdot \dots \cdot X_n^{\nu_{jn}} & \longmapsto & \displaystyle\sum_{j=1}^{m} a_j c_1^{\nu_{j1}} \cdot \dots \cdot c_n^{\nu_{jn}} \end{array}$$

is a homomorphism of rings which acts as the identity on R and maps X_i to c_i $(1 \leq i \leq n)$.

(ii) If $\psi : R \longrightarrow S$ is a homomorphism of rings, then the map

$$\begin{array}{ccc} R[X_1,\ldots,X_n] & \longrightarrow & S[X_1,\ldots,X_n] \\ \sum_{j=1}^{m} a_j X_1^{\nu_{j1}} \cdot \ldots \cdot X_n^{\nu_{jn}} & \longmapsto & \sum_{j=1}^{m} \psi(a_j) X_1^{\nu_{j1}} \cdot \ldots \cdot X_n^{\nu_{jn}} \end{array}$$

is a homomorphism of rings. It is an embedding (surjective, an isomorphism) iff ψ is an embedding (surjective, an isomorphism).

Proof Statement (i) is a straightforward application of Proposition 2.15 with the inclusion of R in S taken for φ. For statement (ii), we apply Proposition 2.15 where φ is ψ followed by the inclusion of S in $S[X_1,\ldots,X_n]$ and $c_i = X_i$ for $1 \leq i \leq n$. The proof of the last statement of the lemma is straightforward. □

The homomorphism of (i) above is called the **substitution homomorphism**. In this case, the image of f is usually denoted by $f(c_1,\ldots,c_n)$. We will also allow ourselves to write $f(\boldsymbol{c})$, where $\boldsymbol{c} = (c_1,\ldots,c_n)$. An n-tuple $\boldsymbol{c} \in S^n$ is called a **zero** of f if $f(\boldsymbol{c}) = 0$.

Next, we discuss zero divisors and units in polynomial rings.

Lemma 2.18 Let R be a ring. Then the following hold:

(i) $R[X_1,\ldots,X_n]$ is a domain if and only if R is a domain.

(ii) If R is a domain, then $\deg(fg) = \deg(f) + \deg(g)$ for all f, $g \in R[X]$ with f, $g \neq 0$.

(iii) Every unit of R is a unit in $R[X_1,\ldots,X_n]$. If R is a domain, then every unit of $R[X_1,\ldots,X_n]$ is a constant and a unit in R.

Proof "$\Longrightarrow$" of (i) follows immediately from $R \subset R[X_1,\ldots,X_n]$. For "$\Longleftarrow$," let R be a domain. We proceed by induction on the number n of variables. Let $n = 1$,

$$f = \sum_{i=0}^{m_1} a_i X^i \quad \text{and} \quad g = \sum_{i=0}^{m_2} b_i X^i$$

with a_{m_1}, $b_{m_2} \neq 0$. Then

$$fg = a_{m_1} b_{m_2} X^{m_1+m_2} + \sum_{i=0}^{m_1+m_2-1} c_i X^i \neq 0 \qquad (c_i \in R)$$

since $a_{m_1} b_{m_2} \neq 0$. If $n > 1$, then $R[X_1,\ldots,X_{n-1}]$ is a domain by induction hypothesis, and by the argument for $n = 1$, so is $R[X_1,\ldots,X_{n-1}][X_n] = R[X_1,\ldots,X_n]$.

Statement (ii) has already been demonstrated in the proof of (i) above. The first statement of (iii) again follows from $R \subset R[X_1,\ldots,X_n]$. Now let

R be a domain. We use induction on n. Let $n = 1$, and suppose f is a unit of $R[X]$. Then there exists $g \in R[X]$ with $fg = 1$. Since f, $g \neq 0$, we may apply (ii) above to conclude that

$$\deg(f) + \deg(g) = \deg(1) = 0,$$

and so both f and g must be constant. This shows that $f \in R$ and that f is a unit of R. Now let $n > 1$, f a unit of $R[X_1, \ldots, X_n] = R[X_1, \ldots, X_{n-1}][X_n]$. Since $R[X_1, \ldots, X_{n-1}]$ is a domain by (i), f is a unit of $R[X_1, \ldots, X_{n-1}]$ by the argument for $n = 1$ above and hence of R by induction hypothesis. □

The condition that R be a domain cannot be dropped in the second part of (iii) above: if $R = \mathbb{Z}/8\mathbb{Z}$, then $f = 1 + 4X$ is a unit of $R[X]$ since $f^2 = 1 + 8X + 16X^2 = 1$. If K is a field, then by (iii) above, the units of $K[X_1, \ldots, X_n]$ are precisely the non-zero constant polynomials. We also note that for any ring R and $f, g \in R[X]$, $\deg(f+g) \leq \max(\deg(f), \deg(g))$.

Exercise 2.19 Show that for any ring R, a constant polynomial which is a unit in $R[X_1, \ldots, X_n]$ is a unit in R.

Exercise 2.20 Show that for any ring R, $\operatorname{char}(R) = \operatorname{char}(R[X_1, \ldots, X_n])$.

We are now in a position to give the long overdue example of a prime ideal that is not maximal.

Lemma 2.21 Let R be a domain, $n \geq 2$. Then the ideal $\operatorname{Id}(X_1)$ generated by X_1 in $R[X_1, \ldots, X_n]$ is prime but not maximal.

Proof $\operatorname{Id}(X_1)$ consists of all polynomials that are multiples of X_1, i.e., all polynomials $f \in R[X_1, \ldots, X_n]$ such that $\deg_{X_1}(t) \geq 1$ for all $t \in T(f)$. We can write an arbitrary $f \in R[X_1, \ldots, X_n]$ in the form $f = f_1 + f_2$, where we have grouped all $t \in T(f)$ with $\deg_{X_1}(t) \geq 1$ into f_1, and we see that $f \in \operatorname{Id}(X_1)$ iff $f_2 = 0$. Now if f is a product of two polynomials g and h, then

$$f = gh = g_1h_1 + g_1h_2 + g_2h_1 + g_2h_2,$$

and it is easy to see that f_1 is the sum of the first three summands, while $f_2 = g_2h_2$. So, in this case, $f_2 = 0$ implies that $g_2 = 0$ or $h_2 = 0$, i.e., $f \in \operatorname{Id}(X_1)$ implies $g \in \operatorname{Id}(X_1)$ or $h \in \operatorname{Id}(X_1)$ We have proved that $\operatorname{Id}(X_1)$ is prime. To see that it is not maximal, we first note that $\operatorname{Id}(X_1)$ is properly contained in $\operatorname{Id}(X_1, X_2)$ since $X_1 \in \operatorname{Id}(X_1, X_2)$ but $X_2 \notin \operatorname{Id}(X_1)$. It remains to show that $\operatorname{Id}(X_1, X_2)$ is proper. $\operatorname{Id}(X_1, X_2)$ consists of all sums of multiples of X_1 and X_2, and such a sum can clearly not contain a non-zero constant monomial. In particular, $1 \notin \operatorname{Id}(X_1, X_2)$. □

One reason why we have defined polynomial rings as a special case of the more general concept of monoid rings is the fact that in Section 7.2 we will need "polynomial rings in infinitely many variables," a concept that we will now make precise. The proof of the following lemma is straightforward from the definitions.

Lemma 2.22 Let I be a (possibly infinite) set. Then the set $\mathbb{N}^I$ of all functions from I to $\mathbb{N}$ is an Abelian monoid under pointwise addition of functions whose neutral element is the zero function. The same holds true for the set

$$M = \{\, \kappa \in \mathbb{N}^I \mid \kappa(i) \neq 0 \text{ for only finitely many } i \in I \,\}. \quad \square$$

Let now R be a ring, I a set, and M the additive monoid as defined in the lemma above. Then we may form the monoid ring RM according to Definition 2.7. (In order to visualize the construction, it is suggested to think of the elements of M as "infinitely long tuples of natural numbers with only finitely many non-zero entries.") For each $i \in I$, we define a canonical element $\varepsilon_i \in M$ by setting

$$\varepsilon_i(j) = \begin{cases} 1_{\mathbb{N}} & \text{iff } j = i \\ 0 & \text{otherwise,} \end{cases}$$

and a canonical monomial $X_i \in RM$ by setting

$$X_i(\nu) = \begin{cases} 1_R & \text{iff } \nu = \varepsilon_i \\ 0 & \text{otherwise.} \end{cases}$$

In analogy to the discussion following the proof of Proposition 2.13, we may now argue that every monomial in RM is of the form

$$a \cdot X_{i_1}^{\nu_1} \cdot \dots \cdot X_{i_k}^{\nu_k}$$

for some $a \in R$ and some $k \in \mathbb{N}$, and that every element of RM is a sum of such monomials. We see that each element of RM looks like an element of a polynomial ring over R in certain variables, and tracing the definitions of the operations in RM, it is not hard to see that these operations are performed in the same way as in the polynomial ring over R in the finitely many variables that are involved in each instance of the respective operation. This phenomenon can actually be given a more precise formulation. Whenever $k \in \mathbb{N}$ and $i_1, \dots, i_k \in I$, then, according to Proposition 2.15 with φ of that proposition taken as the natural embedding of R in RM, the map

$$\begin{array}{rccc} \overline{\varphi}: & R[X_1, \dots, X_k] & \longrightarrow & RM \\ & \displaystyle\sum_{j=1}^{m} a_j X_1^{\nu_{j1}} \cdot \dots \cdot X_k^{\nu_{jk}} & \longmapsto & \displaystyle\sum_{j=1}^{m} a_j X_{i_1}^{\nu_{j1}} \cdot \dots \cdot X_{i_k}^{\nu_{jk}} \end{array}$$

is a homomorphism, and in view of the discussion following the proof of Proposition 2.15, it is not hard to see that $\overline{\varphi}$ is actually an embedding. All this shows that RM behaves like a multivariate polynomial ring over R, except that there is an unlimited supply of variables in case I is an infinite set. The ring RM is therefore called the **polynomial ring in the variables $\{\, X_i \mid i \in I \,\}$ over R.**

We close this section with a few remarks on the possibility of effective computations with polynomials. It is rather obvious that polynomials of $R[X_1, \ldots, X_n]$ can be represented on a computer and their addition, subtraction, and multiplication can be effectively performed if and only if the same is true for the elements of R. In the latter case, we will call the ring R **computable**. A field K is called a **computable field** if it is a computable ring and the inverse of every non-zero element can be effectively computed. The notion of computability will be discussed more rigorously in Section 4.6. For the time being, let us note that $\mathbb{Z}$ is a computable ring since exact integer arithmetic can be implemented on a computer. $\mathbb{Q}$ is easily seen to be a computable field, because exact integer arithmetic together with the possibility of reducing to lowest terms by means of integer gcd's (the next section has a rigorous discussion) allows us to compute with rational numbers. $\mathbb{Z}/m\mathbb{Z}$ is a computable ring for $m \in \mathbb{Z}$ because the arithmetic in $\mathbb{Z}/m\mathbb{Z}$ is defined in terms of the arithmetic in $\mathbb{Z}$, and each element $k + m\mathbb{Z}$ of $\mathbb{Z}/m\mathbb{Z}$ has a unique representation as $r + m\mathbb{Z}$ where r is the remainder of k upon division by m. If p is a prime number, then $\mathbb{Z}/p\mathbb{Z}$ is a field (see Proposition 1.98). Since it is also finite, inverses can in principle be found by trial and error, and we see that $\mathbb{Z}/p\mathbb{Z}$ is a computable field. One of the results of the next section is a better way of finding inverses in $\mathbb{Z}/p\mathbb{Z}$.

2.2 Euclidean Domains

Univariate polynomial rings over a field stand out in the class of all polynomials rings as having particularly nice ring theoretic properties. This is due to the fact that they allow long division of polynomials. *Euclidean domains* are classically defined as domains with a property that is modeled after division with remainder of integers and long division of polynomials. Here, we use a seemingly weaker property which imitates a single step in the division process. We then prove that this is equivalent to the classical definition.

Definition 2.23 A ring R is called a **Euclidean domain** if it is a domain and there exists a map $\varphi : R \setminus \{0\} \longrightarrow \mathbb{N}$ with the following properties.

(i) $\varphi(ab) \geq \varphi(a)$ for all a, $b \in R$ with a, $b \neq 0$.

(ii) For all a, $b \in R$ with a, $b \neq 0$ and $\varphi(a) \geq \varphi(b)$, there exist s, $t \in R$ such that $a - sb = t$, and $\varphi(t) < \varphi(a)$ or $t = 0$.

If, in addition, R is a computable ring and s and t as above can be computed effectively from a and b, then R is called a **computable** Euclidean domain.

We will refer to the function φ as the **abstract degree function** of R, a terminology explained by the next proposition.

Proposition 2.24 *Let K be a field. If we take for $\varphi : K[X] \longrightarrow \mathbb{N}$ the degree function, then $K[X]$ is a Euclidean domain. If, in addition, K is computable, then $K[X]$ is even a computable Euclidean domain.*

Proof Condition (i) of the definition above is immediate from Lemma 2.18. Now let $0 \neq f, g \in K[X]$ with $\deg(f) \geq \deg(g)$, say

$$f = \sum_{i=0}^{m} a_i X^i \quad \text{and} \quad g = \sum_{i=0}^{n} b_i X^i$$

with a_m, $b_n \neq 0$ and $m \geq n$. Then we can satisfy condition (ii) by writing

$$f - \underbrace{\frac{a_m}{b_n} \cdot X^{m-n}}_{s} \cdot g = t$$

because the monomial $a_m X^m$ cancels out and hence t either equals zero or has a degree less than m. This can clearly be done effectively if we can compute with elements of K. □

Proposition 2.25 *$\mathbb{Z}$ becomes a computable Euclidean domain if we take for $\varphi : \mathbb{Z} \longrightarrow \mathbb{N}$ the absolute value function.*

Proof Condition (i) of the definition is an elementary property of the absolute value function. Now let $0 \neq m, n \in \mathbb{Z}$ with $|m| \geq |n|$. If m and n both have the same sign, then we may take $s = 1$ and $t = m - n$: then $m - sn = t$, and it is easy to see that $|m - n| < |m|$. If m and n have opposite signs, then it is equally easy to see that $s = -1$ and $t = m + n$ have the required properties. □

Next, we prove that our definition of Euclidean domains is equivalent to the classical one. The ring elements q and r of condition (ii) below are called, respectively, the **quotient** and **remainder** of a upon division by b.

Proposition 2.26 *Let R be a domain, and suppose there exists a map $\varphi : R \setminus \{0\} \longrightarrow \mathbb{N}$ such that $\varphi(ab) \geq \varphi(a)$ for all $0 \neq a, b \in R$. Then the following are equivalent:*

(i) R is a Euclidean domain with abstract degree function φ.

(ii) For all $a, b \in R$ with $b \neq 0$, there exist $q, r \in R$ such that $a = qb + r$, and $\varphi(r) < \varphi(b)$ or $r = 0$.

Moreover, if R is a computable Euclidean domain, then q and r as described in (ii) can be computed effectively from a and b.

Proof (ii)$\Longrightarrow$(i) is trivial: we may simply take, for s and t, the elements q and r whose existence is guaranteed by (ii). For the proof of (i)$\Longrightarrow$(ii) and the additional statement concerning computability, we give an algorithm DIV (Table 2.1) that computes q and r from a and b. For a general

TABLE 2.1. Algorithm DIV

```
Specification: $(q, r) \leftarrow \mathrm{DIV}(a, b)$
               Computation of quotient and remainder in a
               Euclidean domain
Given: $a, b \in R$ with $b \neq 0$
Find: $q, r \in R$ with $a = qb + r$, and $\varphi(r) < \varphi(b)$ or $r = 0$
begin
REM $\leftarrow a$;   QUOT $\leftarrow 0$
while REM $\neq 0$ and $\varphi(\mathrm{REM}) \geq \varphi(b)$ do
    – choose $s, t \in R$ with REM $- sb = t$ and $\varphi(t) < \varphi(\mathrm{REM})$ or $t = 0$
    REM $\leftarrow t$
    QUOT $\leftarrow$ QUOT $+ s$
end
return((QUOT, REM))
end DIV
```

Euclidean domain R we may interpret the assignments of the algorithm as mathematical constructions and thus read the algorithm together with the proof of its correctness and termination as an existence proof. The algorithm terminates since after each execution of the **while**-loop, either REM $= 0$ or the abstract degree of REM is less than before the loop was entered. The equation $a = \mathrm{QUOT} \cdot b + \mathrm{REM}$ is a loop invariant: it is trivially true after initalization, and during each execution of the **while**-loop, a certain ring element s is added to QUOT, while sb is subtracted from REM. Hence the equation holds for the output values of QUOT and REM. It follows immediately from the **while**-clause that QUOT and REM have the required properties upon termination. □

If we apply the proposition above to the integers, then we see that we could also have quoted Proposition 0.6 to show that $\mathbb{Z}$ is a Euclidean domain: the quotient and remainder of Proposition 0.6 clearly satisfy (ii) above. Note that their uniqueness is not guaranteed by that condition, because it is only required that the remainder r of m upon division by n satisfies $-|n| < r < |n|$. The algorithm DIV provides a (rather crude) method of dividing with remainder under the assumption that integer arithmetic is available.

If we take $R = K[X]$ with computable field K, then it is clear that DIV becomes long division of polynomials. In order to do the outstanding importance of this algorithm justice we have formulated it explicitly in Table 2.2. The algorithm DIVPOL cannot in general be applied to polynomials over an arbitrary ring because it involves division of coefficients. We will later see that, indeed, $R[X]$ is not a Euclidean domain unless R is a field. However, inspection of DIVPOL shows that all divisions occurring are by the head coefficient b_m of the divisor g. So if g is monic, i.e., $b_m = 1$, then

TABLE 2.2. Algorithm DIVPOL

```
Specification: $(q, r) \leftarrow$ DIVPOL$(f, g)$
               Divide $f$ by $g$ with remainder
Given: $f, g \in K[X]$ with $g \neq 0$
Find: $q, r \in K[X]$ with $f = qg + r$, $\deg(r) < \deg(g)$ or $r = 0$
begin
$R \leftarrow f$;  $G \leftarrow g$;  $Q \leftarrow 0$
while $R \neq 0$ and $\deg(R) \geq \deg(G)$ do
    $R \leftarrow R - (a_n/b_m)X^{n-m}G$,
    where $R = \sum_{i=0}^{n} a_i X^i$ and $G = \sum_{i=0}^{m} b_i X^i$ with $a_n, b_m \neq 0$
    $Q \leftarrow Q + (a_n/b_m)X^{n-m}$
end
return$((Q, R))$
end DIVPOL
```

no division at all is required. We have proved the following lemma.

Lemma 2.27 Let R be a ring and f, $g \in R$ with $g \neq 0$ and g monic. Then there exist q, $r \in R[X]$ with $f = qg + r$, and $\deg(r) < \deg(g)$ or $r = 0$. Moreover, if R is a computable ring, then q and r can be computed from f and g by means of the algorithm DIVPOL. □

In contrast to the situation in $\mathbb{Z}$, quotient and remainder of polynomials are automatically unique.

Proposition 2.28 *Let K be a field and f, $g \in K[X]$ with $g \neq 0$. Then quotient and remainder of f upon division by g are uniquely determined by f and g.*

Proof Let q, r, q', $r' \in K[X]$ such that q, r and q', r' satisfy (ii) of Proposition 2.26. Then we have

$$(q' - q)g = r - r'.$$

Since $g \neq 0$ and $K[X]$ is a domain, we conclude that $q' - q \neq 0$ iff $r - r' \neq 0$. Assume for a contradiction that they were both different from 0. Then

$$\begin{aligned} \deg(r - r') &< \deg(g), \quad \text{and} \\ \deg(r - r') &= \deg(q' - q) + \deg(g) \\ &\geq \deg(g), \end{aligned}$$

a contradiction. □

The next lemma shows that, in a manner of speaking, a remainder zero is always unique.

Lemma 2.29 Let R be a Euclidean domain with abstract degree function φ, and let $a, b \in R$ with $b \neq 0$. Then the following are equivalent:

(i) a lies in the ideal generated by b.

(ii) Zero is a remainder of a upon division by b.

(iii) Every possible remainder of a upon division by b equals zero.

Proof The implications (i)$\Longrightarrow$(ii) and (iii)$\Longrightarrow$(i) are trivial. For the proof of (ii)$\Longrightarrow$(iii), suppose $a = q_1 b$ with $q_1 \in R$, and assume for a contradiction that there exist $q_2, r \in R$ with $r \neq 0$ and $\varphi(r) < \varphi(b)$ such that $a = q_2 b + r$. Then $(q_1 - q_2)b = r$, and so

$$\varphi(r) = \varphi\big((q_1 - q_2) \cdot b\big) \geq \varphi(b),$$

a contradiction. □

Euclidean domains have practically all the pleasant properties that a ring can have.

Proposition 2.30 *Every Euclidean domain is a PID.*

Proof Let R be a Euclidean domain with abstract degree function φ, I an ideal of R. If $I = 0$, then $I = 0 \cdot R$ is principal. Otherwise, the set

$$\{\varphi(r) \mid 0 \neq r \in I\} \subseteq \mathbb{N}$$

is not empty and thus has a least element, say m. Let $a \in I$ with $\varphi(a) = m$. We claim that $I = aR$. The inclusion $aR \subseteq I$ follows from $a \in I$. Now let $b \in I$. Then there exist $q, r \in R$ such that $b = qa + r$, and $\varphi(r) < \varphi(a) = m$ or $r = 0$. Since $r = b - aq \in I$, we must have $r = 0$ by the minimality of m. We see that $b \in aR$. □

The proof of the proposition actually shows a little more: if I is a non-trivial ideal of a Euclidean domain, then *every* element of I of minimal degree generates I. But it was one of the more elementary results of Section 1.7 that any two generators of a principal ideal are associated, i.e., differ by a unit factor. In the case of a univariate polynomial ring over a field, the units are precisely the constants, and we have proved the following corollary.

Corollary 2.31 *Let K be a field, I a non-trivial ideal of $K[X]$. Then I contains a unique monic polynomial f of minimal degree, and $I = \mathrm{Id}(f)$.* □

From Propositions 1.74 and 2.30 we conclude that any two elements a and b of a Euclidean domain R have a gcd d in R, and there exist $s, t \in R$ with $d = sa + tb$. The eminent importance of Euclidean domains stems from the fact that in the computable case, we can compute d, s, and t from a and b. For a computable Euclidean domain R, we will denote by DIV an algorithm that returns, after input of any pair a, $b \in R$ with $b \neq 0$, a pair consisting of quotient and remainder of a upon division by b. Note that by Exercise 1.68 (v), the computation of $\gcd(a, b)$ requires no effort if one of a and b equals zero.

Theorem 2.32 *Let R be a computable Euclidean domain with abstract degree function φ. Then the algorithm* EXTEUC *of Table* 2.3 *computes, for given a, $b \in R$ with a, $b \neq 0$, a gcd d of a and b, and s, $t \in R$ with $d = sa + tb$.*

TABLE 2.3. Algorithm EXTEUC

```
Specification: $(d, s, t) \leftarrow$ EXTEUC$(a, b)$
               Extended Euclidean Algorithm
Given: $0 \neq a, b \in R$
Find: a gcd $d$ of $a$ and $b$ in $R$, and $s, t \in R$ with $d = sa + tb$
begin
$A \leftarrow a$;  $B \leftarrow b$
$S \leftarrow 1$;  $T \leftarrow 0$
$U \leftarrow 0$;  $V \leftarrow 1$
while $B \neq 0$ do
    (QUOT, REM) $\leftarrow$ DIV$(A, B)$
    $A \leftarrow B$;  $B \leftarrow$ REM
    $S1 \leftarrow S$;  $T1 \leftarrow T$
    $S \leftarrow U$;  $T \leftarrow V$
    $U \leftarrow S1 - \mathrm{QUOT} \cdot U$;  $V \leftarrow T1 - \mathrm{QUOT} \cdot V$
end
return$(A, S, T)$
end EXTEUC
```

Proof *Termination:* During one execution of the **while**-loop, the value of B is replaced by the remainder of A upon division by B. This means that either the abtract degree of the value of B decreases, or B is set to 0. It is now immediate from Corollary 0.4 that the **while**-condition $B = 0$ must be reached eventually.

Correctness: Let the value of any of the variables after n executions of the loop be denoted by that variable with a subscript n, and assume that there are N executions altogether. We claim that the ideal $I_n = A_n \cdot R + B_n \cdot R$ is a loop invariant. Indeed, the equations

$$A_n = B_{n-1} \quad \text{and} \quad B_n = A_{n-1} - \mathrm{QUOT}_n \cdot B_{n-1} \qquad (n \geq 1)$$

show that A_n, $B_n \in I_{n-1}$ and thus $I_n \subseteq I_{n-1}$. If we rewrite the second equation as

$$A_{n-1} = B_n + \mathrm{QUOT}_n \cdot A_n,$$

then we see that A_{n-1}, $B_{n-1} \in I_n$, and so $I_{n-1} \subseteq I_n$. We have $I_0 = aR+bR$ and $I_N = A_N \cdot R$, and we get

$$A_N \cdot R = aR + bR.$$

Lemma 1.70 now tells us that A_N is indeed a gcd of a and b. It remains to show that S_N and T_N have the desired property. We claim that the two equations

$$A = S \cdot a + T \cdot b \quad \text{and} \quad B = U \cdot a + V \cdot b$$

are invariants of the **while**-loop. They are trivially true after initialization. Let $1 \leq n \leq N$. Assuming that the stated equations are true after $n-1$ executions of thc loop, we see that they remain true after the next one:

$$\begin{aligned} A_n &= B_{n-1} \\ &= U_{n-1} \cdot a + V_{n-1} \cdot b \\ &= S_n \cdot a + T_n \cdot b, \end{aligned}$$

and

$$\begin{aligned} B_n &= A_{n-1} - \mathrm{QUOT}_n \cdot B_{n-1} \\ &= S_{n-1} \cdot a + T_{n-1} \cdot b - \mathrm{QUOT}_n \cdot (U_{n-1} \cdot a + V_{n-1} \cdot b) \\ &= U_n \cdot a + V_n \cdot b. \end{aligned}$$

In particular, we have $A_N = S_N \cdot a + T_N \cdot b$. □

The algorithm of Theorem 2.32 is called the **extended Euclidean algorithm**. If one computes just d but not s and t, it is called the **Euclidean algorithm**.

Exercise 2.33 (i) Compute the gcd of 124 and 56 in $\mathbb{Z}$, and integers s and t with $\gcd(124, 56) = 124s + 56t$.

(ii) Compute the gcd of $f = X^4 + X^2 + 1$ and $g = 2X^3 + X^2 + 2X + 1$ in $\mathbb{Z}/3\mathbb{Z}[X]$, and p, $q \in \mathbb{Z}/3\mathbb{Z}[X]$ with $\gcd(f, g) = pf + qg$.

Exercise 2.34 Let R be a computable Euclidean domain, $2 \leq m \in \mathbb{N}$, and let $a_1, \dots, a_m \in R$. Combine Theorem 2.32 and Lemma 1.79 to show how one can compute a gcd d of $a_1, \dots, a_m$ in R and $s_1, \dots, s_m \in R$ with $d = s_1a_1 + \cdots + s_ma_m$.

Exercise 2.35 Let K be a field and f, $g \in K[X]$ both non-zero. Show the following:

(i) If $0 \neq h \in \mathrm{Id}(f, g)$ is such that $\deg(h) < \deg(fg)$, then there exist s, $t \in K[X]$ with

$$h = sf + tg, \quad \deg(s) < \deg(g), \quad \text{and} \quad \deg(t) < \deg(f).$$

In particular, this holds for $h = \gcd(f, g)$. Moreover, polynomials s and t with these properties can be computed from f and g in case K is computable. (Hint: Use the extended Euclidean algorithm, then divide s by g with remainder.)

(ii) Show that s and t as in (i) are uniquely determined by h, f, and g if $\gcd(f, g) = 1$. (Hint: Use the fact that here, $\operatorname{lcm}(f, g) = fg$.)

(iii) Make up a counterexample to the claim of (ii) in case $\gcd(f, g)$ is not a constant.

Exercise 2.36 Let $2 \leq m \in \mathbb{N}$, and let $\boldsymbol{a}$ be an m-tuple of positive integers. Consider an algorithm which non-deterministically performs one of the following actions as long as this is possible.

(i) Find two entries a_i and a_j of $\boldsymbol{a}$ with $i \neq j$ and $a_i \geq a_j$, and replace a_i by $a_i - a_j$.

(ii) Drop a zero entry from $\boldsymbol{a}$.

Show that the algorithm always terminates and outputs a 1-tuple whose entry is the integer gcd of the entries of $\boldsymbol{a}$.

An important application of EXTEUC is the following method to compute inverses in the field $K = \mathbb{Z}/p\mathbb{Z}$ where p is a prime number. For $m \in \mathbb{Z}$, let us denote the residue class $m + \mathbb{Z}$ by $\overline{m}$. If $\overline{m} \neq 0$, then there exists $s \in \mathbb{Z}$ with $\overline{sm} = 1$ in K, i.e., $1 - sm \in p\mathbb{Z}$, which means that there exists $t \in \mathbb{Z}$ with $1 = sm + tp$. It follows that $\gcd(m, p) = 1$ (see the remark following Lemma 1.70), and we can thus compute s, $t \in \mathbb{Z}$ with $1 = sm + tp$ by means of the extended Euclidean algorithm. We see that $1 - sm \in p\mathbb{Z}$, which means that $\overline{s}$ is the inverse of $\overline{m}$ in K.

Note that by Lemma 1.69, Example 1.17, and Lemma 2.18, integer gcd's are unique up to a sign, whereas gcd's in $K[X]$, where K is a field, are unique up to a non-zero constant factor. The following corollary summarizes our present knowledge about gcd's in polynomial rings.

Corollary 2.37 *Let K be a field, f, $g \in K[X]$. Then f and g have a gcd d in $K[X]$, and there exist s, $t \in K[X]$ with $d = sf + tg$. Here, d is uniquely determined by f and g up to a non-zero constant factor. Moreover, if K is a computable field, then d, s, and t can be computed from f and g.* □

Apart from the fact that the Euclidean algorithm provides a way to effectively compute gcd's, it also has the remarkable theoretical consequence that polynomial gcd's are invariant under extensions of the ground field.

Proposition 2.38 *Let K' be a field, K a subfield of K', and f, $g \in K[X] \subseteq K'[X]$. Then the gcd of f and g in the ring $K[X]$ equals the one in the ring $K'[X]$.*

Proof We have stated the Euclidean algorithm only for computable field, because we already had a general existence proof for gcd's in PID's. But it is clear that the algorithm can also be viewed as an abstract mathematical construction that arrives at a gcd of f and g by means of a finite number of divisions with remainder starting with division of f by g. These divisions involve only addition, multiplication, and division of coefficients, and so the construction will be the same regardless of whether we view f and g as elements of $K[X]$ or of $K'[X]$. □

The reader should note that there are other constructions involving polynomials whose outcome depends strongly on the ground field: if, for example, we wish to factor $f = X^2 + 1$, then the result will depend on whether we view f as an element of $\mathbb{Q}[X]$ or of $\mathbb{C}[X]$.

We may now combine our results on computable Euclidean domains to obtain the following *ideal membership test.*

Proposition 2.39 *Let R be a computable Euclidean domain, and suppose $a, b_1, \ldots, b_m \in R$ are given. Then we can effectively decide whether or not $a \in \operatorname{Id}(b_1, \ldots, b_m)$.*

Proof By Theorem 2.32 and Lemma 1.79, it is possible to compute $d = \gcd(b_1, \ldots, b_m)$. By Exercise 1.80, d generates the ideal $\operatorname{Id}(b_1, \ldots, b_m)$, and Lemma 2.29 allows us to decide whether or not $a \in \operatorname{Id}(d)$. □

There are PID's that are not Euclidean. The class of non-Euclidean PID's is of little interest to us, though: we are now going to show that if we drop either one of the conditions that R be a field or that $n = 1$, then $R[X_1, \ldots, X_n]$ is no longer a PID (and hence, of course, not Euclidean).

Proposition 2.40 *Let R be a domain. Then $R[X_1, \ldots, X_n]$ is a PID iff R is a field and $n = 1$.*

Proof The implication "$\Longleftarrow$" is immediate from Proposition 2.24 and Proposition 2.30. For "$\Longrightarrow$," assume that R is not a field and $n = 1$. Then there exists a non-unit $a \neq 0$ of R, and we consider the ideal I generated by a and X in $R[X]$:

$$I = \operatorname{Id}(a, X) = \{\, af_1 + Xf_2 \mid f_1, f_2 \in R[X] \,\}.$$

Then I is a proper ideal, for if there would exist $f_1, f_2 \in R[X]$ with $1 = af_1 + Xf_2$, then the constant monomial of f_1 would be an inverse of a in R. Now assume for a contradiction that $I = \operatorname{Id}(g)$ for some $g \in R$. Then g is constant since $g \mid a$. It is easy to see that a constant g satisfying $g \mid X$ must be a unit of R. But then $\operatorname{Id}(g) = R[X]$, a contradiction. Finally, if $n > 1$, then $R[X_1, \ldots, X_n] = R[X_1, \ldots, X_{n-1}][X_n]$ and $R[X_1, \ldots, X_{n-1}]$ is not a field by Lemma 2.18 (iii), so $R[X_1, \ldots, X_n]$ is not a PID by the above argument. □

Exercise 2.41 Let R be a domain, $n \geq 2$. Show that any ideal generated by more than one of the inderminates is a non-principal ideal of $R[X_1, \ldots, X_n]$.

So all polynomial rings except those of the form $K[X]$ are not Euclidean and not PID's. Much of the theory of polynomial rings is concerned with the question of how bad it really is, i.e., how much of the nice properties of Euclidean domains (more of which we will soon discuss) can be saved for non-Euclidean polynomial rings. One classical result is the *Hilbert basis theorem*, which implies that every ideal of $K[X_1, \dots, X_n]$, where K is a field, is still finitely generated (see Definition 1.36). Rings with this property are called **noetherian**. Although Hilbert's theorem was obviously known long before the arrival of Gröbner bases, we will obtain a proof of it in the course of the development of Gröbner basis theory. Gröbner basis theory is actually a step further in the same direction: it shows that the division algorithm for $K[X]$ can be generalized to a kind of division of one polynomial by finitely many others in $K[X_1, \dots, X_n]$ in such a way that one still obtains an ideal membership test as in Proposition 2.39 (and many more nice algorithms). In the remaining sections of this chapter, we will pursue the theory of polynomial rings in a slightly different direction. We will discuss the theoretical foundations and some rudimentary algorithms concerning gcd computations, squarefree decomposition, and factoring of polynomials.

2.3 Unique Factorization Domains

Definition 2.42 Let R be a domain, $0 \neq a$ a non-unit of R. Then a is called

(i) **irreducible** if $a = bc$ implies that either b or c is a unit for all b, $c \in R$,

(ii) **prime** if $a \mid bc$ implies $a \mid b$ or $a \mid c$ for all b, $c \in R$.

Note that if $a \in R$ and u is a unit of R, then we can always write $a = u(u^{-1}a)$. An irreducible element of R is thus one that allows no factorizations other than such *trivial* ones. A non-trivial factorization is also called *proper*.

Lemma 2.43 Let R be a domain. Then every prime element of R is irreducible.

Proof Let $a \in R$ be prime, and assume for a contradiction that a is reducible, i.e., $a = bc$ with non-units b and c of R. Since $a \mid a$ and a is prime, we must have $a \mid b$ or $a \mid c$, say $a \mid b$. Then $b = ad$ for some $d \in R$, and thus $a = bc = adc$ which implies $dc = 1$, contradicting the fact that c is not a unit. □

Exercise 2.44 Let D be as in Exercise 1.24. Show that 2 is irreducible but not prime in D. (Hints: $2 \mid 6$, use the square of the norm. Cf. also the discussion following Lemma 1.69.)

Exercise 2.45 Let R be a domain, $a \in R$ prime. Show that if a divides a product of finitely many elements of R, then it divides one of the factors.

Exercises 2.46 Let R be a domain, $a \in R$ irreducible. Show the following:

(i) au is irreducible for every unit u of R.

(ii) For all $b \in R$, $b \mid a$ implies that b is a unit or a and b are associated.

(iii) For all $b \in R$, $a \nmid b$ implies that a and b are relatively prime.

Proposition 2.47 *Let R be a PID. Then every irreducible element of R is prime.*

Proof Let $a \in R$ be irreducible, b, $c \in R$ such that $a \mid bc$, and assume that $a \nmid b$. Then 1 is a gcd of a and b by Exercise 2.46 (iii). Since R is a PID, there exist s, $t \in R$ with $1 = sa + tb$. It follows that $c = sac + tbc$. From this and $a \mid bc$ we conclude that $a \mid c$. □

We see that for PID's, the notions of prime element and irreducible element coincide. In the case of the integers, it is customary to speak of **prime numbers**, or **primes**, whereas for univariate polynomials over a field, the expression **irreducible polynomials** is preferred. It is also important to note that from the point of view of ring theory, 2 is just as prime an integer as -2, but it is customary to require primes to be positive. The next lemma relates primeness and irreducibility of a ring element to properties of the ideal generated by the element.

Lemma 2.48 Let R be a domain, $a \in R$. Then the following hold:

(i) a is prime iff aR is a prime ideal.

(ii) If R is a PID, then a is irreducible iff aR is maximal.

Proof (i) The proof is the same as that of Lemma 1.91:

$$\begin{aligned} a \text{ prime} &\iff a \mid bc \text{ implies } a \mid b \text{ or } a \mid c \text{ for all } b, c \in R \\ &\iff bc \in aR \text{ implies } b \in aR \text{ or } c \in aR \text{ for all } b, c \in R \\ &\iff aR \text{ prime.} \end{aligned}$$

(ii) Lemmas 2.43 and 2.47 say that a is irreducible iff it is prime. By (i), a being prime is equivalent to aR being a prime ideal, and this is equivalent to aR being maximal by Proposition 1.97. □

Exercise 2.49 Give a direct proof of (ii) of the lemma above.

The above results together with those of the previous section show that as far as ideals and residue class rings are concerned, the polynomial rings $K[X]$, where K is a field, behave just like the integers: every ideal is principal, and a non-trivial ideal is prime iff it is maximal iff it is generated by an irreducible polynomial. Moreover, the residue class ring $K[X]/\mathrm{Id}(g)$ is computable for all $g \in K[X]$ whenever K is a computable field: if $f + \mathrm{Id}(g)$ is an arbitrary element of the residue class ring $K[X]/\mathrm{Id}(g)$, then we may divide f by g with unique quotient q and remainder r satisfying $r = 0$ or $\deg(r) < \deg(g)$, and conclude that $r + \mathrm{Id}(g) = f + \mathrm{Id}(g)$ since $f - r = qg \in \mathrm{Id}(g)$. This together with the definition of addition and multiplication in residue class rings rather obviously allows us to represent the residue classes on a computer in a unique way and to perform computations with them. Finally, if K is a computable field and $g \in K[X]$ is irreducible, then we claim that $K[X]/\mathrm{Id}(g)$ is even a computable field. Indeed, if $0 \neq f + \mathrm{Id}(g) \in K[X]$, then $g \nmid f$, so f and g are relatively prime by Exercise 2.46 (iii). By Corollary 2.37, we can compute polynomials s and t with $1 = sf + gt$, which means that $1 - fs \in \mathrm{Id}(g)$ and thus

$$1 + \mathrm{Id}(g) = \big(s + \mathrm{Id}(g)\big)\big(f + \mathrm{Id}(g)\big).$$

If we now replace s by its remainder upon division by g, then we have computed the inverse of f+Id(g). (Cf. the remarks following Theorem 2.32.)

Lemma 2.50 Let R be a Euclidean domain with abstract degree function φ, and let $0 \neq a \in R$. Then the following hold:

(i) If $a = bc$ is a proper factorization of a, i.e., b and c non-units of R, then $\varphi(b)$, $\varphi(c) < \varphi(a)$.

(ii) If $\varphi(a) = 0$, then a is a unit of R.

(iii) If $\varphi(a) = 1$ and a is not a unit, then it is irreducible.

Proof (i) By symmetry, it suffices to show $\varphi(b) < \varphi(a)$. By the definition of Euclidean domains, we have $\varphi(b) \leq \varphi(a)$. Assume for a contradiction that $\varphi(b) = \varphi(a)$. There exist q, $r \in R$ with

$$b = qa + r, \qquad \varphi(r) < \varphi(a) \quad \text{or} \quad r = 0.$$

From $r = b - qa = b - qbc$ we conclude that $b \mid r$ and thus $\varphi(r) \geq \varphi(b) = \varphi(a)$. This means that we must have $r = 0$, and hence $a = bc = qac$. It follows that $qc = 1$, contradicting the fact that c is not a unit.

(ii) Let q, $r \in R$ be a quotient and remainder of 1 upon division by a. Then we must have $r = 0$ because $\varphi(r) < \varphi(a) = 0$ is impossible, and thus $1 = qa$.

(iii) Assume for a contradiction that a is not a unit and has a proper factorization $a = bc$ in R. Then $\varphi(b) = \varphi(c) = 0$ by (i) above, and thus both b and c are units by (ii) above, a contradiction. □

Theorem 2.51 *Let R be a Euclidean domain, $0 \neq a$ a non-unit of R. Then a can be written as a product of irreducible elements (possibly consisting of just one factor). Moreover, this factorization is unique up to order and unit factors, i.e., whenever $p_1 \cdot \cdots \cdot p_k = q_1 \cdot \cdots \cdot q_m$ with p_i and q_j irreducible for $1 \leq i \leq k$ and $1 \leq j \leq m$, then $k = m$, and, possibly after renumbering, p_i and q_i are associated for $1 \leq i \leq m$.*

Proof We begin by proving the existence of the factorization. Let $0 \neq a$ be a non-unit of R. We proceed by induction on $m = \varphi(a)$. By (ii) of the last lemma, the induction starts at $m = 1$. If this is the case, then a is itself irreducible by (iii) of the last lemma. Now let $m > 1$. If a is irreducible, we are done. If not, then $a = bc$ with non-units b, $c \neq 0$ of R. By (i) of the lemma above, we have $\varphi(b)$, $\varphi(c) < m$. By induction hypothesis, both b and c have factorizations of the desired kind, and their product is clearly such a factorization of a. It remains to show uniqueness. Let $p_1, \ldots, p_k$, $q_1, \ldots, q_m \in R$ be irreducible with

$$p_1 \cdot \cdots \cdot p_k = q_1 \cdot \cdots \cdot q_m.$$

We prove our claim by induction on k. If $k = 1$, then we must have $m = 1$ too, since otherwise $p_1 = (q_1 \cdot \cdots \cdot q_{m-1})q_m$ would be a proper factorization of p_1 (Lemma 1.18 (iii)). It follows that $p_1 = q_1$. Now let $k > 1$. Obviously, $p_1 \,|\, p_1 \cdot \cdots \cdot p_k$, so $p_1 \,|\, q_1 \cdot \cdots \cdot q_m$. Since p_1 is prime by Proposition 2.47, we obtain $p_1 \,|\, q_j$ for some $1 \leq j \leq m$. Renumbering, we may assume $j = 1$. Since q_1 is irreducible and p_1, being irreducible, is not a unit by definition, p_1 and q_1 must be associated by Exercise 2.46 (ii), say $q_1 = up_1$ with u a unit. Substituting up_1 for q_1 and then cancelling p_1, our original equation becomes

$$p_2 \cdot \cdots \cdot p_k = uq_2 \cdot \cdots \cdot q_m.$$

By Exercise 2.46 (i), uq_2 is again irreducible. We may thus apply the induction hypothesis to conclude that $m = k$, and that, possibly after renumbering, p_i and q_i $(2 \leq i \leq m)$ are associated too. (We have used Exercise 1.68 (xiii).) □

Recall that a prime number, or prime, is a positive integer that is a prime element of $\mathbb{Z}$. We have thus proved that every integer other than 1, -1, or 0 can be written as a product of primes and a possible factor of -1, and that this *prime factor decomposition* is unique up to the order of the factors. For univariate polynomial rings over a field, we have proved that every non-zero non-constant polynomial can be expressed as a product of irreducible polynomials, and this decomposition into irreducible factors is unique up to unit factors and the order of the factors.

Definition 2.52 A domain R is called a **unique factorization domain**, or **UFD** for short, if the following two conditions hold:

(i) Every non-zero non-unit of R can be written as a product of irreducible elements, and

(ii) any such factorization is unique up to order and unit factors, i.e., whenever $p_1 \cdot \cdots \cdot p_k = q_1 \cdot \cdots \cdot q_m$ with p_i and q_j irreducible for $1 \leq i \leq k$ and $1 \leq j \leq m$, then $k = m$, and, possibly after renumbering, p_i and q_i are associated for $1 \leq i \leq m$.

We have already seen that every Euclidean domain is a unique factorization domain. It is even true that every PID is a unique factorization domain. The proof of this, although not hard, requires the use of the axiom of choice which will be discussed briefly in Chapter 3. We will give the proof there as an example of an application of the axiom of choice. More interestingly for us, we will show in the next section, by means of the *Gaussian lemma*, that a large class of non-Euclidean polynomial rings retains the unique factorization property. In view of this, it is interesting that UFD's still have a number of properties which we only know for PID's thus far.

Proposition 2.53 *In a UFD, every irreducible element is prime.*

Proof Let R be a UFD, $a \in R$ irreducible, and assume that $a \mid bc$ with b, $c \in R$. Then $bc = ad$ for some $d \in R$. Taking unique factorization into irreducible elements everywhere, we obtain

$$p_1 \cdot \cdots \cdot p_k \cdot q_1 \cdot \cdots \cdot q_l = a \cdot r_1 \cdot \cdots \cdot r_m,$$

where all factors are irreducible. By 2.52 (ii), there must exist a p_i $(1 \leq i \leq k)$ or a q_j $(1 \leq j \leq l)$ such that a and p_i or a and q_j are associated. One concludes easily that $a \mid b$ or $a \mid c$. □

The proposition above explains the fact that the factorization into irreducible elements in a UFD is often referred to as the **unique prime factor decomposition**.

Lemma 2.54 Let R be a UFD, $0 \neq a$ a non-unit of R. Then there exists a unit $u \in R$, irreducible, pairwise non-associated elements $p_1, \ldots, p_k \in R$, and positive natural numbers $\nu_1, \ldots, \nu_k$ such that

$$a = u \prod_{i=1}^{k} p_i^{\nu_i}.$$

Moreover, if

$$a = v \prod_{i=1}^{m} q_i^{\mu_i}$$

is another such representation, then $m = k$, and, possibly after renumbering, p_i and q_i are associated and $\mu_i = \nu_i$ for $1 \leq i \leq m$.

Proof Let $r_1 \cdot \cdots \cdot r_l$ be a factorization of a into irreducible elements. Whenever r_i and r_j are associated for some $1 \leq i < j \leq l$, then we can replace the product $r_i r_j$ by ur_i^2 with some unit $u \in R$. Since the product of units is again a unit, it is clear that we can arrive at the desired representation in this way. The stated uniqueness property is immediate from 2.52 (ii) and Exercise 1.68 (xiii). □

Exercise 2.55 Let R be a UFD. Show the following:

(i) If $0 \neq a$ is a non-unit of R, $p_1 \cdot \cdots \cdot p_k$ a factorization of a into irreducible elements, q an irreducible element of R, and $0 < \lambda \in \mathbb{N}$ with $q^\lambda \,|\, a$, then, possibly after renumbering, q is associated to p_i for $1 \leq i \leq \lambda$.

(ii) Let $0 \neq a$, b be non-units of R. Then the following condition is equivalent to $a \,|\, b$: whenever $p^\lambda \,|\, a$ for some irreducible $p \in R$ and $\lambda \in \mathbb{N}$, then $p^\lambda \,|\, b$.

Proposition 2.56 *Any two elements of a UFD have a gcd.*

Proof Let R be a UFD and a, $b \in R$. If one of a and b is 0 or a unit, then the claim follows from Exercise 1.68 (i), (iv), and (v). Otherwise, we produce a gcd seventh-grade style. Let

$$a = u \prod_{i=1}^{k} p_i^{\nu_i} \quad \text{and} \quad b = v \prod_{i=1}^{m} q_i^{\mu_i}$$

be the representations of a and b, respectively, as described in Lemma 2.54. W.l.o.g., we may assume that $k \leq m$. Using the same technique as in the proof of Lemma 2.54, we can, at the cost of getting a different u, change the representation of a in such a way that p_i and q_j are either equal or not associated for all $1 \leq i \leq k$ and $1 \leq j \leq m$. Possibly after renumbering, we can thus find an l with $1 \leq l \leq k$ such that $p_i = q_i$ for $1 \leq i \leq l$, and p_i and q_j are not associated for $l + 1 \leq i \leq k$ and $l + 1 \leq j \leq m$. We claim that

$$d = \prod_{i=1}^{l} p_i^{\lambda_i}$$

is a gcd of a and b, where $\lambda_i = \min(\nu_i, \mu_i)$ for $1 \leq i \leq l$. It is obvious that $d \,|\, a$ and $d \,|\, b$. Now let $d' \in R$ be any common divisor of a and b. If $r^\lambda \,|\, d'$ for some irreducible $r \in R$ and $\lambda \in \mathbb{N}$, then by Exercise 2.55 (i), r must be associated to some p_i and to some q_j ($1 \leq i \leq k$, $1 \leq j \leq m$). By Exercise 1.68 (xiii), we must have $1 \leq i \leq l$. Again by Exercise 2.55 (i), we conclude that $\lambda \leq \lambda_i$. Since r and λ were arbitrary, we finally get, by Exercise 2.55 (ii), $d' \,|\, d$. □

2.4 The Gaussian Lemma

The Gaussian lemma is a rather technical result on polynomials whose depth will only become apparent in its applications. We remind the reader of Lemma 2.18 (iii), which will be used frequently from now on.

Definition 2.57 Let R be a UFD and

$$0 \neq f = \sum_{i=0}^{m} a_i X^i \in R[X]$$

a univariate polynomial with coefficients in R. Then the gcd of $a_0, \dots, a_m$ in R is called the **content** of f and denoted by $\mathrm{c}(f)$. Since $\mathrm{c}(f)$ divides every coefficient of f, it is clear that f can be written in the form $f = \mathrm{c}(f) \cdot g$ with $g \in R[X]$, and this g is called the **primitive part** of f, denoted by $\mathrm{pp}(f)$. f is called **primitive** if $\mathrm{c}(f) = 1$.

With the obvious convention that the gcd of one ring element be that element, the content of a polynomial exists by Proposition 2.56 and Lemma 1.79 (ii). Being a gcd, it is unique only up to a unit factor, but as with gcd's, there will be no harm in speaking of *the* content and *the* primitive part. Note that by Lemma 1.79 (iii), the gcd of $a_0, \dots, a_m$ in the definition of the content is the same as the gcd of those a_i that are non-zero. The statements of the following exercise are easy consequences of Lemma 1.79 (iv). They will be of utmost importance in the rest of this chapter.

Exercise 2.58 Let R be a UFD, $0 \neq f \in R[X]$. Show the following:

(i) $\mathrm{pp}(f)$ is a primitive polynomial.

(ii) The decomposition of f into the product of content and primitive part is unique up to unit factors in the following sense: whenever $f = ag = bh$ with a, $b \in R$ and g, $h \in R[X]$ primitive, then a and b are associated, and so are g and h.

Exercises 2.59 (i) What are $\mathrm{c}(f)$ and $\mathrm{pp}(f)$ if $f = 12X^3 - 3X^2 + 15 \in \mathbb{Z}[X]$, and what are they if $f = (X^2-2X+1)Y^5+(X^3-1)Y^2-(X^2-1) \in \mathbb{Z}[X][Y]$?

(ii) Show that if K is a field, then every non-zero polynomial in $K[X]$ is primitive.

(iii) If $0 \neq f \in R[X]$ and $0 \neq d \in R$, then $\mathrm{c}(df) = d \cdot \mathrm{c}(f)$.

Theorem 2.60 (Gaussian Lemma) *Let R be a UFD. Then the product of two primitive polynomials in $R[X]$ is again primitive.*

Proof Let $0 \neq f$, $g \in R[X]$ be primitive, and assume for a contradiction that their product fg was not primitive. We can write

$$f = \sum_{i=0}^{k} a_i X^i, \quad g = \sum_{i=0}^{m} b_i X^i, \quad \text{and} \quad fg = \sum_{i=0}^{k+m} c_i X^i$$

where all coefficients are in R. By our assumption, $c_1, \dots, c_{k+m}$ have a gcd in R which is not a unit. By Exercise 2.55 (ii), there must exist an irreducible $p \in R$ with $p \,|\, c_i$ for $1 \leq i \leq k+m$. Since f and g are primitive, there exist $1 \leq i \leq k$ and $1 \leq j \leq m$ such that $p \nmid a_i$ and $p \nmid b_j$, and we may assume that i and j are each minimal with that respective property. Now

$$c_{i+j} = a_i b_j + a_{i-1} b_{j+1} + a_{i+1} b_{j-1} + a_{i-2} b_{j+2} + \cdots.$$

Because of the minimality of i and j, each summand on the right except for the first one is divisible by p, and so is c_{i+j}. It follows (Exercise 1.68 (vi)) that $p \mid a_i b_j$. Since p is prime by Proposition 2.53, we obtain $p \mid a_i$ or $p \mid b_j$, a contradiction. □

Recall from Theorem 1.117 that Q_R stands for the field of fractions of a domain R. Note that $R[X_1, \ldots, X_n]$ is a subring of $Q_R[X_1, \ldots, X_n]$ for all $1 \leq n \in \mathbb{N}$. If K is a field, then the field of fractions of the domain $K[X_1, \ldots, X_n]$ is also called the **rational function field** over K in X_1, $\ldots$, X_n and is denoted by $K(X_1, \ldots, X_n)$.

Exercise 2.61 Let R be a domain. Show that $Q_{R[X_1,\ldots,X_n]}$ equals the rational function field $Q_R(X_1, \ldots, X_n)$.

The last statement of the following corollary is sometimes also referred to as the Gaussian lemma.

Corollary 2.62 *Let R be a UFD, $0 \neq f \in R[X]$. Then the following hold:*

(i) If $f = gh$ where $g \in R[X]$ is primitive and $h \in Q_R[X]$, then $h \in R[X]$.

(ii) If g, $h \in Q_R[X]$ with $f = gh$, then there exist a, $b \in Q_R$ with ag, $bh \in R[X]$ and $f = (ag)(bh)$. In particular, if f is irreducible in $R[X]$, then it is irreducible in $Q_R[X]$.

Proof (i) Let d be the product of all denominators of coefficients of h. Then $dh \in R[X]$, and we may write

$$df = d \cdot \mathrm{c}(f) \cdot \mathrm{pp}(f) = gdh = \mathrm{c}(dh) \cdot g \cdot \mathrm{pp}(dh).$$

By Exercise 2.58 (i), Theorem 2.60 and our assumption on g, $\mathrm{pp}(f)$ and $g \cdot \mathrm{pp}(dh)$ are primitive. By Exercise 2.58 (ii), $\mathrm{c}(dh) = ud \cdot \mathrm{c}(f)$ for some unit u of R. Substituting this into the above equation and cancelling d, we obtain

$$f = g \cdot \big(u \cdot \mathrm{c}(f) \cdot \mathrm{pp}(dh)\big).$$

But we already know that $f = gh$, so it follows that $h = u \cdot \mathrm{c}(f) \cdot \mathrm{pp}(dh) \in R[X]$.

(ii) Let d be the product of all denominators of coefficients of g. Then $dg \in R[X]$, and we obtain the equation

$$f = dg \cdot \frac{1}{d} \cdot h = \mathrm{pp}(dg) \cdot \frac{\mathrm{c}(dg)}{d} \cdot h.$$

By (i) above, $(\mathrm{c}(dg)/d)\cdot h \in R[X]$. Since $\mathrm{pp}(dg) = (d/\mathrm{c}(dg))\cdot g$, we see that

$$a = \frac{d}{\mathrm{c}(dg)} \quad \text{and} \quad b = \frac{\mathrm{c}(dg)}{d}$$

have the desired properties. □

In elementary mathematics, statement (ii) above is rarely mentioned but often confirmed by experience: if a polynomial with integer coefficients cannot be factored over the integers (i.e., into proper factors with integer coefficients), then it cannot be factored over the rationals either. The next best shot is then a factorization over the reals.

Exercise 2.63 Generalize Corollay 2.62 (ii) to factorizations of $f \in R[X]$ into more than two factors.

Exercise 2.64 Prove the "trivial directions" of Theorem 2.60 and Corollary 2.62:

(i) if f, g, $h \in R[X]$ with f primitive and $f = gh$, then both g and h are primitive, and

(ii) if $f \in R[X]$ is primitive and irreducible in $Q_R[X]$, then it is irreducible in $R[X]$.

The condition that f be primitive cannot be dropped in (ii) above: $2X + 2 \in \mathbb{Z}[X]$ is irreducible in $\mathbb{Q}[X]$ since it is linear and can therefore only be factored into a constant (i.e., a unit of $\mathbb{Q}[X]$) and a linear polynomial. In $\mathbb{Z}[X]$, however, $2(X + 1)$ is a proper factorization.

Theorem 2.65 *If R is a UFD, then so is $R[X]$.*

Proof Suppose $0 \neq f$ is a non-unit of $R[X]$. Let

$$\mathrm{c}(f) = p_1 \cdot \cdots \cdot p_k \qquad (p_1, \ldots, p_k \in R)$$

and

$$\mathrm{pp}(f) = p_{k+1} \cdot \cdots \cdot p_m \qquad (p_{k+1}, \ldots, p_m \in Q_R[X])$$

be the prime factor decompositions of $\mathrm{c}(f)$ and $\mathrm{pp}(f)$ in the UFD's R and $Q_R[X]$, respectively. Since f is not a unit, at most one of $\mathrm{c}(f)$ and $\mathrm{pp}(f)$ can be a unit, in which case we can choose that factor to be 1 and disregard it in the above decompositions and the following discussion. If one of the p_i $(1 \leq i \leq k)$ had a proper factorization in $R[X]$, this would have to be a factorization into constants, and it would be proper in R too, which is impossible. Hence p_i is irreducible in $R[X]$ for $1 \leq i \leq k$. As for the remaining p_i, we may, by Exercise 2.63, lift them into $R[X]$ without changing their product by means of multiplication with constants from Q_R. Since they were only unique up to unit factors (i.e., constant factors from Q_R) anyway, we may as well assume that they are already in $R[X]$. They are primitive by Exercise 2.64 (i) because their product equals the primitive polynomial $\mathrm{pp}(f)$, and irreducible in $Q_R[X]$, so they are irreducible in $R[X]$ by Exercise 2.64 (ii). Hence

$$f = p_1 \cdot \cdots \cdot p_m$$

is a decomposition of f into irreducible elements of $R[X]$.

It remains to show that this decomposition is unique up to order and unit factors. Let

$$p_1 \cdot \cdots \cdot p_k = q_1 \cdot \cdots \cdot q_m,$$

where all factors are irreducible elements of $R[X]$. Let $p_1, \ldots, p_{k'}, q_1, \ldots, q_{m'}$ be the constants among the factors in the above equation. If one of k' and m' or both are 0, or $k' = k$ and $m' = m$ (just one of these is obviously impossible), the argument below must be modified accordingly. Each of $p_{k'+1}, \ldots, p_k, q_{m'+1}, \ldots, p_m$ must be primitive, since otherwise the factorization into content and primitive part would be proper. With the Gaussian lemma, we obtain

$$\underbrace{p_1 \cdot \cdots \cdot p_{k'}}_{\in R} \cdot \underbrace{p_{k'+1} \cdot \cdots \cdot p_k}_{\text{primitive}} = \underbrace{q_1 \cdot \cdots \cdot q_{m'}}_{\in R} \cdot \underbrace{q_{m'+1} \cdot \cdots \cdot q_m}_{\text{primitive}} .$$

We conclude from Exercise 2.58 (ii) that there exists a unit u of R with

$$p_1 \cdot \cdots \cdot p_{k'} = uq_1 \cdot \cdots \cdot q_{m'} \quad \text{and} \quad p_{k'+1} \cdot \cdots \cdot p_k = u^{-1}q_{m'+1} \cdot \cdots \cdot q_m.$$

Since we claim uniqueness only up to unit factors, we may replace q_1 by uq_1 and $q_{m'+1}$ by $u^{-1}q_{m'+1}$. Since R is a UFD, the first equation implies that $k' = m'$ and that, possibly after renumbering, p_i and q_i are associated in R and thus in $R[X]$ for $1 \le i \le m'$. For $k'+1 \le i \le k$ and $m'+1 \le j \le m$, p_i and q_j are irreducible in $Q_R[X]$ by Corollary 2.62 (ii). Since the latter ring is a UFD by Theorem 2.51, it follows that $k = m$ and, possibly after renumbering, p_i and q_i are associated in $Q_R[X]$ for $m' + 1 \le i \le m$. This means that there are $a_i, b_i \in R$ with $0 \ne b_i$ $(m' + 1 \le i \le m)$ such that

$$p_i = \frac{a_i}{b_i} q_i, \quad \text{i.e.,} \quad b_i p_i = a_i q_i.$$

We have already observed that the p_i and q_i $(m' + 1 \le i \le m)$ must be primitive. Exercise 2.58 (ii) allows us to conclude that $a_i = u_i b_i$ with units u_i of R. This means that $a_i/b_i = u_i \in R$, and we see that p_i and q_i are actually associated in $R[X]$. □

Corollary 2.66 *If R is a UFD, then so is $R[X_1, \ldots, X_n]$ for all $1 \le n \in \mathbb{N}$.*

Proof We use induction on n. If $n = 1$, then the claim is identical with Theorem 2.65. If $n > 1$, then $R[X_1, \ldots, X_{n-1}]$ is a UFD by induction hypothesis, so again by Theorem 2.65, $R[X_1, \ldots, X_n] = R[X_1, \ldots, X_{n-1}][X_n]$ is a UFD too. □

Corollary 2.67 *If K is a field, then $K[X_1, \dots, X_n]$ is a UFD for all $1 \leq n \in \mathbb{N}$.*

Proof The proof is the same as the proof of Corollary 2.66, except that Theorem 2.51 is used for the case $n = 1$. □

Note that by the last two corollaries, all polynomial rings over $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{Z}/p\mathbb{Z}$ are UFD's. So in all of these, every polynomial has a unique factorization into irreducible ones, and any two polynomials have a gcd. Considering that these polynomial rings are computable, the question arises if we can actually compute gcd's and factorizations in these cases. Thus far, we have a positive answer only for gcd's in the Euclidean polynomial rings $\mathbb{Q}[X]$ and $\mathbb{Z}/p\mathbb{Z}[X]$. The rest of this chapter provides the missing algorithms.

2.5 Polynomial Gcd's

There are a variety of techniques for the fast computation of polynomial gcd's. They are based on the extended Euclidean algorithm combined with the following two lemmas.

Lemma 2.68 Let R be a UFD and f, $g \in R[X]$ with f, $g \neq 0$. Suppose d is a gcd of $\mathrm{c}(f)$ and $\mathrm{c}(g)$ in R and h a gcd of $\mathrm{pp}(f)$ and $\mathrm{pp}(g)$ in $R[X]$. Then dh is a gcd of f and g in $R[X]$.

Proof Clearly, d divides $\mathrm{c}(f)$ and $\mathrm{c}(g)$ in R and hence in $R[X]$, and h divides $\mathrm{pp}(f)$ and $\mathrm{pp}(g)$ in $R[X]$. Consequently, dh is a common divisor of f and g by Exercise 1.68 (xiv). Now let q be any common divisor of f and g in $R[X]$. Then there exist s, $t \in R[X]$ with $f = qs$ and $g = qt$. Removing contents, we obtain

$$\begin{aligned} \mathrm{c}(f) \cdot \mathrm{pp}(f) &= \mathrm{c}(q) \cdot \mathrm{c}(s) \cdot \mathrm{pp}(q) \cdot \mathrm{pp}(s), \quad \text{and} \\ \mathrm{c}(g) \cdot \mathrm{pp}(g) &= \mathrm{c}(q) \cdot \mathrm{c}(t) \cdot \mathrm{pp}(q) \cdot \mathrm{pp}(t). \end{aligned}$$

The products of the primitive parts are again primitive by the Gaussian lemma; hence there exist units u, $v \in R$ with

$$\begin{aligned} \mathrm{c}(f) &= u \cdot \mathrm{c}(q) \cdot \mathrm{c}(s) \\ \mathrm{c}(g) &= v \cdot \mathrm{c}(q) \cdot \mathrm{c}(t) \end{aligned} \quad \text{and} \quad \begin{aligned} \mathrm{pp}(f) &= u^{-1} \cdot \mathrm{pp}(q) \cdot \mathrm{pp}(s) \\ \mathrm{pp}(g) &= v^{-1} \cdot \mathrm{pp}(q) \cdot \mathrm{pp}(t). \end{aligned}$$

We see that $\mathrm{c}(q)$ is a common divisor of $\mathrm{c}(f)$ and $\mathrm{c}(g)$ in R, and that $\mathrm{pp}(q)$ is a common divisor of $\mathrm{pp}(f)$ and $\mathrm{pp}(g)$ in $R[X]$. Since d and h were gcd's, $\mathrm{c}(q)$ must divide d in R and hence in $R[X]$, and $\mathrm{pp}(q)$ must divide h in $R[X]$. Again by Exercise 1.68 (xiv), it follows that $q = \mathrm{c}(q) \cdot \mathrm{pp}(q)$ divides dh in $R[X]$. □

Lemma 2.69 Let R be a UFD, $0 \neq f$, $g \in R[X]$ primitive polynomials, h a gcd of f and g in $Q_R[X]$. Let d be the product of all denominators of coefficients of h, so that $dh \in R[X]$. Then the primitive part $\mathrm{pp}(dh)$ of dh is a gcd of f and g in $R[X]$.

Proof It is clear that $\mathrm{pp}(dh) \in R[X]$. Moreover, h and $\mathrm{pp}(dh)$ are associated in $Q_R[X]$ since $\mathrm{pp}(dh) = (d/\mathrm{c}(dh)) \cdot h$, and hence $\mathrm{pp}(dh)$ is still a gcd of f and g in $Q_R[X]$. It remains to show that this is the case even in $R[X]$. $\mathrm{pp}(dh)$ is a common divisor of f and g in $Q_R[X]$, f, g, and $\mathrm{pp}(dh)$ are in $R[X]$, and $\mathrm{pp}(dh)$ is primitive. By Corollary 2.62 (i), it follows that $\mathrm{pp}(dh)$ is a common divisor of f and g in $R[X]$. Now let q be any common divisor of f and g in $R[X]$. Then trivially, q is a common divisor of f and g in $Q_R[X]$, so $q \mid \mathrm{pp}(dh)$ in $Q_R[X]$. q is primitive by Exercise 2.64 (i) since it divides the primitive polynomial f in $R[X]$, so by Corollary 2.62 (i), $q \mid \mathrm{pp}(dh)$ in $R[X]$. □

Theorem 2.70 *Let R be a computable ring that is a UFD and for which an algorithm is known that computes the gcd of any two elements. Then one can find an algorithm that computes the gcd of any two polynomials in $R[X]$.*

Proof Let $0 \neq f$, $g \in R[X]$. By the two previous lemmas, it suffices to compute gcd's of the contents and the primitive parts of f and g in R and $Q_R[X]$, respectively, then lift the latter to $R[X]$ by multiplying it by the product of all denominators of coefficients and then taking its primitive part, and finally multiplying those two gcd's together. Factoring f and g into content and primitive part is a gcd computation in R which can be performed by assumption. The same is true for finding the gcd of the contents. Now $Q_R[X]$ is a Euclidean domain. To see that we can actually compute gcd's in this domain by means of the Euclidean algorithm, we must, by Corollary 2.37, convince ourselves that Q_R is a computable field. It is clear that if we can add and multiply elements of R, then we can add and multiply fractions, and we can certainly invert them simply by turning them upside down. We may even, if we wish, reduce them to lowest terms (i.e., make numerator and denominator relatively prime) by a gcd computation in R. Even when reduced to lowest terms, though, two fractions that are actually the same may still look very different due to the presence of unit factors other than -1 in numerator and denominator. We can, however, effectively *decide* whether or not two fractions are equal: p/q equals r/s iff $ps = rq$, and the latter condition can be decided by a computation in R. This decidability of equality is good enough for effective computations, even though there is no longer a computable unique representthe for each element. (Sections 4.5 and 4.6 have a more rigorous discussion of the phenomenon.) The lifting back to $R[X]$ of the gcd computed in $Q_R[X]$ and the final multiplication can obviously be performed effectively. □

Corollary 2.71 *Let R be either a computable field, or a computable ring that is a UFD and for which an algorithm is known that computes the gcd of any two elements. Then one can find an algorithm that computes the gcd of any two polynomials in $R[X_1, \dots, X_n]$.*

Proof We proceed by induction on n. If $n = 1$, then the claim is identical with Corollary 2.37 or Theorem 2.70. For $n > 1$,

$$R[X_1, \dots, X_n] = R[X_1, \dots, X_{n-1}][X_n],$$

and $R[X_1, \dots, X_{n-1}]$ is a computable ring that is a UFD and allows effective gcd computations by the induction hypothesis. Again our claim follows from Theorem 2.70. □

By the above corollary, we can effectively compute gcd's in $R[X_1, \dots, X_n]$ when R is $\mathbb{Z}$, $\mathbb{Q}$, or $\mathbb{Z}/p\mathbb{Z}$. The inductive proof of the corollary of course translates into a recursive algorithm.

Exercise 2.72 Write a programming-style version of the algorithm that is implicit in the proof of Corollary 2.71.

Exercise 2.73 Find the gcd in $\mathbb{Z}[X, Y]$ of

$$f = X^2Y^2 - XY^2 + 2X^2Y - 2Y^2 - 2XY + X^2 - 4Y - X - 2 \quad \text{and}$$

$$g = XY^2 + X^2Y + Y^2 + 2XY + X^2 + Y + X.$$

Do this one by hand. You won't have to do a Euclidean algorithm. You will get an opportunity to use your nifty computer algebra system in the next section.

It is noteworthy that in the multivariate case, we can not in general write the gcd of f and g in the form $sf + tg$. The gcd of X and Y in $R[X, Y]$, for example, equals 1, but we cannot write $1 = tX + sY$. This is of course due to the fact that $\mathrm{Id}(X, Y)$ is not a principal ideal and hence is not generated by the gcd of X and Y (cf. Lemma 1.70 and Exercise 2.41).

2.6 Squarefree Decomposition of Polynomials

In this section, we will frequently make use of the fact that a polynomial ring over a field is a UFD, and that the notions of prime and irreducible elements coincide in UFD's.

Let R be a UFD and $0 \neq a \in R$ a non-unit of R. By Lemma 2.54, there exists a unit $u \in R$, pairwise non-associated, irreducible elements p_1, $\dots, p_k \in R$, and positive natural numbers $\nu_1, \dots, \nu_k$ with $a = up_1^{\nu_1} \cdot \dots \cdot p_k^{\nu_k}$. We may now combine all factors that carry the same exponent and fill up the product with factors of the form 1^ν ($\nu \in \mathbb{N}$) to arrive at a representation of the form

$$a = ub_1b_2^2b_3^3 \cdot \dots \cdot b_m^m,$$

where $b_1, \dots, b_m$ are pairwise relatively prime, and each of them is a product of irreducible elements that are pairwise relatively prime. This latter property can obviously also be expressed by saying that whenever $p^s \mid b_i$ with $p \in R$ irreducible, then $s \leq 1$. Yet another way of saying it is that $p^2 \nmid b_i$ whenever $p \in R$ is irreducible.

Definition 2.74 Let R be a UFD. An element $a \in R$ is called **squarefree** if $p^2 \nmid a$ whenever $p \in R$ is irreducible. Now let $0 \neq a$ be a non-unit of R. A **squarefree decomposition** of a is a representation of the form

$$a = ub_1b_2^2b_3^3 \cdot \dots \cdot b_m^m,$$

where $u \in R$ is a unit and $b_1, \dots, b_m \in R$ are squarefree and pairwise relatively prime. The product $b_1 \cdot \dots \cdot b_m$ is then called a **squarefree part** of a.

Note that by our definition, every unit of R is squarefree, because it is not divisible by any non-unit at all. The zero element of R, however, is not squarefree because it is divided by everything. The label "squarefree part of a" for the product $b_1 \cdot \dots \cdot b_m$ in the definition above suggests, quite strongly, that this product is squarefree. The next lemma says that this is indeed so.

Lemma 2.75 Let R be a UFD, and let $b_1, \dots, b_m \in R$. Then the following are equivalent:

(i) $b_1, \dots, b_m$ are squarefree and pairwise relatively prime.

(ii) The product $b = b_1 \cdot \dots \cdot b_m$ is squarefree.

Proof Using the unique prime factor decomposition, it is easy to see that any $a \in R$ is squarefree iff it is either a unit or a product of pairwise relatively prime irreducible elements. Now if this is true for $b = b_1 \cdot \dots \cdot b_m$, then it is clearly true for each factor b_i. Conversely, if it is true for each b_i and the b_i are pairwise relatively prime, then it is true for their product b: if we write down the product of the prime factor decompositions of those b_i that are not units, then there cannot occur a pair of associated prime factors, because these would either have to come from one factor b_i or from two factors b_i and b_j. □

Proposition 2.76 *Let R be a UFD, and let $0 \neq a \in R$. Then the following hold:*

(i) a has a squarefree decomposition in R.

(ii) Whenever a is not a unit and

$$a = ub_1b_2^2b_3^3 \cdot \dots \cdot b_m^m = vc_1c_2^2c_3^3 \cdot \dots \cdot c_n^n$$

are squarefree decompositions of a with non-units b_m and c_n, then $m = n$, and b_i and c_i are associated for $1 \leq i \leq n$.

Proof (i) If a is a unit, then it is its own squarefree decomposition. Otherwise, the discussion at the beginning of this section shows how to arrive at a squarefree decomposition of a.

(ii) We manipulate each of the two given squarefree decompositions as follows.

- Combine all unit factors to one unit and move it up front.
- Write out each of the remaining factors, which must now be of the form f^s with f squarefree and not a unit, in the form $f^s = p_1^s \cdot \cdots \cdot p_r^s$ with $p_1, \ldots, p_r$ irreducible and pairwise relatively prime.

It is easy to see that we obtain two represenations of a as discussed in Lemma 2.54. This means that up to order and unit factors, we are looking at the same prime factors and exponents, and this easily implies our claim. □

We will continue our bad habit of speaking of *the* such-and-such if such-and-such is unique up to unit factors.

In this section, we will show how one can compute squarefree decompositions (meaning find the individual b_i) in certain polynomial rings. From a theoretical point of view, one could show how to do the complete factorization into irreducible elements in these polynomial rings and then get squarefree decompositions as a by-product. However, the discussion of effective squarefree decompositions will provide us with quite a bit of mathematical insight that will be useful later. Moreover, it turns out that squarefree decompositions can be computed much faster directly. There are situations where one needs no more than the squarefree decomposition, and even if one is looking for the complete factorization, it is much more efficient to do the squarefree decomposition first and then factor each of the b_i, which then of course provides the complete factorization of the original input polynomial.

Definition 2.77 Let R be a ring,

$$f = \sum_{i=0}^{m} a_i X^i$$

a polynomial in $R[X]$. Then the **derivative** f' of f is defined as

$$f' = \sum_{i=1}^{m} i \cdot a_i X^{i-1}.$$

(See Section 1.9 for the meaning of $i \cdot a_i$.)

The following exercise is tedious but straighforward.

Exercise 2.78 Let R be a ring and f, $g \in R[X]$. Show the following:

(i) $(f+g)' = f' + g'$.

(ii) $(fg)' = f'g + fg'$.

(iii) $(f^m)' = m \cdot f' f^{m-1}$ for $m \in \mathbb{N}$. (Hint: This is easily proved from (ii).)

(iv) $(f(g))' = f'(g) \cdot g'$. (Hint: Use (i)–(iii).)

Note that for $R = \mathbb{R}$, the algebraic definition of the derivative is the same as the one that is used in calculus. However, derivatives have a way of behaving somewhat strangely when $\text{char}(R) \neq 0$, as the following two lemmas show.

Lemma 2.79 Let R be a domain and $f \in R[X]$. Then the following hold:

(i) If $\text{char}(R) = 0$, then $f' = 0$ iff f is constant.

(ii) If $\text{char}(R) = p \neq 0$, then $f' = 0$ iff there exists $g \in R[X]$ with $f = g(X^p)$.

Proof Let $f = \sum_{i=0}^{m} a_i X^i$. Then we have

$$f' = \sum_{i=1}^{m} i \cdot a_i X^{i-1}.$$

We see that in the characteristic zero case, $f' = 0$ is equivalent to $a_i = 0$ for all $i > 0$. If $\text{char}(R) = p \neq 0$, then we conclude with Lemma 1.101 that $f' = 0$ is equivalent to $a_i = 0$ for all i with $p \nmid i$, and so it is equivalent to f being of the form

$$f = \sum_{i=0}^{m'} a_{ip} X^{ip} = \sum_{i=0}^{m'} a_{ip} (X^p)^i. \quad \square$$

Lemma 2.80 Let R be a domain with $\text{char}(R) = p \neq 0$, and let $f \in R[X]$. If $f = g^p$ for some $g \in R[X]$, then $f' = 0$. If, in addition, R is finite (and hence a field), then the converse is true too: $f' = 0$ implies that there exists $g \in R[X]$ with $g^p = f$.

Proof If $f = g^p$, then $f' = p \cdot g' g^{p-1} = 0$ by Lemma 1.101 and Exercise 2.20. Next, let R be a finite field and assume that $f' = 0$. Then $f = g(X^p)$ for some $g \in R[X]$ by the previous lemma, say

$$f = \sum_{i=0}^{m} a_i (X^p)^i.$$

Since every element of R has a pth root by Lemma 1.107, we can write

$$f = \sum_{i=0}^{m} b_i^p (X^i)^p = \left(\sum_{i=0}^{m} b_i X^i\right)^p,$$

the latter equation being true by Lemma 1.106. $\square$

Lemma 2.81 Let K be a field, $\operatorname{char}(K) = p$, and assume that K is finite or $p = 0$. Let $f, q \in K[X]$ with q irreducible, k a positive integer such that $q^k \mid f$ but $q^{k+1} \nmid f$. Then the following hold:

(i) If $p \nmid k$ (in particular, if $p = 0$), then $q^{k-1} \mid f'$ and $q^k \nmid f'$.

(ii) If $p \mid k$, then $q^k \mid f'$.

Proof From $q^k \mid f$ and $q^{k+1} \nmid f$ we conclude that $f \neq 0$ and that there exists $g \in K[X]$ with $f = gq^k$ and $q \nmid g$. We obtain

$$f' = (gq^k)' = g'q^k + k \cdot gq'q^{k-1}.$$

If $p \mid k$, then the last summand is zero by Lemma 1.101, and we see that $q^k \mid f'$. Now assume that $p \nmid k$. It is obvious that q^{k-1} must divide f'. If $p = 0$, then $q' \neq 0$ since q is not a constant. If $p \neq 0$, then $q' \neq 0$ because q, being irreducible, is not a pth power. We see that the last summand is not zero in either case. Assume for a contradiction that $q^k \mid f'$. Then $q \mid g'q + k \cdot gq'$. It follows that $q \mid k \cdot gq'$. Viewing $k \cdot gq'$ as $g(k \cdot q')$ and using the fact that q is prime, we see that $q \mid g$ or $q \mid k \cdot q'$. The former contradicts an earlier conclusion, the latter is impossible since $\deg(k \cdot q') < \deg(q)$. □

Recall that by Proposition 2.56, the gcd of two elements a and b of a UFD can be produced by collecting all prime factors that a and b have in common, where associated ones are treated as equal. From this together with the last two lemmas, we obtain the following two results.

Lemma 2.82 Let K be a field with $\operatorname{char}(K) = 0$, f a non-constant polynomial in $K[X]$, $f = cg_1g_2^2 \cdot \dots \cdot g_m^m$ the squarefree decomposition of f. Then

$$\gcd(f, f') = g_2g_3^2 \cdot \dots \cdot g_m^{m-1},$$

and $f/\gcd(f, f') = cg_1g_2g_3 \cdot \dots \cdot g_m$ is the squarefree part of f. □

Lemma 2.83 Let K be a finite field with $\operatorname{char}(K) = p \neq 0$, f a non-constant polynomial in $K[X]$, $f = cg_1g_2^2 \cdot \dots \cdot g_m^m$ the squarefree decomposition of f.

(i) If $g_i = 1$ for all $i \notin p\mathbb{Z}$, i.e., if f is a pth power, then $f' = 0$.

(ii) Otherwise,

$$\gcd(f, f') = \prod_{\substack{1 \le i \le m \\ i \in p\mathbb{Z}}} g_i^i \cdot \prod_{\substack{1 \le i \le m \\ i \notin p\mathbb{Z}}} g_i^{i-1}. \ \square$$

Before we show how to compute squarefree decompositions, we point out that—as so often happens with algorithmic problems—it is much easier to *decide* whether a given decomposition $f = cg_1g_2^2 \cdot \dots \cdot g_m^m$ of a polynomial is the squarefree one or not. What we must be able to decide is whether

$g_1 \cdot \cdots \cdot g_m$ is squarefree. We will show that over the kind of field that we are considering here, squarefreeness of f is equivalent to $\gcd(f, f') = 1$, a condition that can be decided by means of the Euclidean algorithm. One direction is actually true for an arbitrary field.

Lemma 2.84 Let K be a field, and let $f \in K[X]$ with $\gcd(f, f') = 1$. Then f is squarefree.

Proof Assume for a contradiction that f is not squarefree. Then there exist g, $h \in K[X]$ with g non-constant and $f = g^2h$. It follows that

$$f' = 2 \cdot gg'h + g^2h',$$

and we see that g is a common divisor of f and f'. □

Using Lemmas 2.82 and 2.83, it is not hard to prove the following partial converse to the lemma above. (The case where f is constant is trivial.) Recall from Section 1.9 that fields of characteristic zero are necessarily infinite.

Lemma 2.85 Let K be a field such that either $\mathrm{char}(K) = 0$ or K is finite, and let f be a squarefree polynomial in $K[X]$. Then $\gcd(f, f') = 1$. □

It will be proved in Section 7.3 that this last lemma actually holds for a larger class of fields, and it will also be shown that it is not true over every field.

Proposition 2.86 *Let K be a computable field such that either* $\mathrm{char}(K) = 0$ *or K is finite. Then one can find an algorithm that computes the squarefree decomposition of any non-constant polynomial in $K[X]$.*

Proof Since constant factors, being units of $K[X]$, are irrelevant to the problem, we may divide the input polynomial by its highest coefficient and assume henceforth that all polynomials involved have highest coefficient 1. Let $f \in K[X]$ with squarefree decomposition $f = g_1g_2^2 \cdot \cdots \cdot g_m^m$. Assume first that $\mathrm{char}(K) = 0$. Set $F_0 = f$, and for $1 \leq i \in \mathbb{N}$,

$$F_i = \gcd(F_{i-1}, F'_{i-1})$$

It is easy to see that there exists $1 \leq s \in \mathbb{N}$ with $F_{s+1} = 0$ and $F_i \neq 0$ for $0 \leq i \leq s$. Now if we set

$$H_i = F_{i-1}/F_i \quad \text{for} \quad 1 \leq i \leq s,$$

then, in view of Lemma 2.82, the following self-explanatory diagram exhibits the desired algorithm for the computation of $g_1, \ldots, g_m$.

$$\begin{tikzcd}
F_0 = g_1g_2^2g_3^3g_4^4\cdots \arrow[r] \arrow[d] & F_1 = g_2g_3^2g_4^3\cdots \arrow[r] \arrow[d] \arrow[dl] & F_2 = g_3g_4^2\cdots \arrow[r] \arrow[d] \arrow[dl] & {} \arrow[dl] \\
H_1 = g_1g_2g_3g_4\cdots \arrow[d] & H_2 = g_2g_3g_4\cdots \arrow[d] \arrow[dl] & H_3 = g_3g_4\cdots \arrow[d] \arrow[dl] & {} \arrow[dl] \\
g_1 & g_2 & g_3 &
\end{tikzcd}$$

Now assume that K is finite and $\operatorname{char}(K) = p \neq 0$. We will define the F_i as before, but in order to understand what we obtain, we arrange the factors g_i of the squarefree decomposition not by ascending indices, but by the residue classes mod p that these belong to:

$$f = (g_p^p g_{2p}^{2p}\cdots)(g_1 g_{p+1}^{p+1} g_{2p+1}^{2p+1}\cdots)(g_2^2 g_{p+2}^{p+2} g_{2p+2}^{2p+2}\cdots)\cdots(g_{p-1}^{p-1} g_{p+(p-1)}^{p+(p-1)}\cdots).$$

Let $0 \leq s \leq p-1$ be maximal with the property that there exists an index $i \in s + p\mathbb{Z}$ with $g_i \neq 1$. Then we have

$$f = (g_p^p g_{2p}^{2p}\cdots)(g_1 g_{p+1}^{p+1} g_{2p+1}^{2p+1}\cdots)(g_2^2 g_{p+2}^{p+2} g_{2p+2}^{2p+2}\cdots)\cdots(g_s^s g_{p+s}^{p+s}\cdots),$$

and the rightmost expression is not equal to 1. Lemma 2.83 tells us that upon taking the gcd of f and f', those exponents that are a multiple of p will remain the same, while all others will go down by 1. If we now define F_i for $0 \leq i \leq s$ as before, we obtain

$$\begin{aligned}
F_0 &= (g_p^p g_{2p}^{2p}\cdots)(g_1 g_{p+1}^{p+1} g_{2p+1}^{2p+1}\cdots)(g_2^2 g_{p+2}^{p+2} g_{2p+2}^{2p+2}\cdots)\cdots(g_s^s \; g_{p+s}^{p+s} \; \cdots)\\
F_1 &= (g_p^p g_{2p}^{2p}\cdots)\;\;(g_{p+1}^p g_{2p+1}^{2p}\cdots)(g_2 g_{p+2}^{p+1} g_{2p+2}^{2p+1}\cdots)\cdots(g_s^{s-1} g_{p+s}^{p+(s-1)}\cdots)\\
F_2 &= (g_p^p g_{2p}^{2p}\cdots)\;\;(g_{p+1}^p g_{2p+1}^{2p}\cdots)\;\;(g_{p+2}^p g_{2p+2}^{2p}\cdots)\cdots(g_s^{s-2} g_{p+s}^{p+(s-2)}\cdots)\\
&\;\;\vdots\\
F_s &= (g_p^p g_{2p}^{2p}\cdots)\;\;(g_{p+1}^p g_{2p+1}^{2p}\cdots)\;\;(g_{p+2}^p g_{2p+2}^{2p}\cdots)\cdots\;\;(g_{p+s}^p \;\;\cdots).
\end{aligned}$$

Since F_s is a pth power, we get $F_s' = 0$, while $F_i \neq 0$ for $0 \leq i \leq s$. In particular, when computing the F_i, we will find out what s is. If $s = 0$ (i.e., if f itself is a pth power), then we continue at the paragraph marked $(*)$ below, with $f_L = f$. Otherwise, we perform essentially the same computations as before in the characteristic zero case, setting $H_i = F_{i-1}/F_i$ for $1 \leq i \leq s$, and it is obvious that we get

$$\begin{aligned}
H_1 &= (g_1 g_{p+1} g_{2p+1}\cdots)(g_2 g_{p+2} g_{2p+2}\cdots)\cdots(g_s g_{p+s} g_{2p+s}\cdots)\\
H_2 &= \qquad\qquad\qquad\quad\;\; (g_2 g_{p+2} g_{2p+2}\cdots)\cdots(g_s g_{p+s} g_{2p+s}\cdots)\\
&\;\;\vdots\\
H_s &= \qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\;\; (g_s g_{p+s} g_{2p+s}\cdots).
\end{aligned}$$

Setting $G_i = H_i/H_{i+1}$ for $1 \leq i \leq s-1$ and $G_s = H_s$ now gives us

$$\begin{aligned}
G_1 &= g_1 g_{p+1} g_{2p+1}\cdots\\
G_2 &= g_2 g_{p+2} g_{2p+2}\cdots\\
&\;\;\vdots\\
G_s &= g_s g_{p+s} g_{2p+s}\cdots.
\end{aligned}$$

Going back to $F_0, \ldots, F_s$, we see that in passing from F_i to F_{i+1}, it is precisely the factor g_{i+1} that drops out, while everything else remains there, albeit with a possibly lower exponent. This means that if we set $Q_i = \gcd(G_i, F_i)$ for $1 \leq i \leq s$, then the result will be

$$\begin{aligned} Q_1 &= g_{p+1}g_{2p+1}\cdots \\ Q_2 &= g_{p+2}g_{2p+2}\cdots \\ &\vdots \\ Q_s &= g_{p+s}g_{2p+s}\cdots . \end{aligned}$$

If we finally divide $Q_1, \ldots, Q_s$ out of $G_1, \ldots, G_s$, respectively, then we have isolated $g_1, \ldots, g_s$, and since we already knew that

$$g_{s+1} = \cdots = g_{p-1} = 1,$$

we have in fact isolated $g_1, \ldots, g_{p-1}$. Next, we form the product

$$P_1 = g_1 Q_1^p g_2^2 Q_2^p \cdots g_s^s Q_s^p$$

to obtain

$$P_1 = (g_1 g_{p+1}^p g_{2p+1}^p \cdots)(g_2^2 g_{p+2}^p g_{2p+2}^p \cdots)\cdots(g_s^s g_{p+s}^p g_{2p+s}^p \cdots).$$

Now if we divide this out of the original f, setting $f_1 = F_0/P_1$, then we get

$$f_1 = (g_p^p g_{2p}^{2p}\cdots)(g_{p+1} g_{2p+1}^{p+1}\cdots)(g_{p+2}^2 g_{2p+2}^{p+2}\cdots)\cdots(g_{p+s}^s g_{2p+s}^{p+s}\cdots).$$

We see that $\deg(f_1) < \deg(f)$, and we find ourselves in a position to make a recursive call of the procedure on f_1. Referring to the original call as the zeroth call, let us denote the polynomial on which the kth call is made by f_k. It is clear that the kth call will isolate the factors $g_{kp+1}, \ldots, g_{kp+(p-1)}$. We will thus eventually have isolated all those $g_i \neq 1$ with $p \nmid i$, say after $L-1$ calls. Then

$$f_L = g_p^p g_{2p}^{2p} g_{3p}^{3p} \cdots ,$$

and this will be made obvious to us by the fact that $f_L' = 0$. If $f_L = 1$, we are done. It remains to treat the case $f_L \neq 1$.

(*) Since $f_L = g_p^p g_{2p}^{2p} g_{3p}^{3p} \cdots$, we must have

$$f_L = \sum_{i=0}^{r} a_i X^i$$

with $a_i = 0$ unless $p \mid i$. Since K is a finite field of characteristic p, we can effectively extract a pth root from each a_i (Lemma 1.107) and thus, by Lemma 1.106, write

$$f_L = \sum_{i=0}^{s} b_i^p X^{ip} = \left(\sum_{i=0}^{s} b_i X^i\right)^p = h^p$$

with $h = g_p g_{2p}^2 g_{3p}^3$. Since the degree of h is less than that of f_L, a recursive call of the entire procedure on h must eventually terminate with a constant. □

Exercises 2.87 (i) Write a programming-style version of the algorithm displayed in the proof of Proposition 2.86 for the case $\operatorname{char}(K) = 0$.

(ii) If you really enjoy writing computer programs, do the same for non-zero characteristic.

Exercise 2.88 If you have a computer algebra system at your disposal which computes derivatives and univariate polynomial gcd's, then compute the squarefree decomposition in $\mathbb{Q}[X]$ of

$$\begin{aligned} f = {} & 2X^{17} - 9X^{16} - 38X^{15} + 329X^{14} - 390X^{13} - 2{,}898X^{12} + 11{,}700X^{11} - \\ & 9{,}320X^{10} - 44{,}792X^9 + 149{,}900X^8 - 187{,}976X^7 + 36{,}840X^6 + \\ & 191{,}040X^5 - 230{,}384X^4 + 59{,}680X^3 + 74{,}592X^2 - 62{,}208X + 13{,}824. \end{aligned}$$

Our results thus far allow us to compute squarefree decompositions in $\mathbb{Q}[X]$ and $\mathbb{Z}/p\mathbb{Z}[X]$. The treatment of multivariate polynomials parallels that for gcd's.

Lemma 2.89 Let R be a UFD, f a non-constant polynomial in $R[X]$. Let

$$\mathrm{c}(f) = ua_1a_2^2 \cdot \cdots \cdot a_k^k \quad \text{and} \quad \mathrm{pp}(f) = vg_1g_2^2 \cdot \cdots \cdot g_m^m$$

be the squarefree decompositions of $\mathrm{c}(f)$ and $\mathrm{pp}(f)$ in R and $R[X]$, respectively. Filling up with factors of the form 1^i, we may assume that $k = m$. Then

$$(uv)(a_1g_1)(a_2g_2)^2 \cdot \cdots \cdot (a_mg_m)^m$$

is a squarefree decomposition of f in $R[X]$.

Proof The unique prime factor decomposition of $\mathrm{c}(f)$ in R is the same as that in $R[X]$ since $\mathrm{c}(f)$ is a constant. The decomposition of $\mathrm{pp}(f)$ in $R[X]$ does not contain any constants since otherwise $\mathrm{pp}(f)$ would not be primitive. The claim now follows from the definition of the squarefree decomposition as a partially combined prime factor decomposition. □

Lemma 2.90 Let R be a UFD, f a non-constant, primitive polynomial in $R[X]$. Let $f = ug_1g_2^2 \cdot \cdots \cdot g_m^m$ be the squarefree decomposition of f in $Q_R[X]$. For $1 \le i \le m$, let d_i be the product of all denominators of coefficients of g_i, and $h_i = \mathrm{pp}(d_ig_i) \in R[X]$. Then there exists a unit $v \in R$ such that $f = vh_1h_2^2 \cdot \cdots \cdot h_m^m$ is a squarefree decomposition of f in $R[X]$.

Proof Since $g_1 \cdot \cdots \cdot g_m$ is squarefree in $Q_R[X]$ and g_i is associated to h_i in $Q_R[X]$, $h_1 \cdot \cdots \cdot h_m$ is still squarefree in $Q_R[X]$. By Corollary 2.62 (ii), every prime factor decomposition of $h_1 \cdot \cdots \cdot h_m$ in $R[X]$ is one in $Q_R[X]$, and $h_1 \cdot \cdots \cdot h_m$, being primitive by the Gaussian lemma, cannot have any constant prime factors, so it is still squarefree in $R[X]$. It now suffices to find the unit v to make the equation $f = vh_1h_2^2 \cdot \dots \cdot h_m^m$ hold. We have $f = ug_1g_2^2 \cdot \dots \cdot g_m^m$, and if we multiply this by $d = d_1d_2^2 \cdot \dots \cdot d_m^m$, we get

$$\begin{aligned} df &= ud_1g_1(d_2g_2)^2 \cdot \dots \cdot (d_mg_m)^m \\ &= uch_1h_2^2 \cdot \dots \cdot h_m^m, \end{aligned}$$

where

$$c = \mathrm{c}(d_1g_1) \cdot \left(\mathrm{c}(d_2g_2)\right)^2 \cdot \dots \cdot \left(\mathrm{c}(d_mg_m)\right)^m.$$

By the Gaussian lemma, $h_1h_2^2 \cdot \dots \cdot h_m^m$ is primitive. Since f is primitive too, there must be a unit v of R such that $uc = vd$. Substituting this in the above equation and cancelling d yields the desired result. □

Theorem 2.91 *Let R be a computable ring which is a UFD, has characteristic zero, allows effective gcd computations, and for which an algorithm is known that computes the squarefree decomposition of any non-zero non-unit. Then one can find an algorithm that does the same in $R[X]$.*

Proof Let f be a non-constant polynomial in $R[X]$. By the previous two lemmas, we may proceed as follows. First, we decompose f into content and primitive part, which is a gcd computation in R. Then we compute the squarefree decomposition of $\mathrm{c}(f)$ in R and the one of $\mathrm{pp}(f)$ in $Q_R[X]$. The former is possible by assumption. The latter can be done by Proposition 2.86 since we have already convinced ourselves in the proof of Theorem 2.70 that Q_R is a computable field, and it is easy to see that $Q_R[X]$ has again characteristic zero. The lifting of the squarefree decomposition of $\mathrm{pp}(f)$ from $Q_R[X]$ to $R[X]$ as described in the previous lemma and the final combination of the two decompositions are trivially computable. □

Corollary 2.92 *Let R be a computable ring with* $\mathrm{char}(R) = 0$. *Assume further that R is either a field or a UFD which allows gcd computations and effective squarefree decompositions. Then one can find an algorithm that computes squarefree decompositions in $R[X_1, \dots, X_n]$.*

Proof We proceed by induction on n. If $n = 1$, then the claim is Proposition 2.86 or Theorem 2.91. If $n > 1$, then

$$R[X_1, \dots, X_n] = R[X_1, \dots, X_{n-1}][X_n],$$

and the claim follows from Theorem 2.91, Corollary 2.71, and Exercise 2.20 together with the induction hypothesis. □

Exercise 2.93 Write down a programming-style version of the recursive algorithm that is implicit in the proof of Corollary 2.92.

By the above corollary, we can now do squarefree decompositions in all polynomial rings over $\mathbb{Z}$ and $\mathbb{Q}$. In the case of $\mathbb{Z}$, the algorithm is usually applied to primitive polynomials only, because the derivative-gcd technique that makes squarefree decompositions interesting cannot be applied to constants. We already mentioned that we can do squarefree decompositions in $\mathbb{Z}/p\mathbb{Z}[X]$ (Proposition 2.86). There is a problem though with multivariate polynomials over $\mathbb{Z}/p\mathbb{Z}$. The recursive method of Theorem 2.91 and Corollary 2.92 requires us to compute squarefree decompositions in $\mathbb{Z}/p\mathbb{Z}(X_1, \ldots, X_{n-1})[X_n]$. But the rational function field

$$\mathbb{Z}/p\mathbb{Z}(X_1, \ldots, X_{n-1})$$

is an *infinite* field of characteristic p to which Proposition 2.86 does not apply.

Exercise 2.94 Is there anything one can do in the way of squarefree decompositions in $\mathbb{Z}/p\mathbb{Z}[X_1, \ldots, X_n]$?

2.7 Factorization of Polynomials

In this section, we will discuss the historically earliest method for the complete factorization of polynomials over the integers and rationals which is due to Kronecker. For gcd's and squarefree decompositions, the fast, industrial strength algorithms that are implemented in today's computer algebra systems are refined and more sophisticated versions of the basic algorithms that we have given. The methods employed for fast factorizations, however, are radically different from the classical algorithm. Kronecker's algorithm is still worth looking at because it provides by far the easiest way to see that the problem of effective factorization is solvable at all. We begin with some elementary facts about univariate polynomials. Recall that if R is a ring, $f \in R[X]$, and $a \in R$ with $f(a) = 0$ (cf. Lemma 2.17 (i)), then we call a a zero of f in R.

Proposition 2.95 *Let R be a domain, $f \in R[X]$, and $a \in R$. Then a is a zero of f iff $(X - a) \mid f$ in $R[X]$.*

Proof From $f = q(X - a)$ with $q \in R[X]$ we conclude that $f(a) = 0$ by substituting a for X. Conversely, assume that a is a zero of f. By Lemma 2.27, there exist $q, r \in R[X]$ with $f = q(X-a)+r$, and $\deg(r) < \deg(X-a)$ or $r = 0$. The condition on the degrees says that r is a constant, and substituting a for X in the equation $f = q(X - a) + r$ yields $r = 0$. □

If $0 \neq f \in R[X]$ and $a \in R$ is a zero of f, then by the lemma above and an easy degree consideration, there must exist $0 < m \in \mathbb{N}$ with $(X-a)^m \mid f$ and $(X - a)^{m+1} \nmid f$. This number m is called the **multiplicity** of the zero a of f.

Proposition 2.96 *Let R be a domain and $0 \neq f \in R[X]$, and suppose that $a_1, \ldots, a_k \in R$ are the pairwise different zeroes of f. Then there exist*

$m_1, \ldots, m_k \in \mathbb{N}^+$ and a polynomial $g \in R$ such that $g(a) \neq 0$ for all $a \in R$,

$$f = g \cdot \prod_{i=1}^{k} (X - a_i)^{m_i},$$

and $m_1, \ldots, m_k$ are the respective multiplicities of $a_1, \ldots, a_k$. If R is computable and the zeroes of f in R are known, then the multiplicities and the polynomial g can be computed from f.

Proof We proceed by induction on $m = \deg(f)$. If $m = 0$, then f has no zeroes in R and we may take $g = f$. Now let $m > 0$. If f does not have any zeroes in R, then we may again take $g = f$. If $a_1 \in R$ is a zero of f, then by the previous lemma, we can write $f = (X - a_1)q$ with $q \in R[X]$. The induction hypothesis applied to q yields a representation of the desired form. Now let $1 \leq j \leq k$. To see that m_j is the multiplicity of a_j, assume for a contradiction that $(X - a_j)^{m_j+1} \mid f$. Then

$$(X - a_i) \mid g \cdot \prod_{\substack{i=1 \\ i \neq j}}^{k} (X - a_i)^{m_i},$$

but a_j is not a zero of the polynomial on the right-hand side, a contradiction. The statement on computability is obvious from the fact that we can effectively divide in $R[X]$. □

It is immediate from the proposition above that the sum s of the multiplicities of the zeroes of a polynomial $0 \neq f \in R[X]$ satisfies $s \leq \deg(f)$. Since every zero has multiplicity at least 1, there can be at most $\deg(f)$ many of them.

Corollary 2.97 *Let R be a domain, $0 \neq f \in R[X]$ with $\deg(f) = m$. Then f has at most m different zeroes in R.* □

The next proposition is a result that is often used in numerical mathematics. The proof given here is also known as the *Lagrange interpolation method.*

Proposition 2.98 *Let K be a field, $a_0, \ldots, a_m \in K$ pairwise different and $b_0, \ldots, b_m \in K$. Then there is a unique polynomial $f \in K[X]$ with $\deg(f) \leq m$ and $f(a_i) = b_i$ for $0 \leq i \leq m$. If K is computable, then f can be computed from the a_i and b_i $(0 \leq i \leq m)$.*

Proof We begin by proving uniqueness. Assume that $f, g \in K[X]$ both have the indicated properties. Then $0 = f(a_i) - g(a_i) = (f - g)(a_i)$ for $0 \leq i \leq m$. It follows that $f - g = 0$ as a polynomial, since otherwise it would be a non-zero polynomial of degree less than or equal to m with more than m different zeroes. To prove the existence of f, we define, for $0 \leq i \leq m$,

$$f_i = \prod_{\substack{0 \leq k \leq m \\ k \neq i}} (X - a_k).$$

Note that $f_i(a_j) = 0$ and $f_i(a_i) \neq 0$ whenever $0 \leq i, j \leq m$ with $i \neq j$. Now if we set

$$f = \sum_{i=0}^{m} \frac{b_i}{f_i(a_i)} f_i,$$

then it is easy to see that f has the desired properties. Moreover, if K is computable, then f as above can obviously be effectively computed. □

The following definition describes those rings over which we will be able to factor first univariate—and then, by induction, also multivariate—polynomials.

Definition 2.99 Let R be a computable ring. We call R a **computable unique factorization domain**, or **computable UFD** for short, if it satisfies the following conditions.

(i) R is a UFD, and the unique prime factor decomposition of any non-zero non-unit can be effectively computed.

(ii) R has infinitely many elements, but only finitely many units, and these can be algorithmically determined.

Note that condition (ii) above requires not just that we can decide whether a given ring element is a unit; we must be able to actually list the finite set of units.

Theorem 2.100 (KRONECKER FACTORIZATION ALGORITHM) *If R is a computable UFD, then so is $R[X]$.*

Proof Condition (ii) of the definition of a computable UFD carries over to $R[X]$ because $R \subseteq R[X]$ and the units of the latter are precisely the units of the former. We know from Theorem 2.65 that $R[X]$ is again a UFD. It remains to show that $R[X]$ allows effective factorization.

Let $0 \neq f \in R[X]$. Since we know what the units of R are, we can decide whether or not f is a unit of $R[X]$. If it is not, we proceed as follows. Since R allows the effective computation of unique prime factor decompositions, it allows effective gcd computations. So we can factor f into content and primitive part, and by assumption we can factor the content in R. Since the irreducible factors in R of the content obviously remain irreducible in $R[X]$, it now suffices to find the (non-constant) irreducible factors of a non-constant primitive polynomial f. Let $0 < m = \deg(f)$. If f has a proper factor at all, then it must, by the degree formula for products, have one whose degree is less than or equal to $m/2$. We let s be the greatest integer that is less than or equal to $m/2$. Regarding f as a polynomial in Q_R, we see that f can have at most m different zeroes in Q_R, and hence a fortiori in R. Since R has infinitely many elements, we can thus, by trial and error, find pairwise different $a_0, \ldots, a_s \in R$ with $f(a_i) \neq 0$ for $0 \leq i \leq s$. Since R is a UFD with only finitely many units, each $0 \neq a \in R$ has only finitely many divisors in R, and we can effectively find them by computing a prime factor decomposition of a and then forming all possible products of units and combinations of the prime factors. For $0 \leq i \leq s$, we now set

$$T_i = \left\{ d \in R \mid d \mid f(a_i) \right\}.$$

Since each T_i is finite $(0 \leq i \leq s)$, there are only finitely many $(s+1)$-tuples $(b_0, \ldots, b_s)$ of elements of R with $b_i \in T_i$ for $0 \leq i \leq s$, and we can actually list them all. For each of them, we can compute the unique polynomial $g \in Q_R[X]$ with $\deg(g) \leq s$ and $g(a_i) = b_i$ $(0 \leq i \leq s)$ by means of the Lagrange interpolation method since Q_R is a computable field. Now if $g \in R[X]$ divides f, then it is easy to see, by substituting a for X, that $g(a) \mid f(a)$ for all $a \in R$. In particular, $g(a_i) \mid f(a_i)$ for $0 \leq i \leq s$, so g must be among the finitely many polynomials computed above. Testing f for a zero remainder upon division by all of these in $Q_R[X]$ (which can be done since Q_R is a computable field), we can thus find out if f has proper factors at all, and if so, we can recursively call the entire procedure on factor and quotient. □

A few words on the occurrence of $Q_R[X]$ at the end of the above procedure are in order. By the theory, the proper factors of f in $R[X]$ must be among the g computed by the Lagrange method. So we may, if we wish, disregard those that come out to be in $Q_R[X] \setminus R[X]$, even if they divide f in $Q_R[X]$. If one of the remaining ones divides f in $Q_R[X]$ with a quotient not in $R[X]$, we may again disregard it and keep on trying. In both of these cases, however, we know by the Gaussian lemma that it is only a matter of shifting a constant factor between factor and quotient to lift the factorization to $R[X]$. This is illustrated by the exercise below. Let us first note that the above theorem allows us to factor univariate polynomials with integer coefficients, because $\mathbb{Z}$ has infinitely many elements and the two units 1 and -1, and it certainly allows effective unique prime factor decomposition since $m \in \mathbb{Z}$ can only have prime factors p with $p \leq |m|$ (in fact, with $p \leq \sqrt{|m|}$).

Exercise 2.101 Use Kronecker's method to factor

$$f = X^5 - 5X^4 + 10X^3 - 6X^2 - 6X + 8$$

over the integers. (Suggestion: Choose $a_0 = -1$, $a_1 = 0$, and $a_2 = 1$. Then T_0, T_1, and T_2 will have 8, 8, and 4 elements, respectively. That leaves you with 256 Lagrange interpolations. First, give a reason why you need not consider combinations of the form (a, a, a). Then try (1,2,1), which will miss. Next, do (4,2,1). You should find a factor of f in $\mathbb{Q}[X]$. Lift it to $\mathbb{Z}[X]$. Which triple (a_0, a_1, a_2) would have given you that factor directly?)

An easy induction on n yields the following corollary.

Corollary 2.102 *If R is a computable UFD, then so is $R[X_1, \ldots, X_n]$.* □

Our results thus far do not apply to polynomials with rational coefficients, since there are infinitely many units in $\mathbb{Q}$. Given any polynomial over $\mathbb{Q}$, we can, since units are irrelevant to the factorization problem, multiply it by the product (or, more cleverly, the least common multiple) of the denominators of its coefficients, thus lifting it to a polynomial over $\mathbb{Z}$, and then factor it over $\mathbb{Z}$. In the univariate case, we know by the Gaussian lemma that the result is the desired factorization. In the multivariate case, we need the following "multivariate version of the Gaussian lemma," which is just a trifle more tedious to prove than one would think.

Lemma 2.103 Let R be a UFD, and suppose f is an irreducible polynomial in $R[X_1, \ldots, X_n]$. Then f is irreducible in $Q_R[X_1, \ldots X_n]$.

Proof If $n = 1$, then we are looking at the Gaussian lemma. Now let $n > 1$. The following two observations will be used in the proof below.

(i) f is primitive as an element of the univariate polynomial ring

$$R[X_1, \ldots, X_{n-1}][X_n]$$

since otherwise we would get a proper factorization in $R[X_1, \ldots, X_n]$.

(ii) Since f is irreducible as an element of

$$R[X_1, \ldots, X_n] = R[X_1, \ldots, X_{n-1}][X_n],$$

it remains irreducible as an element of $Q_R(X_1, \ldots, X_{n-1})[X_n]$ by the Gaussian lemma.

Now assume for a contradiction that $f = gh$ were a proper factorization of f in $Q_R[X_1, \ldots, X_n]$. Viewing f, g, and h as elements of

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

we conclude from (ii) above that one of f and g must be a unit in this ring, say

$$g \in Q_R(X_1, \ldots, X_{n-1}).$$

But we also had $g \in Q_R[X_1, \ldots, X_n]$, and so

$$g \in Q_R[X_1, \ldots, X_{n-1}].$$

Let d_1, $d_2 \in R$ be the product of all denominators of coefficients in Q_R of g and h, respectively. Multiplying by $d_1 d_2$, we can lift the equation $f = gh$

to the univariate polynomial ring $R[X_1, \dots, X_{n-1}][X_n]$ and then take out the content of d_2h as a univariate polynomial in X_n:

$$d_1d_2f = d_1g \cdot \mathrm{c}(d_2h) \cdot \mathrm{pp}(d_2h).$$

Now f is primitive in the univariate polynomial ring $R[X_1, \dots, X_{n-1}][X_n]$ by (i) above, and d_1g is a constant in this ring. Hence there must be a unit u of $R[X_1, \dots, X_{n-1}]$ (and thus of R) with

$$ud_1d_2 = d_1g \cdot \mathrm{c}(d_2h).$$

The left-hand side of this equation is in R, and so we must have $d_1g \in R$, from which it follows that $g \in Q_R$, contradicting the fact that g was not a constant in $Q_R[X_1, \dots, X_n]$. □

We can now prove the following second corollary to Theorem 2.100.

Corollary 2.104 *Let R be a computable UFD and $0 \leq i < n$. Then one can effectively compute the prime factor decomposition of any non-zero non-unit of*

$$Q_R(X_1, \dots, X_i)[X_{i+1}, \dots, X_n],$$

with the obvious convention that $Q_R(X_1, \dots, X_0) = Q_R$.

Proof Recall from Exercise 2.61 that $Q_R(X_1, \dots, X_i) = Q_{R[X_1,\dots,X_i]}$. Now if

$$f \in Q_{R[X_1,\dots,X_i]}[X_{i+1}, \dots, X_n],$$

then we may multiply f by a unit of that polynomial ring—which is clearly irrelevant to the factorization problem—to obtain an element of $R[X_1, \dots, X_n]$. Corollary 2.102 tells us that we can perform a factorization in this latter polynomial ring, and the previous lemma, applied to the UFD $R[X_1, \dots, X_i]$ and the variables $X_{i+1}, \dots, X_n$, says that this is the desired factorization in

$$Q_{R[X_1,\dots,X_i]}[X_{i+1}, \dots, X_n]. \quad \square$$

We can now factor all polynomials, univariate and multivariate, over $\mathbb{Z}$ and $\mathbb{Q}$. Kronecker's method does not apply to factorization over $\mathbb{Z}/p\mathbb{Z}$ since the latter is finite. However, since there are only finitely many univariate polynomials of a fixed degree with coefficients in $\mathbb{Z}/p\mathbb{Z}$, it is not hard to see that all factorizations over $\mathbb{Z}/p\mathbb{Z}$ can in principle be done by trial and error. It is precisely the improvement of this crude method of factoring modulo p that modern factorization algorithms focus on.

2.8 The Chinese Remainder Theorem

One of the stepping stones towards improved versions of many polynomial algorithms is the *Chinese remainder theorem* (CRT). Although we will not pursue these improvements here, we give the CRT here because it is a classic of number theory and algebra. Furthermore, we will later use Gröbner bases to obtain a Chinese remainder theorem in multivariate polynomial rings (Proposition 6.23). Since the CRT is classically a result on integer division with remainder, we use the congruence notation explained at the end of Section 1.5. Applied to the integers, the CRT states that one can always find an integer that leaves prescribed remainders upon division by each one out of a set of prescribed, pairwise relatively prime integers.

Theorem 2.105 (Chinese Remainder Theorem) *Let R be a PID. Assume that $m_1, \ldots, m_k \in R$ are pairwise relatively prime and $r_1, \ldots, r_k \in R$. Then the system*

$$x \equiv r_i \bmod m_i \qquad (1 \le i \le k)$$

of congruences has a solution $a \in R$. The set of all solutions equals the residue class $a + mR$, where $m = m_1 \cdot \cdots \cdot m_k$. If R is a computable Euclidean domain, then the solution can be effectively computed.

Proof We set, for $1 \le i \le k$,

$$n_i = \prod_{\substack{1 \le j \le k \\ j \ne i}} m_j.$$

We have $\gcd(n_i, m_i) = 1$ by Lemma 1.88, and thus there must exist s_i, $t_i \in R$ with $1 = s_i n_i + t_i m_i$ for $1 \le i \le k$. We set

$$a = \sum_{j=1}^{k} n_j s_j r_j.$$

It is easy to see from the definitions of n_i and s_i that $n_j \equiv 0 \bmod m_i$ and $n_i s_i \equiv 1 \bmod m_i$ for $1 \le j, i \le k$ with $j \ne i$. We thus obtain

$$a \equiv \sum_{j=1}^{k} n_j s_j r_j \equiv n_i s_i r_i \equiv r_i \mod m_i$$

for $1 \le i \le k$ as desired. If $b \in a + mR$, then obviously $b \in a + m_i R$ and thus

$$b \equiv a \equiv r_i \mod m_i \quad \text{for} \quad 1 \le i \le k.$$

Conversely, assume that $b \in R$ satisfies $b \equiv r_i \bmod m_i$ for $1 \le i \le k$. Then

$$b - a \in \bigcap_{i=1}^{k} m_i R,$$

and the latter ideal equals mR by Proposition 1.89, so $b \in a + mR$. Inspection of the above proof shows that if R is a computable Euclidean ring, then a can be effectively computed. □

Corollary 2.106 *Let R be a PID, assume that $m_1, \ldots, m_k \in R$ are pairwise relatively prime, and set $m = m_1 \cdot \cdots \cdot m_k$. Then the map*

$$\begin{array}{rccc} \varphi : & R/\mathrm{Id}(m) & \longrightarrow & R/\mathrm{Id}(m_1) \times \cdots \times R/\mathrm{Id}(m_k) \\ & a + \mathrm{Id}(m) & \longmapsto & (a + \mathrm{Id}(m_1), \ldots, a + \mathrm{Id}(m_k)) \end{array}$$

is well-defined, and it is an isomorphism of rings.

Proof If we apply the homomorphism theorem to the homomorphism of Lemma 1.116 and observe that

$$\bigcap_{i=1}^{k} m_i R = mR,$$

then we see that φ is a well-defined embedding of rings. Surjectivity of φ is precisely the statement of the Chinese remainder theorem. □

Exercise 2.107 Show that the Lagrange interpolation method of the proof of Proposition 2.98 is actually an application of the CRT in $K[X]$. (Hint: Take $r_i = b_i$ and $m_i = (X - a_i)$.)

Very roughly speaking, the CRT is applied in practice as follows. One wishes to do some computation with polynomials with coefficients in $\mathbb{Z}$. Moreover, it is known that for the particular input in question, all integer coeffcients that occur in the output have absolute value less than some bound $B \in \mathbb{N}$. (Obtaining such bounds is mathematically hard, but it is often possible.) One then performs the entire algorithm modulo p for a couple of different primes p whose product m exceeds $2B$. The Chinese remainder theorem provides a way to combine these solutions modulo the primes p to a solution modulo m. Now if one chooses representatives of residue classes to be between $-m/2$ and $m/2$, then one may conclude from $m > 2B$ that these are the actual coefficients in $\mathbb{Z}$ of the output. A similar technique may be applied with $\mathbb{Z}$ replaced by $\mathbb{Q}[X]$ and the prime numbers p replaced by linear polynomials $X - q$ with pairwise different $q \in \mathbb{Q}$.

Notes

The concept of a polynomial is essentially as old as algebra itself: an algebraic equation with one or more unknowns is by definition an equation between univariate or multivariate polynomials. In the context of real-valued functions, polynomials with real coefficients arise naturally as functions that are most easily differentiated and integrated. As a matter of fact, the mathematical literature up to about 1900 does not distinguish between a polynomial as a formal expression and a polynomial as the description of a function. This does not cause any trouble as long as the coefficients belong to an infinite domain, but it requires some care otherwise, for the following reason. As a consequence of the universal property of polynomial rings (Lemma 2.17 (i)), a polynomial in $R[X_1, \ldots, X_n]$ can always be viewed as a function from S to S for any extension ring S of R. Conversely, the function thus associated with a polynomial determines the polynomial uniquely in case the coefficient domain is infinite. (For univariate polynomials, this is an easy consequence of Corollary 2.97; induction on the number of variables shows that it holds in the multivariate case as well.) Over a finite coefficient domain R, it may well happen that two different polynomials represent the same function from R to R; examples are easily constructed for $R = \mathbb{Z}/2\mathbb{Z}$. The definition of a polynomial f in $R[X_1, \ldots, X_n]$ as a function $f : \mathbb{N}^n \longrightarrow R$ is the set-theoretically rigorous version of a polynomial as formal expression.

The Euclidean algorithm for positive integers appears in Book VII of Euclid's *Elements* (4th century B.C.) in the form of iterated subtraction (cf. Exercise 2.36). His verification of the correctness of the algorithm is remarkably close to the method of finding and verifying loop invariants. He also proves $\gcd(a, b, c) = \gcd(\gcd(a, b), c)$ for positive integers. Books VII and IX of the *Elements* also develop the theory of prime factor decomposition in the integers to an extraordinary degree of mathematical rigor.

The Gaussian lemma for the ring $\mathbb{Z}[X]$ can be found in Gauss's *Disquisitiones arithmeticae*, Paragraph 42 (1801). (Carl Friedrich Gauss was perhaps the greatest of the German mathematicians, which is why his portrait appears on the recently redesigned 10-deutschmark bill.)

Kronecker published his factorization algorithm in 1882. It seems that he had in fact rediscovered a much earlier result found by the astronomer F. von Schubert in 1793. It is noteworthy that Kronecker viewed his algorithm as mathematically essential rather than just a computational gimmick; in Kronecker (1882), he writes: "The definition of irreductibility [sic!] is void of a secure foundation so long as a method has not been stated by means of which it can be decided of a specific, given function whether or not it is irreductible according to the stated definition." Efficient factorization algorithms for polynomials are based on recent work of Berlekamp, Hensel, and Zassenhaus. For the factorization in $\mathbb{Z}[X]$, they employ what is called a modular method, i.e., factorization in $\mathbb{Z}/p\mathbb{Z}[X]$ for a suitable prime p, then a lifting to the rings $\mathbb{Z}/p^k\mathbb{Z}[X]$ for increasing exponent k, and finally the transition to a factorization in $\mathbb{Z}[X]$. For the last step one needs an a priori bound on the size of the coefficients of a factor of a polynomial; such a bound was given by Landau and Mignotte (see, e.g., Mignotte, 1982). Modular methods can also be used to improve the computation of polynomial gcd's and squarefree decompositions. For more information and guidance on all these improvements, we refer the reader to Knuth (1969), Buchberger et al. (1982), Davenport et al. (1988), and the landmark paper of Lenstra et al. (1982).

We have already pointed out at the end of the last section that the Chinese remainder theorem is a key ingredient in many improved versions of polynomial algorithms. The first written account of the Chinese remainder theorem is in the book *Arithmetic* by the 3rd century Chinese mathematician Sun-Tsu. It begins to appear in the writings of Indian, Arabic, and European mathematicians in the 11th century.

3

Vector Spaces and Modules

3.1 Vector Spaces

The theory of vector spaces—also referred to as linear algebra—is as important and widespread in higher mathematics as calculus. Most notably, it provides a complete understanding of the solvability of systems of linear equations. For our purposes, we will need no more than the basic features of the theory.

Definition 3.1 Let K be a field. A **K-vector space** V is an additive Abelian group with an additional operation $\circ : K \times V \longrightarrow V$, called **scalar multiplication**, such that for all λ, $\mu \in K$ and v, $w \in V$, the following hold:

(i) $\lambda \circ (v + w) = \lambda \circ v + \lambda \circ w$,

(ii) $(\lambda + \mu) \circ v = \lambda \circ v + \mu \circ v$,

(iii) $(\lambda \cdot \mu) \circ v = \lambda \circ (\mu \circ v)$, and

(iv) $1 \circ v = v$.

Note that (iii) involves both the multiplication of K and the scalar multiplication. In the following, we denote field multiplication $\lambda \cdot \mu$ by $\lambda\mu$ and scalar multiplication by a dot. It is also customary—and possible without creating confusion—to write λv for scalar multiplication too. The elements of V are referred to as **vectors**, whereas the elements of the field K are called **scalars**. The zero element of K and the zero vector (i.e., the neutral element of the group V) are of course different objects, but we will denote them both by 0. A sum of two vectors of the form $v + (-w)$ will be denoted by $v - w$, and this will be referred to as subtracting w from v.

In each of the following examples, verification of the vector space axioms is a simple matter of checking them off one after another.

Examples 3.2 (i) Let K be a field, $V = \{0\}$ the trivial Abelian group. Then V is a K-vector space with scalar multiplication $\lambda \cdot 0 = 0$ for all $\lambda \in K$.

(ii) Let K be a field, $1 \leq n \in \mathbb{N}$. Define an addition on K^n by setting

$$(v_1, \ldots, v_n) + (w_1, \ldots, w_n) = (v_1 + w_1, \ldots, v_n + w_n).$$

Then K^n is an additive Abelian group with neutral element $(0, \ldots, 0)$ and inverses $-(v_1, \ldots, v_n) = (-v_1, \ldots, -v_n)$. If, in addition, we define a scalar multiplication $K \times K^n \longrightarrow K^n$ by setting

$$\lambda \cdot (v_1, \ldots, v_n) = (\lambda v_1, \ldots \lambda v_n),$$

then K^n becomes a K-vector space.

(iii) Let R be a ring, K a subring of R which happens to be a field. If we define scalar multiplication $K \times R \longrightarrow R$ as multiplication in R and then view R as just an additive Abelian group, then R is a K-vector space. Moreover, if I is a proper ideal of R, then R/I, when viewed as just an Abelian group, becomes a K-vector space under the scalar multiplication $(\lambda, a+I) \longmapsto (\lambda a+I)$. This situation is given whenever R is a polynomial ring over K.

Lemma 3.3 Let V be a K-vector space, $v \in V$, and $\lambda \in K$. Then the following hold:

(i) $0 \cdot v = 0$ and $\lambda \cdot 0 = 0$.

(ii) $(-1) \cdot v = -v$.

Proof For (i), it suffices to note that $0 \cdot v = (0+0) \cdot v = 0 \cdot v + 0 \cdot v$ implies $0 = 0 \cdot v$, and $\lambda \cdot 0 = \lambda \cdot (0 + 0) = \lambda \cdot 0 + \lambda \cdot 0$ implies $0 = \lambda \cdot 0$. For (ii), consider the equation

$$0 = 0 \cdot v = (1 + (-1)) \cdot v = 1 \cdot v + (-1)v = v + (-1) \cdot v.$$

It now follows readily that $-v = (-1) \cdot v$. □

A **subspace** of a K-vector space V is a non-empty subset U of V that is closed under addition and scalar multiplication, i.e., v, $w \in U$ and $\lambda \in K$ imply $v + w \in U$ and $\lambda \cdot v \in U$. Then for each $v \in U$, we have

$$-v = (-1) \cdot v \in U,$$

and since U contains at least one element u, we get $0 = 0 \cdot u \in U$. We see that in particular, U is a subgroup of V. An easy example is as follows: if we view $\mathbb{C}$ as a $\mathbb{Q}$-vector space in the sense of Example 3.2 (iii) above, then $\mathbb{R}$ is a subspace of $\mathbb{C}$.

Let V and W be K-vector spaces. A map $\varphi : V \longrightarrow W$ is called a **homomorphism of K-vector spaces**, or a **linear map**, if it satisfies

$$\begin{aligned} \varphi(u+v) &= \varphi(u) + \varphi(v), \quad \text{and} \\ \varphi(\lambda \cdot v) &= \lambda \cdot \varphi(v) \end{aligned}$$

for all u, $v \in V$ and $\lambda \in K$. Note that in the second equation, the scalar multiplication is in V on the left-hand side and in W on the right. A homomorphism of vector spaces is called an **embedding** if it is injective, an **isomorphism** if it is bijective.

Exercises 3.4 (i) Let R be a ring containing the field K. Let I be an ideal of R, and view R and R/I as K-vector spaces as explained in Example 3.2. Show that the canonical map $R \longrightarrow R/I$ (where $a \longmapsto a + I$) is a homomorphism of K-vector spaces.

(ii) Imitate the proofs of Lemmas 1.29 and 1.32 to show that a homomorphism of vector spaces satisfies $\varphi(0) = 0$ and $\varphi(-v) = -\varphi(v)$, and that it is injective if and only if $\varphi(v) = 0$ implies $v = 0$.

(iii) Let $\varphi : V \longrightarrow W$ be a homomorphism of K-vector spaces. Show that $\varphi(V)$ is a subspace of W and $\ker(\varphi)$ is a subspace of V, where $\ker(\varphi)$, the *kernel* of φ, is defined as the set of all $v \in V$ with $\varphi(v) = 0$.

If $v_1, \ldots, v_n$ are pairwise different elements of a K-vector space V, then any sum of the form

$$\sum_{i=1}^{n} \lambda_i \cdot v_i \qquad (\lambda_i \in K \text{ for } 1 \leq i \leq n)$$

is also called a **linear combination** of the v_i with coefficients λ_i. It will be convenient from now on to define the empty linear combination $\sum_{i\in\emptyset}$ to be the zero vector.

Definition 3.5 Let V be a K-vector space and B a subset of V.

(i) B is called **linearly independent** if for all $n \in \mathbb{N}^+$, $v_1, \ldots, v_n \in B$ pairwise different, and $\lambda_1, \ldots, \lambda_n \in K$,

$$\sum_{i=1}^{n} \lambda_i \cdot v_i = 0 \quad \text{implies} \quad \lambda_1 = \cdots = \lambda_n = 0.$$

A set that is not linearly independent is called **linearly dependent**.

(ii) B is called a **generating system** for V if for all $v \in V$, there exist $n \in \mathbb{N}^+$, $v_1, \ldots, v_n \in B$, and $\lambda_1, \ldots, \lambda_n \in K$ with

$$v = \sum_{i=1}^{n} \lambda_i \cdot v_i.$$

(iii) B is called a **basis** of V if it is a linearly independent generating system.

It is easy to see that the empty set is linearly independent, that every subset of a linearly independent set is again linearly independent, and that any superset of a generating system is again a generating system. Moreover, by (ii) of Definition 3.1, we can always combine like summands in the representation of (ii) above, thus turning it into a linear combination.

Exercise 3.6 Let V be a K-vector space and $B = \{v_1, \ldots, v_n\}$ a *finite* subset of V. Show the following:

(i) B is linearly independent iff for all $\lambda_1, \ldots, \lambda_n \in K$,

$$\sum_{i=1}^{n} \lambda_i \cdot v_i = 0 \quad \text{implies} \quad \lambda_1 = \cdots = \lambda_n = 0.$$

(ii) B is a generating set for V if for all $v \in V$, there exist $\lambda_1, \ldots, \lambda_n \in K$ with

$$v = \sum_{i=1}^{n} \lambda_i \cdot v_i.$$

(Hint: Argue that a linear combination $\sum_{j=1}^{m} \lambda_j \cdot v_{i_j}$ (with $1 \le i_j \le n$ for all j) can be rewriten in the form $\sum_{i=1}^{n} \mu_i \cdot v_i$ by adding summands of the form $0 \cdot v_i$.)

Example 3.7 Let V be the $\mathbb{R}$-vector space $\mathbb{R}^2$, $B = \{(1,2),(3,4)\}$. We claim that B is a basis of V. Any equation $\lambda_1 \cdot (1,2) + \lambda_2 \cdot (3,4) = (0,0)$ with $\lambda_1, \lambda_2 \in \mathbb{R}$ is equivalent to the system

$$\begin{aligned} \lambda_1 + 3\lambda_2 &= 0 \\ 2\lambda_1 + 4\lambda_2 &= 0 \end{aligned}$$

of linear equations, and this implies $\lambda_1 = \lambda_2 = 0$. We have proved linear independence of B. To see that it is also a generating system of V, let $(a_1, a_2) \in V$ be arbitrary. Converting the equation

$$\lambda_1 \cdot (1,2) + \lambda_2 \cdot (3,4) = (a_1, a_2)$$

to a system of linear equations as above, we see that $\lambda_1 = (-4a_1 + 3a_2)/2$ and $\lambda_2 = (2a_1 - a_2)/2$ are (unique) solutions.

Exercise 3.8 Let V be a K-vector space and $v,\, w \in V$.

(i) Show that $\{v\}$ is linearly dependent iff $v = 0$.

(ii) Show that $\{v, w\}$ is linearly dependent iff there exists $\lambda \in K$ with $v = \lambda \cdot w$ or $w = \lambda \cdot v$.

Exercise 3.9 Let K be a field. Show the following:

(i) $\emptyset$ is a basis of the zero K-vector space (Example 3.2 (i)) for any field K.

(ii) For $1 \leq n \in \mathbb{N}$, the set $\{ e_i \mid 1 \leq i \leq n \}$ is a basis of the K-vector space K^n, where e_i is the n-tuple with ith entry 1 and all other entries 0.

A linearly independent set B in a K-vector space V is called **maximal** if $B \cup \{v\}$ is linearly dependent for all $v \in V \setminus B$. A generating system C for V is called **minimal** if $C \setminus \{v\}$ is no longer a generating system for V for all $v \in C$. The next proposition provides three important characterizations of bases of vector spaces.

Proposition 3.10 *Let V be a K-vector space, B a subset of V. Then the following are equivalent:*

(i) B is a basis of V.

(ii) B is generating system of V, and if we disregard zero summands, then the representation of each $v \in V$ as a linear combination of elements of B is uniquely determined by v up to the order of the summands.

(iii) B is a minimal generating system for V.

(iv) B is a maximal linearly independent system.

Proof (i)$\Longrightarrow$(ii): B is a generating system of V by the definition of a basis. Now assume for a contradiction that there exists $v \in V$ and two representations of v as linear combinations of elements of B that cannot be made identical by dropping zero summands and/or reordering the summands. Adding in summands of the form $0 \cdot v_i$, we may assume that the elements of B occurring in the two representations are the same:

$$v = \sum_{i=1}^{n} \lambda_i \cdot v_i = \sum_{i=1}^{n} \mu_i \cdot v_i,$$

where λ_i, $\mu_i \in K$ and $v_i \in B$ for $1 \leq i \leq n$. We must have $\lambda_i \neq \mu_i$ for at least one index $1 \leq i \leq n$, and thus

$$\sum_{i=1}^{n} (\lambda_i - \mu_i) \cdot v_i = 0$$

contradicts the linear independence of B.

(ii)$\Longrightarrow$(iii): Assume that there exists $v \in B$ such that $B' = B \setminus \{v\}$ is still a generating system for V. Then in particular, v has a representation as a linear combination of elements of B'. This is also a representation in terms of B since $B' \subseteq B$, and it is essentially different from the representation $v = v$ of v as a linear combination of elements of B, contradicting (ii).

(iii)$\Longrightarrow$(iv): Assume for a contradiction that B is linearly dependent. Then there exists a linear combination

$$\sum_{i=1}^{n} \lambda_i \cdot v_i = 0$$

where not all λ_i equal zero. Renumbering, we may assume w.l.o.g. that $\lambda_1 \neq 0$, and multiplying the equation by $1/\lambda_1$, we may even assume that $\lambda_1 = 1$. We see that

$$v_1 = -\sum_{i=2}^{n} \lambda_i \cdot v_i.$$

So whenever v_1 occurs in a linear combination of elements of B, we may replace it by the above expression, which, possibly after combining like summands, results in a linear combination of elements of $B' = B \setminus \{v_1\}$. This means that B' is still a generating system for V, a contradiction. It remains to show that B is maximal as a linearly independent set. Let $v \in V \setminus B$. Then v has a representation

$$v = \sum_{i=1}^{n} \lambda_i \cdot v_i$$

as a linear combination of elements of B, and the equation

$$v - \sum_{i=1}^{n} \lambda_i \cdot v_i = 0$$

shows that $B \cup \{v\}$ is linearly dependent.

(iv)$\Longrightarrow$(i): It remains to prove that B is a generating system for V. Let $v \in V$. If $v \in B$, then $v = v$ is the desired representation. If not, then $B \cup \{v\}$ is linearly dependent, and thus there is an equation

$$\lambda \cdot v + \sum_{i=1}^{n} \lambda_i \cdot v_i = 0$$

with $v_1, \ldots, v_n \in B$ and $\lambda, \lambda_1, \ldots, \lambda_n \in K$ not all zero. Now λ cannot be zero since otherwise B would be linearly dependent. Hence we may set $\mu_i = -(\lambda_i/\lambda)$ for $1 \leq i \leq n$ to obtain

$$v = \sum_{i=1}^{n} \mu_i \cdot v_i. \quad \square$$

The proofs of the following two exercises are similar to the one above. Their statements are not needed in our setup of the theory, but they will be instrumental in two algorithms in the next section.

Exercise 3.11 Let V be a K-vector space, B a basis of V, $v \in V$, and let

$$v = \sum_{i=1}^{n} \lambda_i \cdot v_i$$

be a representation of v as a linear combination of elements of B. Show that for $1 \leq i \leq n$, the set $(B \setminus \{v_i\}) \cup \{v\}$ is again a basis of V iff $\lambda_i \neq 0$.

Exercise 3.12 Let V be a K-vector space, A a generating system of V, and $v \in A$. Show that $A \setminus \{v\}$ is still a generating system for V iff there exists a linear combination

$$\lambda \cdot v + \sum_{i=1}^{n} \lambda_i \cdot v_i = 0$$

with $v_1, \ldots, v_n \in A \setminus \{v\}$ and $\lambda \neq 0$.

We have seen (Exercise 3.9) that all vector spaces of the form K^n have bases. We are now in a position to prove the existence of bases in a more general situation.

Theorem 3.13 *Let V be a K-vector space, and assume that V has a finite generating system C. Then V has a basis $B \subseteq C$.*

Proof The set

$$N = \{\, |B| \mid B \subseteq C \text{ and } B \text{ is a finite generating system for } V \,\} \subseteq \mathbb{N}$$

is not empty and thus has a minimal element $n_0 \in \mathbb{N}$. Let $B_0 \subseteq C$ be a finite generating system for V with $|B_0| = n_0$; then B_0 is a minimal generating system and thus a basis of V. □

An algorithmic version of the above theorem will be the subject of an exercise in the next section. We are now going to give an example of a vector space that does not have a finite generating system. In this example, we will be able to find a basis consisting of infinitely many elements. To prove that this is always the case, i.e., that every vector space has a basis, one needs to make a set-theoretic assumption known as *Zorn's lemma* which will be described in Section 4.1. We will give the proof there as an illustration of the use of Zorn's lemma.

Example 3.14 Let K be a field and $R = K[X_1, \ldots, X_n]$ a polynomial ring over R. We may then view R as a K-vector space V as explained in Example 3.2 (iii). A linear combination in V is simply a sum of constant multiples of polynomials. Now if F is a finite subset of V, then there must be a term t in $T = T(X_1, \ldots, X_n)$ (in fact infinitely many) with $t \notin T(f)$ for all $f \in F$. We see that the coefficient of t in any linear combination of elements of F equals zero, and thus the polynomial t cannot be written as such a linear combination. This shows that V cannot have a finite generating system. We claim that the infinite set T of all terms is a basis of V. If $f \in V$, then f has a natural representation

$$f = \sum_{t \in T(f)} a_t t \qquad (a_t \in K)$$

as a sum of monomials, and this can be viewed as a linear combination of elements of T. So T is a generating system for V, and it remains to prove that it is linearly independent. Let T' be a finite subset of T, and let

$$\sum_{t \in T'} \lambda_t \cdot t = 0 \qquad (\lambda_t \in K)$$

be a vanishing linear combination of elements of T'. Lemma 2.14 tells us that the coefficients λ_t are zero for all $t \in T'$.

A K-vector space V is called **computable** if K is a computable field, the elements of V can be represented on a computer, and addition in V, subtraction in V, and scalar multiplication can be effectively performed. When computing in a vector space, one usually needs to know that for any given vector, one can effectively find a representation as a linear combination w.r.t. some specific basis, or any basis, or even any generating system. We do not incorporate any condition of this type into the definition of a computable vector space; rather, we prefer to state the necessary assumptions explicitly in each case.

The standard example of a computable vector space is K^n for computable field K (Example 3.2 (ii)). We have demonstrated in Example 3.7 how the problem of effectively dealing with representations as linear combinations then reduces to handling systems of linear equations. (Systems of linear equation are of course an interesting topic by themselves, both theoretically and computationally; we will treat them from the point of view of Gröbner bases in Section 10.5. Here, we will simply assume that from experience in elementary mathematics, the reader is aware of the fact that solvability of such a system can be effectively dealt with.)

Now assume V is a computable K-vector space, B is a basis of V, and we are given representations of vectors $v_1, \ldots, v_n \in V$ as linear combinations of elements of B:

$$v_j = \sum_{i=1}^{m_j} \lambda_{ij} \cdot b_{ij} \qquad (1 \leq j \leq n,\ \lambda_{ij} \in K,\ b_{ij} \in B).$$

Adding zero summands if necessary, we may assume that the basis vectors occurring in each sum are the same:

$$v_j = \sum_{i=1}^{m} \lambda_{ij} \cdot b_i \qquad (1 \leq j \leq n,\ \lambda_{ij} \in K,\ b_i \in B).$$

We can now constructively decide whether $v_1, \ldots, v_n$ are linearly dependent.

Proposition 3.15 *Let V, B, and $v_1, \ldots, v_n$ be as described above. Then the algorithm* LINDEP *of Table* 3.1 *decides whether $v_1, \ldots, v_n$ are linearly dependent, and if so, produces a non-trivial zero linear combination.*

TABLE 3.1. Algorithm LINDEP

```
Specification: $v \longleftarrow \text{LINDEP}(A, B', \Lambda)$
                    Constructive decision of linear dependence of $A$
Given: $A = \{v_1, \ldots, v_n\} \subseteq V,\ B' = \{b_1, \ldots, b_m\} \subseteq B$, and
       $\Lambda = \{\lambda_{ij} \mid 1 \le i \le m,\ 1 \le j \le n\}$ with
       $v_j = \lambda_{1j} \cdot b_1 + \cdots + \lambda_{mj} \cdot b_m$
Find: $v \in \{\textbf{false}\} \cup (\{\textbf{true}\} \times (K^n \setminus \{(0, \ldots, 0)\}))$
      such that $v = \textbf{false}$ if $A$ is linearly independent, and
      $v = (\textbf{true}, (\mu_1, \ldots, \mu_n))$ with $\mu_1 \cdot v_1 + \cdots + \mu_n \cdot v_n = 0$ otherwise
begin
if the system of linear equations

    $\lambda_{11}x_1 + \cdots + \lambda_{1n}x_n = 0$
       $\vdots$          $\vdots$
    $\lambda_{m1}x_1 + \cdots + \lambda_{mn}x_n = 0$

  has a solution $(\mu_1, \ldots, \mu_n) \in K^n \setminus (0, \ldots, 0)$ then
  return(true, $(\mu_1, \ldots, \mu_n)$)
else return(false)  end
end LINDEP
```

Proof If $\mu_1, \ldots, \mu_n \in K^n$, then

$$\begin{aligned}\sum_{j=1}^{n} \mu_j \cdot v_j &= \sum_{j=1}^{n} \mu_j \cdot \left(\sum_{i=1}^{m} \lambda_{ij} \cdot b_i\right) \\ &= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} \lambda_{ij}\mu_j\right) \cdot b_i\end{aligned}$$

From the fact that $b_1, \ldots, b_m$ are linearly independent we conclude that such a sum equals zero iff

$$\sum_{j=1}^{n} \lambda_{ij}\mu_j = 0$$

for $1 \le i \le m$. With this observation in mind, it is easy to prove the correctness of the algorithm. □

We close this section with a proposition on the behavior of bases under homomorphisms.

Lemma 3.16 Let V and W be K-vector spaces, $\varphi : V \longrightarrow W$ a homomorphism. Then the following hold:

(i) φ is injective iff for every linearly independent subset A of V, the set $\varphi(A)$ is linearly independent in W.

(ii) φ is surjective iff for every generating system C for V, the set $\varphi(C)$ is a generating system for W.

(iii) φ is bijective iff for every basis B of V, the set $\varphi(B)$ is a basis of W.

Proof (i) Assume that φ is injective, let $A \subseteq V$ be linearly independent, $v_1, \ldots, v_n \in A$, and

$$\sum_{i=1}^{n} \lambda_i \cdot \varphi(v_i) = 0 \qquad (\lambda_i \in K \text{ for } 1 \leq \mathrm{i} \leq \mathrm{n}).$$

Then we have

$$0 = \sum_{i=1}^{n} \lambda_i \cdot \varphi(v_i) = \varphi\Bigl(\sum_{i=1}^{n} \lambda_i \cdot v_i\Bigr).$$

From the injectivity of φ we conclude that the sum in parentheses equals zero, and this together with the linear independence of A implies that all λ_i must be zero. For the converse, let $v \in V$ with $\varphi(v) = 0$. Then we must have $v = 0$, since otherwise $A = \{v\}$ would be a linearly independent set with $\varphi(A) = \{0\}$ linearly dependent.

(ii) Assume φ is surjective, let C be a generating system for V, and $w \in W$. Then there exists $v \in V$ with $\varphi(v) = w$, and $v_1, \ldots, v_n \in C$ with

$$v = \sum_{i=1}^{n} \lambda_i \cdot v_i \qquad (\lambda_i \in K \text{ for } 1 \leq i \leq n).$$

It follows that

$$w = \varphi(v) = \varphi\Bigl(\sum_{i=1}^{n} \lambda_i \cdot v_i\Bigr) = \sum_{i=1}^{n} \lambda_i \cdot \varphi(v_i).$$

For the converse, let $w \in W$ be arbitrary. Since all of V is a generating system for V, $\varphi(V)$ is a generating system for W, and thus there exist $v_1, \ldots, v_n \in V$ with

$$w = \sum_{i=1}^{n} \lambda_i \cdot \varphi(v_i) = \varphi\Bigl(\sum_{i=1}^{n} \lambda_i \cdot v_i\Bigr),$$

and the expression in parenthesis is clearly an element of V. Statement (iii) is now an immediate consequence of the definitions. □

3.2 Independent Sets and Dimension

It is an immediate consequence of Exercise 3.11 that if a vector space V has a basis at all, then it will in general have infinitely many different bases. The aim of this section is to prove that if V has a finite basis, then any two bases will have the same number of elements. We perform the argument on a more abstract level because that way, we will be able to use it again in a different context in Section 7.1.

From now on, let X be a set and $\mathcal{U}$ a non-empty collection of subsets of X, i.e.,

$$\emptyset \neq \mathcal{U} \subseteq \mathcal{P}(X).$$

In the applications below, X will be a vector space and $\mathcal{U}$ the collection of all linearly independent sets, so the following terminology is natural: for any $A \subseteq X$, "A is independent" will mean $A \in \mathcal{U}$, and "A is dependent" will mean $A \notin \mathcal{U}$. Assume now that $\mathcal{U}$ satisfies the following two axioms.

(U1) A independent implies B independent for all $B \subseteq A$.

(U2) Whenever $A \subseteq X$ and a, b_1, $b_2 \in X$ such that $A \cup \{b_1, b_2\}$ is independent, $A \cup \{a\}$ is independent, and $b_1 \neq b_2$, then at least one of $A \cup \{b_1, a\}$ and $A \cup \{a, b_2\}$ is independent.

Note that U1 implies that $\emptyset$ is independent. The following theorem states that if two finite independent sets are given, then the smaller one can be enlarged by elements of the larger one to at least the size of the larger one without becoming dependent.

Theorem 3.17 *Let A and B be finite independent subsets of X with $|A| \leq |B|$, and let $B' \subseteq B$ be such that $A \cup B'$ is independent while $A \cup B' \cup \{b\}$ is dependent for all $b \in B \setminus B'$. Then $|A \cup B'| \geq |B|$.*

Proof The claim is trivial if $A \subseteq B$. Else, we proceed by induction on $n = |A|$. If $n = 0$, then necessarily $B' = B$. Now let $n > 0$, and assume for a contradiction that $|A \cup B'| < |B|$. Choose $a \in A \setminus B$. Then

$$|(A \setminus \{a\}) \cup B'| \leq |B| - 2,$$

and by induction hypothesis, there exist b_1, $b_2 \in B \setminus B'$ such that

$$(A \setminus \{a\}) \cup B' \cup \{b_1, b_2\}$$

is still independent. Applying axiom U2 to $(A \setminus \{a\}) \cup B'$, a, b_1, and b_2, we see that at least one of $A \cup B' \cup \{b_1\}$ and $A \cup B' \cup \{b_2\}$ is independent, a contradiction. □

We call an independent set A **maximal** if $A \cup \{a\}$ is dependent for all $a \in X \setminus A$.

Corollary 3.18 *Assume that there exists an independent set $A \subseteq X$ that is finite and maximal. Then every independent set B is finite with $|B| \leq |A|$, and if in addition, B is maximal too, then $|B| = |A|$.*

Proof If there were an infinite independent set, then by axiom U1, there would have to be a finite one with more than $|A|$ elements. It thus suffices to show that every finite independent set B satisfies $|B| \leq |A|$, and $|B| = |A|$ if B is maximal. So let B be finite and independent. If B had more than $|A|$ elements, then A could be enlarged by at least $|B| - |A|$ many elements and still be independent, contradicting its maximality. If B is maximal too, then by the above, we have both $|B| \leq |A|$ and $|A| \leq |B|$. □

Lemma 3.19 Let V be a K-vector space, and let $\mathcal{U}$ be the collection of all linearly independent subsets of V. Then $\mathcal{U}$ satisfies axioms U1 and U2.

Proof U1 is immediate from the definition of linear independence. Now let A, a, b_1, and b_2 be as in U2, and assume for a contradiction that both $A \cup \{b_1, a\}$ and $A \cup \{a, b_2\}$ are linearly dependent. Then we must have b_1, $b_2 \notin A$. Moreover, there exist linear combinations

$$\begin{aligned} \lambda \cdot a + \mu \cdot b_1 + \sum_{i=1}^{m} \lambda_i \cdot v_i &= 0 \\ \lambda' \cdot a + \mu' \cdot b_2 + \sum_{i=1}^{n} \lambda'_i \cdot w_i &= 0, \end{aligned}$$

where $v_1, \ldots, v_m, w_1, \ldots, w_n \in A$, and in each equation, not all coefficients equal 0. We must have λ, μ, λ', $\mu' \neq 0$ since otherwise the above equations would constitute a contradiction to at least one of the premises of U2. Multiplying by $1/\lambda$ and $1/\lambda'$, respectively, we may assume w.l.o.g. that $\lambda = \lambda' = 1$. Subtraction yields

$$\mu \cdot b_1 - \mu' \cdot b_2 + \sum_{i=1}^{m} \lambda_i \cdot v_i - \sum_{i=1}^{n} \lambda'_i \cdot w_i = 0.$$

Combining like summands if necessary, we see that this contradicts the linear independence of $A \cup \{b_1, b_2\}$. □

The proof of the following theorem is now immediate from Proposition 3.10, Theorem 3.13, Corollary 3.18, and Lemma 3.19.

Theorem 3.20 *Let V be a K-vector space, and assume that V has a finite basis B. Then every linearly independent set in V has at most $|B|$ many elements, every generating system for V has at least $|B|$ many elements, and every basis of V has exactly $|B|$ many elements.* □

In the situation of the theorem, the vector space V is called **finite-dimensional** with **dimension** $|B|$. The dimension of V is then denoted by $\dim_K(V)$. If V is not finite-dimensional, then it is called **infinite-dimensional**, and one often writes $\dim_K(V) = \infty$. Examples of finite-dimensional vector spaces are K^n with dimension n for arbitrary field K (Exercise 3.9). Examples of infinite-dimensional vector spaces are the polynomial rings over any field K (Example 3.14). Determining basis and dimension of residue class rings of these modulo an ideal will be one of the main applications of the theory of Gröbner bases.

From Theorem 3.20 together with Proposition 3.10, one easily deduces the following corollary.

Corollary 3.21 *Let V be a finite-dimensional K-vector space and B a finite subset of V with $|B| = \dim_K(V)$. Then B is linearly independent iff it is a generating system for V iff it is a basis of V.* □

We can now describe the behavior of the dimension under homomorphisms.

Lemma 3.22 Let V and W be finite-dimensional K-vector spaces and $\varphi : V \longrightarrow W$ a homomorphism. Then the following hold:

(i) If φ is injective, then $\dim_K(V) \leq \dim_K(W)$.

(ii) If φ is surjective, then $\dim_K(V) \geq \dim_K(W)$.

(iii) If φ is bijective, then $\dim_K(V) = \dim_K(W)$.

Proof Set $n = \dim_K(V)$, and let B be a basis of V. If φ is injective, then $\varphi(B)$ is linearly independent by Lemma 3.16 (i), and $|\varphi(B)| = n$, so $n \leq \dim_K(W)$ by Theorem 3.20. Similarly, surjectivity of φ implies that $\varphi(B)$ is a generating system for W with $|\varphi(B)| \leq n$, and so $n \geq \dim_K(W)$. Statement (iii) is now immediate from the definitions. □

Lemma 3.23 Let V be a finite-dimensional K-vector space, and let $n = \dim_K(V)$. Then the following hold:

(i) Every subspace U of V is finite-dimensional with $\dim_K(U) \leq n$, and the inequality is strict iff $U \neq V$.

(ii) A homomorphism from V to itself is injective iff it is surjective iff it is bijective.

Proof (i) Let U be a subspace of V, and assume for a contradiction that U is not finite-dimensional. U contains a finite linearly independent set, namely, the empty set, but no finite linearly independent set can be maximal, since such a set would be a finite basis by Proposition 3.10. We can thus successively enlarge $C_0 = \emptyset$ to linearly independent sets $C_1, \dots, C_{n+1}$ with $|C_i| = i$ for $1 \leq i \leq n+1$. Then C_{n+1} is linearly independent in V too and has $n + 1$ many elements, contradicting Theorem 3.20. The inequality $\dim_K(U) \leq n$ follows from (i) of the previous lemma together with the fact that the inclusion map from U to V is obviously an injective homomorphism. Now assume that $\dim_K(U) = n$, and let B be a basis of U. Then B is a linearly independent subset of V with n many elements and thus a basis of V by Corollary 3.21. Hence every element of V, being a linear combination of elements of B, is an element of U. Conversely, if $U = V$, then trivially $\dim_K(U) = \dim_K(V) = n$.

(ii) Let $\varphi : V \longrightarrow V$ be a homomorphism. We show that each of injectivity and surjectivity of φ implies bijectivity. To verify the latter, it suffices by Lemma 3.16 to show that $\varphi(B)$ is again a basis of V whenever B is a basis of V. So let B be a basis of V. If φ is injective, then by (i) of Lemma 3.16, $\varphi(B)$ is a linearly independent subset of V with n many elements and thus a basis of V by Corollary 3.21. If φ is surjective, then by (ii) of Lemma 3.16, $\varphi(B)$ is a generating system for V. Theorem 3.20 tells us that $\varphi(B)$ has at least n elements. Being the image of a set with n elements under a map, $\varphi(B)$ cannot have more than n elements, and so we may apply Corollary 3.21 to conclude that $\varphi(B)$ is indeed a basis of V. □

We have now provided all the linear algebra that will be needed in the sequel. For a better understanding, we point out some more consequences of the above results. If we apply Theorem 3.17 to the linear algebra situation and assume that in addition, the larger independent set B is actually maximal, then we obtain the following result: given a finite basis B and a linearly independent set A in a K-vector space V, we can always enlarge A to size $|B|$, i.e., to size $\dim_K(V)$, by adding elements from B, and the result is again a basis of V. This is also known as the *Steinitz exchange theorem.* We are now going to show how the statement of Exercise 3.11 can be used to obtain an algorithmic version of this fact.

Theorem 3.24 (Steinitz Exchange Theorem) *Let V be a finite-dimensional computable K-vector space, and assume that we can effectively express any given vector as a linear combination of vectors from any given basis. Let B be a basis of V and A a linearly independent subset of V. Then the algorithm* EXCHANGE *of Table* 3.2 *replaces $|A|$ many elements of B with the elements of A, thus enlarging A to a basis of V.*

Proof We claim that the tasks of the **for**-loop can always be performed, and that after the kth run through the loop, $\{v_1, \ldots, v_k, b_{k+1}, \ldots, b_n\}$ is still a basis of V. Consider the first run, where $k = 1$. The first task can be performed since B is a basis of V, the second since v_1, being an element of the linearly independent set A, cannot be the zero vector. After switching indices on b_i and b_1, $\{v_1, b_2, \ldots, b_n\}$ is a basis of V by Exercise 3.11. Now assume that $k > 1$, and that after the $(k-1)$-st run,

$$\{v_1, \ldots, v_{k-1}, b_k, \ldots, b_n\}$$

is a basis of V. Let us inspect the kth run. The first task can be performed by the assumption that we just made. If λ_i were 0 for all $k \leq i \leq n$, then

TABLE 3.2. Algorithm EXCHANGE

```
Specification: $C \leftarrow$ EXCHANGE$(A, B)$
               Enlarging $A$ to a basis $C$ using elements of $B$
Given: $A = \{v_1, \dots, v_m\} \subseteq V$ linearly independent,
       $B = \{b_1, \dots, b_n\}$ a basis of $V$
Find: a renumbering of $\{b_1, \dots, b_n\}$ such that
      $C = \{v_1, \dots, v_m, b_{m+1}, \dots, b_n\}$ is again a basis of $V$
begin
for $k = 1$ to $m$ do
    write $v_k$ as a linear combination $\sum_{i=1}^{k-1} \lambda_i \cdot v_i + \sum_{i=k}^{n} \lambda_i \cdot b_i$
    select $k \leq i \leq n$ with $\lambda_i \neq 0$
    switch indices on $b_i$ and $b_k$
end
return($\{v_1, \dots, v_m, b_{m+1}, \dots, b_n\}$)
end EXCHANGE
```

the equation

$$v_k - \sum_{i=1}^{k-1} \lambda_i \cdot v_i = 0$$

would be contradicting the linear independence of A, so the second task can be performed too. It follows again from Exercise 3.11 that after the indicated renumbering, the set $\{v_1, \dots, v_k, b_{k+1}, \dots, b_n\}$ is still a basis of V. □

Exercise 3.25 Extend $A = \{(1, 2, 0), (2, 4, 3)\}$ to a basis of $\mathbb{Q}^3$. (Hint: Apply the algorithm EXCHANGE to A and the basis B of Exercise 3.9.)

Exercise 3.26 Let V be a finite-dimensional computable K-vector space, and assume that we can effectively express any vector in V as a linear combination of elements from any generating system. Use Exercise 3.12 to write an algorithm that shrinks any finite generating system for V to a basis of V.

3.3 Modules

Modules arise naturally in many problems related to the theory of commutative and non-commutative rings. The definition of a module is identical with that of a vector space except that the field K is replaced by an arbitrary ring. Here, we will continue to consider just commutative rings with 1. The concept of a module may also be viewed as a generalization of that of an ideal I in a ring R. (Recall that a subset I of R is an ideal of R if I is non-empty and closed under addition and multiplication with arbitrary elements of R.)

Definition 3.27 Let R be a ring. An **R-module** M is an additive Abelian group with an additional operation $\circ : R \times M \longrightarrow M$, called **scalar multiplication**, such that for all α, $\beta \in R$ and a, $b \in M$, the following hold:

(i) $\alpha \circ (a + b) = \alpha \circ a + \alpha \circ b$,

(ii) $(\alpha + \beta) \circ a = \alpha \circ a + \beta \circ a$,

(iii) $(\alpha \cdot \beta) \circ a = \alpha \circ (\beta \circ a)$, and

(iv) $1 \circ a = a$.

As before with vector spaces, we denote ring multiplication $a \cdot b$ by ab and scalar multiplication by a dot although it would again be possible to write αa for scalar multiplication too. It is easy to see that Lemma 3.3 continues to hold for modules. Verification of the following examples is left to the reader.

Examples 3.28 Let R be a ring.

(i) Let I be an ideal of R. Then I forms an R-module with respect to the addition and multiplication of R. In particular, R itself can be regarded as an R-module, and the zero ideal $\{0\}$ forms an R-module.

(ii) Let $M = \{0\}$ be the trivial additive Abelian group and set $\alpha \cdot 0 = 0$ for all $\alpha \in R$. Then M is an R-module, the **trivial** R-module.

(iii) If R is a field, then the class of R-modules is precisely the class of vector spaces over R.

(iv) Let $M = R^n$ be a finite direct product of R as a ring. If we disregard multiplication on M and define scalar multiplication by $\alpha \cdot (\beta_1, \ldots, \beta_n) = (\alpha\beta_1, \ldots, \alpha\beta_n)$, then M is an R-module. M is called a **free R-module of rank n.**

(v) Every polynomial ring $R[X_1, \ldots, X_n]$ over R is an R-module if we define scalar multiplication as multiplication by a constant in the polynomial ring.

(vi) More generally, let φ be a ring homomorphism from R into some ring S. Then S forms an R-module when scalar multiplication $\circ$ is defined by $\alpha \circ b = \varphi(\alpha) \cdot b$ for $\alpha \in R$ and $b \in S$. In particular, every extension ring S of R forms in a natural way an R-module: take for φ the natural embedding $\iota : R \longrightarrow S$. Moreover, for every ideal I in R, R/I forms an R-module: take for φ the canonical homomorphism $R \longrightarrow R/I$.

(vii) Let M be an R-module and let $\varphi : S \longrightarrow R$ be a ring homomorphism. Then M is also an S-module under the new scalar multiplication $\circ$ defined by $\alpha \circ b = \varphi(\alpha) \cdot b$ for $\alpha \in S$ and $b \in M$.

(viii) Let $R' = R[a_1, \dots, a_n]$ be a finitely generated extension ring of R and let $S = R[X_1, \dots, X_n]$ be the polynomial ring in the indeterminates $X_1, \dots, X_n$ over R. Then by Lemma 2.17 (i), there is a unique ring homomorphism $\varphi : S \longrightarrow R'$ with $\varphi(X_i) = a_i$. So by (vii) above, any R'-module M becomes an S-module with respect to the map φ.

A map $\varphi : M \longrightarrow M'$ between two R-modules M and M' is a **homomorphism of R-modules** if for all a, $b \in M$ and $\alpha \in R$,

$$\begin{aligned} \varphi(a+b) &= \varphi(a) + \varphi(b), \quad \text{and} \\ \varphi(\alpha \cdot a) &= \alpha \cdot \varphi(a). \end{aligned}$$

A homomorphism from M to itself is called an **endomorphism** of M.

Let M be an R-module and let N be an additive subgroup of M. Then N is a **submodule** of M if N is closed under scalar multiplication. We use the notation $N \leq M$ for "N is a submodule of M." Let $M = R[X]$ be a univariate polynomial ring over the ring R, $n \in \mathbb{N}$, and N the subset of M consisting of all polynomials of degree less than or equal to n. Then N is a submodule of M, but not a subring of the ring $R[X]$. Natural examples of submodules are **kernels** of homomorphisms, where of course the kernel $\ker(\varphi)$ of a homomorphism $\varphi : M \longrightarrow M'$ of R-modules consists of all $a \in M$ with $\varphi(a) = 0$.

Whenever B is a subset of M, then there exists a unique smallest submodule N of M that contains B as subset. N consists of all linear combinations

$$\sum_{i=1}^{n} \alpha_i \cdot a_i \qquad (\alpha_i \in R,\ a_i \in B).$$

N is called the submodule **generated by B in M** or the **linear hull of B in M** and will be denoted by $\mathrm{lin}(B)$. A **generating system** for M is a subset B of M with $\mathrm{lin}(B) = M$. M is called a **finitely generated R-module** if M has a finite generating system. M is called a **noetherian** R-module if every submodule of M is finitely generated. Note that this definition is consistent with the one given for rings. As in the case of vector spaces, a subset B of M is called **linearly independent** if for all $n \in \mathbb{N}^+$, $a_1, \dots, a_n \in B$ pairwise different, and $\alpha_1, \dots \alpha_n \in R$,

$$\sum_{i=1}^{n} \alpha_i \cdot a_i = 0 \quad \text{implies} \quad \alpha_1 = \cdots = \alpha_n = 0.$$

B is called a **basis** of M if in addition, B is a generating system for M.

Recall from Section 3.1 that in case R is a field, every R-module that has a finite generating system has a basis. For modules over an arbitrary ring this is false in general; in fact, it is false whenever R is not a field. In that case, there exists an ideal $\{0\} \neq I \neq R$ in R. Let now M be the R-module R/I. Note that $\{1 + I\}$ is a generating system for M, and assume for a contradiction that B is a basis of M. Then B contains some $a+I$ with $a \in R \setminus I$. Now whenever $0 \neq \alpha \in I$, then $\alpha \cdot (a+I) = 0$, a contradiction.

We have already mentioned in Section 3.1 that every vector space has a basis under the set-theoretic assumption of Zorn's lemma. Under that assumption, one can still prove that every module has a maximal linearly independent subset; in the case of vector spaces, any such set is a basis, and it may also be characterized as a minimal generating system. In a module, a maximal linearly independent set need not be a generating system at all: we have just seen an example where $\emptyset$ is a maximal linearly independent set. Minimal generating systems need not even exist in modules.

Exercise 3.29 Let $p \in \mathbb{Z}$ be a prime number, $P = \{p^k \mid k \in \mathbb{N}\}$. Consider $\mathbb{Z}_P$, the ring of quotients of $\mathbb{Z}$ w.r.t. P. Since $\mathbb{Z} \subset \mathbb{Z}_P$, we may regard $\mathbb{Z}_P$ as a $\mathbb{Z}$-module M. Show that M does not contain a minimal generating system.

If M is an R-module and $a_1, \ldots, a_n \in M$, then the set $\operatorname{syz}(a_1, \ldots, a_n)$ of all **syzygies** of the n-tuple $(a_1, \ldots, a_n) \in M^n$ consists of all n-tuples $\alpha = (\alpha_1, \ldots \alpha_n) \in R^n$ such that

$$\sum_{i=1}^{n} \alpha_i \cdot a_i = \alpha \cdot \mathbf{a} = 0.$$

It is easy to verify that $\operatorname{syz}(a_1, \ldots, a_n)$ forms a submodule of the R-module R^n, which is also called the the **(first) module of syzygies** of $(a_1, \ldots, a_n)$.

In Section 1.5, we have explained how one defines the residue class ring R/I of a ring R modulo an ideal I of R. In a similar way, we can form the **factor module** M/N of the module M w.r.t. the submodule N of M: its elements are the residue classes $a+N$ of elements $a \in M$, and the operations on M/N are defined by

$$\begin{aligned} (a+N)+(b+N) &= (a+b)+N, \quad \text{and} \\ \alpha \cdot (a+N) &= \alpha \cdot a + N. \end{aligned}$$

Again, the map $\kappa : M \longrightarrow M/N$ with $\kappa(a) = a+N$ is a homomorphism of R-modules, called the **canonical homomorphism**. It will not come as a surprise now that the homomorphism theorem for rings has a perfect analogue for modules. We leave it up to the reader to make sure that the following theorem can be proved just like Theorem 1.55.

Theorem 3.30 (HOMOMORPHISM THEOREM) *Let $\varphi : M \longrightarrow M'$ be a homomorphism of R-modules, N a submodule of M with $N \subseteq \ker \varphi$. Denote the canonical homomorphism from M to M/N by χ. Then the map $\psi : M/N \longrightarrow M'$, $\psi(a+N) = \varphi(a)$ is well-defined. ψ is a homomorphism satisfying $\psi \circ \chi = \varphi$.*

$$\begin{tikzcd} M \arrow[r, "\varphi"] \arrow[d, "\chi"'] & M' \\ M/N \arrow[ur, "\psi"'] & \end{tikzcd}$$

The map ψ is surjective iff φ is surjective, and it is injective iff $N = \ker \varphi$. □

The following universal property of the module R^n is easy to prove.

Proposition 3.31 *Let M be an R-module generated by $\{b_1, \ldots, b_n\}$. Then the map*

$$\begin{array}{cccc} \varphi : & R^n & \longrightarrow & M \\ & (a_1, \ldots, a_n) & \longmapsto & \sum_{i=1}^{n} a_i b_i \end{array}$$

is a surjective homomorphism of R-modules. In particular, M is isomorphic to $R^n / \ker(\varphi)$. □

Proposition 3.32 *(i) Let M and M' be R-modules, let $\varphi : M \longrightarrow M'$ be a surjective homomorphism, and assume that M is noetherian. Then M' is noetherian too.*

(ii) Let M be a noetherian R-module and let $N \leq M$. Then N and M/N are noetherian.

(iii) Let $M = R^n$ be a free R-module of rank n over a noetherian ring R. Then M is noetherian.

(iv) Let M be a finitely generated R-module over a noetherian ring R. Then M is noetherian.

Proof (i) Let $N' \leq M'$. Then $N = \varphi^{-1}(N') \leq M$, and so N has a finite generating system C. Then $\varphi(C)$ is a generating system for $\varphi(N)$, and by the surjectivity of φ, $\varphi(N) = N'$.

(ii) If $N' \leq N \leq M$, then $N' \leq M$ and so N' is finitely generated. Let $\kappa : M \longrightarrow M/N$ be the canonical homomorphism; then κ is surjective, and so by (i), M/N is noetherian.

(iii) The proof is by induction on n. The case $n = 1$ being trivial, let $n > 1$, assume that R^{n-1} is noetherian, and let N be a submodule of R^n. Let π be the projection

$$\begin{array}{cccc} \pi : & R^n & \longrightarrow & R^{n-1} \\ & (a_1, \ldots, a_n) & \longmapsto & (a_1, \ldots, a_{n-1}), \end{array}$$

and set

$$I = \{ r \in R \mid (0, \ldots, 0, r) \in N \}.$$

It is easy to see that $\pi(N)$ is a submodule of R^{n-1} and I is an ideal of R. It follows that $\pi(N)$ and I have finite generating systems B and C, respectively. Let D be a finite subset of N with $\pi(D) = B$, and set

$$E = \{ (0, \ldots, 0, r) \mid r \in C \}.$$

It is clear that E too is a subset of N, and we claim that $D \cup E$, which is clearly a finite subset of N, is in fact a generating system for N. To see this, let $a \in N$. Then $\pi(a) \in \pi(N) = \operatorname{lin}(B)$, and so there exist $\alpha_1, \ldots, \alpha_k \in R$ and $b_1, \ldots, b_k \in B$ with

$$\pi(a) = \sum_{i=1}^{k} \alpha_i \cdot b_i .$$

By the choice of D, we can find $d_1, \ldots, d_k \in D$ with $\pi(d_i) = b_i$ for $1 \leq i \leq k$. It is now easy to see that the element

$$b = a - \sum_{i=1}^{k} \alpha_i \cdot d_i$$

of N satisfies $\pi(b) = 0$, so that $b = (0, \ldots, 0, r)$ for some $r \in R$. This means that actually $r \in I$, and thus there exist $\beta_1, \ldots, \beta_k \in R$ and $c_1, \ldots, c_k \in C$ with

$$r = \sum_{i=1}^{l} \beta_i c_i ,$$

and thus, setting $e_i = (0, \ldots, 0, c_i) \in E$ for $1 \leq i \leq l$, we have

$$b = (0, \ldots, 0, r) = \sum_{i=1}^{l} \beta_i \cdot e_i .$$

Together, we obtain

$$a = b + \sum_{i=1}^{k} \alpha_i \cdot d_i = \sum_{i=1}^{l} \beta_i \cdot e_i + \sum_{i=1}^{k} \alpha_i \cdot d_i \in \operatorname{lin}(D \cup E).$$

(iv) Let $\{b_1, \ldots, b_n\}$ be a generating system for M. Then by Proposition 3.31 above, the map

$$\begin{array}{rccc} \varphi : & R^n & \longrightarrow & M \\ & (a_1, \ldots, a_n) & \longmapsto & \displaystyle\sum_{i=1}^{n} a_i b_i \end{array}$$

is a surjective homomorphism. The claim now follows immediately from (i) and (iii). □

Notes

The concept of an abstract finite-dimensional vector space over the field of real numbers was introduced by H.G. Grassmann in his *Ausdehnungslehre*

(1844/1862). In 1888, G. Peano presented an axiomatic approach to real vector spaces that included infinite-dimensional spaces. An entirely different motivation, namely, the study of hypercomplex number systems, had led W.R. Hamilton to the four-dimensional space of quaternions (1844), and A. Cayley to the eight-dimensional space of octaves (1845). Infinite-dimensional vector spaces provided the algebraic framework for the development of functional analysis through the interpretation of integral operators as linear maps on a real or complex vector space of functions.

An axiomatic treatment of linear independence was given by B.L. van der Waerden in the 1930s in his *Modern Algebra.* Our axioms U1 and U2 in Section 3.2 are based on a different approach that is due to Whitney (1935), who observed that a multitude of algebraic, geometric, and combinatorial situations can be handled using such simple axioms. The general study of these axioms and their equivalents and applications constitutes the theory of *matroids* (see Welsh, 1976 and White, 1986).

The Steinitz exchange theorem appears in Steinitz's *Algebraische Theorie der Körper* (1910), where it is used in connection with algebraic independence in field extensions (cf. Theorem 7.23).

The concept of a module is a common generalization of a vector space and an ideal of a ring; as we have mentioned before, the older algebraic literature often uses "module" as a synonym of "ideal." One may also view the module concept as a generalization of Abelian groups: every additive Abelian group is a $\mathbb{Z}$-module under a natural scalar multiplication, where $n \cdot a$ is, loosely speaking, the n-fold sum of a with itself. The algebraic structure of R-modules for a ring R is intimately related to the structure of the ring R. This fact plays an important role in the structure theory of non-commutative rings. Noetherian rings and noetherian modules are named after E. Noether, who gave the first thorough study of ideals and modules in a purely axiomatic manner.

4

Orders and Abstract Reduction Relations

The theory of Gröbner bases deals with ideals in polynomial rings and is thus part of commutative algebra. However, the concept of *binary relations*, and in particular, of *orders*, is instrumental in making Gröbner basis theory work. This chapter provides the necessary results by discussing binary relations on an abstract set M. Our treatment centers around the study of various kinds of finiteness properties such as *well-foundedness.* These properties will later be used in a number of ways; eventually, however, their relevance lies in the fact that they provide termination proofs for certain algorithms. We will frequently encounter the *axiom of choice* which we discuss in an introductory section to this chapter.

4.1 The Axiom of Choice and Some Consequences in Algebra

The axiom of choice is concerned with infinite products of sets. In Section 0.2, we defined the Cartesian product of of n sets $A_1, \ldots, A_n$. The product of infinitely many sets, however, cannot be obtained in this way: a tuple of infinite length is not a mathematically meaningful object. To arrive at such a generalization, we must change our point of view slightly. An element $(a_1, \ldots, a_n)$ of the product $\prod_{i=1}^{n} A_i$ can obviously be described by a function

$$f : \{1, \ldots, n\} \longrightarrow \bigcup_{i=1}^{n} A_i$$

with $f(i) = a_i \in A_i$ for $1 \leq i \leq n$.

Definition 4.1 Let I be a set, $\{A_i\}_{i\in I}$ a family of sets indexed by elements of I. Then we define

$$\prod_{i\in I} A_i = \left\{ f : I \longrightarrow \bigcup_{i\in I} A_i \;\middle|\; f(i) \in A_i \text{ for all } i \in I \right\}.$$

We are now in a position to state the axiom of choice.

Axiom of choice (AC) *If* $\{A_i\}_{i\in I}$ *is a family of non-empty sets, then* $\prod_{i\in I} A_i$ *is not empty.*

The point about **AC** that deserves some comment is of course the fact that from a naive point of view, its statement is completely trivial. If each of the sets A_i is non-empty, then we can choose (!) an element $a_i \in A_i$ for each $i \in I$ and define an element f of the product by setting $f(i) = a_i$. Unfortunately, assuming the existence of even the most innocuous looking mathematical objects can lead to surprisingly bizarre consequences, and even to contradictions. The formal system of axiomatic set theory, which is now commonly accepted as the foundation of modern mathematics, therefore proceeds as follows. One starts out with a set of axioms that require the existence of those mathematical objects that are certainly indispensable, such as the existence of the empty set, or the existence of the power set of any existing set. The formal system thus obtained is called Zermelo–Fraenkel set theory, denoted by **ZF**. **ZF** does not yet include the axiom of choice. Interestingly, we do not at present know for sure whether **ZF** is consistent, i.e., leads to contradictions or not, but we know that if we drop any one of its axioms, then we can hardly do any mathematics at all.

The first axiom that is somewhat questionable is **AC**. On the one hand, a great deal of mathematics can be done without it, and adding it in leads to some conclusions that are much less plausible than **AC** itself. On the other hand, there are some desirable mathematical results that require the axiom of choice, e.g., the fact that every vector space has a basis. In the early days of axiomatic set theory, the choice between using **AC** and dropping it was made difficult by the possibility of **ZF** being consistent and **ZF+AC** being inconsistent. It is now known that if **ZF** is consistent, then so is **ZF+AC**. Mathematicians therefore often use **AC** tacitly when they need it, but it is still considered good practice by many to indicate any use of **AC**.

As a matter of fact, we have already encountered an application of **AC**. In Lemma 0.21, we have given a naive proof of the fact that for a surjective map $\varphi : A \longrightarrow B$, there exists a map $\psi : B \longrightarrow A$ with $\varphi \circ \psi = \mathrm{id}_B$. The map ψ was defined by letting $\psi(b)$ be any preimage of $b \in B$ under φ. Really what we have to do here is to consider the family

$$\{\varphi^{-1}(\{b\})\}_{b\in B}$$

of subsets of A, and then to apply the axiom of choice to obtain a function

$$\psi : B \longrightarrow \bigcup_{b\in B} \varphi^{-1}(\{b\}) \qquad (*)$$

with $\psi(b) \in \varphi^{-1}(\{b\})$ for all $b \in B$.

We will now show how **AC** implies that every PID is a UFD. (Cf. the remarks following Definition 2.52.) We precede the proof with a lemma on PID's that does not use **AC**.

Lemma 4.2 Let R be a PID. Then there are no infinite sequences $\{a_n\}_{n\in\mathbb{N}}$ of elements of R such that a_{n+1} properly divides a_n for all $n \in \mathbb{N}$.

Proof Assume for a contradiction that $\{a_n\}_{n\in\mathbb{N}}$ is such a sequence. Let I be the ideal generated by the set $\{\, a_n \mid n \in \mathbb{N} \,\}$, i.e., the set of all finite sums of multiples of elements of this set. Since R is a PID, $I = \mathrm{Id}(b)$ for some $b \in R$. We conclude that $b \mid a_n$ for all $n \in \mathbb{N}$. Furthermore b, being an element of I, can be written in the form

$$b = \sum_{i=1}^{k} r_i a_{n_i}$$

for certain $r_1, \ldots, r_k \in R$ and $n_1, \ldots, n_k \in \mathbb{N}$. From this and the divisibility property of the a_n, we see that $a_n \mid b$ for all $n \geq n_0 = \max\{n_1, \ldots, n_k\}$. Together, we obtain that $a_{m_1} \mid a_{m_2}$ for all m_1, $m_2 \geq n_0$, contradicting the fact that each a_n is properly divided by its successor. □

Note that the statement of the lemma is easy to understand for Euclidean domains such as $\mathbb{Z}$ or $K[X]$: here, an infinite sequence with proper divisibilities would give rise to a strictly descending sequence in $\mathbb{N}$ according to Lemma 2.50.

Proposition 4.3 (**AC**) *Every PID is a UFD.*

Proof Let R be a PID. Uniqueness of the prime factor decomposition can be proved in the exact same way as for Euclidean domains in Theorem 2.51. To prove existence, let S be the set of all non-zero non-units of R that do not have a factorization into irreducible elements. Assume that $S \neq \emptyset$. For each $a \in S$, set

$$D_a = \{\, b \in S \mid b \text{ properly divides } a \,\}.$$

We claim that D_a is not empty for any $a \in S$. Indeed, if a had no proper divisors at all, then a would itself be irreducible, contradicting $a \in S$; and if no proper divisor of a were in S, then we could write $a = bc$ with b, $c \notin S$, and a factorization of a into irreducible elements could be obtained by combining two such factorizations of b and c, respectively. By the axiom of choice, there exists a function

$$f : S \longrightarrow \bigcup_{a\in S} D_a \subseteq S$$

such that $f(a)$ properly divides a for all $a \in S$. We now recursively define a sequence $\{a_n\}_{n\in\mathbb{N}}$ of elements of S by taking for a_0 an arbitrary element of S and setting $a_{n+1} = f(a_n)$. Then a_{n+1} properly divides a_n for all $n \in \mathbb{N}$, contradicting the lemma above. □

The statement of the proposition above is actually of little relevance to us: we have proved in Chapter 2 that a polynomial ring over a ring R is either Euclidean (namely, in the univariate case with R a field), in which case we can prove unique factorization without the axiom of choice (Theorem 2.51), or else it is no longer a PID (namely, in the multivariate case or when R is not a field). One of the most important results on multivariate polynomial rings over a field is that—assuming the axiom of choice—one can prove that every ideal is still finitely generated. This is an immediate consequence of the *Hilbert basis theorem*, which we prove next.

Definition 4.4 A ring R is called **noetherian** if every ideal of R is finitely generated.

Lemma 4.5 (AC) Let R be a ring and let $\mathcal{I}(R)$ be the set of all ideals of R. Then the following are equivalent:

(i) R is noetherian.

(ii) For every $B \subseteq R$ there exists a finite subset C of B with $\mathrm{Id}(C) = \mathrm{Id}(B)$.

(iii) Whenever $\{a_i\}_{i\in\mathbb{N}}$ is a sequence of elements of R, then there exists $m \in \mathbb{N}$ with $a_{m+1} \in \mathrm{Id}(a_0, \dots, a_m)$.

(iv) There does not exist a strictly ascending $\subseteq$-chain of ideals of R, i.e., a family $\{I_i\}_{i\in\mathbb{N}}$ of ideals of R with $I_j \subseteq I_k$ and $I_j \neq I_k$ for $j < k$.

Proof (i)$\Longrightarrow$(ii): If $B \subseteq R$, then, since $\mathrm{Id}(B)$ is finitely generated, there exists a finite subset $D = \{d_1, \dots, d_m\}$ of $\mathrm{Id}(B)$ with $\mathrm{Id}(D) = \mathrm{Id}(B)$. In particular, we may write

$$d_i = \sum_{j=1}^{k_i} r_{ij} b_{ij} \qquad (r_{ij} \in R,\ b_{ij} \in B)$$

for $1 \leq i \leq m$. If we let $C = \{\, b_{ij} \mid 1 \leq i \leq m,\ 1 \leq j \leq k_i \,\}$, then C is a finite subset of B which generates $\mathrm{Id}(B)$ because for every $b \in \mathrm{Id}(B)$, there exist $s_1, \dots, s_m \in R$ with

$$\begin{aligned} b &= \sum_{i=1}^{m} s_i d_i \\ &= \sum_{i=1}^{m} s_i \sum_{j=1}^{k_i} r_{ij} b_{ij} \\ &= \sum_{i=1}^{m} \sum_{j=1}^{k_i} s_i r_{ij} b_{ij} \,. \end{aligned}$$

(ii)$\Longrightarrow$(iii): Let $\{a_i\}_{i\in\mathbb{N}}$ be a sequence of elements of R. Setting

$$I = \mathrm{Id}(\{\, a_i \mid i \in \mathbb{N} \,\}),$$

we conclude from (ii) that there exists $m \in \mathbb{N}$ with $I = \mathrm{Id}(a_0, \dots, a_m)$, and so $a_{m+1} \in \mathrm{Id}(a_0, \dots, a_m)$ as desired.

(iii)$\Longrightarrow$(iv): Assume for a contradiction that $\{I_i\}_{i\in\mathbb{N}}$ is a strictly ascending $\subseteq$-chain in $\mathcal{I}(R)$. By the axiom of choice, there exists a function φ that assigns to each $\emptyset \neq A \subseteq R$ an element of A. We can now recursively define a sequence $\{a_i\}_{i\in\mathbb{N}}$ of elements of R by setting $a_0 = \varphi(I_0)$ and $a_{i+1} = \varphi(I_{i+1} \setminus I_i)$. By (iii), there exists $m \in \mathbb{N}$ with

$$a_{m+1} \in \operatorname{Id}(a_0, \ldots, a_m) \subseteq I_m,$$

a contradiction.

(iv)$\Longrightarrow$(i): Assume for a contradiction that I is an ideal of R that is not finitely generated. Then we can recursively define a sequence $\{I_i\}_{i\in\mathbb{N}}$ of ideals of R as follows. As before, let φ be a choice function which assigns to each $\emptyset \neq A \in \mathcal{P}(R)$ an element of A, and let $a_0 \in I$ be arbitrary. Then we set $a_{i+1} = \varphi(I \setminus \operatorname{Id}(a_0, \ldots, a_i))$, and $I_i = \operatorname{Id}(a_0, \ldots, a_i)$ for all $i \in \mathbb{N}$. Now $\{I_i\}_{i\in\mathbb{N}}$ forms a strictly ascending $\subseteq$-chain in $\mathcal{I}(R)$, contradicting (iv). □

Theorem 4.6 (Hilbert Basis Theorem) **(AC)** *Let R be a noetherian ring. Then the polynomial ring $R[X]$ is again noetherian.*

Proof Assume for a contradiction that I is an ideal of $R[X]$ that is not finitely generated. Then I is not the zero ideal. Using the axiom of choice in a similar way as in the proof of the lemma above, we may now define a sequence $\{f_i\}_{i\in\mathbb{N}}$ of elements of I as follows. Let f_0 be a non-zero element of I of minimal degree, and let f_{i+1} be an element of minimal degree of $I \setminus \operatorname{Id}(f_0, \ldots, f_i)$. It is clear that then $\deg(f_j) \leq \deg(f_k)$ for $j < k$. For $i \in \mathbb{N}$, we denote the head coefficient of f_i by a_i. By (iii) of the previous proposition, there exists $m \in \mathbb{N}$ with $a_{m+1} \in \operatorname{Id}(a_0, \ldots, a_m)$. Let r_0, $\ldots$, $r_m \in R$ be such that $a_{m+1} = r_0 a_0 + \cdots + r_m a_m$, and consider the polynomial

$$f^* = f_{m+1} - \sum_{i=0}^{m} X^{\deg(f_{m+1})-\deg(f_i)}\, r_i f_i\,.$$

We must have $f^* \in I \setminus \operatorname{Id}(f_0, \ldots, f_m)$ since otherwise the equation above would imply that $f_{m+1} \in \operatorname{Id}(f_0, \ldots, f_m)$. Moreover, it is easy to see that $\deg(f^*) < \deg(f_{m+1})$, and thus the existence of f^* contradicts the choice of f_{m+1}. □

Corollary 4.7 *If R is a noetherian ring, then $R[X_1, \ldots, X_n]$ is again noetherian for every $n \geq 1$. In particular, $K[X_1, \ldots, X_n]$ is noetherian if K is a field.*

Proof The proof is by induction on n. If $n = 1$, then the claim is identical with the Hilbert basis theorem as stated above. If $n > 1$, it follows from

$$R[X_1, \ldots, X_n] = R[X_1, \ldots, X_{n-1}][X_n]$$

together with the induction hypothesis. The rest of the corollary is immediate from the fact that a field K has only two ideals, namely, $\{0\}$ and K, both of which are finitely generated. □

We point out that the last theorem and its corollary will not be used in the sequel; for the special case of polynomial rings over fields, in which our main interest lies, the existence proof of Gröbner bases in Section 5.2 will constitute an independent proof of the Hilbert basis theorem.

The axiom of choice is often used in an equivalent form called *Zorn's lemma.* Let X be a set and $\mathcal{A} \subseteq \mathcal{P}(X)$, where $\mathcal{P}(X)$ denotes the power set of X. Then $Y \in \mathcal{P}(X)$ is called a **maximal element** of $\mathcal{A}$ if $Y \in \mathcal{A}$, and $Y \subseteq Z$ implies $Y = Z$ for all $Z \in \mathcal{A}$. A subset $\mathcal{B}$ of $\mathcal{A}$ is called a **chain** in $\mathcal{A}$ if for all $Y, Z \in \mathcal{B}$, we have $Y \subseteq Z$ or $Z \subseteq Y$. We say that $\mathcal{A}$ is closed under unions of chains if $\bigcup_{Y \in \mathcal{B}} Y \in \mathcal{A}$ for every chain $\mathcal{B}$ in $\mathcal{A}$.

Zorn's Lemma *Let X be a set, $\mathcal{A}$ a non-empty subset of $\mathcal{P}(X)$ which is closed under unions of chains. Then $\mathcal{A}$ has a maximal element.*

The proof of **AC** from Zorn's lemma is actually fairly easy. Given $\{A_i\}_{i \in I}$ as in the premise of **AC**, take for X the set of all ordered pairs (i, a) with $a \in A_i$. Then let

$$\mathcal{A} = \left\{ f : J \longrightarrow \bigcup_{j \in J} A_j \;\middle|\; J \subseteq I \text{ and } f(j) \in A_j \text{ for all } j \in J \right\}.$$

$\mathcal{A}$ is not empty because it contains all singletons $\{(i, a)\}$ with $a \in A_i$. (Note that we are using the precise definition of a function given in Section 0.2.) It is now easy to see that $\mathcal{A}$ is closed under unions of chains. Hence it contains a maximal element, and one easily proves that any maximal element of $\mathcal{A}$ is an element of $\prod_{i \in I} A_i$.

The proof of the reverse implication is more tedious. The general idea is to start with any element of $\mathcal{A}$ and then to look for a maximal one by choosing supersets as long as this is possible. This gives a chain whose union is then the desired maximal element. A precise formulation of the definition of this chain requires the use of ordinal numbers and thus a considerable amount of set-theoretical work.

We will now look at some results in ring theory that are dependent upon Zorn's lemma and thus indirectly upon the axiom of choice. Recall that by a ring, we always mean a commutative ring with 1. Let R be a ring and $\mathcal{A} \subseteq \mathcal{I}(R)$ where $\mathcal{I}(R)$ denotes the set of all ideals of R. Then $\mathcal{A}$ is called a **chain** of ideals if it is a chain of sets, i.e., I, $J \in \mathcal{A}$ implies $I \subseteq J$ or $J \subseteq I$ for all I, $J \in \mathcal{A}$.

Exercise 4.8 Show that if $\mathcal{A}$ is a chain of ideals of a ring R, then $\bigcup_{I \in \mathcal{A}} I$ is an ideal of R.

Let R be a ring, M any subset of R and I an ideal of R. Then we say that I is **maximally disjoint** from M if $I \cap M = \emptyset$, and for all ideals J of R such that $I \subseteq J$, we have $I = J$ or $J \cap M \neq \emptyset$.

Lemma 4.9 (**AC**) Let R be a ring, $M \subset R$, and I an ideal of R with $I \cap M = \emptyset$. Then I is contained in an ideal J of R which is maximally disjoint from M. In particular, every proper ideal of R is contained in a maximal ideal.

Proof Let $\mathcal{A}$ be the set of all ideals J of R satisfying $I \subseteq J$ and $J \cap M = \emptyset$. Then $\mathcal{A} \neq \emptyset$ since $I \in \mathcal{A}$. $\mathcal{A}$ is closed under the union of chains by Exercise 4.8 and the obvious facts that a union of elements of $\mathcal{A}$ is again disjoint from M and contains I. By Zorn's lemma, $\mathcal{A}$ has a maximal element J. J contains I, and so does any ideal which contains J. From the maximality of J in $\mathcal{A}$ one now easily concludes that J is maximally disjoint from M. In order to extend a proper ideal I to a maximal one, apply the general result to I and $M = \{1\}$. □

Lemma 4.10 Let R be a ring, M a non-empty multiplicative subset of R (i.e., a non-empty subset that is closed under multiplication), and I an ideal of R that is maximally disjoint from M. Then I is a prime ideal.

Proof The ideal I is proper because $M \neq \emptyset$ and $I \cap M = \emptyset$. For the argument below, we note that for $a \in R$, the ideal $\mathrm{Id}(I, a)$ consists of all ring elements of the form $s + ra$ with $s \in I$ and $r \in R$. Now assume that I were not prime. Then there exist a, $b \in R$ with $ab \in I$ but neither $a \in I$ nor $b \in I$. The ideals $\mathrm{Id}(I, a)$ and $\mathrm{Id}(I, b)$ then both properly contain I. Since I is maximally disjoint from M, we can thus find s_1, $s_2 \in I$ and r_1, $r_2 \in R$ with $s_1 + r_1a$, $s_2 + r_2b \in M$. Since M is multiplicative, it follows that

$$(s_1 + r_1a)(s_2 + r_2b) = s_1s_2 + s_1r_2b + s_2r_1a + r_1r_2ab \in M.$$

But this is also an element of I, contradicting the fact that that I is disjoint from M. □

The following proposition is now obvious.

Proposition 4.11 (**AC**) *Let R be a ring, M a multiplicative subset of R, I an ideal of R with $I \cap M = \emptyset$. Then I is contained in a prime ideal P of R which is still disjoint from M.* □

Proposition 4.11 can also be used to obtain a characterization of the *radical* of an ideal.

Definition 4.12 Let R be a ring, I an ideal of R. Then the set

$$\{\, a \in R \mid a^s \in I \text{ for some } s \in \mathbb{N} \,\}$$

is called the **radical** of I and is denoted by $\mathrm{rad}(I)$. I is called a **radical ideal** if $I = \mathrm{rad}(I)$.

Proposition 4.13 (**AC**) *Let R be a ring, I an ideal of R. Then* $\mathrm{rad}(I)$ *equals the intersection of all prime ideals containing I.*

Proof Let $a \in \mathrm{rad}(I)$, P a prime ideal of R with $I \subseteq P$. Then $a^s \in P$ for some $s \in \mathbb{N}$ and thus $a \in P$. Conversely, assume that $a \notin \mathrm{rad}(I)$. Then $M \cap I = \emptyset$ where M is the multiplicative set $\{ a^s \mid s \in \mathbb{N} \}$. By Proposition 4.11, I is contained in a prime ideal P of R with $M \cap P = \emptyset$. In particular, $a \notin P$. □

It is an immediate consequence of the proposition above that $\mathrm{rad}(I)$ is an ideal of R.

Exercise 4.14 Give a direct proof of the fact that $\mathrm{rad}(I)$ is an ideal of R. Show that $\mathrm{rad}(\mathrm{rad}(I)) = \mathrm{rad}(I)$, meaning that $\mathrm{rad}(I)$ is in fact a radical ideal.

Another important consequence of Zorn's lemma is that every vector space has a basis.

Theorem 4.15 (AC) *Let V be a K-vector space. Then V has a basis.*

Proof If $V = \{0\}$, then $\emptyset$ is a basis of V. Otherwise, let $\mathcal{U}$ be the collection of all linearly independent subsets of V. Then $\mathcal{U} \neq \emptyset$ since $\{v\} \in \mathcal{U}$ for every $0 \neq v \in V$. Let $\mathcal{C}$ be a chain in $\mathcal{U}$, and set

$$C = \bigcup_{U \in \mathcal{C}} U.$$

Then it is easy to see that C is again linearly independent: if a finite linear combination of elements of C equals zero, then each of the finitely many vectors occurring in the sum lies in some $U \in \mathcal{C}$, so they all lie in the one that contains all others ($\mathcal{C}$ is a chain!). But the latter set is linearly independent, so all coefficients of the linear combination must be zero. By Zorn's lemma, $\mathcal{U}$ has a maximal element which is a basis of V by Proposition 3.10. □

To conclude this section, let us take another look at the proof of Proposition 4.3. The argument exemplifies a construction that we will encounter frequently: if one has assigned to each element a of a set A a non-empty subset A_a of A, then one may obtain a sequence $\{a_n\}_{n \in \mathbb{N}}$ of elements of A with $a_{n+1} \in A_{a_n}$ for all $n \in \mathbb{N}$ by choosing $a_0 \in A$ arbitrarily and setting $a_{n+1} = f(a_n)$, where f is a choice function on A that satisfies $f(a) \in A_a$ for all $a \in A$. In Theorem 7.29 of Section 7.2, we will need an interesting variant of this construction which requires a little more thought. Let J be a set, and for any field L, let us denote by $L[\boldsymbol{X}]$ the polynomial ring in the variables $\{ X_j \mid j \in J \}$ as defined in the discussion following Lemma 2.22. Suppose that we have a mathematical construction that defines, for any given field L, a certain proper ideal I_L of the polynomial ring $L[\boldsymbol{X}]$. Now we are given a field K, and we wish to construct a sequence $\{K_n\}_{n \in \mathbb{N}}$ of fields such that $K_0 = K$ and for all $n \in \mathbb{N}$,

$$K_{n+1} = K_n[\boldsymbol{X}]/M \qquad (I_L \subseteq M \text{ a maximal ideal of } K_n[\boldsymbol{X}]).$$

The example is somewhat artificial at the moment, but the construction in Theorem 7.29 is quite similar; the only difference is that J will depend on L. Naively, one would proceed as follows. Assign to each field L the set

$$A_L = \{\, L[X]/M \mid I_L \subseteq M \text{ a maximal ideal of } L[X] \,\}.$$

Then A_L is not empty by Lemma 4.9, and one would set $K_0 = K$ and $K_{n+1} = f(K_n)$, where f is a choice function satisfying $f(L) \in A_L$ for every field L. Unfortunately, this argument is *not* legitimate because the "choice function" f operates on the collection of all fields which is not a set in the framework of **ZF**. (If you are not familiar with **ZF**, just recall that **ZF** takes an extremely conservative attitude towards allowing things to call themselves sets. Also, you will just have to accept the arguments of the rest of this paragraph at face value.) However, a fairly simple proof shows that the construction of the sequence $\{K_n\}_{n\in\mathbb{N}}$ given above is perfectly legitimate in **ZF+AC**. It is clear that A_L as defined above is a set for every field L. Using the axioms of **ZF**, it is not hard to show that for every $n \in \mathbb{N}$, the collection

$$A_n = \{\, L \mid \text{there exist fields } K_0, \dots, K_n \text{ with } K_0 = K,\ K_n = L, \text{ and } K_i \in A_{K_{i-1}} \text{ for } 1 \le i \le n \}$$

is actually a set, where K is still the given field that is to be the starting point of our sequence. Finally, the union $A = \bigcup_{n\in\mathbb{N}} A_n$ is a set, and it is easy to see that $K \in A$, and A_L is a non-empty subset of A for all $L \in A$. So if we now consider a choice function g on A with the property that $g(L) \in A_L$ for all $L \in A$, then g is legitimately defined in **ZF+AC**. Moreover, it is not hard to see that g can replace f in the construction of the sequence $\{K_n\}_{n\in\mathbb{N}}$ above: we are now looking at a construction of the type that was described at the beginning of this paragraph.

Exercise 4.16 If you are familiar with **ZF**, then formulate and prove a theorem in **ZF+AC** that shows that every construction like the one of $\{K_n\}_{n\in\mathbb{N}}$ above is possible in **ZF+AC**.

From now on, we will no longer mark results that need the axiom of choice by an (**AC**), but we will maintain awareness of the problem.

4.2 Relations

Definition 4.17 Let M be a non-empty set. Recall that $M \times M$ denotes the set of all ordered pairs (a, b) of elements $a, b \in M$. A **(binary) relation** on M is a subset r of $M \times M$. The relation

$$\Delta(M) = \{\, (a, a) \mid a \in M \,\}$$

is called the **diagonal** of M. If r and s are relations on M, then

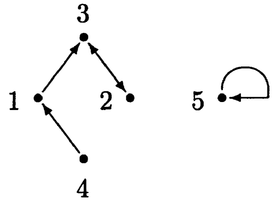
$$r^{-1} = \{\, (a, b) \mid (b, a) \in r \,\}$$

is the **inverse relation** of r, and

$$s \circ r = \{ (a,c) \mid \text{there exists } b \in M \text{ with } (a,b) \in r \text{ and } (b,c) \in s \}$$

is the **product** of the two relations r and s. If $r \subseteq s$, then s is called an **extension** of r.

It may appear more natural to denote $s \circ r$ by $r \circ s$; our notation is, however, chosen in order to conform with the product of two maps f, $g : M \to M$ when regarded as relations on M. To simplify the notation, relations are frequently written in infix notation, i.e., $a\ r\ b$ stands for $(a,b) \in r$. If the name of the relation is irrelevant, r is often simply denoted by an arrow; then $a \to b$ means $(a,b) \in r$. This notation suggests a convenient way of describing a relation $\to$ on a small finite set M: it suffices to draw the elements of M as points in the plane and connect appropriate points by arrows. Consider e.g. the relation $r = \{(1,3),(2,3),(3,2),(4,1),(5,5)\}$ on $M = \{1,2,3,4,5\}$. Then r is represented by the following diagram:



Exercise 4.18 Show that the product of relations is associative, but in general non-commutative, i.e., $r \circ (s \circ t) = (r \circ s) \circ t$ but in general $r \circ s \neq s \circ r$ for relations r, s, and t on M.

Definition 4.19 Let r be a relation on M. Then r is called

(i) **reflexive** if $\Delta(M) \subseteq r$,

(ii) **symmetric** if $r \subseteq r^{-1}$,

(iii) **transitive** if $r \circ r \subseteq r$,

(iv) **antisymmetric** if $r \cap r^{-1} \subseteq \Delta(M)$,

(v) **connex** if $r \cup r^{-1} = M \times M$,

(vi) **irreflexive** if $\Delta(M) \cap r = \emptyset$,

(vii) **strictly antisymmetric** if $r \cap r^{-1} = \emptyset$,

(viii) an **equivalence relation on** M if r is reflexive, symmetric, and transitive,

(ix) a **quasi-order on** M if r is reflexive and transitive,

(x) a **partial order on** M if r is reflexive, transitive and antisymmetric,

(xi) a **(linear) order on** M if r is a connex partial order on M, and

(xii) a **linear quasi-order on** M if r is a connex quasi-order on M.

The infix notation for an equivalence relation, a quasi-order, and a partial order on M is frequently $\sim$ or $\equiv$, $\preceq$, and $\leq$, respectively. The inverse of $\preceq$, $\leq$, and $\rightarrow$ is denoted by $\succeq$, $\geq$, and $\leftarrow$, respectively.

Exercise 4.20 (i) For each of the definitions (i)–(vii) above, write an equivalent version like the following one for symmetry: $a\ r\ b$ implies $b\ r\ a$ for all a, $b \in M$.

(ii) Show that if r is a relation that is irreflexive, symmetric, and transitive, then $r = \emptyset$.

(iii) Let X be a set. Show that $\subseteq$ is a partial order on $\mathcal{P}(X)$.

(iv) Let R be a ring, I an ideal of R. Show that the relation $\equiv_I$ on R defined by

$$a \equiv_I b \quad \text{iff} \quad a + I = b + I$$

is an equivalence relation on R.

(v) Let R be a domain. Show that the divisibility relation $\mid$ on R is a quasi-order which is not linear in general.

(vi) Let $\boldsymbol{I}$ be an interval on the real line and $x_0 \in \boldsymbol{I}$. Define a relation $\preceq$ on $\mathrm{C}(\boldsymbol{I}, \mathbb{R})$ by

$$f \preceq g \quad \text{iff} \quad f(x_0) \leq g(x_0).$$

Show that $\preceq$ is a linear quasi-order.

(vii) Convince yourself that $\leq$ on $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$ are linear orders.

For a thorough understanding of the theory below, let us explain exactly how the different kind of orderings are obtained from each other by specialization.

$$\begin{tikzcd}
 & \text{quasi-orders} \arrow[dl] \arrow[dr] & \\
\text{linear quasi-orders} \arrow[dr] & & \text{partial orders} \arrow[dl] \\
 & \text{(linear) orders} &
\end{tikzcd}$$

Quasi-orders are reflexive and transitive, but they allow non-comparable elements (i.e., $a \npreceq b$ and $b \npreceq a$) and the situation $a \preceq b$ and $b \preceq a$ with $a \neq b$. The natural example to think of is divisibility on a domain such as $\mathbb{Z}$. Passing to a linear quasi-order means to require, in addition, comparability of any two elements, i.e., $a \preceq b$ or $b \preceq a$ for all a and b in the underlying set. The set of functions of (vi) above is an easy example. On the other hand, passing to a partial order means to require antisymmetry, i.e., $a \leq b$ and $b \leq a$ is now possible only if $a = b$. The set inclusion of (iii) above is the natural example. Finally, orders combine all these properties, and they are exemplified by the natural orders on $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$. Let us emphasize again that we use the terms "order" and "linear order" as synonyms. If $\leq$ is a linear order on the set M, then it is also common to express this by saying that "M is totally ordered by $\leq$."

Exercise 4.21 Using finite sets and diagrams as described above, exhibit examples of relations that are

(i) reflexive and transitive but not symmetric,

(ii) reflexive and symmetric but not transitive,

(iii) symmetric and transitive but not reflexive.

A **partition** of M is a subset Π of the power set $\mathcal{P}(M)$ of M such that the elements of Π are pairwise disjoint, non-empty subsets of M whose union equals M. A subset S of M is called a **system of unique representatives** for the partition Π of M if $P \cap S$ contains exactly one element for each $P \in \Pi$. Whenever $\sim$ is an equivalence relation on M, then

$$[a] = \{ b \in M \mid b \sim a \}$$

is called the **equivalence class** of a with respect to $\sim$, and the set

$$\{ [a] \mid a \in M \}$$

of all equivalence classes is denoted by $M/{\sim}$. If, for example, we take for $\sim$ the equivalence relation $\equiv_I$ of Exercise 4.20 (iv), then, by the results of Section 1.5,

$$\begin{aligned} [a] &= \{ b \in R \mid b \equiv_I a \} \\ &= \{ b \in R \mid b + I = a + I \} \\ &= a + I \end{aligned}$$

for all $a \in R$. We already know that the set of residue classes

$$R/{\equiv_I} = R/I = \{ a + I \mid a \in R \}$$

forms a partition of R. We will now show that this phenomenon occurs with every equivalence relation and its equivalence classes.

Lemma 4.22 (i) If $\sim$ is an equivalence relation on M and $a, b \in M$, then $[a] = [b]$ iff $a \sim b$.

(ii) Whenever $\sim$ is an equivalence relation on M, then $M/{\sim}$ is a partition of M.

(iii) Whenever Π is a partition of M, and the relation $\sim_\Pi$ on M is defined by "$a \sim_\Pi b$ iff there exists $B \in \Pi$ with a, $b \in B$," then $\sim_\Pi$ is an equivalence relation on M.

Proof (i) Assume that $[a] = [b]$. We always have, by the reflexivity of $\sim$, $a \in [a]$, and so in this case, $a \in [b]$, from which it follows that $a \sim b$. Now suppose $a \sim b$. To prove $[a] \subseteq [b]$, let $c \in [a]$. Then $c \sim a$, which together with $a \sim b$ implies $c \sim b$ by the transitivity of $\sim$. We see that $c \in [b]$. The proof of the reverse inclusion is similar if we observe that $\sim$ is symmetric.

(ii) We have $[a] \subseteq M$ for each equivalence class, so their union is contained in M. Conversely, each $a \in M$ satisfies $a \sim a$ and thus $a \in [a]$, which shows that

$$M \subseteq \bigcup_{a \in M} [a].$$

It remains to show that the equivalence classes are pairwise disjoint. Assume for a contradiction that a, b, $c \in M$ with $[a] \neq [b]$ but $c \in [a] \cap [b]$. Then, by the definition of $[a]$ and $[b]$ and the symmetry of $\sim$, we have $a \sim c$ and $c \sim b$ and so $a \sim b$. Now (i) yields the desired contradiction.

(iii) Reflexivity and symmetry of $\sim_\Pi$ are immediate consequences of its definition. Now let a, b, $c \in M$ with $a \sim_\Pi b$ and $b \sim_\Pi c$. Then there exist B_1, $B_2 \in \Pi$ with a, $b \in B_1$ and b, $c \in B_2$. We see that $b \in B_1 \cap B_2$, and so $B_1 = B_2$ because different elements of Π are disjoint. It follows that $a \sim c$. □

Exercise 4.23 Let $\sim$ and Π be as in the lemma above. Then

$$\sim_{(M/\sim)} \;=\; \sim \quad \text{and} \quad M/(\sim_\Pi) \;=\; \Pi.$$

Recall that the difference between a quasi-order and a partial order on a set M is that the former allows the situation $a \preceq b$ and $b \preceq a$ with $a \neq b$. The next lemma shows that a quasi-order can be coerced into becoming a partial order by "lumping together" into equivalence classes those elements of M that cause the trouble.

Lemma 4.24 Let $\preceq$ be a quasi-order on M, and let us denote by $\sim$ the relation $\preceq \cap (\preceq)^{-1}$ on M. Then the following hold:

(i) The relation $\sim$ is an equivalence relation on M.

(ii) The relation $\leq$ on $M/\sim$ given by $[a] \leq [b]$ iff $a \preceq b$ is well-defined, and it is a partial order on $M/\sim$.

(iii) If $\preceq$ is a linear quasi-order on M, then $\leq$ is a linear order on $M/\sim$.

Proof (i) If $a \in M$, then $a \preceq a$, and so $a \sim a$. If $a \sim b$, then $a \preceq b$ and $b \preceq a$, which is also the definition of $b \sim a$. Now assume that $a \sim b$ and $b \sim c$. Then $a \preceq b$ and $b \preceq c$, and so $a \preceq c$. A similar argument shows that $c \preceq a$, and we see that $a \sim c$.

(ii) To prove that $\leq$ is well-defined, we must prove that $a \preceq b$ iff $a' \preceq b'$ whenever $[a] = [a']$ and $[b] = [b']$. Assume that a, a', b, $b' \in M$ have the latter property and $a \preceq b$. Using Lemma 4.22 (i), we see that $a' \preceq a$ and $b \preceq b'$, and so $a' \preceq b'$ by the transitivity of $\preceq$. The reverse implication is immediate from the symmetry of the problem. Reflexivity and transitivity of $\leq$ are now immediate consequences of the corresponding properties of $\preceq$. For antisymmetry, suppose $[a] \leq [b]$ and $[b] \leq [a]$. Then $a \preceq b$ and $b \preceq a$, whence $a \sim b$ and so $[a] = [b]$.

(iii) If $a, b \in M$, then $a \preceq b$ or $b \preceq a$, and so $[a] \leq [b]$ or $[b] \leq [a]$. □

The relation $\sim$ of the lemma above is called the equivalence relation **associated with** $\preceq$, and $\leq$ is called the partial order **associated with** $\preceq$.

Example 4.25 Let us look at the special case of the divisibility relation on a domain R. Here, $a \sim b$ means that a and b divide each other and thus are associated (Exercise 1.68 (xii)). The equivalence class $[a]$ of $a \in R$ thus collects all elements of R that are associated to a; in particular, $[0] = \{0\}$ and $[1]$ is the set of all units of R. The resulting partial order $\leq$ on $R/\sim$ is defined by $[a] \leq [b]$ iff $a \mid b$, and we get $[1] \leq [a] \leq [0]$ for all $a \in R$. (Cf. the remarks preceding Exercise 1.68.) In the special case $R = \mathbb{Z}$, this means that $[0] = \{0\}$ and $[m] = \{m, -m\}$ for $m \neq 0$. A system of unique representatives for $\mathbb{Z}/\sim$ is then given by the natural numbers.

Exercise 4.26 Describe the equivalence relation associated with the divisibility relation on $R = K[X]$ where K is a field, and find a system of unique representatives for $R/\sim$.

Exercise 4.27 Assume that $\preceq$ on M is already a partial order. Show that each equivalence class w.r.t. the associated equivalence relation contains exactly one element.

Let r be a relation on M. Then $r \cup r^{-1}$ is obviously the smallest relation extending r that is symmetric on M. It is called the **symmetric closure** of r. In order to get the smallest transitive relation on M extending r, we first define powers r^n of r for $n \in \mathbb{N}$ recursively by

$$r^0 = \Delta(M) \quad \text{and} \quad r^{n+1} = r \circ r^n.$$

Then we call

$$r^+ = \bigcup_{n=1}^{\infty} r^n$$

the **transitive closure** of r, and

$$r^* = \bigcup_{n=0}^{\infty} r^n$$

the **reflexive-transitive closure** of r. So if $a, b \in M$, then $a\ r^*\ b$ if either $a = b$, or there exists a chain

$$a = a_1\ r\ a_2\ r \cdots r\ a_n = b \qquad (n \geq 2,\ a_1, \ldots, a_n \in M).$$

For a symmetric relation r, this shows that r^* is also symmetric.

Exercise 4.28 Let r be a relation on M. Show the following:

(i) r^+ is the smallest transitive relation on M extending r.

(ii) r^* is the smallest reflexive and transitive relation on M extending r. If r is symmetric, then it is in fact the smallest equivalence relation on M extending r.

(iii) Let r_1 be the symmetric closure of the reflexive-transitive closure of r, and let r_2 be the reflexive-transitive closure of the symmetric closure of r. Show that $r_1 \subseteq r_2$, and give an example showing that the reverse inclusion does not hold in general.

With any relation r on M one may associate the **strict part** $r_s = r \setminus r^{-1}$ of r. Here, r will usually be a quasi-order $\preceq$ or a partial order $\leq$ on M, in which case r_s is denoted by $\prec$ or $<$, respectively; the inverse of $\prec$ or $<$ is then denoted by $\succ$ or $>$, respectively.

The most natural example to visualize the strict part of a relation is the natural order on $\mathbb{N}$, where $m < n$ means $m \leq n$ and $m \neq n$. If r is the divisibility relation on a domain R, then $a\ r_s\ b$ means that $a \mid b$ but not $b \mid a$, i.e., a is a proper divisor of b.

Exercise 4.29 Show the following:

(i) Let r be a relation on M. Then $r = r_s$ iff r is strictly antisymmetric.

(ii) If $\preceq$ is a quasi-order on M, then $\prec$ is strictly antisymmetric and transitive.

(iii) Let $\leq$ be a partial order. Whenever $a, b \in M$, then $a < b$ holds iff $a \leq b$ and $a \neq b$.

In view of the above exercise, the strict part of a quasi-order $\preceq$ is called the **strict partial order** associated with $\preceq$.

4.3 Foundedness Properties

Definition 4.30 Let r be a relation on M with strict part r_s, and let $N \subseteq M$. Then an element a of N is called **r-minimal (r-maximal)** in N if there is no $b \in N$ with $b\ r_s\ a$ (with $a\ r_s\ b$). For $N = M$ the reference to N is omitted. A **strictly descending (strictly ascending) r-chain** in

M is an infinite sequence $\{a_n\}_{n\in\mathbb{N}}$ of elements of M such that $a_{n+1}\ r_s\ a_n$ (such that $a_n\ r_s\ a_{n+1}$) for all $n \in \mathbb{N}$. The relation r is called **well-founded (noetherian)** if every non-empty subset N of M has an r-minimal (an r-maximal) element. r is a **well-order** on M if r is a well-founded linear order on M.

When it is clear from the context what relation r is being referred to, we will often speak of just minimal (maximal) elements and chains. In this section, these concepts will be applied to quasi-orders only, but they will also be relevant for other types of relations later on. It is an easy observation that minimality of a in N can be expressed by the condition "$b\ r\ a$ implies $a\ r\ b$ for all $b \in N$," and its maximality by "$a\ r\ b$ implies $b\ r\ a$ for all $b \in N$." Examples for the concepts defined above will be easier to give after the following proposition.

Proposition 4.31 *Let r be a relation on M. Then r is well-founded (noetherian) iff there are no strictly descending (no strictly ascending) r-chains in M.*

Proof If $\{a_n\}_{n\in\mathbb{N}}$ is a strictly descending r-chain in M, then the set

$$N = \{\, a_n \mid n \in \mathbb{N} \,\}$$

has no r-minimal element. Conversely, suppose $\emptyset \neq N \subseteq M$ and N has no r-minimal element. Then the set

$$A_a = \{\, b \in N \mid b\ r_s\ a \,\}$$

is not empty for each $a \in N$. The axiom of choice, applied to the family $\{A_a\}_{a\in N}$, provides a function

$$f : N \longrightarrow \bigcup_{a\in N} A_a \subseteq N$$

with $f(a)\ r_s\ a$ for all $a \in N$. Let $a_0 \in N$. Now the sequence $\{a_n\}_{n\in\mathbb{N}}$ defined recursively by $a_{n+1} = f(a_n)$ forms a strictly descending r-chain. The case in parentheses is handled analogously. □

Examples 4.32 (i) Corollary 0.4 states that the natural order on $\mathbb{N}$ is a well-order.

(ii) Lemma 4.2 says that the divisibility relation | on a PID is well-founded.

(iii) $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$ with their natural orders are not well-founded. In the case of $\mathbb{Z}$, the entire set $\mathbb{Z}$, or also the set of all even integers, do not have minimal elements. In $\mathbb{Q}$ and $\mathbb{R}$, there are many more examples of subsets without minimal elements, e.g., half-open intervals of the form $(a, b]$. For similar reasons, none of these orders is noetherian.

(iv) We claim that the relation $\subseteq$ on the set $\mathcal{I}(R)$ of ideals of a PID R is noetherian. Indeed, suppose there was an infinite sequence $\{I_n\}_{n\in\mathbb{N}}$ of ideals of R such that I_n was properly contained in I_{n+1} for all $n \in \mathbb{N}$. The axiom of choice provides us with a sequence $\{a_n\}_{n\in\mathbb{N}}$ of generators of the I_n, which, according to Exercise 1.68 (xi), would be an infinite sequence in which each member is properly divided by its successor, contradicting Lemma 4.2.

Exercise 4.33 Let R be a domain. Show the following:

(i) $a \in R$ is $|$-minimal in R iff it is a unit of R.

(ii) A non-zero non-unit $a \in R$ is irreducible iff it is a $|$-minimal element of $R \setminus U_R$.

(iii) If there exists $0 \neq b \in R$ which is $|$-maximal in $R \setminus \{0\}$, then R is a field.

Before we discuss the theory of well-founded and noetherian relations, we describe two applications that should give an idea of the relevance of these concepts. Suppose first r is a well-founded (noetherian) relation on the set M, and P is a property that an element of M may or may not have. Then the claim "$P(a)$ for all $a \in M$" can often be proved as follows. One assumes for a contradiction that the set

$$N = \{\, a \in M \mid P(a) \text{ does not hold} \,\}$$

is not empty. Then there exists an r-minimal (an r-maximal) element $b \in N$, and one then tries to achieve the desired contradiction by proving the existence of an element $c \in N$ with $c \; r_s \; b$ (with $b \; r_s \; c$). This type of argument is often referred to as **noetherian induction**, a terminology that is further explained by the following variant of the argument.

Exercise 4.34 Let r be a well-founded (noetherian) relation on the set M, and let P be a property that elements of M may or may not have. Suppose that the following can be proved: whenever $a \in M$ and $P(b)$ holds for all $b \in M$ with $b \, r_s \, a$ (with $a \, r_s \, b$), then $P(a)$ holds. Show that then $P(a)$ for all $a \in M$.

The second application concerns a fundamental problem of computer science and computational mathematics, namely, termination proofs for algorithms. To prove that an algorithm terminates for any input that meets certain specifications, one must essentially proceed as follows. Suppose that the algorithm employs m different variables. Let us call an m-tuple s of values, one for each variable, a *state* of the algorithm if there is a possible course of the computation that starts with an input as specified and encounters the configuration s of values for the variables at some point. One must now try to find a set M with a well-founded relation r and a map φ from the set S of all states to M such that the following holds: whenever s_1 and s_2 are states such that the algorithm, being in state s_1, may move to s_2 as its next state, then $\varphi(s_2) \; r_s \; \varphi(s_1)$. It is clear that the algorithm must then terminate, because an infinite run would give rise to a strictly descending r-chain in M. The argument can of course be modified to employ a noetherian instead of a well-founded relation.

Recall that for a quasi-order $\preceq$ on a set M, the equivalence relation $\sim$ associated with $\preceq$ is defined by $a \sim b$ iff $a \preceq b$ and $b \preceq a$. In order to show that a given quasi-order is well-founded, the following lemma is sometimes useful.

Lemma 4.35 Let $\preceq$ be a quasi-order on M with associated equivalence relation $\sim$, let $\leq$ be a well-founded partial order on N, and let $\varphi : M \longrightarrow N$ be a map such that for all $a, b \in M$, the following hold:

(i) $a \preceq b$ implies $\varphi(a) \leq \varphi(b)$, and

(ii) $\varphi(a) = \varphi(b)$ implies $a \sim b$.

Then $\preceq$ is well-founded.

Proof Assume for a contradiction that $\{a_n\}_{n \in \mathbb{N}}$ is a sequence of elements of M such that $a_j \prec a_i$ for $i < j$. Then $\varphi(a_j) \leq \varphi(a_i)$, and $\varphi(a_i) \neq \varphi(a_j)$ since otherwise $a_i \sim a_j$. So $\{\varphi(a_n)\}_{n \in \mathbb{N}}$ forms a strictly descending $\leq$-chain, a contradiction. □

Let us take another look at the equivalence relation $\sim$ associated with a quasi-order $\preceq$ on a set M. It is defined in such a way that the $\sim$-equivalence class $[a]$ of $a \in M$ collects all $b \in M$ that satisfy $a \preceq b$ and $b \preceq a$. If we are given a subset N of M, then we may of course consider the restriction of $\preceq$ to N and combine elements into such equivalence classes within N. It is easy to see that for $a \in N$, this "restricted equivalence class" equals $[a] \cap N$, where, as before, $[a]$ is the equivalence class of a w.r.t. $\sim$.

Lemma 4.36 If, in the situation described above, the intersection $[a] \cap N$ contains one element b which is $\preceq$-minimal in N, then every element of $[a] \cap N$ is $\preceq$-minimal in N.

Proof If an element c of N is in the same $\sim$-equivalence class as the minimal element b, then $c \sim b$, and so in particular, $c \preceq b$. To prove that c is minimal in N, suppose $d \preceq c$ for some $d \in N$. Then $d \preceq b$ by transitivity. From the minimality of b, it follows that $b \preceq d$, and so $c \preceq d$ again by transitivity. □

The intersections with N of $\sim$-equivalence classes of minimal elements of a set N will turn out to be an important tool in describing and visualizing various kinds of quasi-orders; we therefore call such an intersection a **min-class** in N.

There is now an obvious description of well-foundedness in terms of min-classes: $\preceq$ on M is well-founded iff every non-empty subset of M has at least one min-class. There can, however, be anything from one to infinitely many such min-classes. If we take, for example, the divisibility relation on $\mathbb{Z}$ and consider the subset $N = \mathbb{Z} \setminus \{1, -1\}$, then N has the infinitely many different min-classes $\{p, -p\}$, where p runs through all prime numbers (cf. Example 4.25). The next lemma characterizes those quasi-orders where the number of min-classes is always one.

Lemma 4.37 Let $\preceq$ be a quasi-order on M. Then the following are equivalent:

(i) Each non-empty subset N of M has exactly one min-class in N.

(ii) $\preceq$ is linear and well-founded.

Proof (i)$\Longrightarrow$(ii): To see that $\preceq$ is linear, let a, $b \in M$. If we had $a \not\preceq b$ and $b \not\preceq a$, then, as one easily proves, the set $N = \{a, b\}$ would have the two different min-classes $\{a\}$ and $\{b\}$. Well-foundedness of $\preceq$ is now immediate from the fact that min-classes are not empty by definition.

(ii)$\Longrightarrow$(i): Let $N \subseteq M$. Well-foundedness of $\preceq$ clearly implies the existence of at least one min-class in N. Now let a and b be two minimal elements in N. Linearity of $\preceq$ implies that at least one of $a \preceq b$ and $b \preceq a$ holds, and from the minimality of a and b, it follows that each of these implies the other. We see that $a \sim b$, and so the two are in the same $\sim$-equivalence class. □

We have already noted in Exercise 4.27 that for a partial order, each min-class contains exactly one element. We thus obtain the following important lemma.

Lemma 4.38 A partial order on M is a well-order iff every non-empty subset of M contains a unique minimal element. □.

The direction "$\Longrightarrow$" of the lemma above will be used frequently throughout this book. It is reflected in a common choice of terminology: the unique minimal element of a subset w.r.t. a well-order is called its **least** element. An example is of course $\mathbb{N}$ with its natural order. For an example of a linear, well-founded quasi-order that is really "quasi," i.e., not antisymmetric, we can modify Exercise 4.20 (vi): take for M the set of those functions in $C(\boldsymbol{I}, \mathbb{R})$ that satisfy $f(x_0) \in \mathbb{N}$. Then if $N \subseteq M$, the unique min-class in N equals

$$\{ f \in N \mid f(x_0) = m \},$$

where m is the least element of $\{ g(x_0) \mid g \in N \}$.

What we have not seen is an example where there are finitely many, but more than one min-classes. This can be achieved in a trivial way with a finite set: let $M = \{a, b, c\}$, and take $\preceq \; = \Delta M$. Then M itself has the three min-classes $\{a\}$, $\{b\}$, and $\{c\}$. Infinite sets that have this property in a non-trivial way are not easily constructed. They are, in fact, our next object of study.

Definition 4.39 Let $\preceq$ be a quasi-order on M and let $N \subseteq M$. Then a subset B of N is called a **Dickson basis**, or simply **basis** of N w.r.t. $\preceq$ if for every $a \in N$ there exists some $b \in B$ with $b \preceq a$. We say that $\preceq$ has the **Dickson property**, or is a **Dickson quasi-order**, if every subset N of M has a finite basis w.r.t. $\preceq$.

In the literature, Dickson quasi-orders are also called *well-quasi-orders.* The authors of this book agree that one should generally make every effort to conform with existing terminology. In this particular case though, using the term well-quasi-order makes it exceedingly difficult for the beginner to understand the ensuing theory: we will often discuss two or more closely related orders and quasi-orders and investigate whether they are well-founded and/or Dickson. It is therefore desirable to make a sharp terminological distinction between well-foundedness and the Dickson property, especially in view of the fact that a well-founded order is called a well-order. Besides, the term well-quasi-order is not universally agreed upon: Dickson quasi-orders have also been called *narrow quasi-orders.*

Note that a finite set always has a finite basis, namely, itself, so the Dickson property is relevant only for infinite $N \subseteq M$. The next lemma provides a natural class of examples of Dickson quasi-orders.

Lemma 4.40 Every well-founded linear quasi-order has the Dickson property. In particular, every well-order has the Dickson property.

Proof Let $\preceq$ be a well-founded linear quasi-order on the set M, and let $\emptyset \neq N \subseteq M$. Then N has a minimal element b, and we claim that $B = \{b\}$ is a basis of N. Indeed, every $a \in N$ satisfies at least one of $a \preceq b$ and $b \preceq a$, and we cannot have the former without the latter by the minimality of b. □

We see that the most obvious example of a Dickson quasi-order is the natural linear order on $\mathbb{N}$, where a basis of a non-empty set N is given by the one-element set consisting of the least element of N. The Dickson property is much more interesting for non-linear quasi-orders, but examples of those are not easy to come by. We will have to do a little theory out in mid-air before we can construct them.

Exercise 4.41 Show the following: If a quasi-order $\preceq$ is Dickson, then so is every quasi-order $\preceq'$ extending $\preceq$.

Condition (iii) of the proposition below is the best we will have for a while to visualize Dickson quasi-orders. Together with Lemma 4.37, it also provides an immediate second proof of Lemma 4.40.

Proposition 4.42 *Let $\preceq$ be a quasi-order on M with associated equivalence relation $\sim$. Then the following are equivalent:*

(i) $\preceq$ is a Dickson quasi-order.

(ii) Whenever $\{a_n\}_{n\in\mathbb{N}}$ is a sequence of elements of M, then there exists $i < j$ with $a_i \preceq a_j$.

(iii) *For every nonempty subset N of M, the number of min-classes in N is finite and non-zero.*

Proof (i)$\Longrightarrow$(ii): Set $N = \{ a_n \mid n \in \mathbb{N} \}$ and let B be a finite basis of N. Pick $j \in \mathbb{N}$ such that $j > i$ for all $i \in \mathbb{N}$ with $a_i \in B$. Then $a_{i_0} \preceq a_j$ for some $a_{i_0} \in B$, and the choice of j implies $i_0 < j$.

(ii)$\Longrightarrow$(iii): Suppose there exist infinitely many min-classes in some nonempty subset N of M. Using the axiom of choice, we get an infinite sequence $\{a_n\}_{n \in \mathbb{N}}$ of pairwise $\sim$-inequivalent minimal elements in N. By our assumption (ii), $a_i \preceq a_j$ for some $i < j$. From the the minimality of a_j, we conclude that $a_j \preceq a_i$ and so $a_i \sim a_j$, a contradiction. If, on the other hand, N has no minimal element, then we can produce a strictly descending $\preceq$-chain as in the proof of Proposition 4.31, contradicting (ii).

(iii)$\Longrightarrow$(i): Let N be a non-empty subset of M. Choosing one element out of each of the finitely many min-classes, we can find a finite subset B of N such that each $b \in B$ is minimal, and such that every minimal $a \in N$ is $\sim$-equivalent to some $b \in B$. We claim that B is a basis of N. Let $a \in N$. Then the set

$$N' = \{ d \in N \mid d \preceq a \}$$

contains a minimal element c. It is easy to see that c is minimal in N too, and so $c \sim b$ for some $b \in B$. We now have $b \preceq c \preceq a$ and hence $b \preceq a$. □

There are two important corollaries to the above proposition. The first one deals with Dickson quasi-orders that are actually partial orders. In its proof, we will use the statement of Exercise 4.27.

Corollary 4.43 *Let $\leq$ be a Dickson partial order on M. Then every non-empty subset N of M has a unique* **minimal finite basis** *B, i.e., a finite basis B such that $B \subseteq C$ for all other bases C of N. B consists of all minimal elements of N.*

Proof Let B be the set of all minimal elements of N. Then by the proposition, B is finite and non-empty. Moreover, for every $a \in N$ there exists some $b \in B$ with $b \leq a$. So B is a basis of N. Let now C be another basis of N. Then for every $b \in B$ there exists some $c \in C$ such that $c \leq b$, and so $c = b$ by the minimality of b. This shows that $B \subseteq C$. □

The next corollary is immediate from condition (iii) of Proposition 4.42.

Corollary 4.44 *Every Dickson quasi-order is well-founded.* □

The converse of this last corollary is false in general: consider the divisibility relation $\mid$ on $\mathbb{Z}$. We know that $\mid$ is well-founded (Example 4.32 (ii)), and we have already mentioned that in $N = \mathbb{Z} \setminus \{1, -1\}$, there are the infinitely many min-classes $\{p, -p\}$ (p prime). This counterexample also nicely illustrates the equivalences of Proposition 4.42: if $\{p_n\}_{n \in \mathbb{N}}$ is a sequence of pairwise different prime numbers, then there are no divisibilities in this sequence, and so $\mid$ does not satisfy Proposition 4.42 (ii). Only in the presence of linearity does well-foundedness indeed imply the Dickson property, because then the number of min-classes in each subset shrinks to one (Lemma 4.37).

The following proposition strengthens condition (ii) of Proposition 4.42.

Proposition 4.45 *Let $\preceq$ be a Dickson quasi-order on M, and let $\{a_n\}_{n\in\mathbb{N}}$ be a sequence of elements of M. Then there exists a strictly ascending sequence $\{n_i\}_{i\in\mathbb{N}}$ of natural numbers such that $a_{n_i} \preceq a_{n_j}$ for all $i < j$.*

Proof We define the sequence $\{n_i\}_{i\in\mathbb{N}}$ recursively, and by simultaneous induction on i we verify the following properties:

(i) $a_{n_i} \preceq a_{n_{i+1}}$ for all $i \in \mathbb{N}$, and

(ii) for all $i \in \mathbb{N}$, the set $\{\, n \in \mathbb{N} \mid a_{n_i} \preceq a_n \,\}$ is infinite.

For $i = 0$, let $\{b_1, \ldots, b_k\}$ be a finite basis of the set $\{\, a_n \mid n \in \mathbb{N} \,\}$, and for each j with $1 \leq j \leq k$, set

$$B_j = \{\, n \in \mathbb{N} \mid b_j \preceq a_n \,\}.$$

Then $\bigcup_{j=1}^{k} B_j = \mathbb{N}$ by the choice of B. Since the union of finitely many finite sets is finite, we can find a B_j which is infinite. Moreover, $b_j = a_m$ for some $m \in \mathbb{N}$, and we set $n_0 = m$. For $i + 1$, we consider the set

$$U_i = \{\, a_n \mid a_{n_i} \preceq a_n,\ n_i < n \,\}.$$

By condition (ii) for i, the set $\{\, n \in \mathbb{N} \mid a_n \in U_i \,\}$ is infinite. Choosing some finite basis of U_i, we can, as before, find an element a_m in this basis such that $a_m \preceq a_n$ for infinitely many different $n \in \mathbb{N}$, and we take $n_{i+1} = m$. Conditions (i) and (ii) obviously continue to hold. It now follows easily from condition (i) and the transitivity of $\preceq$ that $\{n_i\}_{i\in\mathbb{N}}$ has the desired property. □

For greater clarity, we did not mention the axiom of choice explicitly in the above proof. It is important to note, however, that we have used **AC**. This is because whenever a sequence $\{n_i\}_{i\in\mathbb{N}}$ is defined recursively, then really n_{i+1} is defined as $F(n_i)$ (or, more generally, as $F(\{n_0, \ldots, n_i\})$), where F is a function whose existence must be a priori guaranteed. Here, F is a function that assigns to each $k \in \mathbb{N}$ with the property that $a_k \preceq a_n$ for infinitely many $n \in \mathbb{N}$ a natural number $F(k) > k$ such that $a_k \preceq a_{F(k)}$ and $a_{F(k)}$ again has the property that $a_{F(k)} \preceq a_n$ for infinitely many $n \in \mathbb{N}$.

If $\preceq$ is a (Dickson) quasi-order on M, then we call $(M, \preceq)$ a **(Dickson) quasi-ordered set**; similarly for partial orders, orders, and well-orders on M. Recall that if M and N are sets, then the Cartesian product $M \times N$ of M and N is the set of all ordered pairs (a, b) with $a \in M$ and $b \in N$. Let now $(M, \preceq)$ and $(N, \preceq)$ be quasi-ordered sets (where the quasi-orders on M and N may be different). Then we define a quasi-order $\preceq'$ on $M \times N$ as follows:

$$(a, b) \preceq' (c, d) \quad \text{iff} \quad a \preceq c \text{ and } b \preceq d$$

for all (a, b), $(c, d) \in M \times N$. It is easy to see that that $\preceq'$ is indeed reflexive and transitive. $(M \times N, \preceq')$ is called the **direct product** of the quasi-ordered sets $(M, \preceq)$ and $(N, \preceq)$.

It is obvious that direct products of linear quasi-orders are not again linear in general: if we form $\mathbb{N} \times \mathbb{N}$ with the natural order on each copy of $\mathbb{N}$, then the elements $(0, 1)$ and $(1, 0)$ are incomparable. Therefore, the following important theorem finally provides us with a means to construct non-linear Dickson quasi-orders from the linear ones that we had before.

Theorem 4.46 *Let $(M, \preceq)$ and $(N, \preceq)$ be Dickson quasi-ordered sets, and let $(M \times N, \preceq')$ be their direct product. Then $(M \times N, \preceq')$ is a Dickson quasi-ordered set.*

Proof We verify (ii) of Proposition 4.42. Let $\{(a_n, b_n)\}_{n \in \mathbb{N}}$ be a sequence of elements of $M \times N$. By Proposition 4.45, there exists a strictly ascending sequence $\{n_i\}_{i \in \mathbb{N}}$ such that $a_{n_i} \preceq a_{n_j}$ for all $i < j$. By (ii) of Proposition 4.42 applied to the sequence $\{b_{n_i}\}_{i \in \mathbb{N}}$, there exist $i < j$ with $b_{n_i} \preceq b_{n_j}$ and thus $(a_{n_i}, b_{n_i}) \preceq' (a_{n_j}, b_{n_j})$. □

The definition of a direct product of two quasi-ordered sets extends naturally to an arbitrary finite number n of factors: if $(M_i, \preceq)$ are quasi-ordered sets for $1 \leq i \leq n$ and $M = M_1 \times \cdots \times M_n$ is their Cartesian product, then the **direct product** of the $(M_i, \preceq)$ is the quasi-ordered set $(M, \preceq')$, where $\preceq'$ is defined by

$$(a_1, \ldots, a_n) \preceq' (b_1, \ldots, b_n) \quad \Longleftrightarrow \quad a_i \preceq b_i \text{ for all } 1 \leq i \leq n.$$

The following corollary can now easily be proved form Theorem 4.46 using induction on the number of factors.

Corollary 4.47 *Let $(M_i, \preceq)$ be Dickson quasi-ordered sets for $1 \leq i \leq n$, and let $(M, \preceq')$ be the direct product of the $(M_i, \preceq)$. Then $(M, \preceq')$ is a Dickson quasi-ordered set.* □

The following special case which is known as **Dickson's lemma** is of utmost importance for the theory of Gröbner bases. (Recall that the natural order on $\mathbb{N}$, being linear and well-founded, is Dickson.)

Corollary 4.48 (Dickson's lemma) *Let $(\mathbb{N}^n, \leq')$ be the direct product of n copies of the natural numbers $(\mathbb{N}, \leq)$ with their natural ordering. Then $(\mathbb{N}^n, \leq')$ is a Dickson partially ordered set. More explicitly, every subset S of $\mathbb{N}^n$ has a finite subset B such that for every $(m_1, \ldots, m_n) \in S$, there exists $(k_1, \ldots, k_n) \in B$ with $k_i \leq m_i$ for $1 \leq i \leq n$.* □

Our proof of Dickson's lemma requires the axiom of choice since the proof of Theorem 4.46 uses the implication (ii) $\Longrightarrow$ (i) of Proposition 4.42 as well as Proposition 4.45. It is worthwhile noting that if one is content with Dickson's lemma itself rather than the more general Corollary 4.47, then one does not need the axiom of choice; the following proposition can obviously replace Theorem 4.46 in the proof of Dickson's lemma.

Proposition 4.49 *Let $(M, \preceq)$ be a Dickson quasi-ordered set, $(\mathbb{N}, \leq)$ the natural numbers with the natural order. Then the direct product $(M \times \mathbb{N}, \preceq')$ is a Dickson quasi-ordered set.*

Proof Let S be a non-empty subset of $(M \times \mathbb{N}, \preceq')$. For every $n \in \mathbb{N}$, we set

$$M_n = \{\, a \in M \mid (a, n) \in S \,\},$$

and we let B_n be a finite basis of M_n and C a finite basis of $\bigcup_{n \in \mathbb{N}} B_n$. Finally, let $r \in \mathbb{N}$ be such that $C \subseteq \bigcup_{i=1}^{r} B_i$, and set

$$B = \{\, (a, i) \in M \times \mathbb{N} \mid 1 \leq i \leq r,\ a \in B_i \,\}.$$

It is clear that B is a finite set, and we claim that it is also a basis of S. Let $(a, n) \in S$. Then $a \in M_n$, and so we can find $b \in B_n$ with $b \preceq a$ and thus $(b, n) \preceq' (a, n)$. If $n \leq r$, then $(b, n) \in B$ and we are done. Otherwise, there exists $c \in C$ with $c \preceq b \preceq a$, and $c \in B_i$ for some $i \leq r < n$. We see that $(c, i) \in B$ and $(c, i) \preceq' (a, n)$. □

Exercise 4.50 Draw a two-dimensional coordinate system (of which you will need the first quadrant only) with the points 1–10 labeled on each axis. Identifying the element (m, n) of $\mathbb{N}^2$ with the point whose x- and y-coordinates are m and n, respectively, use shading to indicate the following subsets of $\mathbb{N}^2$ in your picture.

$$\begin{aligned} A &= \{\, (m, n) \in \mathbb{N}^2 \mid 3 \leq m,\ 4 \leq n \,\} \\ B &= \{\, (m, n) \in \mathbb{N}^2 \mid 5 \leq m,\ 1 \leq n \,\} \\ C &= \{\, (m, n) \in \mathbb{N}^2 \mid 1 \leq m,\ 2 \leq n \,\} \end{aligned}$$

Now indicate the set $D = (A \cup B) \cap C$, and find a finite basis of D according to Dickson's lemma. Do it with common sense first, then try to follow the proof of the proposition above.

We conclude the discussion of the Dickson property by taking another look at the relationship between the Dickson property and well-foundedness. We saw that every Dickson quasi-order is well-founded, but not vice versa (Corollary 4.44 and the discussion following it). Since every quasi-order that extends a given Dickson quasi-order is Dickson too, we immediately obtain the following lemma which will be of great importance in the theory of Gröbner bases.

Lemma 4.51 Let $\preceq$ be a Dickson quasi-order on M. Then every quasi-order $\preceq'$ on M extending $\preceq$ is well-founded. □

The following lemma (which is of much less importance to us than the previous one) shows that the converse is true too.

Lemma 4.52 Let $\preceq$ be a quasi-order on M. Then $\preceq$ has the Dickson property iff all quasi-orders $\preceq'$ on M extending $\preceq$ are well-founded.

Proof The direction "$\Longrightarrow$" is Lemma 4.51. Conversely, suppose $\preceq$ is not Dickson. Then by Proposition 4.42 there exists a sequence $\{a_n\}_{n\in\mathbb{N}}$ of elements of M such that $a_i \not\preceq a_j$ for all $i < j$. We will define a quasi-order $\preceq'$ on M which extends $\preceq$ and is not well-founded. We set $b \preceq' c$ iff $b \preceq c$ or there exist $i < j$ such that $b \preceq a_j$ and $a_i \preceq c$. Then $\preceq'$ clearly extends $\preceq$. Moreover, $\preceq'$ is obviously reflexive, and a straightforward though somewhat tedious argument shows that it is also transitive and hence a quasi-order on M. We claim that $\{a_n\}_{n\in\mathbb{N}}$ is a strictly descending $\preceq'$-chain. It is immediate from the definition of $\preceq'$ that $a_j \preceq' a_i$ for $i < j$, and it remains to prove that $a_i \not\preceq' a_j$ for all $i < j$. Assume for a contradiction that $a_i \preceq' a_j$ with $i < j$. Since $a_i \not\preceq a_j$, there must exist l, $k \in \mathbb{N}$ with $k < l$ such that $a_i \preceq a_l$ and $a_k \preceq a_j$. But this implies that $l \leq i$ and $j \leq k$, which in turn implies $l < k$, a contradiction. □

We will now discuss a condition on partially ordered sets that will turn out to be a weakening of the Dickson property.

Definition 4.53 Let $\leq$ be a partial order on the set M. If $a \in M$, then we call the set

$$U_a = \{ b \in M \mid a < b \}$$

the **upper set** (w.r.t. $\leq$) of a in M. We say that $\leq$ has the **König property**, or is a **König partial order**, if for all $a \in M$, the upper set U_a has a finite basis, and so does the entire set M.

The above definition would also make sense for quasi-orders, but is really relevant only for partial orders. It is clear that every Dickson partial order is König; the example below shows that the converse is not true.

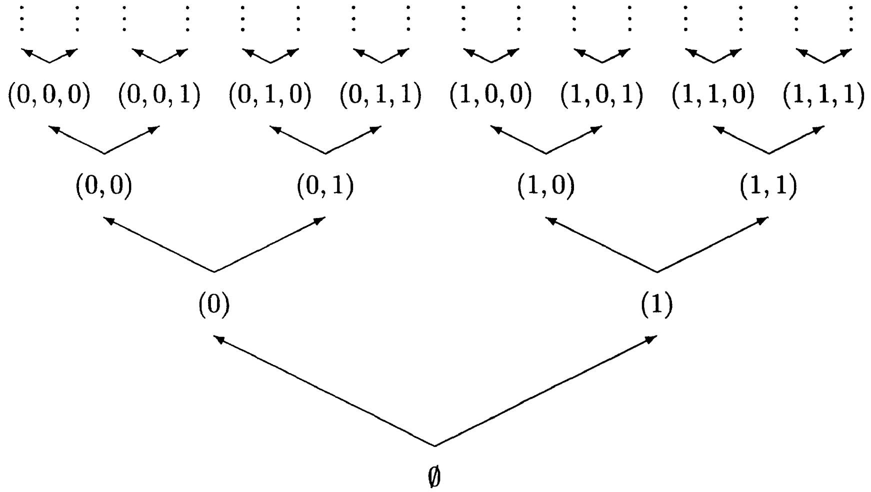
Example 4.54 Let M be the set of all finite tuples whose entries are either 0 or 1, i.e.,

$$M = \bigcup_{n\in\mathbb{N}} \{0,1\}^n .$$

It is easy to see that the relation $\leq$ on M defined by

$$(a_1, \ldots, a_m) \leq (b_1, \ldots, b_n) \quad \Longleftrightarrow \quad m \leq n \text{ and } a_i = b_i \text{ for } 1 \leq i \leq m$$

is a partial order on M. The partially ordered set $(M, \leq)$ is also called the *full binary tree*, and the following diagram indicates how it should be visualized.



From the definition of $\leq$, it is easy to see that a finite basis of the upper set of an element $(a_1, \ldots, a_m) \in M$ is given by the set

$$\{(a_1, \ldots, a_m, 0), (a_1, \ldots, a_m, 1)\}.$$

Moreover, $\{\emptyset\}$ is a basis of all of M, and so $\leq$ has the König property. On the other hand, the elements of the set

$$N = \{\, (a_1, \ldots, a_{n+1}) \in M \mid n \in \mathbb{N},\ a_i = 0 \text{ for } 1 \leq i \leq n, \text{ and } a_{n+1} = 1 \,\}$$

are all incomparable under $\leq$. So N cannot have a finite basis, and we see that $\leq$ does not have the Dickson property.

The main theorem on König partial orders states that the following weaker version of the property of Proposition 4.45 holds.

Theorem 4.55 (König's Lemma) *Let $\leq$ be a König partial order on the infinite set M. Then there exists a sequence $\{a_n\}_{n \in \mathbb{N}}$ of elements of M such that $a_m < a_n$ for all $m < n$.*

Proof We define the desired sequence recursively, and by simultaneous induction on n, we verify the following properties for all $n \in \mathbb{N}$.

(i) $a_n < a_{n+1}$.

(ii) The upper set U_{a_n} of a_n is infinite.

Let B be a finite basis of all of M. It is clear that then

$$M = \bigcup_{b \in B} (\{b\} \cup U_b).$$

Since M is an infinite set, there must exist $b \in B$ such that U_b is infinite, and we take for a_0 such an element of B. Now suppose a_i has been defined for $0 \leq i \leq n$. Let B_n be a finite basis of the upper set U_{a_n} of a_n. Then clearly

$$U_{a_n} = \bigcup_{b \in B_n} (\{b\} \cup U_b),$$

and since U_{a_n} is infinite by induction hypothesis, U_b must be infinite for at least one $b \in B_n$. Now if we take for a_{n+1} such an element of B_n, then a_{n+1} satisfies conditions (i) and (ii). □

Exercise 4.56 Explain how the above proof of König's lemma uses **AC**. (Hint: Cf. the remarks following the proof of Proposition 4.45.)

Exercise 4.57 Use König's lemma to give an alternate proof of Proposition 4.45 for partial orders.

Exercise 4.58 Let $\leq$ be a partial order on the non-empty set M. Show that $\leq$ has the König property iff the number of $\leq$-minimal elements in M and in every non-empty upper set U_a $(a \in M)$ is finite and non-zero. (Hint: Imitate the proof of Proposition 4.42.)

4.4 Some Special Orders

Besides Dickson's lemma, the other essential combinatorial ingredient in the theory of Gröbner bases is the concept of admissible orders. To see what this means, we must first recall from Section 2.1 that $\mathbb{N}^n$ is an additive Abelian monoid in a natural way: the operation is componentwise addition of n-tuples, and $(0, \ldots, 0)$ is the neutral element. If M is any Abelian monoid, we will denote its binary operation by $+$ and its neutral element by 0.

Definition 4.59 Let M be an Abelian monoid and let $\leq$ be a linear order on M. Then we say $\leq$ is **admissible** if for all a, b, $c \in M$,

(i) $0 \leq a$, and

(ii) $a < b$ implies $a + c < b + c$.

If $\leq$ is an admissible order on M, then we call $(M, \leq)$ an **ordered monoid**.

Exercise 4.60 Show that any ordered monoid satisfies the following cancelation law:

$$a + c = b + c \quad \text{implies} \quad a = b$$

for all a, b, $c \in M$.

The following exercise provides the most important examples of admissible orders on the additive monoid $\mathbb{N}^n$.

Exercise 4.61 Show that each of the following is an admissible order on the additive monoid $\mathbb{N}^n$.

(i) $(k_1, \ldots, k_n) \leq (m_1, \ldots, m_n)$ iff the following condition holds: either $(k_1, \ldots, k_n) = (m_1, \ldots, m_n)$, or there exists $1 \leq i \leq n$ with $k_j = m_j$ for $1 \leq j \leq i-1$ and $k_i < m_i$. This admissible order is called the **lexicographical**, or **lexical**, order on $\mathbb{N}^n$ since it orders the elements of $\mathbb{N}^n$ as if they were words in a dictionary, where 0 is the letter A, 1 is the letter B, and so on.

(ii) $(k_1, \ldots, k_n) \leq (m_1, \ldots, m_n)$ iff the following condition holds: either $(k_1, \ldots, k_n) = (m_1, \ldots, m_n)$, or there exists $1 \leq i \leq n$ with $k_j = m_j$ for $i+1 \leq j \leq n$ and $k_i < m_i$. This admissible order is called the **inverse lexicographical**, or **inverse lexical**, order on $\mathbb{N}^n$.

(iii) Let $\leq'$ be an order on $\mathbb{N}^n$ that satisfies condition (ii) of Definition 4.59, e.g., an admissible order or the inverse of an admissible order. Set $(k_1, \ldots, k_n) \leq (m_1, \ldots, m_n)$ iff the following condition holds:

$$\begin{aligned} \sum_{i=1}^{n} k_i &< \sum_{i=1}^{n} m_i, \quad \text{or} \\ \sum_{i=1}^{n} k_i &= \sum_{i=1}^{n} m_i \quad \text{and} \quad (k_1, \ldots, k_n) \leq' (m_1, \ldots, m_n). \end{aligned}$$

(iv) Let $1 \leq i < n$, let $\leq_1$ and $\leq_2$ be admissible orders on $\mathbb{N}^i$ and $\mathbb{N}^{n-i}$, respectively, and set $(k_1, \ldots, k_n) \leq (m_1, \ldots, m_n)$ iff the following condition holds:

$$\begin{aligned} &(k_1, \ldots, k_i) <_1 (m_1, \ldots, m_i), \text{ or} \\ &(k_1, \ldots, k_i) = (m_1, \ldots, m_i) \text{ and } (k_{i+1}, \ldots, k_n) \leq_2 (m_{i+1}, \ldots, m_n). \end{aligned}$$

This type of order is often referred to as a **block order** on $\mathbb{N}^n$.

In the following discussion, we use the notation (m) for the n-tuple $(m_1, \ldots, m_n)$.

Theorem 4.62 *Let $0 \neq n \in \mathbb{N}$. Let $\leq$ be an admissible order on the additive monoid $\mathbb{N}^n$ and $\leq'$ the partial order on $\mathbb{N}^n$ as a direct product of n copies of $\mathbb{N}$ with its natural order. Then $\leq$ is a well-order, and it extends $\leq'$, i.e., $(k) \leq' (m)$ implies $(k) \leq (m)$ for all $(k), (m) \in \mathbb{N}^n$.*

Proof If $(k) \leq' (m)$ in $\mathbb{N}^n$, then there exists $(l) \in \mathbb{N}^n$ with $(k) + (l) = (m)$. Since $(0) \leq (l)$, this implies

$$(k) = (k) + (0) \leq (k) + (l) = (m).$$

This shows that $\leq$ extends $\leq'$. By Dickson's lemma, $\leq'$ is a Dickson partial order on $\mathbb{N}^n$, and so by Lemma 4.51, $\leq$ is a well-order on $\mathbb{N}^n$. □

To prove the important fact that every admissible order is a well-order, we have used Lemma 4.51 which in turn really was a corollary to the general theory. The following exercise provides a more hands-on proof which shows how one obtains $\leq$-minimal elements from finite bases w.r.t. $\leq'$.

Exercise 4.63 Let $0 \neq n \in \mathbb{N}$. Let $\leq$ be an admissible order on the additive monoid $\mathbb{N}^n$ and $\leq'$ the partial order on $\mathbb{N}^n$ as a direct product of n copies of $\mathbb{N}$ with its natural order. Suppose N is a non-empty subset of $\mathbb{N}^n$. Let B be a finite basis of N w.r.t. the Dickson quasi-order $\leq'$, and let b be the $\leq$-least element of B. Show that b is the $\leq$-least element of N.

We will now describe a method to construct admissible orders on $\mathbb{N}^n$. Consider the ring $\mathbb{R}[Z]$ of univariate polynomials in Z with real coefficients. For $0 \neq f \in \mathbb{R}[Z]$, we denote by $\mathrm{HC}(f)$ the head coefficient of f, i.e., the coefficient of Z^m in f where $\deg(f) = m$. Define

$$P = \{\, f \in \mathbb{R}[Z] \mid f \neq 0,\ \mathrm{HC}(f) > 0 \,\},$$

and set

$$f \leq g \quad \text{iff} \quad g - f \in P \cup \{0\}.$$

Then one easily verifies that $\leq$ is a linear order on $\mathbb{R}[Z]$ in with $r < Z$ for all $r \in \mathbb{R}$. Moreover, on $\mathbb{R}$ this order coincides with the natural linear order of $\mathbb{R}$. As an extension ring of $\mathbb{Q}$, $\mathbb{R}[Z]$ forms in a natural way a $\mathbb{Q}$-vector space. We will call $a_1, \ldots, a_n \in \mathbb{R}[Z]$ **rationally independent** if they are linearly independent in this $\mathbb{Q}$-vector space.

Lemma 4.64 Let $0 < a_1, \ldots, a_n \in \mathbb{R}[Z]$ be rationally independent, and let the relation $\leq$ on $\mathbb{Q}^n$ be defined by

$$(q_1, \ldots, q_n) \leq (r_1, \ldots, r_n) \quad \text{iff} \quad \sum_{i=1}^{n} a_i q_i \leq \sum_{i=1}^{n} a_i r_i.$$

Then $\leq$ is a linear order, and $\leq \cap\ (\mathbb{N}^n)^2$ is an admissible order on $\mathbb{N}^n$.

Proof The fact that $\leq$ is reflexive, transitive, and connex is obvious. Antisymmetry follows from the fact that $a_1, \ldots, a_n$ are rationally independent. If (q), (r), $(s) \in \mathbb{Q}^n$ and $(q) \leq (r)$, then

$$\sum_{i=1}^{n} a_i q_i \leq \sum_{i=1}^{n} a_i r_i\,,$$

and so

$$\sum_{i=1}^{n} a_i (q_i + s_i) \leq \sum_{i=1}^{n} a_i (r_i + s_i)\,.$$

We see that $(q) + (s) \leq (r) + (s)$. Finally, we have $(0) \leq (q)$ for $(q) \in \mathbb{N}^n$ since

$$\sum_{i=1}^{n} a_i q_i \geq 0. \quad \square$$

It can be proved that in fact every admissible order on $\mathbb{N}^n$ can be obtained in this way, using polynomials $a_1, \ldots, a_n$ of degree less than or equal to n (cf. Section "Term Orders and Universal Gröbner Bases" on p. 514 in the appendix). The following examples, whose verification is left to the reader, demonstrate this for some special cases.

Examples 4.65 (i) The lexicographical order on $\mathbb{N}^n$ is induced by

$$(Z^{n-1}, Z^{n-2}, \ldots, Z, 1).$$

(ii) The inverse lexicographical order on $\mathbb{N}^n$ is induced by

$$(1, Z, \ldots, Z^{n-2}, Z^{n-1}).$$

(iii) The order of Exercise 4.61 (iii) on $\mathbb{N}^n$ is induced by

$$(Z^n + Z^{n-1}, Z^n + Z^{n-2}, \ldots, Z^n + Z, Z^n + 1).$$

A slightly trickier argument shows that it is also induced by

$$(Z^{n-1} + Z^{n-2}, Z^{n-1} + Z^{n-3}, \ldots, Z^{n-1} + 1, Z^{n-1}).$$

(iv) Let $1 \leq i < n$, and $<_1$ and $<_2$ the orders of Exercise 4.61 (iii) on

$$\mathbb{N}^i \quad \text{and} \quad \mathbb{N}^{n-i},$$

respectively. Using the second characterization of (iii) above, it is not hard to see that the order of Exercise 4.61 (iv) is induced by

$$(Z^n + Z^{n-1}, \ldots, Z^n + Z^{n-i+1}, Z^n, Z^{n-i} + Z^{n-i-1}, \ldots, Z^{n-i} + 1, Z^{n-i}).$$

Exercise 4.66 Let $\leq_1$ and $\leq_2$ be the admissible orders on $\mathbb{N}^3$ induced by the triples $(1, e, \pi)$ and $(Z, \pi Z + 1, e)$, respectively, where e is the base of the natural logarithm. How do the triples $(1,1,0)$, $(1,0,1)$, $(1,3,0)$, and $(1,0,2)$ relate to each other in each of these orders?

The crucial connection between the above results on $\mathbb{N}^n$ and polynomials will be made by passing from a term $X_1^{\nu_1} \cdot \cdots \cdot X_n^{\nu_n}$ to its exponent tuple $(\nu_1, \ldots, \nu_n)$. Dickson's lemma will then translate into a result on divisibility of terms, and admissible orders on $\mathbb{N}^n$ will give rise to linear orders of a certain kind on the set T of all terms. We will have to consider a certain quasi-order on the set of all polynomials which is dependent only on the terms occuring in the polynomials and on the linear order on T. Now the set of all terms occuring in a polynomial is just a finite subset of T, so the following definition and theorem will apply.

Let $(M, \leq)$ be an ordered set, and let $\mathcal{P}_{\text{fin}}(M)$ be the set of all finite subsets of M. Every $\emptyset \neq A \in \mathcal{P}_{\text{fin}}(M)$ obviously has a maximal and a minimal element w.r.t. the order $\leq$. We denote these by $\max(A)$ and $\min(A)$, respectively. With $A' = A \setminus \{\max(A)\}$, we define a binary relation $\leq'$ on $\mathcal{P}_{\text{fin}}(M)$ as follows. Let A, $B \in \mathcal{P}_{\text{fin}}(M)$; then $A \leq' B$ is defined by recursion on the number $|A|$: if $A = \emptyset$, then $A \leq' B$. If $A \neq \emptyset$, then $A \leq' B$ iff $B \neq \emptyset$ and the following conditions holds:

$$\begin{array}{rcll} \max(A) & < & \max(B), & \text{or} \\ \max(A) & = & \max(B) & \text{and} \quad A' \leq' B'. \end{array}$$

Lemma 4.67 Let $(M, \leq)$, $\mathcal{P}_{\text{fin}}(M)$, and $\leq'$ be as in the definition above. Then $(\mathcal{P}_{\text{fin}}(M), \leq')$ is an ordered set.

Proof One easily proves by induction on $|A|$ that $A \leq' A$ for all $A \in \mathcal{P}_{\text{fin}}(M)$, i.e., $\leq'$ is reflexive. For transitivity, assume that A, B, $C \in \mathcal{P}_{\text{fin}}(M)$ with $A \leq' B$ and $B \leq' C$. We use induction on $n = |A|$ to prove that $A \leq' C$. This is trivial if $n = 0$. If $n > 0$, then $A \neq \emptyset$, and it is easy to see from the definition of $\leq'$ that the same must be true for B and C. We thus have

$$\max(A) \leq \max(B) \quad \text{and} \quad \max(B) \leq \max(C).$$

If at least one of the inequalities is strict, then it follows immediately that $A \leq' C$. If we have equality in both cases, then

$$A' \leq' B' \leq' C',$$

and so $A' \leq' C'$ by induction hypothesis. This together with $\max(A) = \max(C)$ implies that $A \leq' C$.

For antisymmetry, suppose A, $B \in \mathcal{P}_{\text{fin}}(M)$ with $A \leq' B$ and $B \leq' A$. To prove that $A = B$, we proceed by induction on $n = |A|$. If $n = 0$, then $A = B = \emptyset$. If $n > 0$, then both A and B must be non-empty, and we must have $\max(A) = \max(B)$ and both $A' \leq' B'$ and $B' \leq' A'$. The induction hypothesis now implies that $A = B$.

It remains to prove that $\leq'$ is connex. Let A, $B \in \mathcal{P}_{\text{fin}}(M)$. If $B = \emptyset$, then $B \leq' A$. For the remaining case $B \neq \emptyset$, we use induction on $n = |A|$ to prove that $A \leq' B$ or $B \leq' A$. This is trivial if $n = 0$. Finally, if $n > 0$, then either the maxima of A and B are different, in which case we are done, or they agree, in which case we apply the induction hypothesis to A' and B'. □

Exercise 4.68 With the notation and assumptions of the previous lemma, assume that A, $B \in \mathcal{P}_{\text{fin}}(M)$ and A is a proper subset of B. Show that $A \leq' B$.

Theorem 4.69 *If $(M, \leq)$ is a well-ordered set, then so is $(\mathcal{P}_{\text{fin}}(M), \leq')$.*

The proof of the theorem will employ a method that is also known as *Cantor's second diagonal argument.* Before giving the general proof, we will illustrate the argument using the special case where $(M, \leq)$ is $\mathbb{N}$ with its natural order. Assume that there was a strictly descending $\leq'$-chain $\{A_n\}_{n\in\mathbb{N}}$ in $\mathcal{P}_{\text{fin}}(\mathbb{N})$. We will show how one can arrive at a contradiction by producing a strictly descending chain $\{a_n\}_{n\in\mathbb{N}}$ of natural numbers. To this end, think of the elements of each set A_n as written in decreasing order from left to right, and the sequence of the A_n as written in descending order from the top down, as in the following example.

$$\begin{array}{lcl} A_1 & = & \{17,13,9,5,1,0\} \\ A_2 & = & \{17,12,7,6,5,4,3,2,1\} \\ A_3 & = & \{16,14,12,11,10,3,0\} \\ A_4 & = & \{16,13,12,9,5,4,3,0\} \\ A_5 & = & \{16,13,12,7,6,5,4,1\} \\ A_6 & = & \{16,13,12,7,5,4,3,2,1\} \\ & \vdots & \end{array}$$

From the fact that $A_{n+1} <' A_n$, we conclude that $\max(A_{n+1}) \leq \max(A_n)$ for all $n \in \mathbb{N}$. It follows that there exists $n_0 \in \mathbb{N}$ with $\max(A_n) = \max(A_{n_0})$ for all $n \geq n_0$. We set a_0 equal to this maximum. Now if we drop everything above A_{n_0} from the list, then we still have a strictly descending $\leq'$-chain. Since all elements of the chain have the same maximum a_0, we can cross out a_0 everywhere (i.e., delete the leftmost column) and still have the same kind of chain. Moreover, every natural number that is left on the list must be strictly less than a_0. We can now repeat the game arbitrarily many times to arrive at the desired chain.

Proof of Theorem 4.69 Assume for a contradiction that there is a strictly descending $\leq'$-chain $\{A_n\}_{n\in\mathbb{N}}$ in $\mathcal{P}_{\text{fin}}(M)$. We will show that then there exists a strictly descending $\leq$-chain in M. We first note that $A_n \neq \emptyset$ for all $n \in \mathbb{N}$ by the definition of $\leq'$. We construct by recursion on $k \in \mathbb{N}$ a sequence

$$\{(a_k, \{B_{kn}\}_{n\in\mathbb{N}})\}_{k\in\mathbb{N}}$$

consisting of ordered pairs; the first component of each pair is an element of M, the second component is a sequence of elements of $\mathcal{P}_{\text{fin}}(M)$. By simultaneous induction on k, we verify the following properties of this sequence:

(i) $\{a_k\}_{k\in\mathbb{N}}$ is a strictly descending $\leq$-chain in M,

(ii) whenever n, $k \in \mathbb{N}$, then $b < a_k$ for all $b \in B_{kn}$, and

(iii) $\{B_{kn}\}_{n\in\mathbb{N}}$ is a strictly descending $\leq'$-chain in $\mathcal{P}_{\text{fin}}(M)$ for each $k \in \mathbb{N}$.

Condition (i) will then yield the desired contradiction to the well-foundedness of $\leq$ on M. For $k = 0$, let

$$C = \{\max(A_n) \mid n \in \mathbb{N}\},$$

and set $a_0 = \min(C)$. Now let j be the least index such that $a_0 \in A_j$. Then $a_0 = \max(A_n)$ for all $n \geq j$, and so if we set $B_{0n} = (A_{j+n})'$, then conditions (ii) and (iii) are satisfied. For $k+1$, we let

$$D = \{\max(B_{kn}) \mid n \in \mathbb{N}\},$$

and set $a_{k+1} = \min(D)$. Now let j be the least index such that $a_{k+1} \in B_{kj}$. Then $a_{k+1} = \max(B_{kj})$ for all $n \geq j$, and so if we set

$$B_{(k+1)n} = (B_{k(n+j)})',$$

then conditions (ii) and (iii) are satisfied. Moreover, condition (ii) for k implies that $a_k > a_{k+1}$, which proves (i). □

The following exercise generalizes the theorem above to well-founded linear quasi-orders.

Exercise 4.70 Let $\preceq$ be a linear quasi-order on the set M with associated equivalence relation $\sim$. For any subset A of M we set

$$A/{\sim} = \left\{ [a] \;\middle|\; a \in A \right\} \subseteq M/{\sim}.$$

We denote by $\leq$ the associated order on $M/{\sim}$, and by $\leq'$ the induced order on $\mathcal{P}_{\text{fin}}(M/{\sim})$ of the above theorem. We define a binary relation $\preceq'$ on $\mathcal{P}_{\text{fin}}(M)$ by setting

$$A \preceq' B \quad \text{iff} \quad A/{\sim} \leq' B/{\sim}$$

for $A, B \in \mathcal{P}_{\text{fin}}(M)$. (Note that $A, B \in \mathcal{P}_{\text{fin}}(M)$ implies $A/{\sim}, B/{\sim} \in \mathcal{P}_{\text{fin}}(M/{\sim})$). Show the following:

(i) $\preceq'$ as defined above is a linear quasi-order on $\mathcal{P}_{\text{fin}}(M)$.

(ii) If the linear quasi-order $\preceq$ on M is well-founded, then so is the induced linear quasi-order $\preceq'$ on $\mathcal{P}_{\text{fin}}(M)$.

4.5 Reduction Relations

In the discussion preceding Lemma 4.22, we saw that the elements of the residue class ring of a ring modulo an ideal are equivalence classes with respect to an equivalence relation. This situation occurs frequently in algebra: an algebraic structure (such as a group, a ring, etc.) is given as the set $M/{\sim}$ of equivalence classes of a set M with respect to an equivalence relation $\sim$ on M. The set M is often given in such a way that membership in M can be decided algorithmically. The equivalence relation $\sim$, however, is usually given in an indirect way, e.g., as the least equivalence relation that contains some given relation on M and is closed under certain rules. When one wants to do computations with the elements of $M/{\sim}$, the most fundamental problem that has to be solved in an algorithmic way is the following: given $a, b \in M$, decide whether $[a] = [b]$, i.e., whether $a \sim b$. This problem, the **equivalence problem**, is the set-theoretic version of the *word problem* for the algebraic structure M. If, for example, $M/\sim$ is $K[X]/(f)$, the univariate polynomial ring over the field K modulo the ideal generated by f, then $[g] = [h]$ iff $g - h \in (f)$ iff $f \mid (g - h)$ iff the remainder of $g - h$ upon division by f equals 0, and the latter condition can be decided effectively if we can compute with the elements of K. The central topic of this book, the theory of Gröbner bases, provides a method for solving the equivalence problem in many other types of rings, notably polynomial rings in several variables over a field. The basic strategy for this method at the level of sets without algebraic structure is described in this section.

Let $\sim$ be an equivalence relation on a non-empty set M. We try to find another relation (a *reduction relation*) $\longrightarrow$ on M with the following properties:

(i) $\longrightarrow$ is noetherian and strictly antisymmetric.

(ii) $\longrightarrow$ is a subset of $\sim$.

(iii) Whenever $a, b \in M$, and $a \sim b$ and a, b are both $\longrightarrow$-maximal, then $a = b$.

If this is the case, then the problem whether $a \sim b$ for two arbitrary elements of M can be solved in the following way: reduce both a and b via $\longrightarrow$ as long as possible. By (i), this will result in two finite chains

$$\begin{aligned} &a \longrightarrow a_1 \longrightarrow \cdots \longrightarrow a_m = a' \quad \text{and} \\ &b \longrightarrow b_1 \longrightarrow \cdots \longrightarrow b_n = b', \end{aligned}$$

where a' and b' are $\longrightarrow$-maximal. If $a' = b'$, then $a \sim b$ by (ii). If $a' \neq b'$, then $a \not\sim b$ by (iii).

In the following we discuss various properties of relations $\longrightarrow$ on M that help to find reduction relations with properties (i), (ii), and (iii) on M.

Definition 4.71 Let $\longrightarrow$ be a relation on a non-empty set M. Then $\longrightarrow$ is called a **reduction relation** on M if $\longrightarrow$ is strictly antisymmetric. In connection with a reduction relation $\longrightarrow$ on M, we will write

$\xrightarrow{*}$ for the reflexive-transitive closure of $\longrightarrow$,

$\longleftrightarrow$ for the symmetric closure of $\longrightarrow$, i.e., $a \longleftrightarrow b$ iff $a \longrightarrow b$ or $b \longrightarrow a$ for $a, b \in M$,

$\xleftrightarrow{*}$ for the reflexive-transitive closure of $\longleftrightarrow$, i.e., the smallest equivalence relation on M extending $\longrightarrow$ (cf. Exercise 4.28),

$\xrightarrow{n}$ for $(\longrightarrow)^n$ (where the exponent refers to the definition of powers of a relation given on p. 154),

$\xleftrightarrow{n}$ for $(\longleftrightarrow)^n$, and

$\downarrow$ for the relation on M defined by $a \downarrow b$ iff there exists $c \in M$ with $a \xrightarrow{*} c$ and $b \xrightarrow{*} c$ (or $a \xrightarrow{*} c \xleftarrow{*} b$ for short).

An element $a \in M$ is said to be **in normal form**, or a **normal form**, with respect to $\longrightarrow$ if a is $\longrightarrow$-maximal in M. We say that $b \in M$ is a **normal form of** $a \in M$ with respect to $\longrightarrow$ if $a \xrightarrow{*} b$ and b is in $\longrightarrow$-normal form.

Note that in the situation of the definition above, $a \xrightarrow{*} b$ means that $a \xrightarrow{n} b$ for some $n \in \mathbb{N}$, and this in turn means that either $a = b$, or there exist $a_0, \dots a_n \in M$ with $a_0 = a$, $a_n = b$, and

$$a_0 \longrightarrow a_1 \longrightarrow \cdots \longrightarrow a_n.$$

Here, the natural number n is called the **length** of the reduction chain $a \xrightarrow{n} b$.

Lemma 4.72 If $\longrightarrow$ is a noetherian reduction relation on M, then each $a \in M$ has at least one normal form $a' \in M$ with respect to $\longrightarrow$.

Proof The set

$$N = \{ b \in M \mid a \xrightarrow{*} b \}$$

is non-empty because $a \in N$, and so N contains a $\longrightarrow$-maximal element. □

In order to establish the noetherianity of $\longrightarrow$, the following simple lemma is useful.

Lemma 4.73 Let r be a well-founded relation on M with strict part r_s, and assume that $a \longrightarrow b$ implies $b \; r_s \; a$. Then $\longrightarrow$ is a noetherian reduction relation on M.

Proof $a \longrightarrow b$ and $b \longrightarrow a$ implies $a \; r_s \; b$ and $b \; r_s \; a$, which is impossible. Assume M has a strictly ascending $\longrightarrow$-chain $\{a_n\}_{n \in \mathbb{N}}$. Then $\{a_n\}_{n \in \mathbb{N}}$ is a strictly descending r-chain, a contradiction. □

In most applications, r is a well-order $\leq$, and so the condition of the lemma means that $a \longrightarrow b$ implies $a > b$.

Definition 4.74 Let $\longrightarrow$ be a reduction relation on M. Then $\longrightarrow$ is said

(i) to be **confluent** if $b \xleftarrow{*} a \xrightarrow{*} c$ implies $b \downarrow c$ for all a, b, $c \in M$,

(ii) to be **locally confluent** if $b \longleftarrow a \longrightarrow c$ implies $b \downarrow c$ for all a, b, $c \in M$,

(iii) to have the **Church–Rosser property** if $b \xleftrightarrow{*} c$ implies $b \downarrow c$ for all b, $c \in M$,

(iv) to have **unique normal forms** if $b \xleftarrow{*} a \xrightarrow{*} c$ with b and c in $\longrightarrow$-normal form implies $b = c$ for all a, b, $c \in M$.

The following theorem shows the equivalence of these properties for noetherian reduction relations.

Theorem 4.75 (Newman's Lemma) *Let $\longrightarrow$ be a noetherian reduction relation on M. Then the following are equivalent:*

(i) $\longrightarrow$ *is locally confluent.*

(ii) $\longrightarrow$ *is confluent.*

(iii) $\longrightarrow$ *has unique normal forms.*

(iv) $\longrightarrow$ *has the Church–Rosser property.*

Proof (i)$\Longrightarrow$(ii): Assume for a contradiction that $\longrightarrow$ is locally confluent but that the set

$$N = \{\, a \in M \mid \text{ there exist } b, c \in M \text{ with } b \xleftarrow{*} a \xrightarrow{*} c, \text{ but not } b \downarrow c \,\}$$

is non-empty. Since $\longrightarrow$ is noetherian, N has a $\longrightarrow$-maximal element a. Let b, $c \in M$ with $b \xleftarrow{*} a \xrightarrow{*} c$, but not $b \downarrow c$. If $a = b$ or $a = c$, then we trivially have $b \downarrow c$. Hence there must exist b', $c' \in M$ (possibly $b' = b$ or $c' = c$) with

$$b \xleftarrow{*} b' \longleftarrow a \longrightarrow c' \xrightarrow{*} c.$$

By the local confluence of $\longrightarrow$, there exists $d \in M$ with $b' \xrightarrow{*} d \xleftarrow{*} c'$. By the maximality of a in N and the fact that

$$b \xleftarrow{*} b' \xrightarrow{*} d,$$

there exists $e \in M$ with $b \xrightarrow{*} e \xleftarrow{*} d$. A look at the diagram below shows that now

$$e \xleftarrow{*} c' \xrightarrow{*} c,$$

and thus, again by the maximality of a in N, there must be $f \in M$ with $e \xrightarrow{*} f \xleftarrow{*} c$.

$$\begin{tikzcd}
b \arrow[dr, "*"] & b' \arrow[l, "*"'] \arrow[dr, "*"] & a \arrow[l] \arrow[r] & c' \arrow[r, "*"] \arrow[dl, "*"] & c \arrow[ddll, "*"] \\
& e \arrow[dr, "*"] & d \arrow[l, "*"'] & & \\
& & f & &
\end{tikzcd}$$

The diagram shows that $b \downarrow c$, a contradiction.

(ii)$\Longrightarrow$(iii): Let $b \xleftarrow{*} a \xrightarrow{*} c$ and suppose b and c are in $\longrightarrow$-normal form. By (ii) there exists $d \in M$ with $b \xrightarrow{*} d \xleftarrow{*} c$, and thus $b = d = c$.

(iii)$\Longrightarrow$(iv): We show by induction on $k \in \mathbb{N}$ that for all $a, b \in M$ with $a \overset{k}{\longleftrightarrow} b$ it follows that $a \downarrow b$. The case $k = 0$ is trivial. Let now $a \overset{k+1}{\longleftrightarrow} b$, say

$$a \overset{k}{\longleftrightarrow} c \longleftrightarrow b.$$

Then by induction hypothesis there exists $d \in M$ with $a \xrightarrow{*} d \xleftarrow{*} c$. If $c \longleftarrow b$, this implies $a \xrightarrow{*} d \xleftarrow{*} b$ and so $a \downarrow b$. If $c \longrightarrow b$, let d' be a normal form of d and let b' be a normal form of b with respect to $\longrightarrow$. Then $d' \xleftarrow{*} c \xrightarrow{*} b'$, and so d' and b' are normal forms of c and hence equal. So $a \xrightarrow{*} d' \xleftarrow{*} b$, which means $a \downarrow b$.

(iv)$\Longrightarrow$(i): Let $b \longleftarrow a \longrightarrow c$. Then $b \overset{*}{\longleftrightarrow} c$, and so $b \downarrow c$ by (iv). □

Corollary 4.76 *Let $\longrightarrow$ be a locally confluent, noetherian reduction relation on M and let $a, b \in M$. Then the following assertions are equivalent:*

(i) $a \overset{*}{\longleftrightarrow} b$

(ii) There exists $c \in M$ such that c is a normal form of a and of b with respect to $\longrightarrow$.

(iii) Whenever a' and b' are normal forms of a and b with respect to $\longrightarrow$, respectively, then a' and b' are equal. □

We have now reached a point where it is necessary to make precise what we mean by decidability and computability. Computational algebra, as opposed to pure algebra, is concerned not only with mathematical objects and their structure, but also with physical representations of these objects, and with procedures that manipulate such representations. Our treatment is based on the following basic viewpoint. Our mathematics, quite classically, takes place in the universe of sets of Zermelo–Fraenkel set theory where only sets exist. (Cf. the discussion of **AC** in Section 4.1.) Besides that, we count on the existence of certain physical primordial objects by which these sets are denoted, or given, or represented. These can be letters, or words in an alphabet, or, eventually, physical configurations inside a machine. These objects are finitary in nature and at our disposal in such a way that we can manipulate them or let a machine manipulate them. Furthermore, we assume that we have an understanding of these objects which allows us to decide their equality simply by inspection. If, for example, we use the letters of the alphabet, then in the strictest physical sense, a and a are two different objects. But it is not hard to agree on the basis of common sense that a equals a as a letter.

Even on the most elementary level, it often happens that one and the same element of a set has many different representations: examples are $1/2$ and $2/4$ in $\mathbb{Q}$, or $X + X$ and $2X$ in $\mathbb{Z}[X]$. (See the next section for more examples.) In that case, meaningful computations are possible only if we have an algorithm that recognizes if two objects refer to the same element.

Definition 4.77 A set M is **decidable** if its elements are given in such a way that there is an algorithm which, upon input of a, $b \in M$, decides whether or not $a = b$.

Definition 4.78 Let $\longrightarrow$ be a reduction relation on M and let $\sim$ be an equivalence relation on M. Then we call $\longrightarrow$ **adequate** for $\sim$ if $\sim$ coincides with the reflexive-symmetric-transitive closure $\overset{*}{\longleftrightarrow}$ of $\longrightarrow$. We say $\longrightarrow$ is **decidable** if there is an algorithm which, on input of $a \in M$, decides whether a is reducible with respect to $\longrightarrow$, and if so selects (possibly non-deterministically) some $b \in M$ with $a \longrightarrow b$.

Using this terminology we can now summarize in a mathematically precise way the use of reduction relations for the solution of the equivalence problem sketched at the beginning of this section. The algorithm EQUIV of the proof below is an example of a non-deterministic algorithm: it involves the choice of some b with $a \longrightarrow b$ for reducible a. In practice, there is usually some heuristic criterion for choosing a particular b with this property; from a theoretical point of view, however, the choice could be a random one.

Theorem 4.79 *Let $\sim$ be an equivalence relation on a non-empty set M. Let $\longrightarrow$ be a locally confluent noetherian reduction relation on M that is adequate for $\sim$, and suppose M and $\longrightarrow$ are decidable. Then there exists an algorithm which, on input of a, $b \in M$, decides whether $a \sim b$ or not.*

Proof It is easy to see that the algorithm EQUIV of Table 4.1 terminates and performs the desired task. □

4.6 Computing in Algebraic Structures

In Chapter 2, we already worked with an intuitive definition of computable rings and fields. We are now in a position to make this more precise. Computability is of course a concept that verges on practical problems of conceiving computer algebra systems and implementing algorithms. Neither will we go into any technical details here, nor is it our aim to discuss space–time optimal solutions. On the other hand, every attempt to define computability also raises foundational and philosophical questions on the mathematical side. We will suppress these issues as well. All we are interested in is a reasonably rigorous definition of computability that is based on common sense insofar as it reflects the capabilities of today's computers. (See also the discussion of decidability of sets in the last section.)

Definition 4.80 A monoid M (a ring R) is called **computable** if the elements of M (of R) are given in such a way that M (R) is decidable as a set, and there is an algorithm (there are algorithms) which, upon input of a, $b \in M$ (of a, $b \in R$) computes $ab \in M$ (compute ab, $a + b$, and $-a \in R$).

TABLE 4.1. Algorithm EQUIV

```
Specification: $v \leftarrow$ EQUIV$(a, b)$
                   Test whether $a \sim b$
Given: $a, b \in M$
Find: $v \in \{$true, false$\}$ such that $v =$ true iff $a \sim b$
begin
$A \leftarrow a$;  $B \leftarrow b$
while $A$ is reducible with respect to $\longrightarrow$ do
     - select $C \in M$ with $A \longrightarrow C$
     $A \leftarrow C$
end
while $B$ is reducible with respect to $\longrightarrow$ do
     - select $D \in M$ with $B \longrightarrow D$
     $B \leftarrow D$
end
if $A = B$ return(true)
else return(false) end
end EQUIV
```

A field K is called a **computable field** if it is a computable ring and there is an algorithm which, upon input of $0 \neq a \in K$, computes $a^{-1} \in K$. A **computable Euclidean domain** is a Euclidean domain R which is a computable ring and for which there is an algorithm that computes, possibly non-deterministically, a quotient and remainder of $a \in R$ upon division by $0 \neq b \in R$. An order $\leq$ on a set M is called **decidable** if there is an algorithm that decides, for a, $b \in M$, whether $a \leq b$.

Example 4.81 $\mathbb{Z}$ is a computable ring: integers can be uniquely represented as strings of digits w.r.t. some base in $\mathbb{N}$, and since handling digits and carries is a finite affair, integer arithmetic can then be performed effectively. Moreover, we can certainly, by means of iterated subtraction, effectively divide with remainder in $\mathbb{Z}$; so $\mathbb{Z}$ is even a computable Euclidean domain.

Example 4.82 If R is a computable domain, i.e., a computable ring with no zero divisors, then we claim that the field of quotients Q_R of R is a computable field. The elements of Q_R are given as formal fractions p/q, where p, $q \in R$ with $q \neq 0$, and two such fractions p/q and r/s represent the same element of Q_R iff $ps = rq$. The latter condition can be decided by our assumption on R, and so Q_R is decidable as a set. It is clear that rational arithmetic can be performed on the basis of the ring operations. If $R = \mathbb{Z}$, then the elements of $Q_R = \mathbb{Q}$ can even be uniquely represented as fractions with a positive denominator that are reduced to lowest terms. A gcd computation by means of the Euclidean algorithm must then be invoked after each addition or multiplication to get the result in this normal form.

Example 4.83 Let R be a computable ring, $\{0\} \neq I$ an ideal of R such that there is an algorithm which, upon input of $a \in R$, decides whether or not $a \in I$. We claim that then the residue class ring R/I is a computable ring. The elements $a + I$ of R/I are obviously given as elements of R viewed as residue classes modulo I. Two elements a and b of R represent the same residue class iff $a - b \in I$, which condition we can effectively test by our assumption on I. Addition, subtraction, and multiplication in R/I are performed by doing the respective operation in R on representatives, which is possible by the assumption on R.

Example 4.84 The situation of the previous example is given whenever R is a computable Euclidean domain and an ideal I of R is given by a generating element $0 \neq b \in R$: for $a \in R$, we then have $a \in I$ iff every remainder of a upon division by b is zero iff some such remainder is zero (cf. Proposition 2.39). By assumption, we can compute such a remainder r, and $r = 0$ iff $r + r = r$. If b is a prime element of R, e.g., a prime number in $\mathbb{Z}$, then R/I is even a computable field: if $a + I \neq I$, then $\gcd(a, b) = 1$ (cf. Exercise 2.46), so the extended Euclidean algorithm computes s, $t \in R$ with $1 = as + bt$, and we see that $(a + I)(s + I) = 1 + I$ since $bt \in I$.

If, in the situation of the above example, remainders in R are either unique, or the division algorithm deterministically computes a specific one for each pair of dividend and divisor, then the elements of R/I can even be uniquely represented by the set of these remainders: clearly, $a + I = r + I$ if $a = qb + r$ for some $q \in R$. This is the case in $\mathbb{Z}$, where remainders can be specified to be in a certain range like $[0, b)$ or $[-b/2, b/2)$. It is also the case in univariate polynomial rings over fields, where remainders are unique to begin with. In this case, the set of unique remainders is a system of unique representatives for the partition $\{\, a + I \mid a \in R \,\}$ of R.

Throughout Chapter 2, we have been assuming that polynomial rings over computable rings are again computable. The rest of this section is devoted to a rigorous argument for this on the basis of our definition of computability. Let $X_1, \ldots, X_n$ be indeterminate symbols and

$$T = T(X_1, \ldots, X_n)$$

the set of all terms, i.e., power products, in these indeterminates. Recall that T is a multiplicative Abelian monoid which is isomorphic to the additive monoid $\mathbb{N}^n$ under the exponent map. An element of T can thus be uniquely represented as the n-tuple of its exponents, and under this representation, multiplication of terms is simply componentwise addition of n-tuples of natural numbers. We see that T with this multiplication is a computable monoid.

Now, in addition, let R be a computable ring. We proved that every monomial in the polynomial ring $R[X_1, \ldots, X_n]$ has a unique representation of the form at with $a \in R$ and $t \in T$. This means that the elements of M can be uniquely represented as ordered pairs with one entry from R and one from T, and these are multiplied componentwise: $(a, s)(b, t) = (ab, st)$. We have demonstrated that M is a computable monoid.

Back in Section 2.1, we argued further that every polynomial has a unique representation as the sum of its monomials. So the obvious thing to do is to represent every polynomial as an ordered m-tuple (or a *list*, as the programmer would say) of monomials:

$$(a_1t_1, \dots, a_mt_m) \quad \text{represents} \quad \sum_{i=1}^{m} a_it_i. \tag{$*$}$$

Contrary to the rigorous mathematical definition, we will, in this context, allow "monomials" of the form $0 \cdot t$ with $t \in T$ and refer to these as *dummy monomials.* It is clear that under the representation ($*$), polynomials can be added by appending lists, the negative of a polynomial is obtained by taking the negative of each monomial in the list, and the product of two polynomials

$$(a_1s_1, \dots, a_ks_k) \quad \text{and} \quad (b_1t_1, \dots, b_mt_m)$$

is obtained by appending the lists

$$((a_ib_1)(s_it_1), \dots, (a_ib_m)(s_it_m))$$

for $1 \leq i \leq k$. The obvious problem that remains to be solved is decidability. The repesentation of a polynomial as a sum of monomials is unique only when the monomials have pairwise different terms (see Section 2.1), and since we are using lists, this uniqueness is only up to the order of the summands. So what we need is an algorithm that effectively decides whether or not the polynomials represented by two lists are equal in $R[X_1, \dots, X_n]$. We are going to show that we can even compute normal forms of lists in such a way that two normal forms are equal iff they represent the same polynomial. Rather obviously, a normal form will be a list where all like terms have been combined, all dummy monomials have been dropped, and the monomials are arranged in some specified order by their terms.

Orders on T will play a crucial role in the theory of Gröbner bases, and the more important ones will be decidable. For the moment, it suffices to note that there is at least one decidable order on T: given $s \neq t \in T$, look through their representations $(\mu_1, \dots, \mu_n)$ and $(\nu_1, \dots, \nu_n)$ as n-tuples of exponents by increasing indices, stop at the first index i where $\mu_i \neq \nu_i$, and declare $s < t$ iff $\mu_i < \nu_i$.

Exercise 4.85 Show that $<$ as defined above is the strict part of an order on T.

Let now L be the set of all lists (i.e., ordered tuples) of monomials as described above, and let $\longrightarrow$ be the following relation on L: $l_1 \longrightarrow l_2$ iff l_2 can be obtained from l_1 in one of the following ways:

(i) drop a dummy monomial,

(ii) combine two like terms, i.e., find two entries of the form at and bt, replace at with $(a+b)t$, and drop bt, or

(iii) switch two adjacent entries a_it_i and $a_{i+1}t_{i+1}$ with $t_{i+1} > t_i$.

From the fact that (i) and (ii) shorten the list and that (iii) can be performed at most once on each pair of entries, we easily conclude that $\longrightarrow$ is a noetherian reduction relation. The following tedious exercise is recommended (to be worked at least partly) although we won't need its statement.

Exercise 4.86 Show that $\longrightarrow$ is locally confluent and thus has unique normal forms by Newman's lemma.

The reason why we don't have to invoke Newman's lemma and the theory surrounding it is that the polynomial ring is not defined as L modulo some equivalence relation: we have an a priori definition of $R[X_1, \ldots, X_n]$, we have the correspondence (*), and we know that two sums of monomials are equal in the polynomial ring iff they both add up to the same function from $\mathbb{N}^n$ to R. On the basis of this understanding, we can now argue as follows.

(1) It is immediate from the definition of the polynomial ring that $l_1 \longrightarrow l_2$ implies that l_1 and l_2 represent the same polynomial. It follows by induction on the length of the reduction chain that the same is true with $\longrightarrow$ replaced by $\xrightarrow{*}$.

(2) A list l of monomials is obviously $\longrightarrow$-maximal iff it contains no zero entries and the terms of the entries are pairwise different and in descending order.

(3) Now let $l_1 \neq l_2$ be $\longrightarrow$-maximal. By the above characterization of $\longrightarrow$-maximality, l_1 and l_2 must be different as sets, i.e., there exists an entry at in l_1 such that either t does not occur in l_2 at all, or it occurs exactly once in a monomial bt with $a \neq b$. We see that the corresponding polynomials have different coefficients on the term t and are thus different.

Claim: $\longrightarrow$ has unique normal forms, and l_1 and l_2 have the same normal form iff they represent the same polynomial.

Proof: If l_1 and l_2 are normal forms of l, then by (1), they both represent the same polynomial as l and thus must be equal by (3). If l_1 and l_2 have the same normal form l, then they both represent the same polynomial as l by (1). Conversely, if l_1 and l_2 represent the same polynomial, then so do their respective normal forms by (1), and (3) tells us that these normal forms are equal.

Since the reduction relation $\longrightarrow$ is obviously decidable in the sense of Definition 4.78, the above claim provides the desired decidability of the set of polynomials under the representation as lists of monomials. Moreover, by passing to unique normal forms after each algebraic operation, we can even turn this into a unique representation.

The representation of polynomials as lists of monomials with each monomial consisting of a coefficient and a list of exponents is in fact the most common choice in computer algebra systems. It is often called the *sparse* representation. When working with univariate polynomials, one may consider the *dense* representation consisting of a list $(a_0, \ldots, a_m)$ with the understanding that a_i is the coefficient of X^i. It may also be advantageous to represent multivariate polynomials recursively as univariate ones in one of their variables.

Notes

The axiom of choice is virtually indispensable in analysis. Up to at least the end of the 19th century its validity was considered to be self-evident. Zermelo (1904) proved that the axiom of choice implies that every set can be well-ordered. The well-ordering principle, however, had been vehemently questioned before by a number of mathematicians, and thus the axiom of choice became questionable too. Another somewhat disconcerting consequence of the axiom of choice is the fact that there exist non-Lebesgue-measurable sets of real numbers; in particular, not every subset of $\mathbb{R}^n$ can be assigned a volume. It is now known that **AC** is independent of but consistent with the remaining axioms of Zermelo–Fraenkel set theory (see, e.g., Moore, 1982; Rubin and Rubin, 1963). This means that when **AC** is being used, there is no more danger of running into inconsistencies than there would be otherwise. This is why nowadays many mathematicians, including most calculus textbook authors, choose to use **AC** tacitly. On the other hand, any application of **AC** has a distinctly non-constructive flavor. An algorithmically oriented treatment of algebra is therefore well-advised to maintain a certain awareness of the problem.

The Hilbert basis theorem appears for the first time as Theorem I in Hilbert (1890). It is proved there for homogeneous polynomials (homogeneous meaning that all terms have the same total degree). The proof of the theorem that we have given here is by far the simplest one; interestingly, it has gone unnoticed until quite recently (Sarges, 1976). Hilbert's original interest was not in ideal bases; to him, his theorem was a tool in proving a conjecture in a now somewhat obsolete area of algebra called *invariant theory*. Given a homogeneous polynomial f in n variables, an *invariant* of f is a homogeneous polynomial i in as many variables as f has coefficients such that $i(\underline{b}) = (\det T)^s \cdot i(\underline{a})$ for some $s \in \mathbb{N}$ whenever T is the regular $n \times n$ matrix of a linear transformation of the original n variables, $\underline{a}$ are the coefficients of f, and $\underline{b}$ are the coefficents of $f(T \cdot \underline{x})$. The conjecture that Hilbert proved was that the ring of invariants of a homogeneous polynomial (in fact, of a *system* of homogeneous polynomials) is finitely generated over the ground field. The basis theorem—in the strong form of Lemma 4.5 (ii)—provided a first step towards this goal: the ideal generated by the invariants has a finite basis which itself consists of invariants.

Hilbert's proof of his theorem prompted a major controversy between the Kronecker–Gordan school of algorithmic algebra and the proponents of the budding axiomatic viewpoint; P.A. Gordan is said to have commented on Hilbert's proof with the words, "This is not mathematics, this is theology." A more recent school of constructivism in mathematics (see Mines et al., 1988) has tried to soften the non-constructivity of the theorem by weakening the definition of noetherianity to the requirement that for every ascending chain $\{I_n\}_{n\in\mathbb{N}}$ of ideals, there exists an $n_0 \in \mathbb{N}$ with $I_{n_0} = I_{n_0+1}$. Regardless of whether or not one believes in any kind of constructivism, it is interesting to see how this property is really all that is needed when applying the Hilbert basis theorem. A related constructive aspect of the theorem is the question of the maximal length of a strictly ascending chain of ideals when starting with a principal ideal; this is discussed in Seidenberg (1971), Seidenberg (1972), and Moreno Socías (1991).

Zorn's lemma is one of several related maximality principles discovered by Hausdorff, Kuratowski, Zorn, Teichmüller, Tuckey, and others (see Moore, 1982). Zorn (1935) published his "lemma" as an axiom that he hoped would supersede the use of the well-ordering principle in abstract algebra. Its equivalence to the axiom of choice was established much later (cf. Rubin and Rubin, 1963). For most applications in algebra, Zorn's lemma has indeed turned out to be the most convenient among the multitude of principles that are equivalent to the axiom of choice; as such it was popularized by the Bourbaki school.

The fact that $\mathbb{R}$ has a basis as a vector space over the field $\mathbb{Q}$ of rational numbers was first proved in Hamel (1905); Hamel's proof was based on the well-ordering theorem. Using such a "Hamel basis," he was able to prove the remarkable fact that the functional equation $f(x+y) = f(x) + f(y)$ has infinitely many different non-continuous solutions besides the obvious continuous solutions $f(x) = cx$ with $c \in \mathbb{R}$. A similar result holds for the functional equation $f(x+y) = f(x) \cdot f(y)$. Hamel's proof can be readily adapted to show that every vector space over a field K has a basis.

The theory of relations is formally part of set theory; its foundations were laid by Hausdorff and Sierpinski starting around 1910. The roots of the theory are much older. The motivations for studying relations are manifold; they arise among others from the study of linear orders, of divisibility, of combinatorial problems, of problems in graph theory, and of the problem of termination of iterative mathematical constructions, in particular of algorithms involving loops. For a comprehensive study of relations we refer the reader to Fraïssé (1986).

Dickson's lemma is by nature a result in infinite combinatorics; it has been called "the most frequently rediscovered mathematical theorem." The first reference seems to be Dickson (1913), Lemma A. In Ritt (1950), it is attributed to Riquier. Other references include Higman (1952), Theorem 2.3, who credits Erdös and Rado with part of the equivalences between different versions of Dickson's lemma, Eilenberg and Schützenberger (1969), Proposition 4.2, and Kolchin (1973), Chapter 0, Section 17, Lemma 15.

The general concept of a Dickson quasi-order (often referred to as *well-quasi-order*, recently also as *narrow* quasi-order) arises naturally from Dickson's lemma (see Kruskal, 1972 and the references given there). The paper of Higman mentioned above actually discusses Dickson's lemma in this general framework.

König's lemma is due to D. König (1936), who proved it in a graph theoretical form. It is usually applied to trees and is therefore also known as the König tree lemma.

5

Gröbner Bases

There are many different ways to look at the theory of Gröbner bases. In the context of classical algebra, the natural point of view is as follows. Suppose first we are given univariate polynomials f, g_1, $\ldots$, g_m over a field, and we wish to decide whether f is in the ideal generated by the g_i. According to the results of Section 2.2, the thing to do is to compute the gcd g of the g_i and then perform long division of f by g. The polynomial f will lie in the ideal in question if and only if the remainder of this division equals zero. Moreover, if this is the case, then one also obtains a polynomial q that satisfies $f = qg$, namely, the quotient of the division, which equals the sum of the monomial multipliers that were used in the individual steps of the division.

Gröbner basis theory generalizes these ideas to multivariate polynomials. The attempt to obtain a suitable division with remainder runs into two difficulties. First of all, there will not exist a single generator of the given ideal in general (Proposition 2.40), and so one must come up with a division of one polynomial by a set of polynomials. This will turn out not to be a problem at all: instead of subtracting a monomial multiple of the divisor from the dividend in each step, one subtracts a monomial multiple of a suitable one of the divisors. A more serious problem is the fact that one needs an ordering of the terms to replace the natural ordering of univariate terms by ascending exponents. It will turn out that the theory works if one chooses an admissible order on $\mathbb{N}^n$ and then orders terms by increasing exponent tuples. The first section of this chapter will develop this generalized division procedure; since the theory of reduction relations—Newman's lemma in particular—will come into play, the process will not be called polynomial division but *polynomial reduction.*

If the "remainder" of a (possibly multivariate) polynomial upon this "generalized division" by $g_1, \ldots, g_m$ equals zero, then just as in the classical case, the sums q_i of the multipliers of the g_i that were used in the individual subtraction steps yield a representation

$$f = \sum_{i=1}^{m} q_i g_i \, .$$

In particular, f then lies in the ideal generated by the g_i. What one would hope for is that the converse is true too: if f lies in the ideal in question, then the above "remainder" should be zero. This cannot be true in general, not even in the univariate case, because we have skipped the Euclidean algorithm altogether. The key theorem that makes Gröbner basis theory work states that it is possible to generalize the Euclidean algorithm to a "preprocessing" of the given set $\{g_1, \ldots, g_m\}$ in such a way that one obtains another set which still generates the same ideal and has the desired property to yield "remainder" zero for every "division" with a member of the ideal as the "dividend." Ideal bases with this property are called *Gröbner bases.* The algorithm that achieves the "preprocessing," i.e., the computation of a Gröbner basis from a given basis, is called the *Buchberger algorithm.* According to the discussion above, it is the multivariate analogue to the Euclidean algorithm. It will later become apparent that it can also be viewed as a generalization of the Gaussian elimination algorithm to the non-linear case.

5.1 Term Orders and Polynomial Reductions

This section combines the results on quasi-orders and abstract reduction relations that were proved in Chapter 4 with the theory of multivariate polynomial rings. A rigorous mathematical definition of these polynomial rings has been given in Section 2.1. For the convenience of the reader, we will recall here the relevant terminology and notation pertaining to polynomials. Knowing that a formally sound treatment has already been achieved, we may now feel free to discuss things in a hands-on manner. Throughout, R will be a commutative ring with 1, and the polynomial ring $R[X_1, \ldots, X_n]$ over R will also be denoted by $R[\,\underline{X}\,]$.

A *term* t in the indeterminates, or variables, $X_1, \ldots, X_n$ is a power product of the form $X_1^{e_1} \cdot \cdots \cdot X_n^{e_n}$ with $e_i \in \mathbb{N}$ for $1 \leq i \leq n$; in particular, $1 = X_1^0 \cdot \cdots \cdot X_n^0$ is a term. We denote by $T(X_1, \ldots, X_n)$, or simply by T, the set of all terms in these variables. T forms an Abelian monoid with neutral element 1 under the natural multiplication where two terms are multiplied by adding the respective exponents of each variable.

The crucial connection with the results of the last chapter is now made possible by passing from a term to its exponent tuple, which is an element of $\mathbb{N}^n$, and vice versa. For clarity, we will write $(T, 1, \cdot)$ for the multiplicative monoid of terms and $(\mathbb{N}^n, (0), +)$ for the additive monoid $\mathbb{N}^n$. Two terms are different if and only if their exponent tuples are different, and they are multiplied by adding their exponent tuples componentwise. This means that $(T, 1, \cdot)$ is isomorphic to $(\mathbb{N}^n, (0), +)$: a natural isomorphism $(T, 1, \cdot) \longrightarrow (\mathbb{N}^n, (0), +)$ is given by the **exponent map** η which assigns to any term its exponent tuple, i.e.,

$$\eta(X_1^{e_1} \cdot \cdots \cdot X_n^{e_n}) = (e_1, \ldots, e_n) \in \mathbb{N}^n.$$

The inverse η^{-1} of η is of course the map from $\mathbb{N}^n$ to T given by

$$(e_1, \dots, e_n) \longmapsto X_1^{e_1} \cdot \dots \cdot X_n^{e_n}.$$

We see that although T is a subset of $R[\underline{X}]$, its structure is independent of R. The partial order $\leq'$ on $\mathbb{N}^n$ obtained by forming the product of n copies of $\mathbb{N}$ with its natural order will be called the **natural partial order** on $\mathbb{N}^n$. Recall that here,

$$(k_1, \dots, k_n) \leq' (m_1, \dots, m_n) \quad \text{iff} \quad k_i \leq m_i \text{ for } 1 \leq i \leq n.$$

The **divisibility relation** $\mid$ on T is defined by $s \mid t$ iff there exists $s' \in T$ with $s \cdot s' = t$. It is now easy to see that under the exponential map, these two relations correspond to each other.

Exercise 5.1 Let s, $t \in T$. Show that $s \mid t$ iff $\eta(s) \leq' \eta(t)$, where $\leq'$ is the natural partial order on $\mathbb{N}^n$.

Dickson's lemma states that the natural partial order on $\mathbb{N}^n$ has the Dickson property. Using the observation of the above exercise, this translates into the following important theorem which is often also referred to as Dickson's lemma.

Theorem 5.2 *The divisibility relation $\mid$ on T is a Dickson partial order on T. More explicitly, every non-empty subset S of T has a finite subset B such that for all $s \in S$, there exists $t \in B$ with $t \mid s$.* □

The other important class of relations on $\mathbb{N}^n$ that we studied was the class of admissible orders. Under our natural correspondence between terms and exponent tuples, these translate into the following type of orders on T.

Definition 5.3 A **term order** is a linear order on T that satisfies the following conditions.

(i) $1 \leq t$ for all $t \in T$.

(ii) $t_1 \leq t_2$ implies $t_1 \cdot s \leq t_2 \cdot s$ for all s, t_1, $t_2 \in T$.

Lemma 5.4 Let $\leq$ be an admissible order on $(\mathbb{N}^n, (0), +)$, and define $\leq'$ on T by setting

$$s \leq' t \quad \text{iff} \quad \eta(s) \leq \eta(t).$$

Then $\leq'$ is a term order on T. Moreover, every term order on T is obtained in this way, and the resulting correspondence between term orders on T and admissible orders on $(\mathbb{N}^n, (0), +)$ is one-to-one.

Proof Let $\leq$ be an admissible order on $(\mathbb{N}^n, (0), +)$. Then $\leq'$ is a term order on T since $\eta(1) = (0)$ and $\eta(s \cdot t) = \eta(s) + \eta(t)$. Using the fact that

$$\eta^{-1} : (\mathbb{N}^n, (0), +) \longrightarrow (T, 1, \cdot)$$

is an isomorphism too, one easily proves that every term order on T is obtained in this way, and that the correspondence is one-to-one. □

Theorem 4.62 stated that every admissible order on $\mathbb{N}^n$ is a well-order (no strictly descending chains), and that it extends the natural partial order on $\mathbb{N}^n$. Translating this by means of Exercise 5.1 and the proposition above, we obtain the following theorem.

Theorem 5.5 *(i) If $\leq$ is a term order on T, then $s\,|\,t$ implies $s \leq t$ for all s, $t \in T$.*

(ii) Every term order is a well-order on T. □

Exercise 5.6 Prove (i) of the theorem above directly from the definition of a term order.

Another important property of term orders which will be used frequently is as follows.

Lemma 5.7 Let $\leq$ be a term order on T and $s_1, t_1, s_2, t_2 \in T$ with $s_1 \leq s_2$ and $t_1 < t_2$. Then $s_1t_1 < s_2t_2$.

Proof From $s_1 \leq s_2$ we may conclude that $s_1t_1 \leq s_2t_1$. From $t_1 < t_2$ it follows that $t_1 \leq t_2$ and thus $s_2t_1 \leq s_2t_2$. It is easy to see that $s_2t_1 = s_2t_2$ would imply $t_1 = t_2$, so we even have $s_1t_2 < s_2t_2$. Transitivity of $\leq$ now yields $s_1t_1 < s_2t_2$. □

From Exercise 4.61 together with Lemma 5.4 we get the following induced examples of term orders on T.

Examples 5.8 Each of the following is a term order on T.

(i) $X_1^{d_1} \cdot \dots \cdot X_n^{d_n} \leq X_1^{e_1} \cdot \dots \cdot X_n^{e_n}$ iff the following holds: $(d_1, \dots, d_n) = (e_1, \dots, e_n)$, or there exists $1 \leq i \leq n$ with $d_j = e_j$ for $1 \leq j \leq i-1$ and $d_i < e_i$. This term order is called the **lexicographical**, or **lexical** order on T.

(ii) $X_1^{d_1} \cdot \dots \cdot X_n^{d_n} \leq X_1^{e_1} \cdot \dots \cdot X_n^{e_n}$ iff the following holds: $(d_1, \dots, d_n) = (e_1, \dots, e_n)$, or there exists $1 \leq i \leq n$ with $d_j = e_j$ for $i+1 \leq j \leq n$ and $d_i < e_i$. This term order is called the **inverse lexicographical**, or **inverse lexical** order on T.

(iii) Let $\leq$ be an order on T that satisfies condition (ii) of Definition 5.3, e.g., a term order or the inverse thereof. Set

$$X_1^{d_1} \cdot \dots \cdot X_n^{d_n} \leq' X_1^{e_1} \cdot \dots \cdot X_n^{e_n}$$

iff the following condition holds:

$$\sum_{i=1}^{n} d_i \quad < \quad \sum_{i=1}^{n} e_i, \qquad \text{or}$$

$$\sum_{i=1}^{n} d_i = \sum_{i=1}^{n} e_i \quad \text{and} \quad X_1^{d_1} \cdot \dots \cdot X_n^{d_n} \leq X_1^{e_1} \cdot \dots \cdot X_n^{e_n}.$$

This class of term orders is called the class of **total degree orders**, because one first compares total degrees and then breaks ties by means of some other order. If this other order is the lexicographical one, then the resulting term order is called the **total degree–lexicographical** order.

(iv) Let $1 \leq i < n$, and set

$$T_1 = T(X_1, \dots, X_i), \quad \text{and} \quad T_2 = T(X_{i+1}, \dots, X_n).$$

Let $\leq_1$ and $\leq_2$ be term orders on T_1 and T_2, respectively. Any $t \in T$ may be written uniquely as $t = t_1 t_2$ with $t_i \in T_i$ for $i = 1$, 2. Then a term order $\leq$ on T is defined by $s \leq t$ iff

$$\begin{aligned} s_1 &<_1 t_1, \quad \text{or} \\ s_1 &= t_1 \quad \text{and} \quad s_2 \leq_2 t_2. \end{aligned}$$

This type of order is sometimes called a **block order** on T.

It is clear that all of the above examples are decidable orders on T. Note also that Lemma 4.64 together with Lemma 5.4 provides a uniform method to construct term orders.

A common cause for confusion is to mix up the lexicographical and the inverse lexicographical term order. To help avoid this, we say that X_j is **lexicographically greater** than X_i and write $X_j \gg X_i$ if $X_j > X_i^d$ for all $d \in \mathbb{N}$. Then the lexicographical order satisfies

$$X_1 \gg X_2 \gg \cdots \gg X_n,$$

whereas the inverse lexicographical one satisfies

$$X_n \gg X_{n-1} \gg \cdots \gg X_1.$$

For greater clarity, we will often add these specifications to the appropriate terminology.

Exercise 5.9 Show that the lexicographical term order is the only one that satisfies

$$X_1 \gg X_2 \gg \cdots \gg X_n,$$

and the inverse lexicographical order is the only one satisfying the reverse inequalities.

A *monomial* in the variables $X_1, \dots, X_n$ over R is a polynomial of the form $m = at$ with $0 \neq a \in R$ and $t \in T$. Here, a is called the *coefficient* of m and t the *term* of m. We have proved in Section 2.1 that a monomial uniquely determines its coefficient and term. The set of monomials (in X_1, $\ldots$, X_n over R) is denoted by M. Multiplication on M is defined by

$$a_1t_1 \cdot a_2t_2 = (a_1a_2)(t_1t_2).$$

M actually forms a commutative monoid under this multiplication, but this monoid structure on the set of monomials alone is of little interest to us. It is clear that M contains both $R \setminus \{0\}$ and T.

Now let $\leq$ be a term order on T. We define the relation $\preceq$ on the set M of monomials by setting

$$as \preceq bt \quad \text{iff} \quad s \leq t$$

for $0 \neq a$, $b \in R$ and s, $t \in T$.

Exercise 5.10 Use the fact that $\leq$ is a well-founded linear order to prove that $\preceq$ is a well-founded linear quasi-order.

We will call $\preceq$ the quasi-order on M *induced by* $\leq$. It is clear that $\preceq$ will not be an order in general: whenever m_1 and m_2 are two monomials with the same term but with different coefficients, then $m_1 \neq m_2$ but $m_1 \preceq m_2$ and $m_2 \preceq m_1$. If this is the case, then m_1 and m_2 are equivalent under the equivalence relation associated with $\preceq$ (Lemma 4.24), and we will write $m_1 \sim m_2$. Now T is a subset of M, and clearly $s \leq t$ iff $s \preceq t$ for s, $t \in T$, so there will be no harm in denoting $\preceq$ by $\leq$ too; it is important though to keep in mind that $\leq$ on M is only a quasi-order.

One of the main results of Section 2.1 was that every element f of $R[\underline{X}]$ is a finite sum of monomials, and moreover, there is a unique set N of pairwise inequivalent monomials such that

$$f = \sum_{m \in N} m.$$

Proposition 5.11 *(i) Let $\leq$ be a term order on T. Then every polynomial $f \in R[\underline{X}]$ has a unique representation in the form $\sum_{i=1}^{k} m_i$ with $m_i \in M$ and $m_1 > \cdots > m_k$.*

(ii) If, in addition, R is a computable ring and $\leq$ is a decidable order on T, then there is an algorithm that computes the representation of (i) from any arbitrary representation of f as a sum of monomials.

Proof For (i), take the unique representation of f as a sum of pairwise inequivalent monomials and index them in descending order according to $\leq$. For (ii), take the normal form algorithm described in Section 4.6. □

It is interesting to see what the above proposition says in the univariate case. Whenever $s = X^{\nu}$ and $t = X^{\mu}$ are univariate terms with $\nu < \mu$, then $s \mid t$, and so necessarily $s < t$ in any term order by Theorem 5.5 (i). It follows that there is only one term order, and it orders the terms by ascending degree. The representation of (i) above is then the natural sparse representation (no zero summands displayed) by decreasing exponents. The concept of term orders can thus be interpreted as a generalization of the natural ordering of univariate terms to the multivariate case. Moreover, Theorem 5.5 and Lemma 5.7 guarantee that multivariate term orders behave just like the univariate one.

The zero polynomial in $R[\underline{X}]$ is identified with the empty sum, i.e.,

$$0 = \sum_{m \in \emptyset} m.$$

Let $f \in R[\underline{X}]$ and assume that all like terms in f have been combined, i.e., f is written as a sum of pairwise inequivalent monomials. The set of monomials occurring in such a representation is denoted by $M(f)$ and called the set of *monomials of* f. The set $T(f)$ of *terms of* f is the set of all terms of monomials $m \in M(f)$. The set $C(f)$ of all *coefficients of* f is the set of all coefficients of monomials $m \in M(f)$.

Next, we show how a given term order may be extended to a well-founded quasi-order on all of $R[\underline{X}]$. Let $\leq$ be a term order on T and let $\leq'$ be the induced well-order on $\mathcal{P}_{\text{fin}}(T)$ of Theorem 4.69. Define a relation $\preceq$ on $R[\underline{X}]$ by setting

$$f \preceq g \quad \text{iff} \quad T(f) \leq' T(g).$$

This relation will play a central role in the theory, so let us describe its definition explicitly. If $T(f) = T(g)$, then we have both $f \preceq g$ and $g \preceq f$. We see that this can happen with $f \neq g$, and thus $\preceq$ will be a quasi-order at best. If $T(f) \neq T(g)$, then we must look at the maxima s and t (w.r.t. the term order $\leq$) of $T(f)$ and $T(g)$. If these are different, then their order is decisive: $f \preceq g$ iff $s < t$. If the maxima agree, we must drop them from $T(f)$ and $T(g)$ and repeat the procedure, comparing the maxima of the smaller sets thus obtained. If one of $T(f)$ and $T(g)$ becomes empty before a decision has been reached in this way, then the other, non-empty one "wins."

Theorem 5.12 *Let $\leq$ be a term order on T. Then $\preceq$ is a linear, well-founded quasi-order on $R[\underline{X}]$ which extends $\leq$ and the induced quasi-order on the set M of monomials.*

Proof The order $\leq'$ on $\mathcal{P}_{\text{fin}}(T)$ upon which the definition of $\preceq$ is based is a linear order by Lemma 4.67, i.e., it is reflexive, transitive and connex. It is easy to see that these three properties are inherited by $\preceq$. Moreover, the premise of Lemma 4.35 is satisfied with $M = R[\underline{X}]$, $N = \mathcal{P}_{\text{fin}}(T)$, and $\varphi : R[\underline{X}] \longrightarrow \mathcal{P}_{\text{fin}}(T)$ defined by $\varphi(f) = T(f)$. We see that $\preceq$ is well-founded. The rest of the theorem is obvious from the definitions. □

The quasi-order $\preceq$ as defined above will be called the quasi-order on $R[\underline{X}]$ *induced* by the term order $\leq$. Since it extends $\leq$, there will be no harm in using $\leq$ for the induced quasi-order as well. It is of utmost importance in the theory of Gröbner bases; whenever a term order $\leq$ has been fixed and $f \leq g$ occurs in a theorem or proof, then it is the induced quasi-order on $R[\underline{X}]$ that is being referred to. Again, as with monomials, it is important to keep in mind that the original term order $\leq$ on T is a linear order, whereas the induced quasi-order $\leq$ on $R[\underline{X}]$ is really just a quasi-order and not an order in general. From Lemma 4.38, we see that every non-empty set of terms has a unique $\leq$-minimal element, i.e., a least element. A non-empty set of polynomials has at least one $\leq$-minimal element f, and by Lemma 4.37, the other minimal elements g in that set are precisely those that satisfy $T(g) = T(f)$.

Exercise 5.13 Let $R[\underline{X}] = \mathbb{Q}[X, Y, Z]$, $\leq$ the lexicographical term order with $Z \ll Y \ll X$. How do the polynomials $X^2YZ^3 - 2XY^2Z + 3YZ + 1$, $5X^2YZ^3 + 2XY^2Z - 3YZ + 3$, and $X^2YZ^3 + 2XY^2Z - Y$ relate to each other in the induced quasi-order on $R[\underline{X}]$?

Exercise 5.14 Let $\leq$ be a term order on T. Show that if the ground ring R is an integral domain, then multiplication on M is still monotone w.r.t. $\leq$, i.e., $m_1 \leq m_2$ implies $m_1 \cdot m_3 \leq m_2 \cdot m_3$.

For arbitrary polynomials in $R[\underline{X}]$, multiplication is no longer monotone as the following example shows. Let $R[\underline{X}] = \mathbb{Q}[X]$, and let

$$f = X, \quad g = X + 1, \quad \text{and} \quad h = X - 1.$$

Let $\leq$ be the unique term order on $T = \{ X^n \mid n \in \mathbb{N} \}$. Then $f < g$, but

$$f \cdot h = X^2 - X \quad \text{and} \quad g \cdot h = X^2 - 1,$$

and so $f \cdot h > g \cdot h$.

Definition 5.15 Let $\leq$ be a term order on T. For any finite, non-empty subset A of M consisting of pairwise inequivalent monomials, we let $\max(A)$ be the unique maximal element of A w.r.t. $\leq$. For any non-zero polynomial $f \in R[\underline{X}]$ we define the **head term** $\mathrm{HT}(f)$, the **head monomial** $\mathrm{HM}(f)$, and the **head coefficient** $\mathrm{HC}(f)$ of f w.r.t. $\leq$ as follows:

$$\begin{aligned} \mathrm{HT}(f) &= \max(T(f)), \\ \mathrm{HM}(f) &= \max(M(f)), \quad \text{and} \\ \mathrm{HC}(f) &= \text{the coefficient of } \mathrm{HM}(f). \end{aligned}$$

The **reductum** $\mathrm{red}(f)$ of f w.r.t. $\leq$ is defined as $f - \mathrm{HM}(f)$, i.e., $f = \mathrm{HM}(f) + \mathrm{red}(f)$. A polynomial $f \in R[\underline{X}]$ is called **monic** w.r.t. $\leq$ if $f \neq 0$ and $\mathrm{HC}(f) = 1$.

Exercise 5.16 Show that $\mathrm{red}(f) < \mathrm{HM}(f) \leq f$ for $0 \neq f \in R[\underline{X}]$.

Lemma 5.17 Let R be an integral domain and let f, $g \in R[\underline{X}]$ with f, $g \neq 0$. Then

(i) $\mathrm{HT}(fg) = \mathrm{HT}(f) \cdot \mathrm{HT}(g)$,

(ii) $\mathrm{HM}(fg) = \mathrm{HM}(f) \cdot \mathrm{HM}(g)$,

(iii) $\mathrm{HC}(fg) = \mathrm{HC}(f) \cdot \mathrm{HC}(g)$, and

(iv) $\mathrm{HT}(f+g) \leq \max\{\mathrm{HT}(f), \mathrm{HT}(g)\}$.

Proof (i) We first note that

$$T(fg) \subseteq \{\, st \mid s \in T(f),\ t \in T(g)\,\}.$$

Moreover, $s \in T(f)$ and $t \in T(g)$ with $s \neq \mathrm{HT}(f)$ or $t \neq \mathrm{HT}(g)$ implies $st < \mathrm{HT}(f) \cdot \mathrm{HT}(g)$ by Lemma 5.7. It is now obvious that $\mathrm{HT}(f) \cdot \mathrm{HT}(g)$ is the head term of fg.

(ii) Every monomial $at \in M(fg)$ can be written as

$$at = \sum \left\{ bc \cdot uv \;\middle|\; bu \in M(f),\ cv \in M(g),\ uv = t \right\}.$$

Moreover, as we just saw, $uv = \mathrm{HT}(fg)$ with $u \in T(f)$ and $v \in T(g)$ happens only if $u = \mathrm{HT}(f)$ and $v = \mathrm{HT}(g)$, and we see that $\mathrm{HM}(f) \cdot \mathrm{HM}(g)$ is the head monomial of fg.

(iii) is an immediate consequence of (ii), and (iv) follows easily from the fact that $T(f+g) \subseteq T(f) \cup T(g)$. □

For the rest of this section, we assume that the ground ring is a field K; as before, we will write $K[\underline{X}]$ for $K[X_1, \ldots, X_n]$. Moreover, we fix a term order $\leq$ on T and denote the induced linear quasi-order on $K[\underline{X}]$ by $\leq$ too.

The next definition generalizes the single steps of the division algorithm for univariate polynomials to the multivariate case. The most important difference is that here, we are aiming at an algorithm that "divides" one polynomial by a *set* of polynomials. This algorithm will appear in the proof of Proposition 5.22 under the name REDPOL.

Definition 5.18 Let $f, g, p \in K[\underline{X}]$ with $f, p \neq 0$, and let P be a subset of $K[\underline{X}]$. Then we say

(i) f **reduces** to g **modulo** p **by eliminating** t (notation $f \xrightarrow[p]{} g\ [t]$), if $t \in T(f)$, there exists $s \in T$ with $s \cdot \mathrm{HT}(p) = t$, and

$$g = f - \frac{a}{\mathrm{HC}(p)} \cdot s \cdot p,$$

where a is the coefficient of t in f,

(ii) f **reduces to** g **modulo** p (notation $f \xrightarrow[p]{} g$), if $f \xrightarrow[p]{} g\ [t]$ for some $t \in T(f)$,

(iii) f **reduces to** g **modulo** P (notation $f \xrightarrow[P]{} g$), if $f \xrightarrow[p]{} g$ for some $p \in P$,

(iv) f is **reducible modulo** p if there exists $g \in K[\underline{X}]$ such that $f \xrightarrow[p]{} g$, and

(v) f is **reducible modulo** P if there exists $g \in K[\underline{X}]$ such that $f \xrightarrow[P]{} g$.

If f is not reducible modulo p (modulo P), then we say f is **in normal form modulo** p **(modulo** P**)**. A **normal form** of f **modulo** P is a polynomial g that is in normal form modulo P and satisfies

$$f \xrightarrow[P]{*} g,$$

where $\xrightarrow[P]{*}$ is the reflexive-transitive closure of $\xrightarrow[P]{}$ of Definition 4.71. We call

$$f \xrightarrow[p]{} g \ [t]$$

a **top-reduction** of f if $t = \mathrm{HT}(f)$; whenever a top-reduction of f exists (with $p \in P$), we say that f is **top-reducible modulo** p **(modulo** P**)**.

A polynomial f that is in normal form modulo some set P of polynomials is sometimes also called *irreducible* modulo P. We will avoid that terminology here because it can lead to confusion with the established use of the word "irreducible" in the sense of "not allowing a proper factorization."

Lemma 5.19 If $\leq$ is a decidable term order and $K[\underline{X}]$ is a polynomial ring over a computable field K, then $\xrightarrow[P]{}$ is decidable for every finite $P \subseteq K[\underline{X}]$.

Proof Using the normal form of polynomials of Proposition 5.11, we can clearly decide whether or not f is reducible modulo P. (We can even detect *all* possible reduction steps.) If the answer is positive, then computability of K and $K[\underline{X}]$ certainly allow us to compute g with $f \xrightarrow[P]{} g$. □

The most important property of polynomial reduction $\xrightarrow[P]{}$ which will be proved soon is that it is a noetherian reduction relation. We will thus be able to apply the results of Section 4.5 here. Note that the above definition of normal forms is consistent with the one given in Definition 4.71. We will frequently make use of the notation for the various closures of $\longrightarrow$ introduced in Definition 4.71. Note that $f \xrightarrow[P]{*} g$ means that there is a reduction chain directed from f to g, whereas $f \overset{*}{\underset{P}{\longleftrightarrow}} g$ means that there is a reduction chain between the two in which arrows in both directions may occur. Moreover, $f \downarrow_P g$ means that there exists h with $f \xrightarrow[P]{*} h$ and $g \xrightarrow[P]{*} h$.

The set of terms in a single variable allows only one term order, namely, the one by increasing exponents. Inspection of the algorithm DIVPOL (long division of polynomials) shows that if f and g are univariate polynomials and we divide f by g with remainder r, then this computation gives rise to a reduction chain

$$f \xrightarrow[g]{*} r \qquad (r \text{ in normal form modulo } g).$$

The concept of reduction of multivariate polynomials w.r.t. a term order generalizes this in two ways: firstly, we do not, as in the case of long division, insist on doing top reductions only. Secondly, we also consider reduction modulo a set of polynomials rather than just one polynomial. Still, long division of polynomials suggests a good way of visualizing what a reduction step does. If we write monomials in decreasing order, then the reduction step of (i) in the definition above can be visualized as follows.

$$\begin{array}{rlcl} & a_1t_1 + \cdots + a_it_i + & at & + \text{ (lower monomials)} \\ - & \Big(& \underbrace{\frac{a}{\mathrm{HC}(p)} \cdot s \cdot \mathrm{HM}(p)}_{=at} & + \text{ (lower monomials)}\Big) \\ = & a_1t_1 + \cdots + a_it_i & & + \text{ (lower monomials)} \end{array}$$

Another way of looking at this reduction step is to say that it replaces the monomial at in f by

$$-\frac{a}{\mathrm{HC}(p)} \cdot s \cdot \mathrm{red}(p).$$

It is clear that an algorithm which reduces f modulo P will in general be non-deterministic. It does, however, always terminate: we will now show that $\xrightarrow[P]{}$ is noetherian. This is essentially due to the fact that when a monomial m is eliminated from f by means of a reduction step, then all monomials m' of f with $m' > m$ remain unchanged.

Lemma 5.20 Let f, g, $p \in K[\underline{X}]$ and P a subset of $K[\underline{X}]$. Then the following hold:

(i) f is reducible modulo p iff there exists $t \in T(f)$ such that $\mathrm{HT}(p) \mid t$.

(ii) If $f \xrightarrow[p]{} f - mp$ for some monomial m, then $\mathrm{HT}(mp) \in T(f)$.

(iii) Suppose $f \xrightarrow[p]{} g\ [t]$. Then $t \notin T(g)$, while for all $t' \in T$ with $t' > t$, we have $t' \in T(f)$ iff $t' \in T(g)$. In fact, $m \in M(f)$ iff $m \in M(g)$ for every monomial $m > t$.

(iv) If $f \xrightarrow[p]{} g$, then $g < f$.

(v) If $f \xrightarrow[P]{*} g$, then $g \leq f$, and $g = 0$ or $\mathrm{HT}(g) \leq \mathrm{HT}(f)$.

Proof (i) and (ii) are immediate from the definitions.

(iii) By the definition of reduction, there must exist $s \in T$ and $a \in K$ such that

$$g = f - \frac{a}{\mathrm{HC}(p)} \cdot s \cdot p,$$

where $s \cdot \mathrm{HT}(p) = t$ and $at \in M(f)$. It is clear that

$$\mathrm{HM}\Big(\frac{a}{\mathrm{HC}(p)} \cdot s \cdot p\Big) = a \cdot s \cdot \mathrm{HT}(p) = at.$$

Furthermore, every $u \in T(sp)$ is of the form $u = sv$ with $v \in T(p)$, and so

$$u \leq s \cdot \mathrm{HT}(p) = t.$$

Looking at the way in which polynomials are subtracted, the claims are now obvious from Proposition 5.11 (i).

(iv) From (iii) above, we see that $T(g) < T(f)$ in the well-order of $\mathcal{P}_{\mathrm{fin}}(T)$ induced by $\leq$ (use the discussion preceding Theorem 5.12 and the diagram preceding the lemma to understand why), and so $g < f$.

(v) The first statement follows easily from (iv) by induction on the length of the reduction chain $f \xrightarrow[P]{*} g$. The second one is now obvious from the definition of the induced quasi-order on $K[\underline{X}]$. □

As an immediate consequence of (iv) above, Lemma 4.73, and the fact that $\leq$ on $K[\underline{X}]$ is well-founded, we obtain the following theorem.

Theorem 5.21 *The relation $\xrightarrow[P]{}$ is a noetherian reduction relation on $K[\underline{X}]$ for every $P \subseteq K[\underline{X}]$.* □

We may now conclude from Lemma 4.72 that every $f \in K[\underline{X}]$ has at least one normal form modulo P. The next proposition shows that we can say considerably more than that. The reader should note how the algorithm REDPOL below is a perfect generalization of DIVPOL to the multivariate case.

Proposition 5.22 *Let P be a subset of $K[\underline{X}]$ and $f \in K[\underline{X}]$. Then there exists a normal form $g \in K[\underline{X}]$ of f modulo P and a family $\mathcal{F} = \{q_p\}_{p \in P}$ of elements of $K[\underline{X}]$ with*

$$f = \sum_{p \in P} q_p p + g \quad \text{and} \quad \max\{\, \mathrm{HT}(q_p p) \mid p \in P,\ q_p p \neq 0\,\} \leq \mathrm{HT}(f).$$

If P is finite, the ground field is computable, and the term order on T is decidable, then g and $\{q_p\}_{p \in P}$ can be computed from f and P.

Proof We give an algorithm REDPOL (Table 5.1) for the computation of g and the q_p. For general, possibly non-computable field, non-decidable term order, and infinite P, the steps of the algorithm can be interpreted as mathematical constructions that prove the existence of the q_p. (The existence of g could of course be inferred in the same way, but we already know from Lemma 4.72 that g exists.) Let us denote by g_i the value of g after the ith run through the **while**-loop, with $g_0 = f$.

Termination: An infinite run of the **while**-loop would give rise to an infinite chain $g_0 \xrightarrow[P]{} g_1 \xrightarrow[P]{} \cdots$, contradicting Theorem 5.21.

TABLE 5.1. Algorithm REDPOL

```
Specification: $(\mathcal{F}, g) \leftarrow$ REDPOL$(f, P)$
                        Complete reduction of $f$ modulo $P$
Given: a finite subset $P$ of $K[\underline{X}]$ and $f \in K[\underline{X}]$
Find: a normal form $g$ of $f$ modulo $P$, and a family
      $\mathcal{F} = \{q_p\}_{p \in P}$ of polynomials with $f = \sum_{p \in P} q_p p + g$ and
      $\max\{\mathrm{HT}(q_p p) \mid p \in P,\ q_p p \neq 0\} \leq \mathrm{HT}(f)$
begin
$q_p \leftarrow 0$ (all $p \in P$)
$g \leftarrow f$
while $g$ is reducible modulo $P$ do
      select $p \in P$ such that $g$ is reducible modulo $p$
      determine a monomial $m$ with $g \xrightarrow[p]{} g - mp$
      $g \leftarrow g - mp$
      $q_p \leftarrow q_p + m$
end
$\mathcal{F} \leftarrow \{q_p\}_{p \in P}$
return$(\mathcal{F}, g)$
end REDPOL
```

Correctness: Suppose there are N runs through the **while**-loop. From $g_i \xrightarrow[P]{} g_{i+1}$ for all $0 \leq i < N$, we conclude that $f \xrightarrow[P]{*} g$ is an invariant of the loop. It is easy to see that the equation

$$f = \sum_{p \in P} q_p p + g$$

is also a loop invariant. Finally, we claim that

$$\max\{\mathrm{HT}(q_p p) \mid p \in P,\ q_p p \neq 0\} \leq \mathrm{HT}(f)$$

is an invariant of the loop. It is trivially true upon initialization. Now suppose it is true after the ith run for $1 \leq i < N$. We have $\mathrm{HT}(g_i) \leq \mathrm{HT}(f)$ by Lemma 5.20 (v) and the first invariant. Let mp be the polynomial that is being subtracted from g_i during the next run. Then $\mathrm{HT}(mp) \in T(g_i)$ and so $\mathrm{HT}(mp) \leq \mathrm{HT}(g_i) \leq \mathrm{HT}(f)$. The claim now follows easily from Lemma 5.17 (iv). □

An obvious consequence of the proposition is that $f - g \in \mathrm{Id}(P)$. We will come back to this in Lemma 5.26. The statement in the proposition concerning the head terms will play an important role in the theory: it will lead to the concept of *standard representations.*

When applied to the special case of long division of univariate polynomials, Proposition 5.22 has a rather surprising consequence. It says that when performing long division, we need not necessarily eliminate terms from the dividend in descending order (although it is certainly a good idea to do so). We may well perform the division in such a way that a previously eliminated term reappears; the algorithm will terminate nevertheless.

Exercise 5.23 Let $K[\underline{X}] = \mathbb{Q}[X,Y,Z]$, $f = XYZ - XY^2 + Z$, $P = \{p_1, p_2\}$ with $p_1 = XY + 1$, $p_2 = YZ + 1$. Perform the algorithm REDPOL with input (f, P) in all possible ways. Why is it not necessary to specify a term order here?

The above exercise shows that the reduction relation $\xrightarrow[P]{}$ does not in general have unique normal forms. We will later on define and construct Gröbner bases as finite sets $P \subseteq K[\underline{X}]$ for which $\xrightarrow[P]{}$ does in fact have unique normal forms. We will now provide some more technical results concerning reduction.

Lemma 5.24 Let $P \subseteq K[\underline{X}]$ and f, g, $h \in K[\underline{X}]$, and let $m \in M$.

(i) If $f \in P$, then $hf \xrightarrow[P]{*} 0$.

(ii) If $f \xrightarrow[P]{} g$, then $mf \xrightarrow[P]{} mg$.

(iii) If $f \xrightarrow[P]{*} g$, then $mf \xrightarrow[P]{*} mg$. In particular, $f \xrightarrow[P]{*} 0$ implies $mf \xrightarrow[P]{*} 0$.

Proof (i) Assume for a contradiction that the set

$$H = \left\{ h \in K[\underline{X}] \mid \text{not } hf \xrightarrow[P]{*} 0 \right\}$$

is non-empty. Then H contains a $\leq$-minimal element $h \neq 0$. With $m = \mathrm{HM}(h)$, we obtain $\mathrm{HM}(hf) = m \cdot \mathrm{HM}(f)$, and so

$$hf \xrightarrow[f]{} hf - mf \quad \text{and} \quad hf - mf = \mathrm{red}(h) \cdot f.$$

We have $\mathrm{red}(h) \notin H$ since $\mathrm{red}(h) < h$, and so $\mathrm{red}(h) \cdot f \xrightarrow[P]{*} 0$. It follows that $hf \xrightarrow[P]{*} 0$, contradicting the assumption $h \in H$.

(ii) Suppose $f \xrightarrow[p]{} g$, say $g = f - m'p$ for some $p \in P$. Then $\mathrm{HT}(m'p) \in T(f)$, and it follows easily that $\mathrm{HT}(m'mp) \in T(mf)$. We see that

$$mf \xrightarrow[p]{} mf - m'mp \quad \text{and} \quad mf - mm'p = mg.$$

(iii) This follows from (ii) by induction on the length k of the reduction chain $f \xrightarrow[P]{k} g$. □

Lemma 5.25 (Translation Lemma) Let f, g, h, $h_1 \in K[\underline{X}]$, and let $P \subseteq K[\underline{X}]$.

(i) If $f - g = h$ and $h \xrightarrow[P]{*} h_1$, then there exist f_1, $g_1 \in K[\underline{X}]$ such that $f_1 - g_1 = h_1$, $f \xrightarrow[P]{*} f_1$, and $g \xrightarrow[P]{*} g_1$.

(ii) If $f - g \xrightarrow[P]{*} 0$, then $f \downarrow_P g$, and so in particular $f \xleftrightarrow[P]{*} g$.

Proof For (i), we show by induction on $k \in \mathbb{N}$ that $f - g = h$ and $h \xrightarrow[P]{k} h_1$ implies the existence of f_1, $g_1 \in K[\underline{X}]$ with the indicated properties. For $k = 0$, we take $f_1 = f$ and $g_1 = g$. Let now $h \xrightarrow[P]{k+1} h_1$, say

$$h \xrightarrow[P]{k} h_2 \xrightarrow[P]{} h_1.$$

By the induction hypothesis, there exist f_2, $g_2 \in K[\underline{X}]$ with

$$f_2 - g_2 = h_2, \quad f \xrightarrow[P]{*} f_2, \quad \text{and} \quad g \xrightarrow[P]{*} g_2.$$

It now suffices to find f_1, $g_1 \in K[\underline{X}]$ with

$$f_1 - g_1 = h_1, \quad f_2 \xrightarrow[P]{*} f_1, \quad \text{and} \quad g_2 \xrightarrow[P]{*} g_1,$$

as indicated in the diagram below.

$$\begin{tikzcd}[row sep=small, column sep=small]
f \arrow[d, "P"', "*"] & - & g \arrow[d, "P"', "*"] & = & h \arrow[d, "P"', "k"] \\
f_2 \arrow[d, "P"', "*"] & - & g_2 \arrow[d, "P"', "*"] & = & h_2 \arrow[d, "P"'] \\
f_1 & - & g_1 & = & h_1
\end{tikzcd}$$

Suppose

$$h_1 = h_2 - \frac{c}{b} \cdot u \cdot p,$$

where $p \in P$, $b = \mathrm{HC}(p)$, $u \in T$, and $0 \neq c$ is the coefficient of the monomial in $M(h_2)$ whose term is $u \cdot \mathrm{HT}(p)$. Let c_1 be the coefficient of the monomial in $M(f_2)$ with term $u \cdot \mathrm{HT}(p)$ if $u \cdot \mathrm{HT}(p) \in T(f_2)$, and let c_1 be zero otherwise. Define c_2 in the same way w.r.t. g_2. Set

$$f_1 = f_2 - \frac{c_1}{b} \cdot u \cdot p \quad \text{and} \quad g_1 = g_2 - \frac{c_2}{b} \cdot u \cdot p.$$

Then $c_1 - c_2 = c$ because $h_2 = f_2 - g_2$, and we see that $f_1 - g_1 = h_1$. Furthermore, we have defined f_1 and g_1 in such a way that $f_2 \xrightarrow[P]{*} f_1$ and $g_2 \xrightarrow[P]{*} g_1$. (ii) is the special case $h_1 = 0$ of (i). □

Next, we relate polynomial reduction in $K[\underline{X}]$ to *congruence relations* on $K[\underline{X}]$ induced by ideals in $K[\underline{X}]$. For every $P \subseteq K[\underline{X}]$, we let $\mathrm{Id}(P)$ be the ideal generated by P in $K[\underline{X}]$, i.e., the set of all finite linear combinations $\sum h_i p_i$ with $h_i \in K[\underline{X}]$ and $p_i \in P$ (see Definition 1.36). If I is an ideal in $K[\underline{X}]$, then the equivalence relation $\equiv_I$ defined by

$$f \equiv_I g \quad \text{iff} \quad f - g \in I$$

(cf. Exercise 4.20 (iv)) is called the **congruence relation** modulo I on $K[\underline{X}]$. We thus have $f \equiv_I g$ iff $f + I = g + I$ in the residue class ring $K[\underline{X}]/I$. Furthermore, $f \equiv_I g$ implies that $f \in I$ iff $g \in I$. (Cf. the remarks at the end of Section 1.5 and the discussion preceding Lemma 4.22.)

Lemma 5.26 Let $P \subseteq K[\underline{X}]$ and let f, $g \in K[\underline{X}]$. Then $f \equiv_{\mathrm{Id}(P)} g$ iff $f \xleftrightarrow[P]{*} g$. In particular, $f \xrightarrow[P]{*} g$ implies $f - g \in \mathrm{Id}(P)$, and $f \xrightarrow[P]{*} 0$ implies $f \in \mathrm{Id}(P)$.

Proof "$\Longleftarrow$": We show by induction on $k \in \mathbb{N}$ that $f \xleftrightarrow[P]{k} g$ implies $g - f \in \mathrm{Id}(P)$. If $k = 0$, then $f = g$, and so $g - f = 0 \in \mathrm{Id}(P)$. If $f \xleftrightarrow[P]{k+1} g$, say

$$f \xleftrightarrow[P]{k} h \xleftrightarrow[P]{} g,$$

then $h - f \in \mathrm{Id}(P)$ by the induction hypothesis, and $g - h = mp$ for some $m \in M$ and some $p \in P$ by the definition of $\xrightarrow[P]{}$. Consequently, $g - f = (g - h) + (h - f) \in \mathrm{Id}(P)$.

"$\Longrightarrow$": Let $g - f \in \mathrm{Id}(P)$. Then there exist $p_i \in P$ and $h_i \in K[\underline{X}]$ $(1 \leq i \leq k)$ such that

$$g = f + \sum_{i=1}^{k} h_i p_i.$$

We show by induction on k that $f \xleftrightarrow[P]{*} g$. If $k = 0$, then $f = g$. If

$$g = f + \sum_{i=1}^{k} h_i p_i + h_{k+1} p_{k+1},$$

then $g \xleftrightarrow[P]{*} (f + h_{k+1} p_{k+1})$ by induction hypothesis. It now suffices to show that

$$(f + h_{k+1} p_{k+1}) \xleftrightarrow[P]{*} f.$$

This follows readily from the translation lemma together with the fact that $h_{k+1} p_{k+1} \xrightarrow[P]{*} 0$ by Lemma 5.24. □

The reader should note that in the computable case, the algorithm RED-POL effectively provides the representation of f as a sum of multiples of elements of P when it reduces f to zero modulo P. In the terminology of Definition 4.78, we have proved the following.

Proposition 5.27 *Let $P \subseteq K[\underline{X}]$. Then the reduction relation $\xrightarrow[P]{}$ on $K[\underline{X}]$ is adequate for $\equiv_{\mathrm{Id}(P)}$.* □

Exercise 5.28 Use the results on polynomial reduction to give an alternate proof of the Hilbert basis theorem for polynomial rings over fields. (Hint: Assume that there exists an ideal of $K[\underline{X}]$ that is not finitely generated, define a sequence of non-zero polynomials such that every element of the sequence is in normal form modulo the set of its predecessors, and apply Proposition 4.42 (ii) to the sequence of head terms.)

We close this section with an algorithm that turns an arbitrary finite subset P of $K[\underline{X}]$ into another finite set that generates the same ideal and has the additional property that each of its elements is in normal form modulo the rest.

Definition 5.29 Let $P \subseteq K[\underline{X}]$. Then P is called **monic** if every $p \in P$ is monic; P is called **reduced** (or **autoreduced**) if every $p \in P$ is monic and in normal form modulo $P \setminus \{p\}$.

Note that $0 \notin P$ if P is monic. The following algorithm and the proof of its correctness and termination can actually be interpreted as a mathematical proof of the fact that every ideal in $K[\underline{X}]$ has a finite reduced basis. This, however, turns out to be of little relevance in the theory (cf. Theorem 5.43), whereas the actual computation of a reduced basis from a given basis is often important.

Proposition 5.30 *Let P be a finite subset of $K[\underline{X}]$. Suppose the ground field is computable and the term order on T is decidable. Then the algorithm* REDUCTION *of Table* 5.2 *computes a finite reduced subset Q of $K[\underline{X}]$ such that* $\mathrm{Id}(Q) = \mathrm{Id}(P)$.

TABLE 5.2. Algorithm REDUCTION

```
Specification: $Q \leftarrow \mathrm{REDUCTION}(P)$
               Construction of a finite reduced set $Q$
               such that $\mathrm{Id}(Q) = \mathrm{Id}(P)$
Given: $P$ = a finite subset of $K[\underline{X}]$
Find: $Q$ = a finite reduced set in $K[\underline{X}]$ with $\mathrm{Id}(Q) = \mathrm{Id}(P)$
begin
$Q \leftarrow P$
while there is $p \in Q$ which is reducible modulo $Q \setminus \{p\}$ do
        select $p \in Q$ which is reducible modulo $Q \setminus \{p\}$
        $Q \leftarrow Q \setminus \{p\}$
        $h \leftarrow$ some normal form of $p$ modulo $Q$
        if $h \neq 0$ then $Q \leftarrow Q \cup \{h\}$ end
end
$Q \leftarrow \{ (\mathrm{HC}(q))^{-1} \cdot q \mid q \in Q \}$
end REDUCTION
```

Proof It is an easy exercise to show that $\mathrm{Id}(Q) = \mathrm{Id}(P)$ is a loop invariant of the **while**-loop. Correctness of the algorithm is now immediate from the **while**-clause. To prove termination, let $P = \{p_1, \ldots, p_m\}$ be any input set. We may regard P as an ordered m-tuple $(p_1, \ldots, p_m)$ rather than a set. If some p_i $(1 \leq i \leq m)$ is selected in the **while**-loop, then we replace it by its normal form h even if $h = 0$ (rather than throwing it out). Let now Q_i be the m-tuple thus obtained after the ith run through the loop. Assume that the algorithm does not terminate. Since at least one entry of the m-tuple Q_i is changed when passing from Q_i to Q_{i+1}, there must be $1 \leq k \leq m$ such that the kth entry changes infinitely many times. But a zero entry never changes back to something non-zero, and all other changes replace some p by $h < p$. Hence we are looking at a strictly descending chain w.r.t. the induced quasi-order on $K[\underline{X}]$, which is impossible. □

It is clear that for an actual implementation of the above algorithm, one must find some way to test the **while**-clause without redundance.

It is interesting to see what REDUCTION does when applied to a set of polynomials of total degree at most 1. Let F be a finite subset of $K[\underline{X}]$ where each $f \in F$ has total degree at most 1. If we apply the algorithm REDUCTION to F, then it will subtract constant multiples of polynomials from others until no two polynomials have the same variable as their head term and no variable that is a head term of some polynomial occurs in any other polynomial. This is precisely what (a certain version of) the so-called *Gaussian elimination algorithm* for the solution of a system of linear equations does. This connection will be explained in detail in Section 10.5.

Another special case that is worth mentioning is that of two univariate polynomials f and g. If we modify REDUCTION in such a way that it performs top reduction only, then it is easy to see that when applied to $\{f, g\}$, it will perform the same "back and forth divisions" as the Euclidean algorithm and eventually output $\{\gcd(f, g)\}$.

Exercise 5.31 Let P be a finite set of univariate polynomials over a field. Show that REDUCTION(P) is a one-element set consisting of the gcd of the elements of P.

5.2 Gröbner Bases—Existence and Uniqueness

The main facts on polynomial reduction proved in the previous section can be summarized as follows.

Proposition 5.32 *Let $K[\underline{X}] = K[X_1, \ldots, X_n]$ be a polynomial ring over a field K, let $\leq$ be an arbitrary term order on T, and let polynomial reduction $\xrightarrow[P]{}$ on $K[\underline{X}]$ for $P \subseteq K[\underline{X}]$ be defined w.r.t. $\leq$. Then $\xrightarrow[P]{}$ is a noetherian reduction relation on $K[\underline{X}]$ that is adequate for the equivalence relation $\equiv_{\mathrm{Id}(P)}$. In particular, $f \xrightarrow[P]{*} 0$ implies $f \in \mathrm{Id}(P)$. Moreover, if K is computable, $\leq$ is decidable and P is finite, then $\xrightarrow[P]{}$ is decidable.* □

Unfortunately, $\xrightarrow[P]{}$ is general not locally confluent, and so Theorem 4.79 is not applicable in order to solve the equivalence problem for the relation $\equiv_{\mathrm{Id}(P)}$. Consider the following example in $\mathbb{Q}[X]$: $f = X + 1$ and $P = \{X, X + 1\}$. Then

$$f \xrightarrow[X]{} 1 \quad \text{and} \quad f \xrightarrow[X+1]{} 0,$$

and so 0 and 1 are different normal forms of f w.r.t. $\xrightarrow[P]{}$, contradicting Newman's lemma. The example also shows that it is not true in general that $f \in \mathrm{Id}(P)$ implies $f \xrightarrow[P]{*} 0$: here, $1 = (X + 1) - X \in \mathrm{Id}(P)$, but 1 is clearly in normal form modulo P. In this example, the problem can be resolved easily. Since $1 \in \mathrm{Id}(P)$, we may enlarge P to $P' = \{X, X+1, 1\}$. Then $\mathrm{Id}(P) = \mathrm{Id}(P')$, every $0 \neq f \in K[\underline{X}]$ is reducible modulo P', and so $\xrightarrow[P']{}$ is locally confluent and adequate for $\equiv_{\mathrm{Id}(P)}$. The reason why $\xrightarrow[P']{}$ is locally confluent is that $1 \in P'$ is a generator for the ideal $\mathrm{Id}(P') = \mathrm{Id}(P)$. This will become apparent from the next proposition.

As in the previous section, let K be a field, $K[\underline{X}] = K[X_1, \ldots, X_n]$, and $\leq$ a fixed term order on T.

Proposition 5.33 *Let $0 \neq p \in K[\underline{X}]$. Then $\xrightarrow[p]{}$ is locally confluent.*

Proof Suppose $f \xrightarrow[p]{} f_i\ [t_i]$ for $i = 1, 2$. In order to show that $f_1 \,{}_p{\downarrow}\, f_2$, it suffices by the translation lemma to verify that $f_1 - f_2 \xrightarrow[p]{*} 0$. Let $f_i = f - m_i p$ with $m_i \in M$. Then $f_1 - f_2 = (m_2 - m_1) \cdot p \xrightarrow[p]{*} 0$ by Lemma 5.24. □

Corollary 5.34 *Let $P \subseteq K[\underline{X}]$ such that* $\mathrm{Id}(P) = \mathrm{Id}(p)$ *for some $0 \neq p \in P$. Then $\xrightarrow[P]{}$ is locally confluent.*

Proof Let $f \xrightarrow[P]{} f_i$ for $i = 1, 2$. Then $f_1 \xleftrightarrow[P]{*} f_2$, and so by Lemma 5.26, $f_1 - f_2 \in \mathrm{Id}(P) = \mathrm{Id}(p)$. It follows that

$$f_1 \xleftrightarrow[p]{*} f_2,$$

so $f_1 \,{}_p{\downarrow}\, f_2$ by Newman's lemma and the fact that $\xrightarrow[p]{}$ is locally confluent, and so $f_1 \,{}_P{\downarrow}\, f_2$. □

For the case of univariate polynomials, this shows once more in a somewhat roundabout way that for any finite $P \subseteq K[\underline{X}] = K[X]$, the equivalence problem for $\equiv_{\mathrm{Id}(P)}$ is decidable: it suffices to compute $p = \gcd(P)$ by the Euclidean algorithm, and then to apply Theorem 4.79 to the reduction relation $\xrightarrow[p]{}$ (or to $\xrightarrow[P']{}$, where $P' = P \cup \{p\}$). What is the unique normal form h of $f \in K[\underline{X}]$ obtained in this way? It is simply the remainder of f upon division by p. So for univariate polynomials all these considerations lead straight back to division with remainder and the Euclidean algorithm.

For polynomial rings $K[\underline{X}]$ in several variables the situation is, however, entirely different: on the one hand, the Hilbert basis theorem asserts that every ideal I of $K[\underline{X}]$ is finitely generated; on the other hand, we saw in Section 2.2 that not all ideals of $K[\underline{X}]$ have a single generator. So given a finite set P of polynomials in $K[\underline{X}]$, it will in general be impossible to enlarge P by a generator p of $\mathrm{Id}(P)$ to P' so that $\xrightarrow[P']{}$ becomes locally confluent and thus can be employed to decide the equivalence problem for the relation $\equiv_{\mathrm{Id}(P)}$.

This raises the following *fundamental question*: Given a finite set $P \subseteq K[\underline{X}]$, is it possible to construct another finite set $G \subseteq K[\underline{X}]$ such that $\mathrm{Id}(P) = \mathrm{Id}(G)$ and $\xrightarrow[G]{}$ is locally confluent? Rather surprisingly, the answer is *yes*. The rest of this section and the next are devoted to a proof of this fact and some of its consequences. We begin by relating the local confluence of $\xrightarrow[G]{}$ to more algebraic properties of G.

If $P \subseteq K[\underline{X}]$, then we set $\mathrm{HT}(P) = \{\,\mathrm{HT}(p) \mid 0 \neq p \in P\,\}$; for $S \subseteq T$,

$$\mathrm{mult}(S) = \{\, t \in T \mid \text{there is } s \in S \text{ with } s \,|\, t\,\}$$

denotes the set of all multiples of elements of S.

Theorem 5.35 *Let G be a subset of $K[\underline{X}]$. Then the following are equivalent:*

(i) $\xrightarrow[G]{}$ *is locally confluent.*

(ii) $\xrightarrow[G]{}$ *is confluent.*

(iii) $\xrightarrow[G]{}$ *has unique normal forms.*

(iv) $\xrightarrow[G]{}$ *has the Church–Rosser property.*

(v) $f \xrightarrow[G]{*} 0$ *for all* $f \in \mathrm{Id}(G)$.

(vi) Every $0 \neq f \in \mathrm{Id}(G)$ *is reducible modulo* G.

(vii) Every $0 \neq f \in \mathrm{Id}(G)$ *is top-reducible modulo* G.

(viii) For every $s \in \mathrm{HT}(\mathrm{Id}(G))$ *there exists* $t \in \mathrm{HT}(G)$ *with* $t \,|\, s$.

(ix) $\mathrm{HT}(\mathrm{Id}(G)) \subseteq \mathrm{mult}(\mathrm{HT}(G))$.

(x) The polynomials that are in normal form w.r.t. $\xrightarrow[G]{}$ *form a system of unique representatives for the partition*

$$\{\, f + \mathrm{Id}(G) \mid f \in K[\underline{X}]\,\}$$

of $K[\underline{X}]$.

Proof The equivalence of (i)–(iv) has already been shown in Theorem 4.75 for arbitrary noetherian reduction relations.

(iv)$\Longrightarrow$(v): Let $f \in \mathrm{Id}(G)$. Then $f - 0 \in \mathrm{Id}(G)$ and thus $f \overset{*}{\underset{G}{\longleftrightarrow}} 0$ by Lemma 5.26. Since $\xrightarrow[G]{}$ has the Church–Rosser property, there exists $h \in K[\underline{X}]$ with $f \xrightarrow[G]{*} h$ and $0 \xrightarrow[G]{*} h$. Since 0 is always in normal form, we get $h = 0$.

(v)$\Longrightarrow$(vi): Let $0 \neq f \in \mathrm{Id}(G)$. By (v), there exists $h \in K[\underline{X}]$ with $f \xrightarrow[G]{} h \xrightarrow[G]{*} 0$.

(vi)$\Longrightarrow$(vii): Assume for a contradiction that $0 \neq f \in \mathrm{Id}(G)$ is minimal (w.r.t. the given quasi-order on $K[\underline{X}]$) such that it is not top-reducible w.r.t. $\xrightarrow[G]{}$. Then by (vi), there exists $h \in K[\underline{X}]$ with $f \xrightarrow[G]{} h$. It follows that $h \in \mathrm{Id}(G)$ and $h < f$. Moreover, $\mathrm{HM}(h) = \mathrm{HM}(f)$ since f was not top-reducible. By the minimal choice of f, h is top-reducible w.r.t. $\xrightarrow[G]{}$, say $h \xrightarrow[g]{} h_1$ for some $g \in G$. But then $\mathrm{HT}(g) \,|\, \mathrm{HT}(h)$, and so f is top-reducible w.r.t. $\xrightarrow[g]{}$, a contradiction.

(vii), (viii), and (ix) are simple reformulations of each other.

(ix)$\Longrightarrow$(x): Assume for a contradiction that there exist f_1, $f_2 \in K[\underline{X}]$ both in normal form w.r.t. $\xrightarrow[G]{}$ with $f_1 \neq f_2$ and

$$f_1 + \mathrm{Id}(G) = f_2 + \mathrm{Id}(G).$$

Then $f_1 - f_2 \in \mathrm{Id}(G)$, and so there exists $g \in G$ with $\mathrm{HT}(g) \,|\, \mathrm{HT}(f_1 - f_2)$. But

$$\mathrm{HT}(f_1 - f_2) \in T(f_1) \cup T(f_2),$$

and so f_1 or f_2 is reducible modulo G.

(x)$\Longrightarrow$(iv): Let f_1, $f_2 \in K[\underline{X}]$ with $f_1 \xleftrightarrow[G]{*} f_2$. Then $f_1 - f_2 \in \mathrm{Id}(G)$ by Lemma 5.26, and so

$$f_1 + \mathrm{Id}(G) = f_2 + \mathrm{Id}(G).$$

Let h_1 and h_2 be normal forms of f_1 and f_2, respectively. Then h_1, $h_2 \in f_1 + \mathrm{Id}(G)$ again by Lemma 5.26, and so $h_1 = h_2$ by (x). □

Exercise 5.36 Give direct proofs for the following implications of the above theorem: (v)$\Longleftrightarrow$(vii), (x)$\Longrightarrow$(v), (vi)$\Longrightarrow$(v), and (vi)$\Longrightarrow$(iv).

Definition 5.37 A subset G of $K[\underline{X}]$ is called a **Gröbner basis** (w.r.t. the term order $\leq$) if it is finite, $0 \notin G$, and G satisfies the equivalent conditions of Theorem 5.35. If I is an ideal of $K[\underline{X}]$, then a **Gröbner basis of** I (w.r.t. $\leq$) is a Gröbner basis G (w.r.t. $\leq$) such that $\mathrm{Id}(G) = I$.

The requirement $0 \notin G$ of the definition above is of course not in any way essential. However, we will frequently make assumptions or draw conclusions concerning all non-zero elements of an ideal basis, and so excluding zero in the first place will make many results simpler to formulate. Recall that we have the convention $\mathrm{Id}(\emptyset) = \{0\}$; the empty set is thus a Gröbner basis of the zero ideal.

If G is a Gröbner basis of the ideal I of $K[\underline{X}]$, then in particular, G is a Gröbner basis and thus satisfies conditions (v)–(x) of Theorem 5.35. We also have $\mathrm{Id}(G) = I$ by definition, so these conditions trivially remain valid if we replace $\mathrm{Id}(G)$ by I. The following proposition provides a converse to this.

Proposition 5.38 *Let I be an ideal of $K[\underline{X}]$ and G a finite subset of I with $0 \notin G$. Then each of the following is equivalent to G being a Gröbner basis of I.*

(i) $f \xrightarrow[G]{*} 0$ for all $f \in I$.

(ii) Every $0 \neq f \in I$ is reducible modulo G.

(iii) Every $0 \neq f \in I$ is top-reducible modulo G.

(iv) For every $s \in \mathrm{HT}(I)$ there exists $t \in \mathrm{HT}(G)$ with $t \,|\, s$.

(v) $\mathrm{HT}(I) \subseteq \mathrm{mult}(\mathrm{HT}(G))$.

(vi) The polynomials $h \in K[\underline{X}]$ that are in normal form w.r.t. $\xrightarrow[G]{}$ form a system of unique representatives for the partition $\{\, f+I \mid f \in K[\underline{X}]\,\}$ of $K[\underline{X}]$.

Proof We have just explained how G being a Gröbner basis of I implies each of the listed conditions. Now assume that (i) holds. We have $\mathrm{Id}(G) \subseteq I$ by assumption, and so the condition trivially implies that G is a Gröbner basis. It remains to show that $I \subseteq \mathrm{Id}(G)$. If $f \in I$, then $f \xrightarrow[G]{*} 0$ and thus $f \in \mathrm{Id}(G)$ by Lemma 5.26. The proof can now be finished by showing that (i)–(vi) are equivalent, which can easily be achieved using exactly the same arguments as in the proof of Theorem 5.35 and Exercise 5.36. □

Exercise 5.39 Complete the proof of the above proposition.

Exercise 5.40 Let $G \subseteq K[\underline{X}]$ be a Gröbner basis w.r.t. $\leq$ of the ideal I of $K[\underline{X}]$. Show that $\mathrm{HT}(G)$ is a Gröbner basis w.r.t. $\leq$ of the ideal $\mathrm{Id}(\mathrm{HT}(I))$ of $K[\underline{X}]$. In fact, if $0 \neq f \in \mathrm{Id}(\mathrm{HT}(I))$, then every monomial of f is reducible modulo $\mathrm{HT}(G)$.

We can now give a simple non-constructive existence proof for Gröbner bases.

Theorem 5.41 *Let I be an ideal of $K[\underline{X}]$. Then there exists a Gröbner basis G of I w.r.t. $\leq$.*

Proof By Theorem 5.2, the divisibility relation is a Dickson partial order on T. So the set $\mathrm{HT}(I)$ has a finite basis S w.r.t. divisibility. For each $t \in S$, there exists $f_t \in I$ such that $\mathrm{HT}(f_t) = t$. Let now $G = \{\, f_t \mid t \in S\,\}$. Then G satisfies condition (iv) of the previous proposition, and so G is a Gröbner basis of I. □

Note that we have just found another proof of the Hilbert basis theorem for polynomial rings over a field K: a Gröbner basis of an ideal I is a finite basis of I. A Gröbner basis of an ideal I is of course far from being uniquely determined by I; even if we work with the unique minimal basis of $\mathrm{HT}(I)$ in the above proof, I may still contain many different polynomials with the same head term. We are now going to show how Gröbner bases that are reduced in the sense of Definition 5.29 always exist and are uniquely determined by the ideal they generate. Such a basis will, rather obviously, be called a **reduced Gröbner basis**.

Lemma 5.42 Suppose I is an ideal of $K[\underline{X}]$, m is a monomial, and f and g are minimal polynomials in I such that $\mathrm{HM}(f) = \mathrm{HM}(g) = m$. Then $f = g$.

Proof We must have $T(f) = T(g)$ since otherwise $f < g$ or $g < f$. Note that $f - g \in I$, and $f - g = 0$ or $s = \mathrm{HT}(f - g) < m$. In the latter case,

$s \in T(f) = T(g)$. It follows that there exists $0 \neq c \in K$ such that $s \notin T(h)$, where

$$h = f - c(f - g) \in I.$$

So $\mathrm{HM}(h) = m$ and $h < f$, contradicting the minimality of f. □

For *reduced* Gröbner bases Theorem 5.41 can now be improved as follows.

Theorem 5.43 *Let I be an ideal of $K[\underline{X}]$. Then there exists a unique reduced Gröbner basis G of I w.r.t. $\leq$.*

Proof Let S be the unique minimal basis of $\mathrm{HT}(I)$ w.r.t. the divisibility relation (Lemma 4.43). For each $t \in S$, there is $f_t \in I$ with $\mathrm{HT}(f_t) = t$. Since I is an ideal, we may assume that $\mathrm{HM}(f_t) = t$, and by Lemma 5.42 we may even assume that f_t is the unique minimal member of I with this property. Set

$$G = \{\, f_t \mid t \in S \,\}.$$

Then G is a Gröbner basis of I by the same argument as in the proof of theorem 5.41 above. G is obviously monic, and we claim that it is also reduced. Assume for a contradiction that there exist $g_1, g_2 \in G$ and $f \in K[\underline{X}]$ with $g_1 \xrightarrow[g_2]{} f$ and $g_1 \neq g_2$. If this is a top reduction, then $\mathrm{HT}(g_2) \mid \mathrm{HT}(g_1)$, and so $S' = S \setminus \{\mathrm{HT}(g_1)\}$ was also a basis of $\mathrm{HT}(I)$, contradicting the minimal choice of S. Otherwise, $\mathrm{HM}(f) = \mathrm{HM}(g_1)$ and $f < g_1$, contradicting the minimal choice of g_1.

It remains to prove uniqueness. Assume for a contradiction that H is another reduced Gröbner basis of I. Let g be an element of the symmetric set difference $G \triangle H$ such that $\mathrm{HT}(g)$ is minimal in $\mathrm{HT}(G \triangle H)$, and assume w.l.o.g. that $g \in G \setminus H$. By Theorem 5.35 (viii), there exists $h \in H$ with $\mathrm{HT}(h) \mid \mathrm{HT}(g)$. We must in fact have $h \in H \setminus G$: otherwise G would not be reduced because $h \neq g$. By the minimal choice of g, the divisibility $\mathrm{HT}(h) \mid \mathrm{HT}(g)$ cannot be proper, i.e., we have $\mathrm{HT}(h) = \mathrm{HT}(g)$. Now consider $f = g - h$. Then

$$\mathrm{HT}(f) < \mathrm{HT}(g) = \mathrm{HT}(h)$$

because G and H were monic. Moreover, we must have $\mathrm{HT}(f) \in T(g)$ or $\mathrm{HT}(f) \in T(h)$, say $\mathrm{HT}(f) \in T(g)$. But $f \in I$ implies that there exists $p \in G$ with $\mathrm{HT}(p) \mid \mathrm{HT}(f)$. We see that now a term of g other than the head term is divisible by the head term of some element of G, and this contradicts the fact that G was reduced. □

The results of this section concerning the existence and uniqueness of Gröbner bases for ideals of $K[\underline{X}]$ are of great theoretical importance. However, their proofs are non-constructive. They provide no means to *construct* such bases nor even to *recognize* whether a given set of polynomials in $K[\underline{X}]$ is a Gröbner basis since all the characterizations obtained so far refer to an infinity of tests. The next section will provide algorithmic solutions to these problems.

5.3 Gröbner Bases—Construction

We keep the conventions of the last section: $K[\underline{X}] = K[X_1, \ldots, X_n]$ is a polynomial ring over the field K, and $\leq$ is a term order on T that extends canonically to a well-founded linear quasi-order on $K[\underline{X}]$. Our first goal is to find a characterization of Gröbner bases that involves only finitely many tests. This will show that the property of being a Gröbner basis is algorithmically decidable.

Consider the following example. Let $K[\underline{X}] = \mathbb{Q}[X, Y, Z]$, $P = \{p_1, p_2\}$ where $p_1 = XY + 1$ and $p_2 = YZ + 1$. Then $f = Z - X \in \mathrm{Id}(P)$ since $f = Zp_1 - Xp_2$, but f is in normal form modulo P w.r.t. every term order since XY and YZ are necessarily the head terms of p_1 and p_2, respectively. We have found a member of $\mathrm{Id}(P)$ which does not reduce to 0 modulo P, so P is not a Gröbner basis. The way we have created the problem polynomial was to lift the head terms of p_1 and p_2 to their least common multiple XYZ and then to subtract so that the head monomials drop out. It is clear that there is no reason why the result of this subtraction should be reducible modulo P in general. Gröbner basis algorithms are based on the remarkable fact that if the finitely many differences of the above kind are all benign, i.e., reduce to 0, then *every* $f \in \mathrm{Id}(P)$ reduces to 0 and hence P is a Gröbner basis.

The following lemma provides a characterization of Gröbner bases that still involves infinitely many tests but is an important step towards the reduction to finitely many tests.

Lemma 5.44 Let G be a finite subset of $K[\underline{X}]$ with $0 \notin G$. Assume that whenever $g_1, g_2 \in G$ with $g_1 \neq g_2$ and m_1 and m_2 are monomials such that

$$\mathrm{HM}(m_1 g_1) = \mathrm{HM}(m_2 g_2),$$

it follows that $m_1 g_1 - m_2 g_2 \xrightarrow[G]{*} 0$. Then G is a Gröbner basis.

Proof We show that $\xrightarrow[G]{}$ is locally confluent. Let f, f_1, $f_2 \in K[\underline{X}]$ with $f \xrightarrow[G]{} f_i$, where $f_i = f - m_i g_i$ for some $m_i \in M$ and $g_i \in G$ for $i = 1, 2$. Then by the translation lemma, $f_1 \;{}_G{\downarrow}\; f_2$ provided that

$$m_1 g_1 - m_2 g_2 = f_2 - f_1 \xrightarrow[G]{*} 0.$$

Case 1: $\mathrm{HT}(m_1 g_1) \neq \mathrm{HT}(m_2 g_2)$, say $\mathrm{HT}(m_1 g_1) > \mathrm{HT}(m_2 g_2)$. Then we may reduce $m_1 g_1 - m_2 g_2$ to 0 modulo G by means of two top-reductions:

$$m_1 g_1 - m_2 g_2 \xrightarrow[g_1]{} -m_2 g_2 \xrightarrow[g_2]{} 0.$$

Case 2: $\mathrm{HT}(m_1 g_1) = \mathrm{HT}(m_2 g_2) = t$. Then $\mathrm{HM}(m_1 g_1) = \mathrm{HM}(m_2 g_2)$ since both eliminate the same term t from f. It follows that

$$m_1 g_1 - m_2 g_2 \xrightarrow[G]{*} 0$$

by the hypothesis of the lemma. □

Exercise 5.45 Let s, $t \in T$,

$$s = X_1^{k_1} \cdot \cdots \cdot X_n^{k_n} \quad \text{and} \quad t = X_1^{l_1} \cdot \cdots \cdot X_n^{l_n}.$$

Define $\operatorname{lcm}(s,t)$ as $X_1^{m_1} \cdot \cdots \cdot X_n^{m_n}$, where $m_i = \max(k_i, l_i)$ for $1 \le i \le n$. Show that $\operatorname{lcm}(s,t)$ has the properties of a **least common multiple (lcm)**:

$$s \mid \operatorname{lcm}(s,t) \quad \text{and} \quad t \mid \operatorname{lcm}(s,t),$$

and $\operatorname{lcm}(s,t) \mid u$ whenever $s \mid u$ and $t \mid u$ for $u \in T$.

Definition 5.46 For $i = 1, 2$, let $0 \neq g_i \in K[\underline{X}]$, $t_i = \operatorname{HT}(g_i)$, $a_i = \operatorname{HC}(g_i)$, and $t = s_i t_i = \operatorname{lcm}(t_1, t_2)$ with $s_i \in T$. Then the **S-polynomial** of g_1 and g_2 is defined as

$$\operatorname{spol}(g_1, g_2) = a_2 s_1 g_1 - a_1 s_2 g_2.$$

Exercise 5.47 Let g_1, $g_2 \in K[\underline{X}]$. Show the following:

(i) If $g_1 = g_2$, then $\operatorname{spol}(g_1, g_2) = 0$. Similarly, if both g_1 and g_2 are monomials, then $\operatorname{spol}(g_1, g_2) = 0$.

(ii) If $g_1 \neq g_2$, then either $\operatorname{spol}(g_1, g_2) = 0$, or else

$$\operatorname{HT}(\operatorname{spol}(g_1, g_2)) < \operatorname{lcm}(\operatorname{HT}(g_1), \operatorname{HT}(g_2)).$$

(iii) If $\operatorname{HT}(g_2) \mid \operatorname{HT}(g_1)$, then $\operatorname{HC}(g_2) \cdot g_1 \xrightarrow[g_2]{} \operatorname{spol}(g_1, g_2)$, and this is actually a top reduction.

Theorem 5.48 *Let G be a finite subset of $K[\underline{X}]$ with $0 \notin G$. Then the following are equivalent:*

(i) G is a Gröbner basis.

(ii) Whenever g_1, $g_2 \in G$ and $h \in K[\underline{X}]$ is a normal form of $\operatorname{spol}(g_1, g_2)$ *modulo G, then $h = 0$.*

(iii) $\operatorname{spol}(g_1, g_2) \xrightarrow[G]{*} 0$ *for all g_1, $g_2 \in G$.*

Proof (i)$\Longrightarrow$(ii): $\operatorname{spol}(g_1, g_2)$ is obviously in $\operatorname{Id}(G)$ for all g_1, $g_2 \in G$. So by Theorem 5.35, $\operatorname{spol}(g_1, g_2)$ has 0 as a normal form modulo G. Since moreover normal forms are unique, it follows that $h = 0$.

(ii)$\Longrightarrow$(iii) is trivial.

(iii)$\Longrightarrow$(i): By the last lemma, it suffices to show that polynomials of the form $m_1 g_1 - m_2 g_2$ with $g_1 \neq g_2$ in G, monomials m_1 and m_2, and

$$\operatorname{HM}(m_1 g_1) = \operatorname{HM}(m_2 g_2) \tag{$*$}$$

reduce to 0 modulo G. For $i = 1, 2$, let $t_i = \operatorname{HT}(g_i)$, $a_i = \operatorname{HC}(g_i)$, and $m_i = b_i u_i$ with $b_i \in K$ and $u_i \in T$. Then the equation $(*)$ becomes

$$b_1 a_1 u_1 t_1 = b_2 a_2 u_2 t_2. \tag{$**$}$$

Now let s_1, $s_2 \in T$ such that $s_i t_i = \mathrm{lcm}(t_1, t_2)$ for $i = 1, 2$. From (**) we see that $u_1 t_1 = u_2 t_2$ is a common multiple of t_1 and t_2. It follows that there exists $v \in T$ such that for $i = 1, 2$,

$$u_i t_i = v \cdot \mathrm{lcm}(t_1, t_2) = v s_i t_i.$$

We see that $u_i = v s_i$. Furthermore, (**) implies that $(b_1/a_2) = (b_2/a_1)$, and we obtain

$$\begin{aligned} m_1 g_1 - m_2 g_2 &= b_1 u_1 g_1 - b_2 u_2 g_2 \\ &= b_1 v s_1 g_1 - b_2 v s_2 g_2 \\ &= \frac{b_1}{a_2} \cdot v \cdot (a_2 s_1 g_1 - a_1 s_2 g_2) \\ &= \frac{b_1}{a_2} \cdot v \cdot \mathrm{spol}(g_1, g_2). \end{aligned}$$

From Lemma 5.24 and the fact that $\mathrm{spol}(g_1, g_2) \xrightarrow[G]{*} 0$, we conclude that

$$\frac{b_1}{a_2} \cdot v \cdot \mathrm{spol}(g_1, g_2) \xrightarrow[G]{*} 0. \quad \square$$

The following corollary is now immediate from the fact that the S-polynomial of two monomials equals zero.

Corollary 5.49 *Let $G \subseteq K[\underline{X}]$ be a finite set of monomials. Then G is a Gröbner basis.* □

As another consequence of this theorem, we can now show that the Gröbner basis property is preserved under extensions of the polynomial ring $K[\underline{X}]$. It is clear that for n, $n' \in \mathbb{N}$ with $n < n'$, the restriction of a term order on $T' = T(X_1, \ldots, X_{n'})$ to $T = T(X_1, \ldots, X_n)$ is a term order on T.

Exercise 5.50 Show that every term order on T is the restriction of some term order on T'.

Corollary 5.51 *Let $K[\underline{X}] = K[X_1, \ldots, X_n]$, K a subfield of K', $n' \geq n$, and $K'[\underline{X}'] = K'[X_1, \ldots, X_{n'}]$. Then the following hold:*

(i) Suppose $\leq$ is a term order on T'. Then every Gröbner basis G in $K[\underline{X}]$ w.r.t. the restriction of $\leq$ to T is a Gröbner basis in $K'[\underline{X}']$ w.r.t. $\leq$.

(ii) Let F be a finite subset of $K[\underline{X}]$ and denote by

$$\mathrm{Id}_{K[\underline{X}]}(F) \quad \text{and} \quad \mathrm{Id}_{K'[\underline{X}']}(F)$$

the ideals generated by F in $K[\underline{X}]$ and $K'[\underline{X}']$, respectively. Then

$$\mathrm{Id}_{K'[\underline{X}']}(F) \cap K[\underline{X}] = \mathrm{Id}_{K[\underline{X}]}(F).$$

Proof (i) Let g_1, $g_2 \in G$. Then $\operatorname{spol}(g_1, g_2)$ is the same in $K[\underline{X}]$ and in $K'[\underline{X}']$. So $\operatorname{spol}(g_1, g_2) \xrightarrow[G]{*} 0$ in $K[\underline{X}]$ and hence in $K'[\underline{X}']$. So by the theorem above, G is a Gröbner basis in $K'[\underline{X}']$.

(ii) The inclusion "$\supseteq$" is trivial. Now let

$$f \in \mathrm{Id}_{K'[\underline{X}']}(F) \cap K[\underline{X}].$$

Let $G \subseteq K[\underline{X}]$ be a Gröbner basis of $\mathrm{Id}_{K[\underline{X}]}(F)$ w.r.t. some term order $\leq$. By (i), G is a Gröbner basis in $K'[\underline{X}']$ w.r.t any term order on T' whose restriction to T equals $\leq$. From the fact that $\mathrm{Id}_{K[\underline{X}]}(F) = \mathrm{Id}_{K[\underline{X}]}(G)$ one easily concludes that

$$\mathrm{Id}_{K'[\underline{X}']}(G) = \mathrm{Id}_{K'[\underline{X}']}(F).$$

It follows that $f \xrightarrow[G]{*} 0$. Since $f \in K[\underline{X}]$ and $G \subseteq K[\underline{X}]$, this reduction takes place in $K[\underline{X}]$, and we see that $f \in \mathrm{Id}_{K[\underline{X}]}(G) = \mathrm{Id}_{K[\underline{X}]}(F)$ by Lemma 5.26. □

As an immediate consequence of Theorem 5.48, we obtain an algorithm that decides whether or not a given finite set of polynomials is a Gröbner basis. It is clear that we can test all S-polynomials for reduction to 0 if we have computability of the ground field and decidability of the term order. Termination of the following algorithm is trivial, correctness follows immediately from Theorem 5.48. It is clear that the requirement $0 \notin G$ is not in any way critical when it comes to algorithms involving Gröbner bases; *we will therefore henceforth assume that* 0 *is removed by default from all finite sets of polynomials occurring in algorithms (but not, of course, in theorems).*

Corollary 5.52 *Let G be a finite subset of $K[\underline{X}]$. Suppose the ground field is computable, and the term order on T is decidable. Then the algorithm* GRÖBNERTEST *of Table* 5.3 *decides whether G is a Gröbner basis.* □

A more important consequence of Theorem 5.48 is the following algorithm for the construction of a Gröbner basis from an arbitrary ideal basis. The algorithm of the following theorem that achieves this is also called the **Buchberger algorithm**.

Theorem 5.53 (Buchberger Algorithm) *Let F be a finite subset of $K[\underline{X}]$. Suppose the ground field is computable, and the term order on T is decidable. Then the algorithm* GRÖBNER *of Table* 5.4 *computes a Gröbner basis G in $K[\underline{X}]$ such that $F \subseteq G$ and* $\mathrm{Id}(G) = \mathrm{Id}(F)$.

Proof *Termination:* Assume for a contradiction that the **while**-loop does not terminate. Let $F = G_0 \subset G_1 \subset G_2 \subset \cdots$ be the successive values of G. Considering those runs through the **while**-loop that actually enlarge G, we see that there exists an ascending sequence $\{n_i\}_{i \in \mathbb{N}}$ of natural numbers such that for all $1 \leq i \in \mathbb{N}$, there exist $h_i \in G_{n_i} \setminus G_{n_{i-1}}$ which is in normal

TABLE 5.3. Algorithm GRÖBNERTEST

```
Specification: $v \leftarrow$ GRÖBNERTEST($G$)
               Test whether $G$ is a Gröbner basis
Given: $G =$ a finite subset of $K[\underline{X}]$
Find: $v \in \{\mathbf{true}, \mathbf{false}\}$ such that $v = \mathbf{true}$ iff $G$ is a Gröbner basis
begin
$B \leftarrow \{\, \{g_1, g_2\} \mid g_1, g_2 \in G$ with $g_1 \neq g_2 \,\}$
while $B \neq \emptyset$ do
    select $\{g_1, g_2\}$ from $B$
    $h \leftarrow$ some normal form of spol($g_1, g_2$) modulo $G$
    if $h = 0$ then
      $B \leftarrow B \setminus \{\{g_1, g_2\}\}$
    else return(false)
    end
end
return(true)
end GRÖBNERTEST
```

TABLE 5.4. Algorithm GRÖBNER

```
Specification: $G \leftarrow$ GRÖBNER($F$)
               Construction of a Gröbner basis $G$ of Id($F$)
Given: $F =$ a finite subset of $K[\underline{X}]$
Find: $G =$ a finite subset of $K[\underline{X}]$ such that $G$ is a Gröbner basis in
      $K[\underline{X}]$ with $F \subseteq G$ and Id($G$) = Id($F$)
begin
$G \leftarrow F$
$B \leftarrow \{\, \{g_1, g_2\} \mid g_1, g_2 \in G$ with $g_1 \neq g_2 \,\}$
while $B \neq \emptyset$ do
    select $\{g_1, g_2\}$ from $B$
    $B \leftarrow B \setminus \{\{g_1, g_2\}\}$
    $h \leftarrow$ spol($g_1, g_2$)
    $h_0 \leftarrow$ some normal form of $h$ modulo $G$
    if $h_0 \neq 0$ then
      $B \leftarrow B \cup \{\, \{g, h_0\} \mid g \in G \}$
      $G \leftarrow G \cup \{h_0\}$
    end
end
end GRÖBNER
```

form modulo G_{n_i-1}. Let $t_k = \mathrm{HT}(h_k)$ for all $k \in \mathbb{N}$; then $i < j$ implies that t_i does not divide t_j, for otherwise h_j would be top-reducible modulo $\{h_i\}$ and hence modulo $\{G_{n_j-1}\}$. Since divisibility of terms is a Dickson partial order, this contradicts Proposition 4.42 (ii).

Correctness: We claim that the following are loop invariants of the **while**-loop: G is a finite subset of $K[\underline{X}]$ such that $F \subseteq G \subseteq \mathrm{Id}(F)$, and

$$\mathrm{spol}(g_1, g_2) \xrightarrow[G]{*} 0$$

for all g_1, $g_2 \in G$ such that $\{g_1, g_2\} \notin B$. The first claim follows easily from the fact that a normal form of an S-polynomial of two elements of G is in $\mathrm{Id}(G)$. For the second one, it suffices to note that

$$\mathrm{spol}(g_1, g_2) \xrightarrow[G]{*} h_0 \quad \text{implies} \quad \mathrm{spol}(g_1, g_2) \xrightarrow[G\cup\{h_0\}]{*} 0.$$

Upon termination, we have $B = \emptyset$, and so $\mathrm{spol}(g_1, g_2) \xrightarrow[G]{*} 0$ for all g_1, $g_2 \in G$. It now follows from Theorem 5.48 that G is a Gröbner basis. □

Note that this algorithm is non-deterministic; the resulting Gröbner basis is not uniquely determined by the input F. The pairs that get placed in the set B are often referred to as **critical pairs**. It is clear that the algorithm is potentially rather complex: every newly added reduced S-polynomial enlarges the set B by all its descendants. In the next section, we will see how this combinatorial growth can be controlled to some extent by eliminating unnecessary critical pairs.

Exercise 5.54 Let $K[\underline{X}] = \mathbb{Q}[X, Y, Z]$, $F = \{X+1, Y+1, XY+Z\}$. Compute a Gröbner basis for $\mathrm{Id}(F)$ w.r.t. the lexicographical term order where $Z \ll Y \ll X$. Get a feeling for the complexity of the algorithm by making up your own examples w.r.t. different term orders.

Recall that if R is a ring and I a proper ideal of R , then the *equivalence problem* for R/I is the problem to effectivly decide whether $a \equiv_I b$ (i.e., whether $a - b \in I$) for a, $b \in R$. We can now combine the results obtained thus far in this chapter to prove the following important theorem. (Cf. Proposition 2.39, the remarks following Exercise 2.41, and also the remarks at the end of this section.)

Theorem 5.55 *Let F be a finite subset of $K[\underline{X}]$, let $I = \mathrm{Id}(F)$, and assume that K is computable. Then the following hold:*

(i) The equivalence problem for the ideal I is decidable. In particular, one can decide membership in I.

(ii) The residue class ring $K[\underline{X}]/I$ (which may be formed if I is proper) is computable.

Proof (i) Let $\leq$ be a decidable term order on T. By Theorem 5.53, we may compute a Gröbner basis w.r.t. $\leq$ of I. Now let f, $g \in K[\underline{X}]$. By Theorem 5.21 and Lemma 5.19, we can compute a normal form h of $f - g$ w.r.t. this Gröbner basis. By Theorem 5.35, $h = 0$ if and only if $f - g \in I$. To decide membership in I, simply take $g = 0$.

(ii) By (i), we are in the situation of Example 4.83, so $K[\underline{X}]/I$ is a computable ring. (We are in an even better position since we can represent each residue class *uniquely* by the normal form it contains.) □

Next, we show how the unique reduced Gröbner basis of an ideal $\mathrm{Id}(F)$ can be computed from a given ideal basis $F \subseteq K[\underline{X}]$. One way to achieve this would be to apply the algorithm REDUCTION to the Gröbner basis GRÖBNER(F). It is clear that the output is reduced, and it is possible to show that it is still a Gröbner basis of $\mathrm{Id}(F)$. This procedure, however, is far more costly than necessary in general. The following algorithm first throws away all those polynomials of a given Gröbner basis of $\mathrm{Id}(F)$ whose head terms are multiples of head terms of others and then performs the remaining non-top reductions.

Proposition 5.56 *Let G be a Gröbner basis in $K[\underline{X}]$. Suppose K is computable and the term order on T is decidable. Then the algorithm* REDGRÖBNER *of Table* 5.5 *computes the reduced Gröbner basis of* $\mathrm{Id}(G)$.

TABLE 5.5. Algorithm REDGRÖBNER

```
Specification: $H \leftarrow$ REDGRÖBNER($G$)
                    Construction of the reduced Gröbner basis of Id($G$)
Given: $G =$ a Gröbner basis in $K[\underline{X}]$
Find: $H =$ the reduced Gröbner basis of Id($G$)
begin
$H \leftarrow \emptyset$;   $F \leftarrow G$
while $F \neq \emptyset$ do
        select $f_0$ from $F$
        $F \leftarrow F \setminus \{f_0\}$
        if HT($f$) $\nmid$ HT($f_0$) for all $f \in F$ and
           HT($h$) $\nmid$ HT($f_0$) for all $h \in H$ then
           $H \leftarrow H \cup \{f_0\}$ end
end
$H \leftarrow$ REDUCTION($H$)
end REDGRÖBNER
```

Proof *Termination*: The **while**-loop terminates trivially, and termination of REDUCTION has been proved for an arbitrary finite input set.

Correctness: It is clear that $\mathrm{mult}(\mathrm{HT}(H \cup F))$ is an invariant of the **while**-loop. At the beginning, $H = \emptyset$, while at the end, $F = \emptyset$. It is now immediate from Proposition 5.38 (v) that at the end of the loop, H is still a Gröbner basis of $\mathrm{Id}(G)$. The other loop invariant is that $\mathrm{HT}(f) \nmid \mathrm{HT}(h)$ for all $h \in H$ and $f \in F \cup (H \setminus \{h\})$. We see that when $F = \emptyset$, then $\mathrm{HT}(h_1) \nmid \mathrm{HT}(h_2)$ for all $h_1, h_2 \in H$ with $h_1 \neq h_2$. It is clear that REDUCTION's first reduction step will not be a top reduction in this case. A non-top-reduction does not change any head terms; so none of REDUCTION's reduction steps ever will be a top reduction, and $\mathrm{HT}(H)$ is an invariant of the entire computation. It follows that at the very end, H is still a Gröbner basis of $\mathrm{Id}(G)$, and we already know that the output of REDUCTION is a reduced set. □

Corollary 5.57 *Let F be a finite subset of $K[\underline{X}]$. Suppose the ground field is computable and the term order on T is decidable. Then the composition of the algorithms* GRÖBNER *and* REDGRÖBNER *computes the unique reduced Gröbner basis of* $\mathrm{Id}(F)$. □

Exercise 5.58 Denote by $\leq'$ the linear quasi-order on $\mathcal{P}_{\mathrm{fin}}(K[\underline{X}])$ induced by $\leq$ on $K[\underline{X}]$ according to Exercise 4.70. Let F be a finite subset of $K[\underline{X}]$, and denote by $\mathcal{G}$ the set of all monic Gröbner bases of $\mathrm{Id}(F)$. Show that $G \in \mathcal{G}$ is the unique reduced Gröbner basis of $\mathrm{Id}(F)$ if and only if the following two conditions are satisfied.

(i) G is $\leq'$-minimal in $\mathcal{G}$.

(ii) $|G| \leq |G'|$ for all $G' \in \mathcal{G}$ with $G \leq' G'$ and $G' \leq' G$.

We have already mentioned at the beginning of Section 5.2 that for univariate polynomials, the Gröbner basis problem leads back to gcd computations by means of the Euclidean algorithm. Indeed, if f, $g \in K[X]$, where K is a field, and $h = \gcd(f, g)$, then h is a generator for the ideal generated by f and g (Lemma 1.70). Moreover, $\{h\}$ is trivially a Gröbner basis for this ideal since there are no S-polynomials to be tested (cf. Proposition 5.33), and if we set the head coefficient of h to 1, then it is even a reduced Gröbner basis. Now if we apply the plain algorithm GRÖBNER to the set $F = \{f, g\}$, then F will be enlarged by reduced S-polynomials the last one of which will be h: once h has been added, every member of the ideal, in particular every S-polynomial, reduces to 0 mod h. These considerations remain of course valid for the gcd of more than two univariate polynomials. It is important to note though that in the univariate case, the algorithm GRÖBNER does not provide any improvement—either theoretical or practical—over the Euclidean algorithm: in view of Exercise 5.47 (iii), it is easy to see that the plain algorithm GRÖBNER is nothing but a tremendously blown-up Euclidean algorithm in this case. (Cf. also the discussion at the end of Section 5.1.)

5.4 Standard Representations

In this section, we will provide some more characterizations of Gröbner bases. Although non-algorithmic by nature, these characterizations will be powerful tools for the verification of improved Gröbner basis algorithms. We let $K[\underline{X}]$, T, and $\leq$ be as in the previous section.

Definition 5.59 Let $0 \neq f \in K[\underline{X}]$, P a finite subset of $K[\underline{X}]$. A representation

$$f = \sum_{i=1}^{k} m_i p_i$$

with *monomials* $0 \neq m_i = a_i t_i \in K[\underline{X}]$ and $p_i \in P$ not necessarily pairwise different $(1 \leq i \leq k)$ is called a **standard representation** of f w.r.t. P (and $\leq$) if

$$\max\{\,\mathrm{HT}(m_i p_i) \mid 1 \leq i \leq k\,\} \leq \mathrm{HT}(f).$$

A standard representation w.r.t. P is thus a representation of f as a sum of monomial multiples of elements of P in which there is no cancelation of monomials "protruding beyond $\mathrm{HT}(f)$." Note that we must have $\mathrm{HT}(m_i p_i) = \mathrm{HT}(f)$ for at least one index $1 \leq i \leq k$.

In the literature, standard representations are often defined as representations of the form

$$f = \sum_{p \in P} q_p p$$

with *polynomials* q_p such that $\mathrm{HT}(q_p p) \leq \mathrm{HT}(f)$ for all those $p \in P$ with $q_p \neq 0$. It is clear that using the distributive law, we can convert every sum of polynomial multiples into a sum of monomial multiples and vice versa, and that under this correspondence, the two definitions are equivalent. The monomial version will make some results and proofs easier to visualize. The following lemma is now immediate from Proposition 5.22.

Lemma 5.60 Let P be a finite subset of $K[\underline{X}]$ and $0 \neq f \in K[\underline{X}]$ with $f \xrightarrow[P]{*} 0$. Then f has a standard representation w.r.t. P. □

From the obvious fact that $\mathrm{HT}(m_i p_i) = \mathrm{HT}(f)$ for at least one summand in a standard representation, we immediately obtain the following partial converse to the above lemma.

Lemma 5.61 Let P be a finite subset of $K[\underline{X}]$ and $0 \neq f \in K[\underline{X}]$ such that f has a standard representation w.r.t. P. Then f is top-reducible modulo P. □

We can now combine the last two lemmas to obtain a new characterization of Gröbner bases.

Theorem 5.62 *A finite subset G of $K[\underline{X}]$ with $0 \notin G$ is a Gröbner basis w.r.t. the term order $\leq$ iff every $0 \neq f \in \mathrm{Id}(G)$ has a standard representation w.r.t. G and $\leq$.*

Proof If G is a Gröbner basis, then every $0 \neq f \in \mathrm{Id}(G)$ reduces to zero modulo G and thus has a standard representation w.r.t. G. Conversely, if every $0 \neq f \in \mathrm{Id}(G)$ has a standard representation w.r.t. G, then every such f is top-reducible modulo G, and so G is a Gröbner basis. □

The following exercise demonstrates that the existence of a standard representation does not in general imply reducibility to 0.

Exercise 5.63 Let $P = \{p_1, p_2\} \subseteq \mathbb{Q}[X,Y,Z]$ with $p_1 = XY + 1$, $p_2 = YZ + 1$, $f = XY^2 + X + Y - Z$, $\leq$ the lexicographical term order with $Z \ll Y \ll X$. Show that f has a standard representation w.r.t. P, but not $f \xrightarrow[P]{*} 0$.

We are now going to show that for a finite subset G of $K[\underline{X}]$ to be a Gröbner basis it is sufficient that for all g_1, $g_2 \in G$, $\mathrm{spol}(g_1, g_2)$, unless equal to 0, has a standard representation w.r.t. G. In view of the above exercise, this is a seemingly weaker condition than $\mathrm{spol}(g_1, g_2) \xrightarrow[G]{*} 0$. For the sake of a certain application in the next section, we will actually prove a somewhat more subtle result.

The following terminology will be useful. Let P be a finite subset of $K[\underline{X}]$, $0 \neq f \in K[\underline{X}]$, and $t \in T$. Suppose

$$f = \sum_{i=1}^{k} m_i p_i$$

with monomials $0 \neq m_i \in K[\underline{X}]$ and $p_i \in P$ not necessarily pairwise different $(1 \leq i \leq k)$. Then we say that this is a **t-representation** of f w.r.t. P if

$$\max\{\,\mathrm{HT}(m_i p_i) \mid 1 \leq i \leq k\,\} \leq t.$$

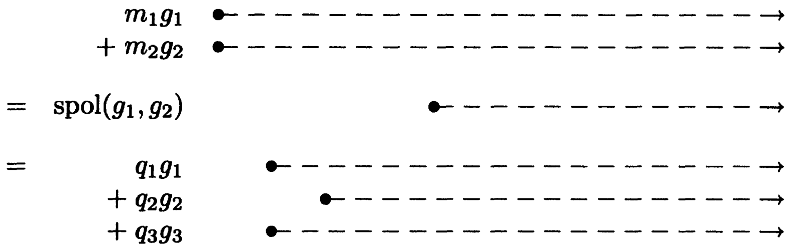
Any $\mathrm{HT}(f)$-representation of f is thus a standard representation of f. In the general case of a t-representation, the term t may be viewed as a measure of how far at most the representation is from being a standard representation. It is clear that—as with standard representations—one can formulate a "polynomial version" of t-representations that is equivalent to the definition above in an obvious sense: a t-representation of f w.r.t. P is then a representation

$$f = \sum_{p \in P} q_p p$$

with polynomials q_p such that $\mathrm{HT}(q_p p) \leq t$ for all those $p \in P$ with $q_p \neq 0$.

Theorem 5.64 *Let G be a finite subset of $K[\underline{X}]$ with $0 \notin G$. Assume that for all g_1, $g_2 \in G$, $\mathrm{spol}(g_1, g_2)$ either equals zero or it has a t-representation w.r.t. G for some $t < \mathrm{lcm}(\mathrm{HT}(g_1), \mathrm{HT}(g_2))$. Then G is a Gröbner basis.*

Before we prove the theorem, we show how one may visualize its hypothesis. In the following picture, the dash lines represent polynomials with their monomials in descending order, so that the dot indicates the head monomial. Here, the S-polynomial of g_1 and g_2 has a t-representation w.r.t. $G = \{g_1, g_2, g_3\}$ as required in the theorem, but this representation is not a standard representation.



Proof of Theorem 5.64 We show that every $0 \neq f \in \mathrm{Id}(G)$ has a standard representation w.r.t. G. Let $0 \neq f \in \mathrm{Id}(G)$. Then f has a representation

$$f = \sum_{g \in G} q_g g$$

with $q_g \in K[\underline{X}]$ for all $g \in G$, which we can of course turn into a representation

$$f = \sum_{i=1}^{k} m_i g_i \tag{$*$}$$

with monomials $0 \neq m_i = a_i t_i \in K[\underline{X}]$ and $g_i \in G$ not necessarily pairwise different $(1 \leq i \leq k)$. We may assume that

$$s = \max\{\, \mathrm{HT}(m_i g_i) \mid 1 \leq i \leq k \,\}$$

is minimal among all such representations of f w.r.t. G. We must prove that $s = \mathrm{HT}(f)$. Assume for a contradiction that $\mathrm{HT}(f) < s$. We will produce an s'-representation of f w.r.t. G for an $s' < s$, contradicting the minimal choice of s. We proceed by induction on the number n_s of indices i with $\mathrm{HT}(m_i g_i) = s$. Since s cancels out, $n_s = 1$ is impossible. Let $n_s = 2$. W.l.o.g., we may assume that $\mathrm{HT}(m_1 g_1) = \mathrm{HT}(m_2 g_2) = s$. This means that

$$s = t_1 \cdot \mathrm{HT}(g_1) = t_2 \cdot \mathrm{HT}(g_2),$$

and so $\mathrm{lcm}(\mathrm{HT}(g_1), \mathrm{HT}(g_2)) \mid s$, say $s = u \cdot \mathrm{lcm}(\mathrm{HT}(g_1), \mathrm{HT}(g_2))$ with $u \in T$. Since $n_s = 2$, we must even have $\mathrm{HM}(m_1 g_1) = -\mathrm{HM}(m_2 g_2)$. It follows that

$$a_1 \cdot \mathrm{HC}(g_1) = -a_2 \cdot \mathrm{HC}(g_2).$$

Now if we set $a = a_1/\mathrm{HC}(g_2) = -a_2/\mathrm{HC}(g_1)$, then it is not hard to prove that

$$m_1 g_1 + m_2 g_2 = au \cdot \mathrm{spol}(g_1, g_2).$$

By assumption, $\operatorname{spol}(g_1, g_2) = 0$, or it has a t-representation

$$\operatorname{spol}(g_1, g_2) = \sum_{i=1}^{k'} m'_i g'_i \qquad (g'_i \in G)$$

for some $t < \operatorname{lcm}(\operatorname{HT}(g_1), \operatorname{HT}(g_2))$. Substituting for $m_1 g_1 + m_2 g_2$ in $(*)$, we obtain a representation

$$f = \sum_{i=3}^{k} m_i g_i + au \sum_{i=1}^{k'} m'_i g'_i, \tag{$**$}$$

where the second sum is not present if the S-polynomial is zero. The maximum of the head terms occurring in the first sum is $< s$ by our assumption $n_s = 2$; the maximum of the head terms in the second sum (if any) is less than or equal to ut, and

$$ut < u \cdot \operatorname{lcm}\bigl(\operatorname{HT}(g_1), \operatorname{HT}(g_2)\bigr) = s.$$

Together, we see that the maximum s' of the head terms in the representation $(**)$ satisfies $s' < s$, which means that $(**)$ is the s'-representation that we were looking for.

Now let $n_s > 2$, and assume w.l.o.g. that $\operatorname{HT}(m_i g_i) = s$ for $i = 1, 2$. Then we write

$$\begin{aligned} f &= \sum_{i=1}^{k} m_i g_i \\ &= m_1 g_1 - \frac{\operatorname{HC}(m_1 g_1)}{\operatorname{HC}(m_2 g_2)} m_2 g_2 + \left(\frac{\operatorname{HC}(m_1 g_1)}{\operatorname{HC}(m_2 g_2)} + 1 \right) m_2 g_2 + \sum_{i=3}^{k} m_i g_i. \end{aligned}$$

The induction hypothesis clearly applies to the first two summands. In the remaining $k-1$ summands, the term s occurs at most $n_s - 1$ times: there are exactly $n_s - 2$ occurrences in the last $k-2$ summands on the right, and the third summand contributes one more occurrence unless it happens to vanish. It follows that the induction hypothesis applies here too, and the sum of the two representations thus obtained is clearly an s'-representation of f for some $s' < s$. □

The following corollary is obvious from the fact that $\operatorname{HT}(\operatorname{spol}(g_1, g_2)) < \operatorname{lcm}(\operatorname{HT}(g_1), \operatorname{HT}(g_2))$.

Corollary 5.65 *Let G be a finite subset of $K[\underline{X}]$ with $0 \notin G$, and assume that for all g_1, $g_2 \in G$,* $\operatorname{spol}(g_1, g_2)$ *equals zero or has a standard representation w.r.t. G. Then G is a Gröbner basis.* □

5.5 Improved Gröbner Basis Algorithms

In this section, we show how the combinatorial complexity of the algorithm GRÖBNER can be reduced by testing out certain S-polynomials which need not be considered. Let $K[\underline{X}]$, T, and $\leq$ be as in the previous section.

We call s, $t \in T$ **disjoint** if s and t have no variable in common; in other words $\gcd(s,t) = 1$ in the monoid T. It is easy to see that this is equivalent to $\operatorname{lcm}(s,t) = st$.

Lemma 5.66 (BUCHBERGER'S FIRST CRITERION) Let $f, g \in K[\underline{X}]$ with disjoint head terms. Then $\operatorname{spol}(f,g) \xrightarrow[\{f,g\}]{*} 0$.

Proof Assume that

$$f = \sum_{i=1}^{k} a_i s_i \quad \text{and} \quad g = \sum_{j=1}^{l} b_j t_j,$$

where a_i, $b_j \in K$ with $a_i, b_j \neq 0$, and $s_i, t_j \in T$ for $1 \leq i \leq k$ and $1 \leq j \leq l$. We may assume that $s_1 > \cdots > s_k$ and $t_1 > \cdots > t_l$. Since $\gcd(s_1, t_1) = 1$, we must have $\operatorname{lcm}(s_1, t_1) = s_1 t_1$ and thus

$$\operatorname{spol}(f,g) = b_1 t_1 f - a_1 s_1 g = b_1 t_1 \sum_{i=2}^{k} a_i s_i - a_1 s_1 \sum_{j=2}^{l} b_j t_j.$$

We claim that the two sums have no terms in common. Indeed, if $s_i t_1 = t_j s_1$ for some $2 \leq i \leq k$ and $2 \leq j \leq l$, then $s_i t_1$, being a common multiple of s_1 and t_1, is divided by $\operatorname{lcm}(s_1, t_1) = s_1 t_1$. It follows that $s_1 t_1 \leq s_i t_1$ and thus $s_1 \leq s_i$, a contradiction. Furthermore, each term in the second sum is a multiple of $\operatorname{HT}(f)$. If we now successively add

$$b_l t_l f,\ b_{l-1} t_{l-1} f,\ \ldots,\ b_2 t_2 f$$

to $\operatorname{spol}(f,g)$, then each of these additions is a reduction step: after adding $b_l t_l f + \cdots + b_j t_j f$ $(2 < j \leq l)$, all terms $t_{j-1} s_1, \ldots, t_2 s_1$ will still be present because each of them is strictly greater than anything in $b_l t_l f + \cdots + b_j t_j f$. Algebraically speaking, reducing by means of f amounts to substituting $-\sum_{i=2}^{k} a_i s_i$ for $a_1 s_1$. We see that

$$\begin{aligned} \operatorname{spol}(f,g) &\xrightarrow[f]{*} b_1 t_1 \sum_{i=2}^{k} a_i s_i + \sum_{j=2}^{l} b_j t_j \sum_{i=2}^{k} a_i s_i \\ &= g \sum_{i=2}^{k} a_i s_i \\ &\xrightarrow[g]{*} 0. \quad \square \end{aligned}$$

Exercise 5.67 Use the results of the previous section on standard representations to give an alternate proof of Buchberger's first criterion.

The proof of the following theorem is now immediate from Theorem 5.48 and the lemma above.

Theorem 5.68 *Let G be a finite subset of $K[\underline{X}]$ with $0 \notin G$. Then the following are equivalent:*

(i) G is a Gröbner basis.

(ii) For all $g_1, g_2 \in G$ with non-disjoint head terms, $\operatorname{spol}(g_1, g_2) \xrightarrow[G]{*} 0$.

(iii) Whenever $g_1, g_2 \in G$ with non-disjoint head terms and $h \in K[\underline{X}]$ such that h is a normal form of $\operatorname{spol}(g_1, g_2)$ modulo G, then $h = 0$. □

Accordingly, the algorithms GRÖBNERTEST and GRÖBNER of Corollary 5.52 and Theorem 5.53 can be improved by placing only those pairs $\{g_1, g_2\}$ in the set B that have non-disjoint head terms.

We are now in a position to see how the computation of Gröbner bases generalizes the Gaussian elimination algorithm of linear algebra. Let F be a finite subset of $K[\underline{X}]$ where each $f \in F$ has total degree at most 1. We have already remarked at the end of Section 5.1 that the algorithm REDUCTION applied to F is precisely the Gaussian elimination algorithm: it will subtract constant multiples of polynomials from others until no two polynomials have the same variable as their head term. By Buchberger's first criterion, the result is already a Gröbner basis.

Buchberger's first criterion is a local criterion. It allows us to skip the testing of certain S-polynomials because we know beforehand that they will reduce to zero. The second criterion, which we discuss next, is considerably deeper. It says that certain S-polynomials can be deleted despite the fact that they might not be reducible to zero at the time when we drop them. The proof will make use of the concept of t-representations as discussed in the previous section.

Exercise 5.69 Let s, t, $u \in T$. Show that the following are equivalent:

(i) $t \mid \operatorname{lcm}(s, u)$

(ii) $\operatorname{lcm}(s, t) \mid \operatorname{lcm}(s, u)$

(iii) $\operatorname{lcm}(t, u) \mid \operatorname{lcm}(s, u)$

The next proposition, together with Theorem 5.64, is the theoretical basis for the second improvement of the algorithm GRÖBNER.

Proposition 5.70 (Buchberger's Second Criterion) *Let F be a finite subset of $K[\underline{X}]$ and g_1, p, $g_2 \in K[\underline{X}]$ such that the following hold:*

(i) $\operatorname{HT}(p) \mid \operatorname{lcm}(\operatorname{HT}(g_1), \operatorname{HT}(g_2))$, and

(ii) $\operatorname{spol}(g_i,p)$ *has a* t_i*-representation w.r.t.* F *with*

$$t_i < \operatorname{lcm}\big(\mathrm{HT}(g_i),\mathrm{HT}(p)\big) \quad \text{for} \quad i=1,2.$$

Then the S-polynomial $\operatorname{spol}(g_1,g_2)$ *has a* t*-representation w.r.t.* F *for some* $t < \operatorname{lcm}(\mathrm{HT}(g_1),\mathrm{HT}(g_2))$.

Proof By assumption (ii), there are representations

$$\operatorname{spol}(g_1,p) = \sum_{i=1}^{k_1} m_{1i} f_{1i}$$

with $\max\{\,\mathrm{HT}(m_{1i}f_{1i}) \mid 1 \le i \le k_1\,\} < \operatorname{lcm}(\mathrm{HT}(g_1),\mathrm{HT}(p))$, and

$$\operatorname{spol}(p,g_2) = \sum_{i=1}^{k_2} m_{2i} f_{2i}$$

with $\max\{\,\mathrm{HT}(m_{2i}f_{2i}) \mid 1 \le i \le k_2\,\} < \operatorname{lcm}(\mathrm{HT}(p),\mathrm{HT}(g_2))$ as sums of monomial multiples of elements of F. By Exercise 5.69, there exist s_1, $s_2 \in T$ with

$$\begin{aligned} s_1 \cdot \operatorname{lcm}\big(\mathrm{HT}(g_1),\mathrm{HT}(p)\big) &= \operatorname{lcm}\big(\mathrm{HT}(g_1),\mathrm{HT}(g_2)\big),\\ s_2 \cdot \operatorname{lcm}\big(\mathrm{HT}(p),\mathrm{HT}(g_2)\big) &= \operatorname{lcm}\big(\mathrm{HT}(g_1),\mathrm{HT}(g_2)\big). \end{aligned}$$

We let $a = \mathrm{HC}(g_1)$, $b = \mathrm{HC}(p)$, $c = \mathrm{HC}(g_2)$, and u_1, v_1, u_2, $v_2 \in T$ such that

$$\begin{aligned} \operatorname{lcm}\big(\mathrm{HT}(g_1),\mathrm{HT}(p)\big) &= u_1 \cdot \mathrm{HT}(g_1) = v_1 \cdot \mathrm{HT}(p),\\ \operatorname{lcm}\big(\mathrm{HT}(p),\mathrm{HT}(g_2)\big) &= u_2 \cdot \mathrm{HT}(p) = v_2 \cdot \mathrm{HT}(g_2). \end{aligned}$$

It is easy to see that $s_1 v_1 = s_2 u_2$, and we obtain

$$\begin{aligned} & cs_1 \cdot \operatorname{spol}(g_1,p) + as_2 \cdot \operatorname{spol}(p,g_2)\\ =\ & cs_1(bu_1g_1 - av_1p) + as_2(cu_2p - bv_2g_2)\\ =\ & cbs_1u_1g_1 - abs_2v_2g_2\\ =\ & b \cdot \operatorname{spol}(g_1,g_2). \end{aligned}$$

Substituting the representations of the first two S-polynomials into the equation yields

$$\operatorname{spol}(g_1,g_2) = \frac{1}{b}\Big(cs_1 \sum_{i=1}^{k_1} m_{1i}f_{1i} + as_2 \sum_{i=1}^{k_2} m_{2i}f_{2i}\Big). \qquad (*)$$

By the choice of these representations, we may conclude that

$$\begin{aligned} s_1 \cdot \mathrm{HT}(m_{1i}f_{1i}) &< s_1 \cdot \operatorname{lcm}\big(\mathrm{HT}(g_1),\mathrm{HT}(p)\big)\\ &= \operatorname{lcm}\big(\mathrm{HT}(g_1),\mathrm{HT}(g_2)\big), \end{aligned}$$

for $1 \leq i \leq k_1$, and similarly,

$$\begin{aligned} s_2 \cdot \mathrm{HT}(m_{2i} f_{2i}) &< s_2 \cdot \mathrm{lcm}(\mathrm{HT}(p), \mathrm{HT}(g_2)) \\ &= \mathrm{lcm}(\mathrm{HT}(g_1), \mathrm{HT}(g_2)) \end{aligned}$$

for $1 \leq i \leq k_2$. Now if we let t be the maximum of all $s_1 \cdot \mathrm{HT}(m_{1i} f_{1i})$ for $1 \leq i \leq k_1$ and $s_2 \cdot \mathrm{HT}(m_{2i} f_{2i})$ for $1 \leq i \leq k_2$, then we see that $(*)$ is a t-representation of $\mathrm{spol}(g_1, g_2)$, and $t < \mathrm{lcm}(\mathrm{HT}(g_1), \mathrm{HT}(g_2))$. □

The above version of Buchberger's second criterion is rather convenient for correctness proofs of improved algorithms. A more general, theoretical version of the criterion will be given at the end of Section 6.1.

There are at least two ways of incorporating the second criterion into the algorithm GRÖBNER. We present first the one that is most easily exhibited. The algorithm GRÖBNERNEW1 of the next theorem is based on the algorithm GRÖBNER, with the following modifications. The algorithm keeps track of which critical pairs have already been selected from the list during an execution of the **while**-loop. If two polynomials g_1 and g_2 in the set G have disjoint head terms, then, in view of Buchberger's first criterion, the algorithm does not even bother to put $\{g_1, g_2\}$ on the critical pair list, but it does mark that pair as having been treated. The selection of elements from the critical pair list during executions of the **while**-loop is governed by the strategy to prefer those pairs where the lcm of the head terms is minimal w.r.t. the term order. This strategy is commonly called the **normal strategy**. (See also the comments following the proof of the next theorem.) When a critical pair $\{g_1, g_2\}$ is selected from the list, the algorithm first checks if it can find $p \in G$ such that

$$\mathrm{HT}(p) \mid \mathrm{lcm}(\mathrm{HT}(g_1), \mathrm{HT}(g_2))$$

and the two pairs $\{g_1, p\}$ and $\{p, g_2\}$ are marked "treated." If this is the case, then nothing is done about $\{g_1, g_2\}$. Else, the pair $\{g_1, g_2\}$ is treated as in the algorithm GRÖBNER.

In the algorithm GRÖBNERNEW1 of the theorem below, the marking of critical pairs already treated is achieved by creating a matrix that contains a Boolean entry for each critical pair that surfaces. This describes actual implementations of this version of the Buchberger algorithm fairly accurately. We mention that from a theoretical point of view, this matrix is superfluous: it is easy to see that a critical pair is marked "TRUE," i.e., "already treated," if and only if it is not on the critical pair list B.

Theorem 5.71 *Let F be a finite subset of $K[\underline{X}]$. Suppose the ground field is computable and the term order on T is decidable. Then the algorithm* GRÖBNERNEW1 *of Table* 5.6 *computes a Gröbner basis G in $K[\underline{X}]$ such that* $\mathrm{Id}(G) = \mathrm{Id}(F)$. *The algorithm eliminates superfluous S-polynomials according to Buchberger's criteria.*

TABLE 5.6. Algorithm GRÖBNERNEW1

```
Specification: $G \leftarrow$ GRÖBNERNEW1($F$)
                    Construction of a Gröbner basis $G$ for Id(F)
Given: $F$ = a finite subset of $K[\underline{X}]$
Find: $G$ = a finite subset of $K[\underline{X}]$ such that $G$ is a
          Gröbner basis in $K[\underline{X}]$ with $\mathrm{Id}(G) = \mathrm{Id}(F)$
begin
$G \leftarrow$ REDUCTION($F$)
$B \leftarrow \{\{g_1, g_2\} \mid g_1, g_2 \in G$ with non-disjoint head terms, $g_1 \neq g_2\}$
create a matrix $M$ with a Boolean entry $M(g_1, g_2)$ for
each $\{g_1, g_2\}$, where $g_1, g_2 \in G$ with $g_1 \neq g_2$
for all $\{g_1, g_2\}$ with $g_1, g_2 \in G$ and $g_1 \neq g_2$ do
          if $\{g_1, g_2\} \in B$ then $M(g_1, g_2) \leftarrow$ false
          else $M(g_1, g_2) \leftarrow$ true end
end
while $B \neq \emptyset$ do
          select $\{g_1, g_2\}$ from $B$ with $\mathrm{lcm}(\mathrm{HT}(g_1), \mathrm{HT}(g_2))$
          minimal among all pairs in $B$
          $B \leftarrow B \setminus \{\{g_1, g_2\}\}$
          $M(g_1, g_2) \leftarrow$ true
          if there does not exist $p \in G$ with:
             $\mathrm{HT}(p) \mid \mathrm{lcm}(\mathrm{HT}(g_1), \mathrm{HT}(g_2))$ and
             $M(g_1, p) = M(p, g_2) =$ true then
             $h \leftarrow \mathrm{spol}(g_1, g_2)$
             $h_0 \leftarrow$ some normal form of $h$ modulo $G$
             if $h_0 \neq 0$ then
                for all $g \in G$ do
                           enlarge $M$ by an entry for $\{h_0, g\}$
                           if $\mathrm{HT}(g), \mathrm{HT}(h_0)$ disjoint then
                              $M(g, h_0) \leftarrow$ true
                           else
                              $B \leftarrow B \cup \{\{g, h_0\}\}$
                              $M(g, h_0) \leftarrow$ false
                           end
                end
                $G \leftarrow G \cup \{h_0\}$
             end
          end
end
end GRÖBNERNEW1
```

Proof The algorithm terminates since an infinite loop would be an infinite loop of the algorithm GRÖBNER. The application of REDUCTION to F at the beginning just produces a possibly different basis of the same ideal and may therefore be disregarded in the correctness proof. For correctness, we first note that as with the algorithm GRÖBNER, an invariant of the **while**-loop is the fact that G is finite with

$$F \subseteq G \subseteq \mathrm{Id}(F),$$

and thus the output G_{out} is a finite basis of the ideal $\mathrm{Id}(F)$. To see that it is in fact a Gröbner basis, we verify the criterion of Theorem 5.64. Assume for a contradiction that there exists a pair $\{g_1, g_2\} \in G_{\mathrm{out}}$ such that $\mathrm{spol}(g_1, g_2)$ does not have a t-representation w.r.t. G_{out} for any $t < \mathrm{lcm}(\mathrm{HT}(g_1), \mathrm{HT}(g_2))$. If g_1 and g_2 had disjoint head terms, then $\mathrm{spol}(g_1, g_2)$ would reduce to 0 modulo G_{out} by Lemma 5.66 and thus even have a standard representation w.r.t. G_{out} by Lemma 5.60. We conclude that $\{g_1, g_2\}$ was placed on the critical pair list B at some point during computation. We may assume w.l.o.g. that among all such problem pairs, $\{g_1, g_2\}$ was the first one to be selected from B during an execution of the **while**-loop. The S-polynomial $\mathrm{spol}(g_1, g_2)$ does not reduce to 0 modulo G_{out} since otherwise it would even have a standard representation w.r.t. G_{out}. It follows that $\{g_1, g_2\}$ must have been tested out by the **if**-condition following the assignment $M(g_1, g_2) \longleftarrow \mathbf{true}$. This means that there exists $p \in G_{\mathrm{out}}$ with

$$\mathrm{HT}(p) \mid \mathrm{lcm}(\mathrm{HT}(g_1), \mathrm{HT}(g_2)), \quad \text{and} \quad M(g_1, p) = M(p, g_2) = \mathbf{true}$$

at that point of the computation. We may now conclude that for $i = 1$, 2, either g_i and p have disjoint head terms, or the pair $\{g_i, p\}$ made the list B and was selected from it at an earlier stage. By Lemma 5.66 and our choice of the pair $\{g_1, g_2\}$, $\mathrm{spol}(g_i, p)$ has a t_i-representation for some $t_i < \mathrm{lcm}(\mathrm{HT}(g_i), \mathrm{HT}(p))$ for $i = 1$, 2, and Proposition 5.70 provides the desired contradiction. □

Exercise 5.72 We mentioned at the end of Section 5.3 that the plain algorithm GRÖBNER, when applied to univariate polynomials, is an unnecessarily blown-up version of the Euclidean algorithm. Show the following:

(i) When GRÖBNERNEW1 is applied to univariate polynomials f and g, it will perform the exact same back-and-forth divisions as the Euclidean algorithm.

(ii) When GRÖBNERNEW1 is applied to a finite set of more than two univariate polynomials, it will act like a recursive application of the Euclidean algorithm in the sense of Lemma 1.79, proceeding by ascending degrees of the input polynomials.

Instead of applying REDUCTION at the beginning GRÖBNERNEW1, one may prefer to do top reductions only (or nothing at all, for that matter), but this decision does not seem to influence the computation in a major way in general.

The normal strategy for the selection of critical pairs from B which GRÖBNERNEW1 employs has turned out to be a good one in practice, but the problem of finding the optimal strategy is not really considered settled. Another one which is often preferred will be described at the end of Section 10.3. A disadvantage of GRÖBNERNEW1 is that one is forced to adhere to the normal strategy in the following sense. The correctness of the algorithm is independent of the strategy, simply because the strategy was never mentioned in the correctness proof. If one switches to a different strategy, however, the algorithm may miss out on instances of Buchberger's second criterion. To see this, assume that at some point during computation, the pair $\{g_1, g_2\}$ is selected from B, and there is, at this time, $p \in G$ with

$$\mathrm{HT}(p) \mid \mathrm{lcm}\big(\mathrm{HT}(g_1), \mathrm{HT}(g_2)\big).$$

Then Buchberger's second criterion tells us that we should treat the pairs (g_1, p) and (p, g_2) and test out (g_1, g_2). If g_i and p have disjoint head terms, then we trivially have $M(g_i, p) = \mathbf{true}$ for $i = 1$, 2 and so $\{g_1, g_2\}$ is tested out. If $\mathrm{lcm}(\mathrm{HT}(g_i), \mathrm{HT}(p))$ *properly* divides $\mathrm{lcm}(\mathrm{HT}(g_1), \mathrm{HT}(g_2))$, then the same is true as one easily sees from the way the normal strategy works. Finally, assume that

$$\mathrm{lcm}\big(\mathrm{HT}(g_1), \mathrm{HT}(p)\big) = \mathrm{lcm}\big(\mathrm{HT}(g_1), \mathrm{HT}(g_2)\big).$$

Then $\mathrm{HT}(g_2) \mid \mathrm{lcm}(\mathrm{HT}(g_1), \mathrm{HT}(p))$, and we may just as well treat the pair $\{g_1, g_2\}$ first because $\{g_1, p\}$ will test out by means of g_2. The same is true for the case

$$\mathrm{lcm}\big(\mathrm{HT}(p), \mathrm{HT}(g_2)\big) = \mathrm{lcm}\big(\mathrm{HT}(g_1), \mathrm{HT}(g_2)\big).$$

It should be clear now how the algorithm misses instances of the second criterion when a different strategy is used; it thus does not allow a fair comparison of different strategies. This problem is overcome by the second implementation GRÖBNERNEW2 of the improved Buchberger algorithm which we discuss next.

The main difference between GRÖBNERNEW1 and GRÖBNERNEW2 is as follows. GRÖBNERNEW1 waits until a critical pair is up for treatment before it makes the decision whether or not that pair may be deleted on the basis of Buchberger's second criterion. GRÖBNERNEW2, by contrast, tries to eliminate critical pairs as early as possible. Moreover, it even deletes certain polynomials from the set G en route, knowing that every critical pair that they will henceforth occur in is superfluous, and that these polynomials themselves will be superfluous in the output set. The mechanisms that achieve these deletions of critical pairs and polynomials are placed in the main **while**-loop at the point where a new non-zero normal form h of an S-polynomial has been found and the sets B and G are about to be updated. We formulate this process as a subalgorithm named UPDATE (Table 5.7) to be called by GRÖBNERNEW2. The reason for this is that the exact same procedure is used at the beginning of GRÖBNERNEW2 for the initialization of B and G: the polynomials from the input set will be treated exactly as if they were new polynomials found by a run through the **while**-loop. (Here, it is of course assumed that K is computable and the term order $\leq$ is decidable.)

Before we give the algorithm GRÖBNERNEW2 and prove its correctness, we will, for a better understanding, discuss *on an intuitive level* why the eliminations performed by UPDATE are appropriate. If g_1, h, $g_2 \in K[\underline{X}]$ are such that the equivalent conditions

$$\begin{aligned}&\mathrm{HT}(h) \mid \mathrm{lcm}\big(\mathrm{HT}(g_1), \mathrm{HT}(g_2)\big),\\ &\mathrm{lcm}\big(\mathrm{HT}(g_1), \mathrm{HT}(h)\big) \mid \mathrm{lcm}\big(\mathrm{HT}(g_1), \mathrm{HT}(g_2)\big), \quad \text{and}\\ &\mathrm{lcm}\big(\mathrm{HT}(h), \mathrm{HT}(g_2)\big) \mid \mathrm{lcm}\big(\mathrm{HT}(g_1), \mathrm{HT}(g_2)\big)\end{aligned}$$

are satisfied, then we will refer to (g_1, h, g_2) as a **Buchberger triple**. Proposition 5.70 together with Theorem 5.64 tells us that if a Buchberger triple (g_1, h, g_2) shows up in a Buchberger algorithm and the pairs $\{g_1, h\}$ and $\{h, g_2\}$ have been taken care of, then the pair $\{g_1, g_2\}$ need not be treated. Now consider the special case where two of the three lcm's of head terms involved are equal, say

$$\mathrm{lcm}\big(\mathrm{HT}(g_1), \mathrm{HT}(h)\big) = \mathrm{lcm}\big(\mathrm{HT}(g_1), \mathrm{HT}(g_2)\big).$$

Then both (g_1, h, g_2) and (h, g_2, g_1) are Buchberger triples, and we can choose either one of $\{g_1, g_2\}$ and $\{h, g_1\}$ for deletion. What makes the implementation of the criterion difficult is the danger of erroneously deleting both of these, first $\{g_1, g_2\}$ on account of h and then $\{h, g_1\}$ on account of g_2. The version GRÖBNERNEW1 solves this "two-out-of-three problem" by convincing itself explicitly, before deleting a pair, that the two other ones involved have been dealt with otherwise, either by reduction of the S-polynomial or by another instance of the criterion.

Now let us look at the first **while**-loop of UPDATE. The loop looks at each element $\{h, g_1\}$ on the list of new critical pairs and tries to find another one $\{h, g_2\}$ still on the list such that

$$\mathrm{lcm}\big(\mathrm{HT}(h), \mathrm{HT}(g_2)\big) \mid \mathrm{lcm}\big(\mathrm{HT}(h), \mathrm{HT}(g_1)\big).$$

If this is the case, then $\{h, g_1\}$ is deleted. This is possible because here, (h, g_2, g_1) is a Buchberger triple. Moreover, the third pair $\{g_1, g_2\}$ is not present on this list at all, and therefore the two-out-of-three trap is disabled at this point. This **while**-loop does, however, keep all pairs (h, g) where h and g have disjoint head terms. These are all thrown out by the next **while**-loop on the basis of Buchberger's first criterion. The reason for keeping all

TABLE 5.7. Subalgorithm UPDATE

```
Specification: $(G_{\text{new}}, B_{\text{new}}) \leftarrow \text{UPDATE}(G_{\text{old}}, B_{\text{old}}, h)$
               Update of critical pair list and ideal basis as
               required by GRÖBNERNEW2
Given: a finite subset $G_{\text{old}}$ of $K[\underline{X}]$, a finite set $B_{\text{old}}$ of
       pairs of elements of $K[\underline{X}]$, and $0 \neq h \in K[\underline{X}]$
Find: updates $G_{\text{new}}$ of $G_{\text{old}}$ and $B_{\text{new}}$ of $B_{\text{old}}$
begin
$C \leftarrow \{\{h,g\} \mid g \in G_{\text{old}}\}$;   $D \leftarrow \emptyset$
while $C \neq \emptyset$ do
        select $\{h,g_1\}$ from $C$;   $C \leftarrow C \setminus \{\{h,g_1\}\}$
        if HT($h$) and HT($g_1$) are disjoint or
          (
          $\text{lcm}(\text{HT}(h),\text{HT}(g_2)) \nmid \text{lcm}(\text{HT}(h),\text{HT}(g_1))$ for all $\{h,g_2\} \in C$
          and
          $\text{lcm}(\text{HT}(h),\text{HT}(g_2)) \nmid \text{lcm}(\text{HT}(h),\text{HT}(g_1))$ for all $\{h,g_2\} \in D$
          )
          then $D \leftarrow D \cup \{\{h,g_1\}\}$ end
end
$E \leftarrow \emptyset$
while $D \neq \emptyset$ do
        select $\{h,g\}$ from $D$;   $D \leftarrow D \setminus \{\{h,g\}\}$
        if HT($h$) and HT($g$) are not disjoint then
          $E \leftarrow E \cup \{\{h,g\}\}$ end
end
$B_{\text{new}} \leftarrow \emptyset$
while $B_{\text{old}} \neq \emptyset$ do
        select $\{g_1,g_2\}$ from $B_{\text{old}}$;   $B_{\text{old}} \leftarrow B_{\text{old}} \setminus \{\{g_1,g_2\}\}$
        if  $\text{HT}(h) \nmid \text{lcm}(\text{HT}(g_1),\text{HT}(g_2))$   or
            $\text{lcm}(\text{HT}(g_1),\text{HT}(h)) = \text{lcm}(\text{HT}(g_1),\text{HT}(g_2))$   or
            $\text{lcm}(\text{HT}(h),\text{HT}(g_2)) = \text{lcm}(\text{HT}(g_1),\text{HT}(g_2))$   then
            $B_{\text{new}} \leftarrow B_{\text{new}} \cup \{\{g_1,g_2\}\}$ end
end
$B_{\text{new}} \leftarrow B_{\text{new}} \cup E$;   $G_{\text{new}} \leftarrow \emptyset$
while $G_{\text{old}} \neq \emptyset$ do
        select $g$ from $G_{\text{old}}$;   $G_{\text{old}} \leftarrow G_{\text{old}} \setminus \{g\}$
        if $\text{HT}(h) \nmid \text{HT}(g)$ then $G_{\text{new}} \leftarrow G_{\text{new}} \cup \{g\}$ end
end
$G_{\text{new}} \leftarrow G_{\text{new}} \cup \{h\}$
return($G_{\text{new}}, B_{\text{new}}$)
end UPDATE
```

of them during the first **while**-loop is this: If two or more pairs in C have the same lcm of head terms, so that there is a choice as to which one(s) should be deleted, then it is advantageous to try and keep one where the head terms are disjoint; that way, one eventually gets rid of all of them.

The third **while**-loop eliminates from the list B_{old} of old pairs those pairs $\{g_1, g_2\}$ where (g_1, h, g_2) is a Buchberger triple. Here, UPDATE protects itself from the two-out-of-three error by dropping only those triples (g_1, h, g_2) where the two divisibilities

$$\begin{aligned}&\operatorname{lcm}(\mathrm{HT}(g_1), \mathrm{HT}(h)) \mid \operatorname{lcm}(\mathrm{HT}(g_1), \mathrm{HT}(g_2)) \quad \text{and}\\ &\operatorname{lcm}(\mathrm{HT}(h), \mathrm{HT}(g_2)) \mid \operatorname{lcm}(\mathrm{HT}(g_1), \mathrm{HT}(g_2))\end{aligned}$$

are proper. Using Exercise 5.69, one easily proves that then there cannot have been a divisibility between

$$\operatorname{lcm}(\mathrm{HT}(g_1), \mathrm{HT}(h)) \quad \text{and} \quad \operatorname{lcm}(\mathrm{HT}(h), \mathrm{HT}(g_2)),$$

and so none of $\{h, g_1\}$ and $\{h, g_2\}$ was tested out by means of the other during the first **while**-loop.

Next, the updated lists of the old and the new pairs are united and assigned to the output B_{new}. Finally, UPDATE eliminates from G_{old} all those polynomials g whose head term is a multiple of the head term of h. This is legitimate for two reasons. Firstly, $\mathrm{HT}(h) \mid \mathrm{HT}(g)$ implies

$$\mathrm{HT}(h) \mid \operatorname{lcm}(\mathrm{HT}(g), \mathrm{HT}(f))$$

for arbitrary $f \in K[\underline{X}]$, and so (g, h, f) is a Buchberger triple for any future arrival f in G. Moreover, g will not be missed at the end because in a Gröbner basis, polynomials whose head terms are multiples of others are superfluous.

Let us emphasize again that the above informal discussion of UPDATE is very far from being a correctness proof of anything.

Theorem 5.73 *Let F be a finite subset of $K[\underline{X}]$. Suppose K is computable and the term order on T is decidable. Then the algorithm* GRÖBNERNEW2 *of Table* 5.8 *computes a Gröbner basis G in $K[\underline{X}]$ such that* $\mathrm{Id}(G) = \mathrm{Id}(F)$. *The algorithm eliminates superfluous critical pairs according to Buchberger's criteria.*

Proof *Termination*: The first **while**-loop terminates trivially. An infinite number of repetitions of the second **while**-loop would, just as with the algorithm GRÖBNER, give rise to an infinite sequence $\{t_k\}_{k \in \mathbb{N}}$ of terms with $t_i \nmid t_j$ for all $i < j$, contradicting Proposition 4.42 (ii) since divisibility of terms is a Dickson partial order.

Correctness: We must prove that G_{out} is a Gröbner basis of $\mathrm{Id}(F)$, where G_{out} is the last value of G. Let us denote by G_{all} the union of all values that G ever held during computation. We claim that it suffices to prove that

TABLE 5.8. Algorithm GRÖBNERNEW2

```
Specification: $G \leftarrow$ GRÖBNERNEW2($F$)
                    Construction of a Gröbner basis $G$ of Id($F$)
Given: $F$ = a finite subset of $K[\underline{X}]$
Find: $G$ = a finite subset of $K[\underline{X}]$ such that $G$ is a Gröbner basis in
      $K[\underline{X}]$ with Id($G$) = Id($F$)
begin
$G \leftarrow \emptyset$;  $B \leftarrow \emptyset$
while $F \neq \emptyset$ do
        select $f$ from $F$
        $F \leftarrow F \setminus \{f\}$
        $(G,B) \leftarrow$ UPDATE($G,B,f$)
end
while $B \neq \emptyset$ do
        select $\{g_1, g_2\}$ from $B$
        $B \leftarrow B \setminus \{\{g_1, g_2\}\}$
        $h \leftarrow$ some normal form of spol($g_1, g_2$) modulo $G$
        if $h \neq 0$ then $(G,B) \leftarrow$ UPDATE($G,B,h$) end
end
end GRÖBNERNEW2
```

(i) $G_{\text{out}} \subseteq \text{Id}(F)$,

(ii) $\text{mult}\big(\text{HT}(G_{\text{out}})\big) = \text{mult}\big(\text{HT}(G_{\text{all}})\big)$, and

(iii) G_{all} is a Gröbner basis of $\text{Id}(F)$.

Indeed, from (ii) and (iii) together with Proposition 5.38, it follows that

$$\text{mult}\big(\text{HT}(G_{\text{out}})\big) = \text{mult}\big(\text{HT}(G_{\text{all}})\big) = \text{mult}\Big(\text{HT}\big(\text{Id}(F)\big)\Big),$$

and this together with (i) implies that G_{out} is a Gröbner basis of $\text{Id}(F)$ again by Proposition 5.38. We are thus left with the task of proving statements (i)–(iii).

For (ii), we first note that the inclusion "$\subseteq$" is trivial because clearly $G_{\text{out}} \subseteq G_{\text{all}}$. For the reverse inclusion, assume for a contradiction that there exists $g \in G_{\text{all}}$ with $\text{HT}(g) \notin \text{mult}(\text{HT}(G_{\text{out}}))$. Then in particular, $g \notin G_{\text{out}}$, and so g must have been tested out of some value of G by the last **while**-loop of some call of UPDATE. We may thus assume that g is the last such problem polynomial to be removed from G. Inspecting the mechanism of the last **while**-loop of UPDATE, we see that there exists $h \in G_{\text{all}}$ with $\text{HT}(h) \mid \text{HT}(g)$. Moreover, h either stayed in G to the very end, or else it was removed at a later point in time than g. By our choice of g, we must have $\mathrm{HT}(h) \in \mathrm{mult}(\mathrm{HT}(G_{\mathrm{out}}))$ and so $\mathrm{HT}(g) \in \mathrm{mult}(\mathrm{HT}(G_{\mathrm{out}}))$ as well, a contradiction.

It remains to prove (i) and (iii). The reader should have no trouble verifying the following observation which will be used below: if a polynomial or a pair are placed in G or B, respectively, at some point during computation, then this is due to a call of UPDATE, and if the pair $\{g_1, g_2\}$ is being placed in B by some call of UPDATE, then each of g_1 and g_2 was placed in G by that same call or an earlier one. We claim that an invariant of both **while**-loops is given by

$$G \subseteq \mathrm{Id}(F). \tag{$*$}$$

The inclusion clearly holds upon initialization. The first **while**-loop places elements of F into G, and so $(*)$ is trivially preserved. An execution of the second **while**-loop places into G, if anything at all, a normal form h of an S-polynomial of a pair (g_1, g_2) taken from the list B. By the remark above, both g_1 and g_2 are or have been members of G before the present execution of the second **while**-loop, and it follows easily that $h \in \mathrm{Id}(F)$. We have proved that $(*)$ is indeed an invariant of the entire computation, and we immediately obtain (i) as well as the inclusion "$\mathrm{Id}(G_{\mathrm{all}}) \subseteq \mathrm{Id}(F)$." For the inclusion "$\mathrm{Id}(F) \subseteq \mathrm{Id}(G_{\mathrm{all}})$," we note that every $f \in F$ is placed into G by the call of UPDATE which follows its selection from F. This means that $F \subseteq G_{\mathrm{all}}$, even though the elements of F need not ever be in G all at the same time. We have now proved (i), (ii), and the equality $\mathrm{Id}(G_{\mathrm{all}}) = \mathrm{Id}(F)$, and thus the correctness proof is reduced to proving the rest of (iii), i.e., to the proof of the claim

"G_{all} is a Gröbner basis."

The following terminology will greatly simplify the wording of the proof. Let g_1, $g_2 \in G$ with $g_1 \neq g_2$. The pair $\{g_1, g_2\}$ will be called *good* if the following holds: the S-polynomial of g_1 and g_2 either equals 0, or else it has a t-representation w.r.t. G_{all} for some $t < \mathrm{lcm}(\mathrm{HT}(g_1), \mathrm{HT}(g_2))$. Theorem 5.64 tells us that it now suffices to prove that every pair $\{g_1, g_2\}$ of elements of G_{all} is good. There are essentially two ways for us to show that such a pair is good. Firstly, if

$$\mathrm{spol}(g_1, g_2) \xrightarrow[G_{\mathrm{all}}]{*} 0,$$

then by the results of the previous section, $\mathrm{spol}(g_1, g_2)$ either equals 0, or else it has a standard representation w.r.t. G_{all}, and in both cases, it follows that it is good. Secondly, if we can find $h \in G_{\mathrm{all}}$ such that (g_1, h, g_2) is a Buchberger triple and both $\{g_1, h\}$ and $\{h, g_2\}$ are good, then Proposition 5.70 asserts that $\{g_1, g_2\}$ is good too. Now assume for a contradiction that

$$V = \{\, \{g_1, g_2\} \mid g_1, g_2 \in G_{\mathrm{all}},\ g_1 \neq g_2,\ \{g_1, g_2\} \text{ not good} \,\} \neq \emptyset.$$

Then the set

$$V_{\min} = \{\, \{g_1, g_2\} \in V \mid \mathrm{lcm}(\mathrm{HT}(g_1), \mathrm{HT}(g_2)) \leq \mathrm{lcm}(\mathrm{HT}(h_1), \mathrm{HT}(h_2)) \\ \text{for all } \{h_1, h_2\} \in V \,\}$$

is not empty either. To arrive at the desired contradiction, we distinguish three cases. We denote by B_{all} the union of all values held by B during computation.

Case 1: $V_{\min} \cap B_{\text{all}} \neq \emptyset$.
Let $\{g_1, g_2\} \in V_{\min} \cap B_{\text{all}}$. Since $B = \emptyset$ upon termination, $\{g_1, g_2\}$ must have been removed from B at some point. There are two ways that this could have happened. If $\{g_1, g_2\}$ was selected from B at the beginning of some run through the second **while**-loop of GRÖBNERNEW2, then, in view of the fact that UPDATE always places the third component of its input into the first one, it is clear that

$$\operatorname{spol}(g_1, g_2) \xrightarrow[G_{\text{all}}]{*} 0,$$

and so $\{g_1, g_2\}$ is good, a contradiction. Else, $\{g_1, g_2\}$ was removed from B by the third **while**-loop of some call of UPDATE. This means that there exists $h \in G_{\text{all}}$ such that (g_1, h, g_2) is a Buchberger triple, and moreover, the divisibilities

$$\begin{aligned} &\operatorname{lcm}(\operatorname{HT}(g_1), \operatorname{HT}(h)) \mid \operatorname{lcm}(\operatorname{HT}(g_1), \operatorname{HT}(g_2)) \quad \text{and} \\ &\operatorname{lcm}(\operatorname{HT}(h), \operatorname{HT}(g_2)) \mid \operatorname{lcm}(\operatorname{HT}(g_1), \operatorname{HT}(g_2)) \end{aligned}$$

are both proper. It follows that

$$\begin{aligned} &\operatorname{lcm}(\operatorname{HT}(g_1), \operatorname{HT}(h)) < \operatorname{lcm}(\operatorname{HT}(g_1), \operatorname{HT}(g_2)) \quad \text{and} \\ &\operatorname{lcm}(\operatorname{HT}(h), \operatorname{HT}(g_2)) < \operatorname{lcm}(\operatorname{HT}(g_1), \operatorname{HT}(g_2)), \end{aligned}$$

and so the pairs $\{g_1, h\}$ and $\{h, g_2\}$ are good because $\{g_1, g_2\} \in V_{\min}$. We see that again $\{g_1, g_2\}$ is good, a contradiction.

For the next case, we denote by C_{all} the union of all values held by C, where the union ranges over all calls of UPDATE made by GRÖBNERNEW2.

Case 2: $V_{\min} \cap B_{\text{all}} = \emptyset$ and $V_{\min} \cap C_{\text{all}} \neq \emptyset$.
Let $\{h, g_1\} \in V_{\min} \cap C_{\text{all}}$. Then $\{h, g_1\}$ was placed in C by a certain call of UPDATE which we will refer to as the *present call*, and we may assume that this is the first call of UPDATE that ever places an element of $V_{\min}$ in C. If $\{h, g_1\}$ was passed on to D and then to E, then it ended up in B_{new} and thus in B_{all}, and so it is good by our assumption $V_{\min} \cap B_{\text{all}} = \emptyset$, a contradiction. If it was passed on to D but not to E, then the head terms of h and g_1 are disjoint, and so

$$\operatorname{spol}(h, g_1) \xrightarrow[G_{\text{all}}]{*} 0$$

by Lemma 5.66, which means that $\{h, g_1\}$ is good, again a contradiction. It remains to treat the case where $\{h, g_1\}$ is tested out by the first **while**-loop of UPDATE. It is not hard to see from the mechanism of this loop that there must exist a pair $\{h, g_2\} \in C$ which is passed on to D and satisfies

$$\operatorname{lcm}(\operatorname{HT}(h), \operatorname{HT}(g_2)) \mid \operatorname{lcm}(\operatorname{HT}(h), \operatorname{HT}(g_1))$$

We see that (h, g_2, g_1) is a Buchberger triple. Since g_2 is in the first component G of the input of the present call of UPDATE, it must be in G_{all}. To arrive at the desired contradiction "$\{h, g_1\}$ good," we must thus prove that $\{h, g_2\}$ and $\{g_2, g_1\}$ are both good.

As for $\{h, g_2\}$, there are two possibilities. If the divisibility

$$\operatorname{lcm}(\mathrm{HT}(h), \mathrm{HT}(g_2)) \mid \operatorname{lcm}(\mathrm{HT}(h), \mathrm{HT}(g_1))$$

is proper, then $\{h, g_2\}$ is good by the same argument that was used in Case 1 above. If the two lcm's are equal, then $\{h, g_2\} \in V$ implies $\{h, g_2\} \in V_{\min}$. But $\{h, g_2\}$ was passed on to D, and we have already argued that no element of $V_{\min} \cap C$ can have been passed on to D in the present case. So again, $\{h, g_2\}$ must be good.

For $\{g_2, g_1\}$, we are looking at the same two possibilities. If the divisibility

$$\operatorname{lcm}(\mathrm{HT}(g_2), \mathrm{HT}(g_1)) \mid \operatorname{lcm}(\mathrm{HT}(h), \mathrm{HT}(g_1))$$

is proper, then $\{g_2, g_1\}$ is good by the same argument as above. If the two lcm's are equal, then as before, $\{g_2, g_1\} \in V$ implies $\{g_2, g_1\} \in V_{\min}$. But g_1 and g_2 are in the value of G that is passed to the present call of UPDATE, and so they must both have been placed in G by earlier calls of UPDATE. These two placements happened in a certain order, and when the second one of g_1 and g_2 arrived, the first one was still there because otherwise they would not now both be in G. This means that when the second one arrived, the pair $\{g_2, g_1\}$ was placed in C by that call of UPDATE. But we are talking about an earlier call of UPDATE than the present one, and so $\{g_2, g_1\}$ cannot be in $V_{\min} \cap C_{\text{all}}$ by the chronologically minimal choice of $\{h, g_1\}$. So $\{g_2, g_1\}$ cannot be in $V_{\min}$ and hence not in V by the remark above, which means that it is good.

Case 3: $V_{\min} \cap B_{\text{all}} = \emptyset$ and $V_{\min} \cap C_{\text{all}} = \emptyset$.

Let $\leq_{\text{c}}$ be the chronological order on G_{all}, where $g_1 \leq_{\text{c}} g_2$ iff $g_1 = g_2$ or g_1 was placed in G by an earlier call of UPDATE than g_2. To every pair in $V_{\min}$, we assign an ordered pair with the same entries such that the first component of the ordered pair is chronologically less than the second one. We denote the set of ordered pairs thus obtained by $V_{\min}^{\text{ord}}$, and we consider the "lexicographical-chronological" order $\leq_{\text{lc}}$ on $V_{\min}^{\text{ord}}$, where

$$\begin{aligned}(g_1, g_2) \leq_{\text{lc}} (h_1, h_2) \quad \text{iff} \quad & g_1 <_{\text{c}} h_1, \quad \text{or} \\ & g_1 = h_1 \text{ and } g_2 \leq_{\text{c}} h_2.\end{aligned}$$

Let now (g_1, g_2) be the $\leq_{\text{lc}}$-last element of $V_{\min}^{\text{ord}}$. Then g_1 must have been tested out of G by the last **while**-loop of some call of UPDATE, and this must have happened before g_2 was placed in G, because otherwise g_2 would have met g_1 when the former was placed in G, and so the pair $\{g_1, g_2\}$ would have been placed in C which is not the case by the assumption

$V_{\min} \cap C_{\text{all}} = \emptyset$. After g_1 was removed from G, the same call of UPDATE placed into G a polynomial h with $\text{HT}(h) \mid \text{HT}(g_1)$, and we see that

$$\text{HT}(h) \mid \text{lcm}\big(\text{HT}(g_1), \text{HT}(g_2)\big),$$

which means that (g_1, h, g_2) is a Buchberger triple. To arrive at the desired contradiction“$\{g_1, g_2\}$ good,” it thus once again suffices to prove that $\{g_1, h\}$ and $\{h, g_2\}$ are both good.

For $\{g_1, h\}$, we first note that properness of the divisibility

$$\text{lcm}\big(\text{HT}(g_1), \text{HT}(h)\big) \mid \text{lcm}\big(\text{HT}(g_1), \text{HT}(g_2)\big)$$

implies that $\{g_1, h\}$ is good by a now familiar argument. If these lcm's are equal, then $\{g_1, h\} \in V$ implies $\{g_1, h\} \in V_{\min}$. But h was the polynomial that was responsible for eliminating g_1 from G, and so the pair $\{g_1, h\}$ was placed in C at the beginning of that same call of UPDATE. The case assumption $V_{\min} \cap C_{\text{all}} = \emptyset$ thus implies that $\{g_1, h\}$ is good.

As for $\{h, g_2\}$, we can argue as before that properness of the divisibility

$$\text{lcm}\big(\text{HT}(h), \text{HT}(g_2)\big) \mid \text{lcm}\big(\text{HT}(g_1), \text{HT}(g_2)\big)$$

implies that $\{h, g_2\}$ is good, and that in the remaining case of equality of these lcm's, $\{h, g_2\} \in V$ implies $\{h, g_2\} \in V_{\min}$. In order to prove that $\{h, g_2\}$ is good in this last case, we need to look at the chronological order of arrivals and departures in the set G:

$$\begin{array}{lllllll} g_1 \text{ in} & \longrightarrow & g_1 \text{ out} & & & & \\ & & h \text{ in} & \longrightarrow & (\text{possibly } h \text{ out}) & \longrightarrow & g_2 \text{ in.} \end{array}$$

We see that if $\{h, g_2\}$ were in $V_{\min}$, then the corresponding element of $V_{\min}^{\text{ord}}$ would be (h, g_2), and moreover, we would have $(g_1, g_2) <_{\text{lc}} (h, g_2)$ which is impossible by the choice of (g_1, g_2) as the $\leq_{\text{lc}}$-last element of $V_{\min}^{\text{ord}}$. We have proved that $\{h, g_2\}$ is good. □

It is clear that the exploitation of Buchberger's second criterion by the algorithm GRÖBNERNEW2 does not have the kind of dependence on the selection strategy of critical pairs that we found in GRÖBNERNEW1. It should be noted though that there is always a certain random dependence: for any pair $\{g_1, g_2\}$ that is on the critical pair list at some point, there may, in the unforeseeable future, appear an h such that (g_1, h, g_2) is a Buchberger triple, and a strategy that happens to hold $\{g_1, g_2\}$ long enough will exploit this instance, while others may not. This phenomenon also makes it hard to make precise a statement like “GRÖBNERNEW1 and GRÖBNERNEW2 exploit the second criterion equally well.”

Exercise 5.74 Find reasons to support the statement “GRÖBNERNEW2 does not blatantly miss instances of the second criterion.” (Hint: Use the fact that if (g_1, h, g_2) is a Buchberger triple and

$$\text{lcm}\big(\text{HT}(g_1), \text{HT}(h)\big) = \text{lcm}\big(\text{HT}(g_1), \text{HT}(g_2)\big),$$

then (h, g_2, g_1) is a Buchberger triple too.)

It is perhaps noteworthy that GRÖBNERNEW2 keeps the set G top-reduced throughout the computation; so if one wishes to pass to the reduced Gröbner basis afterwards, then one may skip the elimination of polynomials whose head terms are multiples of others in the algorithm REDGRÖBNER.

Finally, we mention that the term order relative to which the Gröbner basis is computed can have dramatic effects on the running time of the algorithm; since one often needs just *some* Gröbner basis, it is therefore desirable to be able to make an intelligent choice of the term order. When the normal strategy is used, then the algorithm tends to run faster for total degree orders than for lexicographical orders. Among all total degree orders, a good choice is the one that breaks ties according to an inverted lexicographical order. (This is sometimes referred to as the *Buchberger order.*) If a lexicographical order is desired without an a priori preference for the variable ordering, then the best choice in general seems to be the following. For $1 \le i \le n$, let $D_i = \{\deg_{X_i}(f) \mid f \in F\}$ where F is the input set. Let $\le$ be a lexicographical term order such that $D_i \succeq D_j$ implies $X_i \ll X_j$, where $\succeq$ is the linear order on $\mathcal{P}_{\text{fin}}(\mathbb{N})$ induced by the natural order on $\mathbb{N}$ according to Theorem 4.69. In other words, a variable is placed lexicographically low if it occurs with a high degree in F. In case of a tie, variables that occur in larger numbers should be placed lexicographically lower. Then the algorithm tends to run faster for $\le$ than for any other lexicographical term order.

5.6 The Extended Gröbner Basis Algorithm

Let us once more compare the algorithm GRÖBNER to the Euclidean algorithm. If we use the latter to compute the gcd d of univariate polynomials f and g (or of more than two by an iterated application), then we have found the reduced Gröbner basis of $\text{Id}(f, g)$. But we can do more than that: if we use the *extended* Euclidean algorithm, then we also obtain a representation of d as a sum of multiples of f and g. This raises the question if, in the multivariate case, we can effectively express each element of the (reduced) Gröbner basis that we have computed as a sum of multiples of the input polynomials. In this section, we show how this can be achieved. An important application will be discussed in Section 6.1.

Let $K[\underline{X}]$, T, and $\le$ be as in the previous section. Suppose F is a finite subset of $K[\underline{X}]$ and $G = \text{GRÖBNER}(F)$. We wish to extend the algorithm GRÖBNER in such a way that it provides a family

$$\left\{\{q_{gf}\}_{f\in F}\right\}_{g\in G} \tag{$*$}$$

of families $\{q_{gf}\}_{f\in F}$ of polynomials q_{gf} such that

$$g = \sum_{f\in F} q_{gf} f$$

for all $g \in G$. This is quite easily achieved. The algorithm GRÖBNER starts out with $G = F$, and it is clear that then the polynomials $q_{gf} = \delta_{gf}$ have the required property, where

$$\delta_{gf} = \begin{cases} 1 & \text{if } f = g \\ 0 & \text{otherwise} \end{cases}$$

is the *Kronecker symbol.* Now suppose that we are about to enter a run through the **while**-loop, and the family (∗) has been computed for the current value of G. If nothing is being added to G during that run, then there is nothing that needs to be done. Otherwise, the newly added polynomial h is a normal form modulo G of an S-polynomial of a pair of elements of G. This means that $g \xrightarrow[G]{*} h$ where

$$g = m_1 g_1 - m_2 g_2 = \operatorname{spol}(g_1, g_2)$$

for a pair of elements g_1, $g_2 \in G$ and certain monomials m_1 and m_2. The algorithm has determined what the monomials m_1 and m_2 are, and it has performed the reduction of g to h by means of the algorithm REDPOL. It may therefore provide itself with a family $\{q_f\}_{f \in F}$ of polynomials such that

$$h - g = \sum_{p \in G} q_p p .$$

We then have

$$\begin{aligned} h &= g + \sum_{p \in G} q_p p \\ &= m_1 g_1 - m_2 g_2 + \sum_{p \in G} q_p p \\ &= \sum_{f \in F} m_1 q_{g_1 f} f - \sum_{f \in F} m_2 q_{g_2 f} f + \sum_{p \in G} q_p \sum_{f \in F} q_{pf} f \\ &= \sum_{f \in F} \left(m_1 q_{g_1 f} - m_2 q_{g_2 f} + \sum_{p \in G} q_p q_{pf} \right) f . \end{aligned}$$

If we now set

$$q_{hf} = m_1 q_{g_1 f} - m_2 q_{g_2 f} + \sum_{p \in G} q_p q_{pf} ,$$

then we see that $\{q_{hf}\}_{f \in F}$ is the family by which the family (∗) must be enlarged to achieve our purpose. (Note that h, being a normal form modulo G, is necessarily different from each element of G, and so there is no conflict with existing elements of the family (∗).)

In connection with representations of elements of G as sums of multiples of elements of F, one often needs the reverse transformation too, i.e., representations of elements of F as sums of multiples of elements of G.

This is even easier to achieve: since G is a Gröbner basis of $\mathrm{Id}(F)$, we have $f \xrightarrow[G]{*} 0$ for all $f \in F$, and so the desired representations can be obtained immediately by running the the algorithm REDPOL on (f, G). Moreover, they even come out as standard representations.

We have actually proved the correctness of the algorithm EXTGRÖBNER of the theorem below. Termination is trivial because the mechanism of the **while**-loop is the same as in the algorithm GRÖBNER. Before we state the algorithm, a remark on the assignments involving families is in order. If A is a set, then formally, a family $\mathcal{F} = \{a_i\}_{i \in I}$ of elements of A is a function from the index set I to A, which in turn is a set of ordered pairs, namely,

$$\mathcal{F} = \{a_i\}_{i \in I} = \{\, (i, a_i) \mid i \in I \,\}.$$

So if we enlarge the index set I by a new element j, and we wish to enlarge the family $\mathcal{F}$ accordingly by an element a_j, then this is achieved by the assignment $\mathcal{F} \leftarrow \mathcal{F} \cup \{(j, a_j)\}$.

Theorem 5.75 *Let F be a finite subset of $K[\underline{X}]$. Suppose the ground field is computable, and the term order on T is decidable. Then the algorithm* EXTGRÖBNER *of Table* 5.9 *computes a Gröbner basis G in $K[\underline{X}]$ such that $F \subseteq G$ and* $\mathrm{Id}(G) = \mathrm{Id}(F)$, *and families*

$$\mathcal{G} = \left\{\{q_{gf}\}_{f \in F}\right\}_{g \in G} \quad \text{and} \quad \mathcal{F} = \left\{\{p_{fg}\}_{g \in G}\right\}_{f \in F}$$

such that

$$g = \sum_{f \in F} q_{gf} f \quad \text{and} \quad f = \sum_{g \in G} p_{fg} g$$

for all $g \in G$ and $f \in F$, and the representations of $f \in F$ are in fact standard representations. □

There is no need to state a general theorem for arbitrary field K and term order $\leq$ here, because the mere existence of the families $\mathcal{G}$ and $\mathcal{F}$ is trivial: F and G are bases of the same ideal I, and moreover, G is a Gröbner basis of I, and so every $f \in I$ even has a standard representation w.r.t. G. It is clear that the algorithm GRÖBNERNEW1 can be extended in the exact same way as above, because it differs from GRÖBNER only insofar as it skips certain critical pairs. Extending the second version GRÖBNERNEW2, by contrast, is a slightly more tedious affair because here, polynomials that are employed in the reduction of an S-polynomial may later be removed from the set G.

Exercise 5.76 Discuss an extended version of GRÖBNERNEW2.

Now assume that we wish to find the *reduced* Gröbner basis G' of $\mathrm{Id}(F)$ and families $\mathcal{G}'$ and $\mathcal{F}'$ as in the theorem above. We know that G' can be found by applying REDGRÖBNER to GRÖBNER(F). The family $\mathcal{F}'$ poses no problem: we just perform the **for all**-loop of EXTGRÖBNER at

TABLE 5.9. Algorithm EXTGRÖBNER

```
Specification: $(G, \mathcal{G}, \mathcal{F}) \leftarrow$ EXTGRÖBNER$(F)$
               Construction of a Gröbner basis $G$ of Id$(F)$ and
               back-and-forth transformations between $F$ and $G$
Given: $F =$ a finite subset of $K[\underline{X}]$
Find: $G =$ a finite subset of $K[\underline{X}]$ such that $G$ is a Gröbner
      basis in $K[\underline{X}]$ with $F \subseteq G$ and Id$(G) =$ Id$(F)$, and
      families $\mathcal{G}$ and $\mathcal{F}$ as described in Theorem 5.75
begin
$G \leftarrow F$
$\mathcal{G} \leftarrow \{\{\delta_{gf}\}_{f \in F}\}_{g \in G}$
$B \leftarrow \{\, \{g_1, g_2\} \mid g_1, g_2 \in G$ with $g_1 \neq g_2 \,\}$
while $B \neq \emptyset$ do
      select $\{g_1, g_2\}$ from $B$
      $B \leftarrow B \setminus \{\{g_1, g_2\}\}$
      $g \leftarrow m_1 g_1 - m_2 g_2$, where $m_1 g_1 - m_2 g_2 = \mathrm{spol}(g_1, g_2)$
      $(\{q_p\}_{p \in G}, h) \leftarrow$ REDPOL$(g, G)$
      if $h \neq 0$ then
        $B \leftarrow B \cup \{\, \{p, h\} \mid p \in G \,\}$
        $\mathcal{G} \leftarrow \mathcal{G} \cup \{(h, \{q_{hf}\}_{f \in F})\}$,
           where $q_{hf} = m_1 q_{g_1 f} - m_2 q_{g_2 f} + \sum_{p \in G} q_p q_{pf}$
        $G \leftarrow G \cup \{h\}$
      end
end
$\mathcal{F} \leftarrow \emptyset$
for all $f \in F$ do
   $\mathcal{F} \leftarrow \mathcal{F} \cup \{(f, \{p_{fg}\}_{g \in G})\}$, where $(\{p_{fg}\}_{g \in G}, 0) =$ REDPOL$(f, G)$
end
return$(G, \mathcal{G}, \mathcal{F})$
end EXTGRÖBNER
```

the end of the entire computation. As for $\mathcal{G}'$, we first use EXTGRÖBNER to find a Gröbner basis G of Id(F) and a family $\mathcal{G}$ as described in the theorem above. Then we apply an algorithm EXTREDGRÖBNER which acts just like REDGRÖBNER with the following extension. During the **while**-loop, REDGRÖBNER tries to delete polynomials from G. Everytime that this happens, we simply remove the corresponding element $\{q_{gf}\}_{f \in F}$ from $\mathcal{G}$. REDGRÖBNER then applies REDUCTION, which tries to select g from G and replace it by the result h of a complete reduction modulo $G \setminus \{g\}$. (The normal form h will never be zero here because no top reductions are possible.) Every time this happens, we may let the algorithm REDPOL provide a family $\{q_p\}_{p \in G \setminus \{g\}}$ with

$$h - g = \sum_{p \in G \setminus \{g\}} q_p p .$$

We then have

$$\begin{aligned} h &= g + \sum_{p \in G \setminus \{g\}} q_p p \\ &= \sum_{f \in F} q_{gf} f + \sum_{p \in G \setminus \{g\}} q_p \sum_{f \in F} q_{pf} f \\ &= \sum_{f \in F} \left(q_{gf} + \sum_{p \in G \setminus \{g\}} q_p q_{pf} \right) f . \end{aligned}$$

All we have to do to adjust the family $\mathcal{G}$ to the modified set G is thus to remove the pair $(g, \{q_{gf}\}_{f \in F})$ and replace it by $(h, \{q_{hf}\}_{f \in F})$, where

$$q_{hf} = q_{gf} + \sum_{p \in G \setminus \{g\}} q_p q_{pf}$$

for all $f \in F$.

Exercise 5.77 Write the algorithm EXTREDGRÖBNER as described above.

Exercise 5.78 Use the result of Exercise 5.31 and the ideas of this section to write an algorithm that computes the gcd of a finite set of univariate polynomials and a representation of this gcd as a sum of multiples of the input polynomials.

Notes

Gröbner basis theory originates in the doctoral dissertation of Bruno Buchberger, which was written in 1965 at the University of Innsbruck, Austria, under the supervision of Wolfgang Gröbner. Gröbner had asked if there was an algorithm that computes a vector space basis over K of the residue class ring $K[\underline{X}]/\mathrm{Id}(F)$, where $K[\underline{X}]$ is a multivariate polynomial ring over the field K and F is a given finite subset of $K[\underline{X}]$. The algorithm was to be such that it would also make possible effective computations in the ring $K[\underline{X}]/\mathrm{Id}(F)$. Gröbner's interest was mainly in algebraic geometry; however, he favored an ideal theoretic and thus potentially algorithmic approach that was somewhat beside the mainstream of his time. Interestingly, Buchberger's results received precious little attention until the early seventies, when their relevance and the scope of their applications were finally realized. It was only then that Buchberger introduced the term "Gröbner basis."

There are a number of results in the earlier literature which now, from hindsight, turn out to be reminiscent of and related to Gröbner basis theory and the Buchberger algorithm. Macaulay (1916) introduced the concept of an H-basis of an ideal in $K[\underline{X}]$. An H-basis is a finite subset F of $K[\underline{X}]$ such that every $0 \neq g \in \mathrm{Id}(F)$ has a so-called H-representation, i.e., a representation of the form

$$g = \sum_{f \in F} h_f f \qquad (h_f \in K[\underline{X}])$$

with $\max\{\deg(h_f f) \mid f \in F\} \leq \deg(g)$. This condition bears a strong resemblance to the characterization of Gröbner bases in terms of standard representations (cf. Theorem 5.62). In fact, if the term order in question is a total degree order, then every standard representation is an H-representation; as a consequence, every Gröbner basis is an H-basis. The converse fails in general because the degree condition is strictly weaker than the corresponding condition for standard representations in case of a total-degree term order, and incompatible with the latter for other term orders.

Macaulay proves the existence of an H-basis for a given ideal non-constructively as a simple consequence of the Hilbert basis theorem together with homogenization techniques. For a particular example, he also sketches a construction method for H-bases, which he claims to be "a general one." The idea of the critical pair completion procedure which forms the overall structure of the Buchberger algorithm appears independently in Knuth and Bendix (1970). More details and references can be found in the section "Term Rewriting" on p. 523 in the appendix. There, it is used to enlarge a set of equations between first-order terms in such a way that the corresponding set of rewrite rules gives rise to a confluent reduction relation. Hironaka (1964) proves the existence of a certain kind of ideal bases in rings of power series that have since turned out to be analogues to Gröbner bases. Again, the appendix has more details.

As we have mentioned before, the concept of standard representations is essentially present in Macaulay's work. The Gröbner basis criterion in terms of t-representations of Theorem 5.64 derives from a criterion that goes back to Lazard (1983) and involves "lifting of syzygies" as discussed in Section 6.1 (see also the discussion in the Notes to Chapter 6 on p. 291). We have modified the argument in order to make Buchberger's second criterion accessible without the use of syzygies. The latter was first proved in Buchberger (1979); the algorithm GRÖBNERNEW1 describes Buchberger's original implementation. The version GRÖBNERNEW2 is due to Gebauer and Möller (1988).

6

First Applications of Gröbner Bases

In this chapter, we discuss some of the most immediate and important applications of Gröbner bases. The central part is formed by Sections 6.2 and 6.3, which deal with Gröbner bases in ideal theory. The theory of polynomial ideals plays an important role in *algebraic geometry.* There, one considers polynomials with coefficients in some field K and investigates the behavior of zeroes of these polynomials in an extension field K' of K. (Recall that a zero of $f(X_1, \ldots, X_n)$ is an n-tuple $(a_1, \ldots, a_n)$ of elements of K' with $f(a_1, \ldots, a_n) = 0$; cf. also Lemma 2.17 (i)). This leads to a large number of questions of an algorithmic nature, such as these: given finite bases of two ideals, what is a basis of the intersection of the latter, or, given a polynomial f and an ideal I, is it true or not that some power of f lies in I? It has been known for a long time that all these problems can be algorithmically solved. Before the arrival of Gröbner bases, however, the complexity of these algorithms was out of bounds for all practical purposes. In this chapter, we will demonstrate how Gröbner bases provide rather straightforward solutions to many decision and construction problems in the theory of polynomial ideals. Bringing these computations within the realm of feasibility has of course stimulated vigorous mathematical research on how to further improve them. We do not attempt to capture the state of the art in the field; our aim is to lay firm mathematical foundations and to present algorithms that anyone could implement in today's computer algebra systems. We will also use Gröbner bases in the development of the theory, thereby demonstrating that the theory of Gröbner bases is not only a powerful algorithmic method but also a cornerstone of commutative algebra.

6.1 Computation of Syzygies

The results of this section will be needed for Proposition 6.33 and its corollary, and then again in Section 10.5. Since Proposition 6.33 and its corollary have no further applications in this book, this section can be skipped for now by those who are mainly interested in ideal theory.

We will use some terminology and notation (but not really any theory) from Section 3.3. Let R be any ring. Then a **homogeneous linear equation** over R in the indeterminates $Y_1, \ldots, Y_m$ with **coefficients** $f_1, \ldots, f_m \in R$ is an equation of the form

$$Y_1 f_1 + \cdots + Y_m f_m = 0. \tag{1}$$

A *solution* of the equation (1) is any m-tuple $(h_1, \ldots, h_m)$ of elements of R with

$$h_1 f_1 + \cdots + h_m f_m = 0.$$

Quite obviously, the m-tuple $(0, \ldots, 0)$ is always a solution, called the *trivial* solution; any other solution is called *non-trivial.*

Let $S \subseteq R^m$ denote the set of all solutions of (1). Recall from Section 3.3 that R^m forms an R-module under componentwise addition and scalar multiplication. It is easy to see that a scalar multiple of a solution and a sum of two solutions of (1) is again a solution, and thus S is an R-submodule of R^m. It is clear that S is precisely what we have called the *(first) module of syzygies* of $(f_1, \ldots, f_m)$ in Section 3.3.

The goal of this section is to describe the R-module S by specifying a finite set of generators of S for the case that R is a multivariate polynomial ring $K[\underline{X}] = K[X_1, \ldots, X_n]$ over a field K, and to compute this set of generators from the coefficients of the equation in case that K is computable. We will approach the problem in two steps. For the first step, we assume that the coefficients $f_1, \ldots, f_m$ of (1) form a Gröbner basis in $K[\underline{X}]$ with respect to some term order; in the second step, we reduce the general problem to the first case by passing from the given coefficients $f_1, \ldots, f_m \in K[\underline{X}]$ to a Gröbner basis of the ideal $\mathrm{Id}(f_1, \ldots, f_m)$.

For the first step, let $F = \{f_1, \ldots, f_m\}$ be a Gröbner basis in $K[\underline{X}]$ with respect to some term order $\leq$ on the set $T = T(X_1, \ldots, X_n)$. To be sure that the *set* F determines the equation (1), we assume that the coefficients of the given equation are pairwise different; this is certainly a reasonable supposition. Furthermore, it will turn out to be convenient to assume that F is monic, i.e., $\mathrm{HC}(f_i) = 1$ for $1 \leq i \leq m$. If this is not already the case, then we replace f_i by $f_i/\mathrm{HC}(f_i)$ for $1 \leq i \leq m$. It is obvious that if $(h_1, \ldots, h_m)$ is a solution of the modified equation (1), then

$$(h_1/\mathrm{HC}(f_1), \ldots, h_m/\mathrm{HC}(f_m))$$

is a solution of the original equation.

For $1 \leq i < j \leq m$, we let

$$p_{ij} = \mathrm{spol}(f_i, f_j) = s_{ij} f_i - s_{ji} f_j \tag{2}$$

with s_{ij}, $s_{ji} \in T$ be the S-polynomial of f_i and f_j. Note that the constants appearing in the definition of the S-polynomial equal 1 because F is monic. Each p_{ij} is an element of $\mathrm{Id}(F)$, and so it reduces to 0 modulo the Gröbner basis F of $\mathrm{Id}(F)$. This means that either p_{ij} equals 0, or else there exist, by Proposition 5.22, polynomials q_{ijk} $(1 \le k \le m)$ such that

$$p_{ij} = \sum_{k=1}^{m} q_{ijk} f_k \quad \text{with} \quad \mathrm{HT}(q_{ijk} f_k) \le \mathrm{HT}(p_{ij}) \text{ for } 1 \le k \le m. \tag{3}$$

For simplicity, let us agree to ignore all references to $\mathrm{HT}(q_{ijk} f_k)$ whenever $q_{ijk} = 0$; if $p_{ij} = 0$, then this convention applies to all q_{ijk} for $1 \le k \le m$. Subtracting (2) from (3) yields

$$(q_{iji} - s_{ij}) f_i + (q_{ijj} + s_{ji}) f_j + \sum_{\substack{k=1 \\ k \ne i,j}}^{m} q_{ijk} f_k = 0, \tag{4}$$

and we see that we have found solutions of (1), one for each (i, j) with $1 \le i < j \le m$. We are going to prove that this set of solutions is already the desired generating system for the module $\boldsymbol{S}$ of syzygies in question. To this end, we introduce the following notation:

$$r_{ijk} = \begin{cases} q_{iji} - s_{ij} & \text{if} \quad k = i \\ q_{ijj} + s_{ji} & \text{if} \quad k = j \\ q_{ijk} & \text{otherwise,} \end{cases}$$

and

$$\boldsymbol{r}_{ij} = (r_{ij1}, \ldots, r_{ijm}).$$

Proposition 6.1 *The set $B = \{ \boldsymbol{r}_{ij} \mid 1 \le i < j \le m \}$ generates $\boldsymbol{S}$ as an $K[\underline{X}]$-module.*

Proof Assume for a contradiction that the set $M = \boldsymbol{S} \setminus \mathrm{lin}(B)$ is non-empty. Let $(h_1, \ldots, h_m) \in M$ be such that

$$t = \max\{ \mathrm{HT}(h_k f_k) \mid 1 \le k \le m \}$$

is minimal w.r.t. the term order $\le$ in the set

$$\{ \max\{ \mathrm{HT}(g_k f_k) \mid 1 \le k \le m \} \mid (g_1, \ldots, g_m) \in M \} \subseteq T.$$

Assume further that among all possible choices that satisfy this requirement, $(h_1, \ldots, h_m)$ is such that the cardinality of the set

$$J = \{ k \mid 1 \le k \le m, \ \mathrm{HT}(h_k f_k) = t \}$$

is minimal. We have

$$\sum_{k=1}^{m} h_k f_k = 0 \tag{5}$$

since $(h_1, \ldots, h_m) \in \boldsymbol{S}$, and it follows that the sum of all monomials in this sum that have t as their term equals 0. So the set J must contain at least two different elements, say i and j with $i < j$. This means that

$$t = \mathrm{HT}(h_i) \cdot \mathrm{HT}(f_i) = \mathrm{HT}(h_j) \cdot \mathrm{HT}(f_j)$$

is a common multiple of $\mathrm{HT}(f_i)$ and $\mathrm{HT}(f_j)$. By the definition of p_{ij} as the S-polynomial of f_i and f_j, we know that

$$s = s_{ij} \cdot \mathrm{HT}(f_i) = s_{ji} \cdot \mathrm{HT}(f_j)$$

is the least common multiple of $\mathrm{HT}(f_i)$ and $\mathrm{HT}(f_j)$, and we may conclude that s divides t, say $t = us$. With $a_i = \mathrm{HC}(h_i)$, we now add $a_i u$ times (4) to (5) and obtain

$$\sum_{k=1}^{m} (h_k + a_i u r_{ijk}) f_k = 0.$$

Setting $g_k = h_k + a_i u r_{ijk}$ for $1 \leq k \leq m$, we see that $(g_1, \ldots, g_m) \in \boldsymbol{S}$.

Claim: $\mathrm{HT}(g_k f_k) \leq t$ for $1 \leq k \leq m$, and the number of occurences of t in the sum $g_1 f_1 + \cdots + g_m f_m$ is less than the cardinality of J.

Proof: We first note that (3) together with Exercise 5.47 (ii) implies that $\mathrm{HT}(q_{ijk} f_k) < s$ for all $1 \leq k \leq m$. The claim is an easy consequence of the following three statements. For $k \neq i, j$, we have

$$\begin{aligned} \mathrm{HT}(a_i u r_{ijk} f_k) &= u \cdot \mathrm{HT}(q_{ijk} f_k) \\ &< us = t, \end{aligned}$$

and so $\mathrm{HT}(g_k f_k) \leq t$, and $\mathrm{HT}(h_k f_k) < t$ implies $\mathrm{HT}(g_k f_k) < t$. For $k = j$, we get

$$\begin{aligned} \mathrm{HT}(a_i u r_{ijk} f_k) &= u \cdot \mathrm{HT}(q_{ijj} f_j + s_{ji} f_j) \\ &= us = t, \end{aligned}$$

and so $\mathrm{HT}(g_j f_j) \leq t$. Finally, for $k = i$,

$$\begin{aligned} \mathrm{HM}(a_i u r_{iji} f_i) &= a_i u \cdot \mathrm{HM}(q_{iji} f_i - s_{ij} f_i) \\ &= -a_i us = -a_i t = -\mathrm{HM}(h_i f_i), \end{aligned}$$

and so $\mathrm{HT}(g_i f_i) < t$.

By the minimal choice of $(h_1, \ldots, h_m)$, it now follows that $(g_1, \ldots, g_m) \in \mathrm{lin}(B)$. But

$$(h_1, \ldots, h_m) = (g_1, \ldots, g_m) - a_i u \cdot \boldsymbol{r}_{ij},$$

and so $(h_1, \ldots, h_m) \in \mathrm{lin}(B)$ as well, a contradiction. □

Exercise 6.2 Let $1 \leq i < j \leq m$, and suppose there is $1 \leq k \leq m$ with $k \neq i, j$ and

$$\mathrm{HT}(f_k) \mid \mathrm{lcm}\big(\mathrm{HT}(f_i), \mathrm{HT}(f_j)\big).$$

Show that $B \setminus \{\boldsymbol{r}_{ij}\}$ still generates $\boldsymbol{S}$. (Hint: Cf. the proof of Proposition 5.70.)

The special case where $F = \{f_1, \ldots, f_m\}$ is a Gröbner basis being settled, let now $f_1, \ldots, f_m$ be arbitrary polynomials in $K[\underline{X}]$. Let $g_1, \ldots, g_r \in K[\underline{X}]$ be pairwise different such that $G = \{g_1, \ldots, g_r\}$ is a monic Gröbner basis of the ideal $\mathrm{Id}(F)$, and let us denote by S_F and S_G the module of syzygies of $(f_1, \ldots, f_m)$ and $(g_1, \ldots, g_r)$, respectively. We know how to find a generating system for S_G, and we are now going to show how this can be used to obtain a generating system for S_F. Since F and G generate the same ideal I in $K[\underline{X}]$, we have "forward" and "backward" transformations between these ideal bases, i.e., there exist c_{ij}, $d_{ji} \in K[\underline{X}]$ with

$$\begin{aligned} g_i &= \sum_{j=1}^{m} c_{ij} f_j \quad \text{for} \quad 1 \le i \le r, \quad \text{and} \\ f_j &= \sum_{i=1}^{r} d_{ji} g_i \quad \text{for} \quad 1 \le j \le m. \end{aligned}$$

Composing these transformation both ways, we obtain

$$g_i = \sum_{j=1}^{m} c_{ij} f_j = \sum_{j=1}^{m} c_{ij} \sum_{k=1}^{r} d_{jk} g_k = \sum_{k=1}^{r} \left(\sum_{j=1}^{m} c_{ij} d_{jk} \right) g_k$$

for $1 \le i \le r$, and

$$f_j = \sum_{i=1}^{r} d_{ji} g_i = \sum_{i=1}^{r} d_{ji} \sum_{l=1}^{m} c_{il} f_l = \sum_{l=1}^{m} \left(\sum_{i=1}^{r} d_{ji} c_{il} \right) f_l \tag{6}$$

for $1 \le j \le m$. Let δ_{ij} be the Kronecker symbol, i.e., $\delta_{ii} = 1$ and $\delta_{ij} = 0$ for $i \ne j$. There is a certain temptation to conclude from the equations above that

$$\sum_{j=1}^{m} c_{ij} d_{jk} = \delta_{ik} \quad \text{and} \quad \sum_{i=1}^{r} d_{ji} c_{il} = \delta_{jl}.$$

Unfortunately, this conjecture is false in general, even when the transformations are obtained from a Gröbner basis computation and polynomial reduction.

Exercise 6.3 Let $F = \{f_1, f_2\} \subseteq \mathbb{Q}[X, Y, Z]$ with $f_1 = XY + 1$ and $f_2 = XZ + 1$. Use the algorithms EXTGRÖBNER and REDPOL to compute a Gröbner basis G of F w.r.t. the total degree–lexicographical term order (where $X \gg Y \gg Z$), and back-and-forth transformations between F and G. Show that these refute the above conjecture.

We can, however, rewrite the equation (6) in the form

$$\sum_{l=1}^{m} \left(\delta_{jl} - \sum_{i=1}^{r} d_{ji} c_{il} \right) f_l = 0 \quad \text{for} \quad 1 \le j \le m.$$

Now if we set

$$a_{jl} = \delta_{jl} - \sum_{i=1}^{r} d_{ji}c_{il} \quad \text{for} \quad 1 \leq j, l \leq m,$$

then we see that the m-tuple $\boldsymbol{a}_j = (a_{j1}, \ldots, a_{jm})$ is an element of $\boldsymbol{S}_F$ for $1 \leq j \leq m$.

To obtain the desired generating system for $\boldsymbol{S}_F$, we set $A = \{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m\}$. Let $B = \{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_s\}$ be a generating system for the $K[\underline{X}]$-module $\boldsymbol{S}_G$ of syzygies of $(g_1, \ldots, g_r)$, and let $B^* = \{\boldsymbol{b}_1^*, \ldots, \boldsymbol{b}_s^*\}$ where

$$\boldsymbol{b}_l^* = (b_{l1}^*, \ldots, b_{lm}^*) \quad \text{with} \quad b_{lj}^* = \sum_{i=1}^{r} b_{li}c_{ij}$$

for $1 \leq l \leq s$. The proof of the following theorem is based on the fact that we can use the transformations between F and G to transform elements of $\boldsymbol{S}_F$ into elements of $\boldsymbol{S}_G$ and vice versa. As we will see shortly, B^* is in fact a transformation of B to a set of solutions of $\boldsymbol{S}_F$.

Theorem 6.4 *$A \cup B^*$ is a generating system for the $K[\underline{X}]$-module $\boldsymbol{S}_F$ of syzygies of $(f_1, \ldots, f_m)$.*

Proof From the fact that $\boldsymbol{b}_l = (b_{l1}, \ldots, b_{lr})$ are elements of $\boldsymbol{S}_G$, we conclude that

$$\begin{aligned} 0 &= \sum_{i=1}^{r} b_{li}g_i = \sum_{i=1}^{r} b_{li} \sum_{j=1}^{m} c_{ij}f_j \\ &= \sum_{j=1}^{m} \left(\sum_{i=1}^{r} b_{li}c_{ij} \right) f_j = \sum_{j=1}^{m} b_{lj}^* f_j. \end{aligned}$$

This shows that $B^* \subseteq \boldsymbol{S}_F$, and thus $A \cup B^* \subseteq \boldsymbol{S}_F$. In order to show that $A \cup B^*$ generates $\boldsymbol{S}_F$ as $K[\underline{X}]$-module, let $\boldsymbol{h} = (h_1, \ldots, h_m) \in \boldsymbol{S}_F$ be arbitrary and define $\boldsymbol{h}_* = (h_{*1}, \ldots, h_{*r})$ by

$$h_{*i} = \sum_{j=1}^{m} h_j d_{ji}.$$

Then $\boldsymbol{h} \in \boldsymbol{S}_F$ implies

$$\begin{aligned} 0 &= \sum_{j=1}^{m} h_j f_j = \sum_{j=1}^{m} h_j \sum_{i=1}^{r} d_{ji}g_i \\ &= \sum_{i=1}^{r} \left(\sum_{j=1}^{m} h_j d_{ji} \right) g_i = \sum_{i=1}^{r} h_{*i}g_i, \end{aligned}$$

and so $\boldsymbol{h}_* \in S_G$. Now S_G is generated by B, and so $\boldsymbol{h}_*$ is a linear combination of elements of B, i.e., there exist $\alpha_1, \ldots, \alpha_s \in K[\underline{X}]$ with

$$\boldsymbol{h}_* = \sum_{l=1}^{s} \alpha_l \boldsymbol{b}_l.$$

Define $\boldsymbol{k} = (k_1, \ldots, k_m)$ by setting

$$k_j = \sum_{i=1}^{r} h_{*i} c_{ij} \quad \text{for} \quad 1 \leq j \leq m.$$

Then

$$k_j = \sum_{i=1}^{r} \left(\sum_{l=1}^{s} \alpha_l b_{li} \right) c_{ij} = \sum_{l=1}^{s} \alpha_l \sum_{i=1}^{r} b_{li} c_{ij} = \sum_{l=1}^{s} \alpha_l b_{lj}^*,$$

and so $\boldsymbol{k} = \sum_{l=1}^{s} \alpha_l \boldsymbol{b}_l^* \in \text{lin}(B^*)$. We claim that $\boldsymbol{h} - \boldsymbol{k}$ (where subtraction is performed componentwise) is in $\text{lin}(A)$. Indeed,

$$\begin{aligned} h_j - k_j &= h_j - \sum_{i=1}^{r} h_{*i} c_{ij} \\ &= h_j - \sum_{i=1}^{r} \left(\sum_{l=1}^{m} h_l d_{li} \right) c_{ij} \\ &= h_j - \sum_{l=1}^{m} h_l \left(\sum_{i=1}^{r} d_{li} c_{ij} \right) \\ &= \sum_{l=1}^{m} h_l \left(\delta_{lj} - \sum_{i=1}^{r} d_{li} c_{ij} \right) \\ &= \sum_{l=1}^{m} h_l a_{lj} \end{aligned}$$

for $1 \leq j \leq m$, and so $\boldsymbol{h} - \boldsymbol{k} \in \text{lin}(A)$. Together, we conclude that

$$\boldsymbol{h} = \boldsymbol{k} + (\boldsymbol{h} - \boldsymbol{k}) \in \text{lin}(A \cup B^*). \ \square$$

If the ground field K is computable, then it is a simple matter of going through the constructions described thus far to prove that a generating system for the module of syzygies of any given m-tuple of elements of $K[\underline{X}]$ can actually be computed: all that is needed are the algorithms REDPOL and EXTGRÖBNER (and EXTREDGRÖBNER if one wishes to work with reduced Gröbner bases) w.r.t. a term order of our choice.

Exercise 6.5 Let $(f_1, f_2, f_3) \subseteq (\mathbb{Q}[X, Y, Z])^3$ with $f_1 = XY + 1$, $f_2 = XZ + 1$, and $f_3 = Y^2 - YZ$. Compute a generating system for the first module of syzygies of (f_1, f_2, f_3).

Exercise 6.6 Write an algorithm that computes a generating system for the module of syzygies of any given m-tuple of elements of $K[\underline{X}]$.

Having dealt with the solution of a homogeneous linear equation, we will now discuss the general **linear equation**

$$Y_1 f_1 + \cdots + Y_m f_m = g \qquad (f_1, \ldots, f_m, g \in K[\underline{X}]). \tag{7}$$

As before, we will assume that the f_i are pairwise different. A *solution* of (7) is of course an m-tuple $(h_1, \ldots, h_m)$ of elements of $K[\underline{X}]$ such that

$$h_1 f_1 + \cdots + h_m f_m = g.$$

It is obvious that such a solution exists if and only if $g \in \operatorname{Id}(f_1, \ldots, f_m)$. If $\boldsymbol{h}$ is a solution of (7) and $\boldsymbol{h}_0$ is a solution of the corresponding homogeneous equation

$$Y_1 f_1 + \cdots + Y_m f_m = 0, \tag{8}$$

then it is easy to see that $\boldsymbol{h} + \boldsymbol{h}_0$ (where addition is performed componentwise) is again a solution of (7). Conversely, whenever $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ are solutions of (7), then $\boldsymbol{h}_1 - \boldsymbol{h}_2$ is a solution of (8). We have proved the following lemma.

Lemma 6.7 Let $\boldsymbol{S}$ be the set of solutions of the linear equation (7). Then $\boldsymbol{S} \neq \emptyset$ iff $g \in \operatorname{Id}(f_1, \ldots, f_m)$. In that case, $\boldsymbol{S} = \boldsymbol{h} + \boldsymbol{S}'$, where $\boldsymbol{h}$ is a solution of (7) and $\boldsymbol{S}'$ is the module of syzygies of $(f_1, \ldots, f_m)$. □

Now suppose K is computable. We wish to decide whether (7) has a solution and compute the set of solutions in case of a positive answer. The mere decision of solvability is an ideal membership test which we know how to do (Theorem 5.55). For the computation of a solution it suffices, by the lemma above and the earlier results of this section, to compute one special solution. This is easy enough if $F = \{f_1, \ldots, f_m\}$ is a Gröbner basis w.r.t. some term order. In that case, we simply do REDPOL(g, F). If the resulting normal form h of g is not zero, then $g \notin \operatorname{Id}(F)$ and (7) has no solution. Otherwise, REDPOL provides $h_1, \ldots, h_m \in K[\underline{X}]$ with

$$h_1 f_1 + \cdots + h_m f_m = g,$$

and we have found a solution. If F is not a Gröbner basis, then we may compute one w.r.t. a term order of our choice, say $G = \{g_1, \ldots g_r\}$, and using the algorithm EXTGRÖBNER, we also obtain polynomials c_{ij} with

$$g_i = \sum_{j=1}^{m} c_{ij} f_j \quad \text{for} \quad 1 \leq i \leq r.$$

Now we do REDPOL(g, G). If the resulting normal form h of g is not zero, then $g \notin \operatorname{Id}(G) = \operatorname{Id}(F)$ and (7) has no solution. Otherwise, REDPOL provides $q_1, \ldots, q_m \in K[\underline{X}]$ with

$$q_1 g_1 + \cdots + q_m g_r = g,$$

and a now familiar argument shows that a solution of (7) is given by $(h_1, \ldots, h_m)$ with

$$h_j = \sum_{i=1}^{r} q_i c_{ij} \quad \text{for} \quad 1 \leq j \leq m.$$

Exercise 6.8 Write an algorithm that decides solvability of a linear equation over $K[\underline{X}]$ and computes the set of solutions if the latter is not empty.

As an application, we can now strengthen Theorem 5.55 (ii) as follows.

Lemma 6.9 Let F be a finite subset of $K[\underline{X}]$, and let $I = \mathrm{Id}(F)$. Suppose the ground field is computable and the term order on T is decidable. Then there is an algorithm that decides whether an element of $K[\underline{X}]/I$ is a unit and computes the inverse if it exists. In particular, if I is maximal, then $K[\underline{X}]/I$ is a computable field.

Proof The residue class $g+I$ is a unit in $K[\underline{X}]/I$ iff there exist $h \in K[\underline{X}]$ and $q_f \in K[\underline{X}]$, one for each $f \in F$, such that

$$hg + \sum_{f \in F} q_f f = 1. \tag{9}$$

Solvability in $K[\underline{X}]$ of the equation

$$Yg + \sum_{f \in F} Y_f f = 1$$

can be decided and a solution can be computed if it exists. With the notation of (9), the desired inverse is then $h + I$. □

Note that a different Gröbner basis is needed for each computation of an inverse according to the lemma above. In practice, it would be advantageous to compute a Gröbner basis of F first, so that afterwards, the set $G \cup \{f\}$ is already "close" to being a Gröbner basis.

We close this section with a result that can be viewed as an abstract version of Buchberger's second criterion. For the rest of this section, $F = \{f_1, \ldots, f_m\}$ will be a finite subset of $K[\underline{X}]$, and $\leq$ will be a term order on $T = T(X_1, \ldots, X_n)$. We assume w.l.o.g. that all f_i are monic, and using the same notation as earlier on in this section, we let, for $1 \leq i < j \leq n$, the terms s_{ij} and s_{ji} be such that

$$\mathrm{lcm}(\mathrm{HT}(f_i), \mathrm{HT}(f_j)) = s_{ij} \cdot \mathrm{HT}(f_i) = s_{ji} \cdot \mathrm{HT}(f_j).$$

We then have

$$\mathrm{spol}(\mathrm{HT}(f_i), \mathrm{HT}(f_j)) = s_{ij} \cdot \mathrm{HT}(f_i) - s_{ji} \cdot \mathrm{HT}(f_j) = 0.$$

We see that the set $\mathrm{HT}(F) = \{\mathrm{HT}(f_1), \dots, \mathrm{HT}(f_m)\}$ is a Gröbner basis in $K[\underline{X}]$. (This has in fact already been observed in Corollary 5.49.) Moreover, Proposition 6.1 tells us that the $K[\underline{X}]$-module

$$\boldsymbol{S} = \left\{ (h_1, \dots, h_m) \in \left(K[\underline{X}]\right)^m \;\middle|\; \sum_{i=1}^{m} h_i \cdot \mathrm{HT}(f_i) = 0 \right\}$$

of syzygies of $\left(\mathrm{HT}(f_1), \dots, \mathrm{HT}(f_m)\right)$ is generated by the set

$$B = \{ \boldsymbol{r}_{ij} \mid 1 \leq i < j \leq n \},$$

where $\boldsymbol{r}_{ij} = (r_{ij1}, \dots, r_{ijm})$ with

$$r_{ijk} = \begin{cases} s_{ij} & \text{if } \; k = i \\ -s_{ji} & \text{if } \; k = j \\ 0 & \text{otherwise.} \end{cases}$$

This generating system is not in general a minimal one: whenever we are in the situation of Buchbeger's second criterion, say we have f_i, f_j, $f_k \in F$ satisfying the equivalent conditions (cf. Exercise 5.69)

$$\mathrm{lcm}\left(\mathrm{HT}(f_i), \mathrm{HT}(f_j)\right) = u \cdot \mathrm{lcm}\left(\mathrm{HT}(f_i), \mathrm{HT}(f_k)\right) \quad \text{with} \quad u \in T$$

and

$$\mathrm{lcm}\left(\mathrm{HT}(f_i), \mathrm{HT}(f_j)\right) = v \cdot \mathrm{lcm}\left(\mathrm{HT}(f_k), \mathrm{HT}(f_j)\right) \quad \text{with} \quad v \in T,$$

then, as one easily sees, $\boldsymbol{r}_{ij} = u \cdot \boldsymbol{r}_{ik} + v \cdot \boldsymbol{r}_{kj}$ (cf. the proof of Proposition 5.70). Loosely speaking, the content of Buchberger's second criterion was that in this situation, $\mathrm{spol}(f_i, f_j)$ can be disregarded when testing the original set F for the Gröbner basis property. The aim of the rest of this section is to show that whenever $C \subseteq B$ is a generating system for $\boldsymbol{S}$, then to conclude that F is a Gröbner basis, it suffices to know that $\mathrm{spol}(f_i, f_j) \xrightarrow[F]{*} 0$ for those pairs (i, j) of indices that satisfy $\boldsymbol{r}_{ij} \in C$. This will be proved in two steps. First, we prove a more general proposition on generating systems for $\boldsymbol{S}$ that satisfy a certain hypothesis; then we show that C as described above satisfies this hypothesis. The proof of the proposition will employ standard representations and t-representations; we will be using the "polynomial versions" as opposed to the "monomial versions" (see the remarks following the definitions in Section 5.4).

Proposition 6.10 (Lifting of Syzygies) *Let $\{\boldsymbol{d}_1, \dots, \boldsymbol{d}_s\}$ be a generating system for $\boldsymbol{S}$ with the following property: for all $1 \leq i < j \leq m$, there exist q_1, …, $q_s \in K[\underline{X}]$ with*

$$\boldsymbol{r}_{ij} = \sum_{k=1}^{s} q_k \cdot \boldsymbol{d}_k \tag{10}$$

and

$$\max\{\mathrm{HT}(f_l)\cdot\mathrm{HT}(q_k d_{kl}) \mid 1 \le l \le m\} \le \mathrm{lcm}\big(\mathrm{HT}(f_i),\mathrm{HT}(f_j)\big) \qquad (11)$$

for $1 \le k \le s$, *where* $\boldsymbol{d}_k = (d_{k1}, \ldots d_{km})$ *for* $1 \le k \le s$. *Then the following are equivalent:*

(i) F *is a Gröbner basis.*

(ii) For all $1 \le k \le s$,

$$\sum_{l=1}^{m} d_{kl} f_l \xrightarrow[F]{*} 0.$$

Proof The direction "(i)⟹(ii)" is trivial in view of Theorem 5.35 because the sum in (ii) is in $\mathrm{Id}(F)$. For the direction "(ii)⟹(i)," we verify the hypothesis of Theorem 5.64. Let $1 \le i < j \le m$ and $q_1, \ldots, q_s \in K[\underline{X}]$ satisfying (10) and (11). Recalling the definition of $\boldsymbol{r}_{ij}$, we conclude from (10) that

$$\begin{aligned} \mathrm{spol}(f_i, f_j) &= s_{ij} f_i - s_{ji} f_j \\ &= \sum_{l=1}^{m} r_{ijl} f_l \\ &= \sum_{l=1}^{m}\sum_{k=1}^{s} q_k d_{kl} f_l \\ &= \sum_{k=1}^{s} q_k \sum_{l=1}^{m} d_{kl} f_l\,. \qquad (12) \end{aligned}$$

Condition (ii) together with Lemma 5.60 provides us with representations

$$\sum_{l=1}^{m} d_{kl} f_l = \sum_{l=1}^{m} p_{kl} f_l \qquad (p_{kl} \in K[\underline{X}]) \qquad (13)$$

such that for $1 \le k \le s$,

$$\max\{\mathrm{HT}(p_{kl} f_l) \mid 1 \le l \le m\} \le \mathrm{HT}\Big(\sum_{l=1}^{m} d_{kl} f_l\Big). \qquad (14)$$

Substituting from (13) into (12), we obtain the representation

$$\mathrm{spol}(f_i, f_j) = \sum_{k=1}^{s} q_k \sum_{l=1}^{m} p_{kl} f_l = \sum_{l=1}^{m}\Big(\sum_{k=1}^{s} q_k p_{kl}\Big) f_l\,. \qquad (15)$$

We claim that this is a t-representation of $\mathrm{spol}(f_i, f_j)$ for some $t \in T$ with

$$t < \mathrm{lcm}\big(\mathrm{HT}(f_i),\mathrm{HT}(f_j)\big)$$

as desired. We set

$$t = \max\left\{ \mathrm{HT}\left(\sum_{k=1}^{s} q_k p_{kl} f_l\right) \,\middle|\, 1 \le l \le m \right\}$$

Then (15) is clearly a t-representation of $\mathrm{spol}(f_i, f_j)$. Furthermore, from the fact the $\boldsymbol{d}_k$ are elements of $\boldsymbol{S}$, we conclude that for $1 \le k \le s$,

$$\sum_{l=1}^{m} d_{kl} \cdot \mathrm{HT}(f_l) = 0,$$

and it follows easily that

$$\mathrm{HT}\left(\sum_{l=1}^{m} d_{kl} f_l\right) < \max\{\mathrm{HT}(d_{kl}) \cdot \mathrm{HT}(f_l) \mid 1 \le l \le m\}.$$

From this together with (11), we may conclude that for $1 \le k \le s$,

$$\mathrm{HT}\left(q_k \sum_{l=1}^{m} d_{kl} f_l\right) < \mathrm{lcm}\big(\mathrm{HT}(f_i), \mathrm{HT}(f_j)\big).$$

Combining this last inequality with (14), we see that for $1 \le l \le m$ and $1 \le k \le s$,

$$\mathrm{HT}(q_k p_{kl} f_l) < \mathrm{lcm}\big(\mathrm{HT}(f_i), \mathrm{HT}(f_j)\big),$$

and it now follows easily that indeed

$$t < \mathrm{lcm}\big(\mathrm{HT}(f_i), \mathrm{HT}(f_j)\big). \quad \square$$

The following simple example shows that the equivalence of the proposition above does not hold for arbitrary generating systems for S.

Example 6.11 Let $F = \{X, X+1\} \subseteq \mathbb{Q}[X]$. It is easy to see that the module S of "head term syzygies" is generated by the single element $(1, -1)$ in this case. It is now clear that the the set $\{(X+1, -X-1), (X, -X)\}$ is another generating system, and it is even a minimal one. The "liftings" of condition (ii) of Proposition 6.10 are $-X-1$ and $-X$, both of which reduce to 0 modulo F. But F is obviously not a Gröbner basis.

Lemma 6.12 Let $C = \{\boldsymbol{r}_{i_1 j_1}, \ldots, \boldsymbol{r}_{i_s j_s}\}$ be a subset of B that still generates $\boldsymbol{S}$. Then C satisfies the hypothesis of the previous proposition.

Proof For all $1 \le i < j \le m$, we define $\boldsymbol{p}_{ij} = (p_{ij1}, \ldots, p_{ijm})$ to be $\boldsymbol{r}_{ij}$ multiplied by $\mathrm{HT}(f_l)$ in the lth component, so that

$$p_{ijl} = \begin{cases} \mathrm{lcm}\big(\mathrm{HT}(f_i), \mathrm{HT}(f_j)\big) & \text{if } l = i \text{ or } l = j \\ 0 & \text{otherwise.} \end{cases}$$

Now let $1 \le i < j \le m$. Since C generates $\boldsymbol{S}$ and $\boldsymbol{r}_{ij} \in \boldsymbol{S}$, there exist $q_1, \ldots, q_s \in K[\underline{X}]$ such that

$$\boldsymbol{r}_{ij} = \sum_{k=1}^{s} q_k \cdot \boldsymbol{r}_{i_k j_k} .$$

If we write the equation componentwise and multiply the lth equation by $\operatorname{HT}(f_l)$ for $1 \le l \le m$, we obtain

$$\sum_{k=1}^{s} q_k p_{i_k j_k l} = p_{ijl} = \begin{cases} \operatorname{lcm}(\operatorname{HT}(f_i), \operatorname{HT}(f_j)) & \text{if } \ l = i \text{ or } l = j \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

Next, we define monomials m_k for $1 \le k \le s$ as follows. If there exists a monomial $a_k t_k$ in $M(q_k)$ with

$$t_k p_{i_k j_k i_k} = \operatorname{lcm}(\operatorname{HT}(f_i), \operatorname{HT}(f_j)),$$

then we let $m_k = a_k t_k$, and we set $m_k = 0$ otherwise. For each $1 \le k \le s$, the non-zero entries of the m-tuple $\boldsymbol{p}_{i_k j_k}$ (of which there are exactly two, namely, $p_{i_k j_k i_k}$ and $p_{i_k j_k j_k}$) agree, and so for $1 \le k \le s$ with $m_k \ne 0$,

$$t_k p_{i_k j_k l} = \begin{cases} \operatorname{lcm}(\operatorname{HT}(f_i), \operatorname{HT}(f_j)) & \text{if } \ l = i \text{ or } l = j \\ 0 & \text{otherwise.} \end{cases}$$

Comparing coefficients of $\operatorname{lcm}(\operatorname{HT}(f_i), \operatorname{HT}(f_j))$ in (16), we see that

$$\sum_{k=1}^{s} m_k p_{i_k j_k l} = p_{ijl} = \begin{cases} \operatorname{lcm}(\operatorname{HT}(f_i), \operatorname{HT}(f_j)) & \text{if } \ l = i \text{ or } l = j \\ 0 & \text{otherwise} \end{cases}$$

for $1 \le l \le m$. Now if we divide $\operatorname{HT}(f_l)$ back out of the lth equation, we obtain

$$\boldsymbol{r}_{ij} = \sum_{k=1}^{s} m_k \cdot \boldsymbol{r}_{i_k j_k} .$$

This representation of $\boldsymbol{r}_{ij}$ does indeed satisfy (11) of the previous proposition: if $1 \le k \le s$ with $m_k \ne 0$, then

$$\begin{aligned} \operatorname{HT}(f_l) \cdot \operatorname{HT}(m_k r_{i_k j_k l}) &= \begin{cases} \operatorname{HT}(f_l) \cdot t_k r_{i_k j_k l} & \text{if } \ l = i_k \text{ or } l = j_k \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} t_k p_{i_k j_k l} & \text{if } \ l = i_k \text{ or } l = j_k \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} \operatorname{lcm}(\operatorname{HT}(f_i), \operatorname{HT}(f_j)) & \text{if } \ l = i_k \text{ or } l = j_k \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

for $1 \le l \le m$. □

Combining the lemma and the proposition, we get the following theorem.

Theorem 6.13 *Let C be a subset of B that generates $\boldsymbol{S}$. Then the following are equivalent:*

(i) F is a Gröbner basis.

(ii) $\operatorname{spol}(f_i, f_j) \xrightarrow[F]{*} 0$ *for all pairs (i, j) of indices with $\boldsymbol{r}_{ij} \in C$.* □

6.2 Basic Algorithms in Ideal Theory

We have already seen how Gröbner bases can be used to decide whether $g \equiv h \bmod \mathrm{Id}(f_1, \ldots, f_m)$ for polynomials g, h, f_1, …, f_m with coefficients in a computable field: we must compute a Gröbner basis G of $\mathrm{Id}(f_1, \ldots, f_m)$ and then compare the unique normal forms of g and h modulo G (Theorem 5.55). In particular, this allows us to decide membership in a given ideal I, and to compute in the residue class ring modulo I. We also saw in the previous section how we can decide invertibility in the residue class ring and compute inverses where they exist. As another immediate application, we can decide inclusion of ideals:

$$\mathrm{Id}(f_1, \ldots, f_k) \subseteq \mathrm{Id}(g_1, \ldots, g_m)$$

is true if and only if $f_i \in \mathrm{Id}(g_1, \ldots, g_m)$ for $1 \leq i \leq k$. In this section, we discuss a number of more sophisticated algorithms that make use of Gröbner bases.

Throughout this section, K will be a field. We write $K[\underline{X}]$ for the polynomial ring $K[X_1, \ldots, X_n]$ and $T(\underline{X})$ for the set of all terms in the variables X_1, …, X_n. If $\{U_1, \ldots, U_r\} \subseteq \{X_1, \ldots, X_n\}$, then $T(\underline{U})$ is the set of those terms in $T(\underline{X})$ containing only variables in $\{U_1, \ldots, U_r\}$, with the convention that $T(\emptyset) = \{1\}$. $K[\underline{U}]$ is the subring of $K[\underline{X}]$ consisting of those polynomials $f \in K[\underline{X}]$ that satisfy $T(f) \subseteq T(\underline{U})$. In particular, $K[\emptyset] = K$. Similarly, $T(\underline{X} \setminus \underline{U})$ will denote the set of all those terms in $T(\underline{X})$ that contain only variables in $\{X_1, \ldots, X_n\} \setminus \{U_1, \ldots, U_r\}$.

Elimination ideals and properness

If I is an ideal in $K[\underline{X}]$ and $\{U_1, \ldots, U_r\} \subseteq \{X_1, \ldots, X_n\}$, then it is easy to see that $I \cap K[\underline{U}]$ is an ideal of the ring $K[\underline{U}]$. This ideal is called the **elimination ideal** of I w.r.t. $\{U_1, \ldots, U_r\}$, or w.r.t. $\underline{U}$ for short, and we will denote it by $I_{\underline{U}}$. If a term order $\leq$ on $T(\underline{X})$ is given and $\{U_1, \ldots, U_r\} \subseteq \{X_1, \ldots, X_n\}$, then we write

$$\underline{U} \ll \underline{X} \setminus \underline{U} \quad \text{for} \quad \{U_1, \ldots, U_r\} \ll \{X_1, \ldots, X_n\} \setminus \{U_1, \ldots, U_r\},$$

which means $s < t$ for all $s \in T(\underline{U})$ and $1 \neq t \in T(\underline{X} \setminus \underline{U})$. We see that we can always find a decidable term order $\leq$ on $T(\underline{X})$ satisfying $\underline{U} \ll \underline{X} \setminus \underline{U}$: just take for $\leq$ a lexicographical order where every variable in $\{U_1, \ldots, U_r\}$

is less than every one not in that set. Alternatively, we could take any pair of term orders $\leq_1$ on $T(\underline{U})$ and $\leq_2$ on $T(\underline{X} \setminus \underline{U})$ and combine the two lexicographically as in Example 5.8 (iv).

We remark that the notations involving $\underline{U}$ may be considered questionable from a formalist's point of view; however, we feel that they express the intended meaning with the least degree of possible confusion.

Lemma 6.14 Suppose $\{U_1, \ldots, U_r\} \subseteq \{X_1, \ldots, X_n\}$ and $\leq$ is a term order that satisfies $\underline{U} \ll \underline{X} \setminus \underline{U}$. Then the following hold:

(i) If $s \in T(\underline{X})$ and $t \in T(\underline{U})$ with $s < t$, then $s \in T(\underline{U})$.

(ii) If $f \in K[\underline{U}]$ and p, $g \in K[\underline{X}]$ with $f \xrightarrow[p]{} g$, then $p \in K[\underline{U}]$ and $g \in K[\underline{U}]$.

(iii) If $f \in K[\underline{U}]$ and $G \subseteq K[\underline{X}]$, then every normal form of f modulo G lies in $K[\underline{U}]$.

Proof (i) Assume for a contradiction that $s \notin T(\underline{U})$. Then s can be written as uv with $1 \neq v \in T(\underline{X} \setminus \underline{U})$. We obtain $uv = s < t < v$, a contradiction.

(ii) Since $\mathrm{HT}(p)$ divides some $t \in T(f)$, we must have $\mathrm{HT}(p) \in T(\underline{U})$ and thus $T(p) \subseteq T(\underline{U})$ by (i), i.e., $p \in K[\underline{U}]$. It now follows easily from the definition of reduction that $g \in K[\underline{U}]$ too. Statement (iii) can now easily be proved from (ii) by induction on the length of reduction chains. □

The next proposition will provide a way to compute elimination ideals. Recall that our convention is that $\mathrm{Id}(\emptyset) = \{0\}$, so that the empty set is a Gröbner basis of the zero ideal.

Proposition 6.15 *Let I be an ideal of $K[\underline{X}]$ and $\{U_1, \ldots, U_r\}$ a subset of $\{X_1, \ldots, X_n\}$. Assume further that $\leq$ is a term order on $T(\underline{X})$ that satisfies $\underline{U} \ll \underline{X} \setminus \underline{U}$, and G a Gröbner basis of I w.r.t. $\leq$. Then $G \cap K[\underline{U}]$ is a Gröbner basis of the elimination ideal $I_{\underline{U}}$.*

Proof Set $G \cap K[\underline{U}] = G'$. We show that every $0 \neq f \in I_{\underline{U}}$ is reducible modulo G'. Let $0 \neq f \in I_{\underline{U}}$. Then $f \in I$, and thus f is reducible modulo G, say $f \xrightarrow[g]{} h$ with $g \in G$. Lemma 6.14 (ii) tells us that $g \in G'$, and thus f is reducible modulo G'. □

The proposition above applied with $\{U_1, \ldots, U_r\} = \emptyset$ says that the elimination ideal $I \cap K$ is generated by $G \cap K$ for every Gröbner basis G of I. Since K is a field, this elimination ideal can only be $\{0\}$ or K; in the former case, $\mathrm{Id}(G)$ is proper, whereas in the latter case it is not. We thus obtain the following corollary.

Corollary 6.16 *Let I be an ideal of $K[\underline{X}]$. Then $I = K[\underline{X}]$ iff some Gröbner basis of I contains a constant iff every Gröbner basis of I contains a constant.* □

The following corollary uses the obvious fact that the intersection $F \cap K[\underline{U}]$ can be found by inspection whenever F is a finite subset of $K[\underline{X}]$ and $\{U_1, \ldots, U_r\} \subseteq \{X_1, \ldots, X_n\}$.

Corollary 6.17 *Assume that K is computable, let F be a finite subset of $K[\underline{X}]$, and let $\{U_1, \ldots, U_r\} \subseteq \{X_1, \ldots, X_n\}$. Then the algorithm* ELIMINATION *of Table* 6.1 *computes a Gröbner basis of the elimination ideal* $(\mathrm{Id}(F))_{\underline{U}}$. □

TABLE 6.1. Algorithm ELIMINATION

```
Specification: $G \leftarrow$ ELIMINATION($F, U_1, \ldots, U_r$)
                    Computation of the elimination ideal of Id($F$) w.r.t. $\underline{U}$
Given: $F$ = a finite subset of $K[\underline{X}]$ and $\{U_1, \ldots, U_r\} \subseteq \{X_1, \ldots, X_n\}$
Find: $G$ = a Gröbner basis of $(\mathrm{Id}(F))_{\underline{U}}$
begin
choose a decidable term order $\leq$ on $T(\underline{X})$ with $\underline{U} \ll \underline{X} \setminus \underline{U}$
$G' \leftarrow$ a Gröbner basis of Id($F$) w.r.t. $\leq$
$G \leftarrow G' \cap K[\underline{U}]$
end ELIMINATION
```

In the sequel, we will allow ourselves to write ELIMINATION$(F, \underline{U})$ instead of ELIMINATION$(F, U_1, \ldots, U_r)$. It is clear that by choosing the inverse lexicographical term order (where $X_n \gg X_{n-1} \gg \cdots \gg X_1$) in the above algorithm, we can simultaneously compute elimination ideals w.r.t. $\{X_1, \ldots, X_i\}$ where i ranges from 0 to n.

Corollary 6.18 *Assume that K is computable, and let F be a finite subset of $K[\underline{X}]$. Then the algorithm* PROPER *of Table* 6.2 *decides whether* Id(F) *is proper.* □

TABLE 6.2. Algorithm PROPER

```
Specification: $v \leftarrow$ PROPER($F$)
                    Decision whether or not Id($F$) is proper
Given: $F$ = a finite subset of $K[\underline{X}]$
Find: $v \in$ {true, false} such that
        $v$ = true iff Id($F$) is proper
begin
choose a decidable term order $\leq$ on $T(\underline{X})$
$G \leftarrow$ a Gröbner basis of Id($F$) w.r.t. $\leq$
if $G \cap K = \emptyset$ then return(true)
else return(false) end
end PROPER
```

Intersection of ideals

Let $Y_1, \ldots, Y_m$ be new indeterminates. We will use the obvious notation

$$K[\underline{X},\underline{Y}] = K[X_1, \ldots, X_n, Y_1, \ldots Y_m],$$

and similarly, we write $T(\underline{X},\underline{Y})$ for the set of terms in the variables X_1, $\ldots$, X_n, Y_1, $\ldots$, Y_m. Moreover, $\underline{X} \ll \underline{Y}$ will stand for

$$\{X_1, \ldots, X_n\} \ll \{Y_1, \ldots, Y_m\},$$

which of course means $s < t$ for all $s \in T(\underline{X})$ and $1 \neq t \in T(\underline{Y})$.

Proposition 6.19 *Let I_1, $\ldots$, I_m be ideals of $K[\underline{X}]$, and let*

$$J = \mathrm{Id}\left(\{1 - (Y_1 + \cdots + Y_m)\} \cup \bigcup_{i=1}^{m} Y_i I_i\right)$$

in the ring $K[\underline{X},\underline{Y}]$. Then $\bigcap_{i=1}^{m} I_i$ equals the elimination ideal $J_{\underline{X}}$.

Proof Let $f \in J_{\underline{X}}$. Then in particular, $f \in J$, so that we have a representation

$$f = g\left(1 - \sum_{i=1}^{m} Y_i\right) + \sum_{i=1}^{m} \sum_{j=1}^{k_i} g_{ij} Y_i s_{ij}$$

with $k_1, \ldots, k_m \in \mathbb{N}$, $g, g_{ij} \in K[\underline{X},\underline{Y}]$, and $s_{ij} \in I_i$ for $1 \leq i \leq m$ and $1 \leq j \leq k_i$. Now let $1 \leq k \leq m$. Setting $Y_k = 1$ and $Y_i = 0$ for $i \neq k$ leaves the left-hand side unchanged and turns the right-hand side into an element of I_k. Conversely, let $f \in \bigcap_{i=1}^{m} I_i$. Then the equation

$$f = f\left(1 - \sum_{i=1}^{m} Y_i\right) + \sum_{i=1}^{m} Y_i f$$

shows that $f \in J_{\underline{X}}$. □

If $I_1, \ldots I_m$ are ideals of $K[\underline{X}]$ with finite bases F_i, then it is clear that in the ring $K[\underline{X},\underline{Y}]$,

$$\mathrm{Id}\left(\bigcup_{i=1}^{m} Y_i I_i\right) = \mathrm{Id}\left(\bigcup_{i=1}^{m} Y_i F_i\right).$$

Corollary 6.20 *Assume that K is computable, and let F_1, $\ldots$, F_m be finite subsets of $K[\underline{X}]$. Then the algorithm* INTERSECTION *of Table* 6.3 *computes a Gröbner basis of the ideal $\bigcap_{i=1}^{m} \mathrm{Id}(F_i)$ of $K[\underline{X}]$.* □

Exercise 6.21 Let I_1 and I_2 be ideals of $K[\underline{X}]$, and let

$$J = \mathrm{Id}\big(YI_1, (Y-1)I_2\big).$$

Show that $I_1 \cap I_2$ equals the elimination ideal $J_{\underline{X}}$.

TABLE 6.3. Algorithm INTERSECTION

```
Specification: $G \leftarrow$ INTERSECTION($F_1, \dots, F_m$)
               Computation of the intersection $\bigcap_{i=1}^{m} \mathrm{Id}(F_i)$
Given: $F_1, \dots, F_m$ = finite subsets of $K[\underline{X}]$
Find: $G$ = a Gröbner basis of $\bigcap_{i=1}^{m} \mathrm{Id}(F_i)$
begin
$G \leftarrow$ ELIMINATION($\{1 - \sum_{i=1}^{m} Y_i\} \cup \bigcup_{i=1}^{m} Y_i F_i, \underline{X}$)
end INTERSECTION
```

Exercise 6.22 Compute the intersection of the ideals

$$I_1 = \mathrm{Id}(X_1^2 - 2, X_1 + X_2) \quad \text{and} \quad I_2 = \mathrm{Id}(X_1^2 - 2, X_1 - X_2)$$

in $\mathbb{Q}[X_1, X_2]$.

The intersection of ideals being settled, we now turn to the intersection of residue classes.

Intersection of Residue Classes and Interpolation

Let I_1, ..., I_m be ideals of $K[\underline{X}]$ and $\boldsymbol{f} = (f_1, \dots, f_m)$ an m-tuple of polynomials in $K[\underline{X}]$. As before, we let $Y_1, \dots, Y_m$ bc new indeterminates and set

$$J = \mathrm{Id}\left(\{1 - (Y_1 + \cdots + Y_m)\} \cup \bigcup_{i=1}^{m} Y_i I_i\right).$$

Furthermore, we let $f^* = \sum_{i=1}^{m} Y_i f_i \in K[\underline{X}, \underline{Y}]$ and

$$A_{\boldsymbol{f}} = \bigcap_{i=1}^{m} (f_i + I_i) \subseteq K[\underline{X}].$$

Now looking for an element of $A_{\boldsymbol{f}}$ is tantamount to looking for a solution of the system of congruences

$$f \equiv f_i \bmod I_i \qquad (1 \leq i \leq m).$$

The following theorem is thus a Chinese remainder theorem for $K[\underline{X}]$ (cf. Section 2.8).

Theorem 6.23 *Let $\leq$ be a term order on $T(\underline{X}, \underline{Y})$ that satisfies $\underline{X} \ll \underline{Y}$, let G be a Gröbner basis of the ideal J in $K[\underline{X}, \underline{Y}]$ w.r.t. $\leq$, and let h be the unique normal form of f^* modulo G. Then the following assertions are equivalent:*

(i) $A_{\boldsymbol{f}} \neq \emptyset$.

(ii) $h \in K[\underline{X}]$.

(iii) $h \in A_f$.

Moreover, $A_f = h + \bigcap_{k=1}^m I_k$*,* h *is minimal in* A_f *w.r.t. the quasi-order on* $K[\underline{X}]$ *induced by* $\leq$*, and for all* $g \in K[\underline{X}]$*, we have* $g \in A_f$ *iff* h *is the normal form of* g *modulo the Gröbner basis* $G \cap K[\underline{X}]$*.*

Proof (i)$\Longrightarrow$(ii): Let $g \in A_f$. Then $g - f_i \in I_i$ for $1 \leq i \leq m$, and so

$$g - f^* = \sum_{i=1}^m Y_i(g - f_i) + \left(1 - \sum_{i=1}^m Y_i\right) g \in J.$$

Since G is a Gröbner basis of J, this implies that h is the unique normal form of g modulo G. By Lemma 6.14 (iii), the fact that $g \in K[\underline{X}]$ implies that $h \in K[\underline{X}]$ as well.

(ii)$\Longrightarrow$(iii): From the assumption $h - f^* \in J$, we may conclude that there exist $k_1, \ldots, k_m, \in \mathbb{N}$, q, $q_{ij} \in K[\underline{X}, \underline{Y}]$, and $s_{ij} \in I_i$ $(1 \leq i \leq m,$ $1 \leq j \leq k_i)$ with

$$\begin{aligned} h - f^* &= h - \sum_{i=1}^m Y_i f_i \\ &= q\left(1 - \sum_{i=1}^m Y_i\right) + \sum_{i=1}^m \sum_{j=1}^{k_i} q_{ij} Y_i s_{ij}. \end{aligned}$$

For arbitrary but fixed k, we now set $Y_k = 1$ and $Y_i = 0$ for $i \neq k$. The above equation then shows that $h - f_k \in I_k$, and thus $h \in A_f$.

(iii)$\Longrightarrow$(i) is trivial.

Proving the equality $A_f = h + \bigcap_{k=1}^m I_k$ is a straightforward exercise. The rest of the theorem follows from the fact that $G \cap K[\underline{X}]$ is a Gröbner basis of $\bigcap_{k=1}^m I_k$. □

Corollary 6.24 *Assume that* K *is computable, and let* $\leq'$ *be a decidable term order on* $T(\underline{X})$*. If* $F_1, \ldots, F_m$ *are finite subsets of* $K[\underline{X}]$ *and* $f_1, \ldots, f_m \in K[\underline{X}]$*, then the algorithm* CRT *of Table* 6.4 *decides whether*

$$\bigcap_{i=1}^m f_i + \mathrm{Id}(F_i)$$

is empty, and if it is not, it outputs an element of this intersection that is minimal w.r.t. the quasi-order induced by $\leq$*.* □

Exercise 6.25 Write an algorithm that decides whether or not a given polynomial lies in the intersection of finitely many given residue classes.

TABLE 6.4. Algorithm CRT

```
Specification: $v \leftarrow \mathrm{CRT}(F_1,\ldots,F_m,f_1,\ldots,f_m)$
               Decision whether or not $A_f = \bigcap_{i=1}^m f_i + \mathrm{Id}(F_i) = \emptyset$,
               computation of a minimal element of $A_f$ if one exists
Given: $F_1,\ldots,F_m$ = finite subsets of $K[\underline{X}]$, $f_1,\ldots,f_m \in K[\underline{X}]$,
       $\leq'$ a decidable term order on $K[\underline{X}]$
Find: $v \in \{\mathbf{false}\} \cup (\{\mathbf{true}\} \times K[\underline{X}])$ such that
      $v = \begin{cases} \mathbf{false} & \text{if } A_f = \emptyset \\ (\mathbf{true}, f) \text{ with } f \in A_f \text{ minimal} & \text{otherwise} \end{cases}$
begin
choose a decidable term order $\leq$ on $T(\underline{X},\underline{Y})$ with
  $\leq \cap\, (T(\underline{X}))^2 \;=\; \leq'$ and $\underline{X} \ll \underline{Y}$
$G \leftarrow$ a Gröbner basis of $\mathrm{Id}(\{1 - \sum_{i=1}^m Y_i\} \cup \bigcup_{i=1}^m Y_iF_i)$ in
    $K[\underline{X},\underline{Y}]$ w.r.t. $\leq$
$f \leftarrow \sum_{i=1}^m Y_i f_i$
$h \leftarrow$ the normal form of $f$ w.r.t. $G$
if $h \in K[\underline{X}]$ then return((true,$h$))
else return(false) end
end CRT
```

A noteworthy feature of the algorithm CRT is that the computation of the Gröbner basis involves only the ideals but not the f_i; so it is particularly suitable if one wishes to vary the f_i but not the I_i. Moreover, it decides the non-emptiness of A_f in every specific case regardless of any conditions on the I_i which would guarantee solvability for arbitrary f. The lemma after the next gives such a condition (cf. the Chinese remainder theorem of Section 2.8).

Two ideals I and J of a ring are called **comaximal** if $1 \in I + J$.

Lemma 6.26 Let $I_1, \ldots, I_m$ be pairwise comaximal ideals of a ring R. Then the ideals

$$I_i \quad \text{and} \quad \bigcap_{\substack{j=1 \\ j\neq i}}^{m} I_i$$

are comaximal for each $1 \leq i \leq m$.

Proof If $1 \leq i \leq m$, then by assumption, there exist $p_{ij} \in I_i$ and $q_j \in I_j$ for $1 \leq j \leq m$ and $j \neq i$ with $1 = p_{ij} + q_j$. The equation

$$1 = 1^{m-1} = \prod_{\substack{j=1 \\ j\neq i}}^{m} (p_{ij} + q_j)$$

shows that

$$1 \in I_i + \bigcap_{\substack{j=1\\ j\neq i}}^{m} I_j . \quad \square$$

Lemma 6.27 Let $I_1, \ldots, I_m$ be pairwise comaximal ideals of a ring R. Then $\bigcap_{i=1}^m a_i + I_i \neq \emptyset$ for all $(a_1, \ldots, a_m) \in R^m$.

Proof As before, we use the notation

$$A_{\boldsymbol{a}} = \bigcap_{i=1}^{m} a_i + I_i \qquad \big(\boldsymbol{a} = (a_1, \ldots, a_m) \in R^m\big).$$

For $1 \leq i \leq m$, let $\boldsymbol{e}_i = (e_{i1}, \ldots, e_{im})$ where

$$e_{ij} = \begin{cases} 1 & \text{if } \; i = j \\ 0 & \text{otherwise.} \end{cases}$$

Now if $A_{\boldsymbol{e}_i} \neq \emptyset$, say $b_i \in A_{\boldsymbol{e}_i}$ for $1 \leq i \leq m$, then

$$\sum_{i=1}^{m} b_i a_i \in A_{\boldsymbol{a}}$$

and thus $A_{\boldsymbol{a}} \neq \emptyset$ for arbitrary $\boldsymbol{a} = \{a_1, \ldots, a_m\}$. Let $1 \leq i \leq m$. By the previous lemma,

$$1 \in I_i + \bigcap_{\substack{j=1\\ j\neq i}}^{m} I_j,$$

say $1 = h_1 + h_2$ with h_1 and h_2 in the first and second summand, respectively. We see that $h_2 \in A_{\boldsymbol{e}_i}$. $\square$

Recall from Lemma 2.17 (i) that for $f \in K[\underline{X}]$ and $a_1, \ldots, a_n \in K$, we have given a meaning to $f(a_1, \ldots, a_n)$: $f(a_1, \ldots, a_n) \in K$ is obtained by substituting a_i for X_i $(1 \leq i \leq n)$, and we will also refer to this as **evaluating f at** $(a_1, \ldots, a_n)$. If $\boldsymbol{a} = (a_1, \ldots, a_n) \in K^n$, then we also write $f(\boldsymbol{a})$ for $f(a_1, \ldots, a_n)$. We say that f **vanishes at $\boldsymbol{a}$** $\in K^n$ if $\boldsymbol{a}$ is a zero of f, i.e., $f(\boldsymbol{a}) = 0$. We saw in Section 2.8 that the Lagrange interpolation problem can be viewed as an instance of the Chinese remainder problem in $K[X]$. We will now use Theorem 6.23 to obtain a solution to the interpolation problem in several variables. With every n-tuple $\boldsymbol{a} = (a_1, \ldots, a_n)$ of elements of K we associate the **vanishing ideal $I_{\boldsymbol{a}}$** of $\boldsymbol{a}$:

$$I_{\boldsymbol{a}} = \mathrm{Id}(X_1 - a_1, \ldots, X_n - a_n).$$

Lemma 6.28 Let $\boldsymbol{a}$, $\boldsymbol{b} \in K^n$ with $\boldsymbol{a} \neq \boldsymbol{b}$, and $f \in K[\underline{X}]$. Then the following hold:

(i) The set $G = \{ X_i - a_i \mid 1 \leq i \leq n \}$ is a Gröbner basis of $I_{\boldsymbol{a}}$.

(ii) The unique normal form of f modulo G equals $f(\boldsymbol{a})$.

(iii) f vanishes at $\boldsymbol{a}$ iff $f \in I_{\boldsymbol{a}}$.

(iv) $I_{\boldsymbol{a}}$ and $I_{\boldsymbol{b}}$ are comaximal.

Proof Statement (i) is immediate from the fact that the head terms of any two different elements of G are disjoint.

(ii) We first note that the unique normal form r of f modulo G is a constant since G contains a polynomial with head term X_i for $1 \le i \le n$. The equation

$$f = \sum_{i=1}^{n} q_i(X_i - a_i) + r \qquad (q_i \in K[\underline{X}])$$

shows that $r = f(\boldsymbol{a})$.

(iii) This is immediate from (ii) and the fact that f reduces to 0 modulo G iff $f \in I_{\boldsymbol{a}}$.

(iv) From $\boldsymbol{a} \neq \boldsymbol{b}$ it follows that $a_i \neq b_i$ for some $1 \le i \le m$, and so

$$1 = \frac{1}{b_i - a_i}\Big((X_i - a_i) - (X_i - b_i)\Big) \in I_{\boldsymbol{a}} + I_{\boldsymbol{b}}. \quad \square$$

Combining the lemma with Theorem 6.23, Corollary 6.24, and Lemma 6.27, we obtain the following corollary.

Corollary 6.29 *Let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m \in K^n$ be pairwise different, $r_1, \ldots, r_m \in K$. Then there exists $f \in K[\underline{X}]$ with $f(\boldsymbol{a}_i) = r_i$ for $1 \le i \le m$. If K is computable, then the algorithm* CRT *applied to $I_{\boldsymbol{a}_1}, \ldots, I_{\boldsymbol{a}_m}$ and $r_1, \ldots, r_m$ computes such an f with the additional property that f is minimal w.r.t. the quasi-order induced by a chosen decidable term order on $T(\underline{X})$.* □

Exercise 6.30 Use the corollary above to compute a polynomial $f \in \mathbb{Q}[X_1, X_2]$ with $f(0,0) = 0$, $f(0,1) = -1$, $f(0,-1) = 1$, and $f(1,-1) = 3$. (Hint: Set up

$$J = \mathrm{Id}\left(\left\{1 - \sum_{i=1}^{4} Y_i\right\} \cup Y_1 \cdot I_{(1,-1)} \cup Y_2 \cdot I_{(0,-1)} \cup Y_3 \cdot I_{(0,1)} \cup Y_4 \cdot I_{(0,0)}\right),$$

and use the lexicographical term order with $Y_1 \gg Y_2 \gg Y_3 \gg Y_4 \gg X_1 \gg X_2$. Do not go for the full Gröbner basis. Just reduce completely; this will be enough to reduce f^* into $\mathbb{Q}[X_1, X_2]$.)

Ideal quotients and radical membership

Let I be an ideal of $K[\underline{X}]$, $F \subseteq K[\underline{X}]$. Then we define the **quotient** $I : F$ **of** I **by** F as

$$I : F = \{ g \in K[\underline{X}] \mid gf \in I \text{ for all } f \in F \}.$$

For ideals I and J, $I : J$ is also called the **ideal quotient of** I **by** J. If $F = \{f\}$, then we also write $I : f$ instead of $I : F$.

Exercise 6.31 Let I be an ideal of $K[\underline{X}]$, $F \subseteq K[\underline{X}]$. Show the following:

(i) $I : F$ is an ideal of $K[\underline{X}]$.

(ii) $I : F = \bigcap_{f \in F} I : f$.

(iii) $I : F = I : \mathrm{Id}(F)$, so only ideal quotients need to be considered.

Exercise 6.32 Let I_1, I_2, J_1, J_2 be ideals of $K[\underline{X}]$. Show the following:

(i) $(I_1 \cap I_2) : (J_1 + J_2) = (I_1 : J_1) \cap (I_2 : J_2) \cap (I_1 : J_2) \cap (I_2 : J_1)$.

(ii) If $I_1 \subseteq I_2$ and $J_2 \subseteq J_1$, then $I_1 : J_1 \subseteq I_2 : J_2$.

We now show how Gröbner bases can be used to compute $I : J$ from finite bases of I and J. If F_J is a finite basis of J, then we already know that

$$I : J = I : \mathrm{Id}(F_J) = I : F_J = \bigcap_{f \in F_J} I : f,$$

and so it will suffice to show how to compute $I : f$ for an ideal I and an element f of $K[\underline{X}]$.

Proposition 6.33 *Let* $F = \{f_1, \dots, f_m\} \subseteq K[\underline{X}]$, *and let* $f \in K[\underline{X}]$. *Assume that* $\boldsymbol{G} \subseteq (K[\underline{X}])^{m+1}$ *is a generating set for the module* $\boldsymbol{S}$ *of syzygies of* $(f, f_1, \dots, f_m)$, *and let*

$$H = \left\{ h \in K[\underline{X}] \;\middle|\; \text{there are } h_1, \dots, h_m \text{ with } (h, h_1, \dots, h_m) \in \boldsymbol{G} \right\}.$$

Then H *is a basis of the ideal* $\mathrm{Id}(F) : f$.

Proof If $h \in H$, then there are $h_1, \dots, h_m \in K[\underline{X}]$ with

$$hf + h_1 f_1 + \cdots + h_m f_m = 0,$$

and so clearly $hf \in \mathrm{Id}(F)$. We see that $H \subseteq \mathrm{Id}(F) : f$. To show that H generates $\mathrm{Id}(F) : f$, let $g \in \mathrm{Id}(F) : f$. Then there are $g_1, \dots, g_m \in K[\underline{X}]$ with

$$gf = g_1 f_1 + \cdots + g_m f_m.$$

This means that $\boldsymbol{g} = (g, -g_1, \dots, -g_m)$ is an element of $\boldsymbol{S}$. It follows that there exist $(m+1)$-tuples $\boldsymbol{g}_1, \dots, \boldsymbol{g}_k \in \boldsymbol{G}$ and $q_1, \dots, q_k \in K[\underline{X}]$ with

$$\boldsymbol{g} = \sum_{i=1}^{k} q_i \boldsymbol{g}_i,$$

where addition and multiplication are performed componentwise. Looking at the first component only, we recognize g as a sum of polynomial multiples of elements of H, i.e., as an element of $\mathrm{Id}(H)$. □

Together with the remarks preceding the proposition, we have proved the following.

Corollary 6.34 *Assume that K is computable, and let F_1 and F_2 be finite subsets of $K[\underline{X}]$. Then the algorithm* IDEALDIV1 *of Table* 6.5 *computes a Gröbner basis of the ideal* $\mathrm{Id}(F_1) : \mathrm{Id}(F_2)$. □

TABLE 6.5. Algorithm IDEALDIV1

```
Specification: $G \leftarrow$ IDEALDIV1($F_1, F_2$)
               Computation of the ideal $\mathrm{Id}(F_1) : \mathrm{Id}(F_2)$
Given: $F_1, F_2 =$ finite subsets of $K[\underline{X}]$
Find: a Gröbner basis $G$ of the ideal $\mathrm{Id}(F_1) : \mathrm{Id}(F_2)$
begin
for all $f \in F_2$ do
    $G_f \leftarrow$ a generating system for the module of syzygies
        of $(f, f_1, \dots, f_m)$, where $\{f_1, \dots, f_m\} = F_1$
    $H_f \leftarrow \{ h \in K[\underline{X}] \mid$ there exist $h_1, \dots, h_m$
        with $(h, h_1, \dots, h_m) \in G_f \}$
end
$G \leftarrow$ INTERSECTION($\{H_f\}_{f \in F_2}$)
end IDEALDIV1
```

Exercise 6.35 Find an algorithm for the computation of the intersection of ideals that uses the computation of syzygies.

Next we discuss a construction similar to the division of ideals that is needed for a number of algorithms in ideal theory.

Lemma 6.36 Let I be an ideal of $K[\underline{X}]$, $f \in K[\underline{X}]$. Then there exists $s \in \mathbb{N}$ with

$$I : f^s = I : f^{s+1} = \bigcup_{i \in \mathbb{N}} I : f^i.$$

Proof By Exercise 6.32 (ii), we have

$$I = I : f^0 \subseteq I : f \subseteq I : f^2 \subseteq \cdots .$$

Since $K[\underline{X}]$ is noetherian, this chain of ideals must eventually become constant. The claim of the lemma is now obvious. □

We will now show how the natural number s of the lemma and $I : f^s$ can be computed from I and f. We write $I : f^\infty$ for $\bigcup_{i \in \mathbb{N}} I : f^i$.

Proposition 6.37 *Let I be an ideal of $K[\underline{X}]$, $0 \neq f \in K[\underline{X}]$, and let J be the ideal $\mathrm{Id}(I, 1 - Yf)$ of $K[\underline{X}, Y]$. Then $I : f^\infty$ equals the elimination ideal $J_{\underline{X}}$. If $\{f_1, \dots, f_k\}$ is a basis of I and $\{g_1, \dots, g_m\}$ is a basis of $J_{\underline{X}}$ with*

$$g_i = h_i(1 - Yf) + \sum_{j=1}^{k} h_{ij} f_j \qquad (1 \leq i \leq m,\ h_i, h_{ij} \in K[\underline{X}, Y]),$$

then

$$s = \max\{ \deg_Y(h_{ij}) \mid 1 \le i \le m,\ 1 \le j \le k \}$$

satisfies $I : f^s = I : f^\infty$.

Proof If $g \in J_{\underline{X}}$, then $g = q_1 p + q_2(1 - Yf)$ with q_1, $q_2 \in K[\underline{X}, Y]$ and $p \in I$. Temporarily passing to the field of quotients $K(\underline{X}, Y)$ of $K[\underline{X}, Y]$, we may replace Y by $1/f$ and then multiply the equation by f^d, where $d = \deg_Y(q_1)$. We thus obtain an equation $f^d g = qp$ with $q \in K[\underline{X}]$. Conversely, let $g \in I : f^\infty$, say $f^d g \in I \subseteq J$. From $1 \equiv Yf \bmod J$ we conclude $1 \equiv (Yf)^d \bmod J$. We then have

$$g \equiv Y^d f^d g \equiv 0 \mod J.$$

It remains to show that $I : f^\infty = I : f^s$. Let $g \in I : f^\infty$. Then

$$\begin{aligned} g &= \sum_{i=1}^m q_i g_i \qquad (q_i \in K[\underline{X}]) \\ &= \sum_{i=1}^m q_i \left(h_i(1 - Yf) + \sum_{j=1}^k h_{ij} f_j \right) \end{aligned}$$

If we now once again replace Y by $1/f$ and multiply the equation by f^s, then we see that $f^s g \in I$. □

Recall that the algorithm EXTGRÖBNER is capable of producing representations of the polynomials of the Gröbner basis it computes in terms of the input basis.

Corollary 6.38 *Assume that K is computable. Let F be a finite subset of $K[\underline{X}]$ and $0 \neq f \in K[\underline{X}]$. Then the algorithm* IDEALDIV2 *of Table* 6.6 *computes a Gröbner basis of the ideal* $\mathrm{Id}(F) : f^\infty$ *as well as* $s \in \mathbb{N}$ *with* $\mathrm{Id}(F) : f^\infty = \mathrm{Id}(F) : f^s$. □

The exponent s that IDEALDIV2 computes need of course not be minimal because we have not placed any requirements on the h_{gp} at all. It is clear, however, that the least possible s could be found by trial and error using IDEALDIV1.

Exercise 6.39 Let $F = \{X_1^2 + X_3, X_2^2 + X_3\} \subseteq \mathbb{Q}[X_1, X_2, X_3]$ and $f = X_1 - X_2 \in \mathbb{Q}[X_1, X_2, X_3]$. Compute $\mathrm{Id}(F) : f^\infty$ and the least $s \in \mathbb{N}$ with $\mathrm{Id}(F) : f^\infty = \mathrm{Id}(F) : f^s$.

Exercise 6.40 Let $p_1, \ldots, p_m \in K[\underline{X}]$ be irreducible, and suppose

$$g = p_1^{\nu_1} \cdot \cdots \cdot p_m^{\nu_m} \qquad (\nu_1, \ldots, \nu_m \in \mathbb{N}),$$

and

$$f = p_1^{\mu_1} \cdot \cdots \cdot p_k^{\mu_k} \qquad (1 \le k \le m,\ \mu_1, \ldots, \mu_k \in \mathbb{N}).$$

Show that $\mathrm{Id}(g) : f^\infty = \mathrm{Id}(p_{k+1}^{\nu_{k+1}} \cdot \cdots \cdot p_m^{\nu_m})$ (where the empty product is defined to be 1), and that the least s with $\mathrm{Id}(g) : f^\infty = \mathrm{Id}(g) : f^s$ is given by

$$s = \min\{ \sigma \in \mathbb{N} \mid \sigma \cdot \mu_i \ge \nu_i \text{ for } 1 \le i \le k \}.$$

TABLE 6.6. Algorithm IDEALDIV2

```
Specification: $(G, s) \leftarrow$ IDEALDIV2$(F, f)$
               Computation of the ideal $\mathrm{Id}(F) : f^\infty$
               and $s \in \mathbb{N}$ with $\mathrm{Id}(F) : f^\infty = \mathrm{Id}(F) : f^s$
Given: $F$ = a finite subset of $K[\underline{X}]$ and $0 \neq f \in K[\underline{X}]$
Find: $(G, s)$ where $G$ = a Gröbner basis of $\mathrm{Id}(F) : f^\infty$,
      $s \in \mathbb{N}$ with $\mathrm{Id}(F) : f^\infty = \mathrm{Id}(F) : f^s$
begin
$G \leftarrow$ ELIMINATION$(F \cup \{1 - Yf\}, \underline{X})$
$s \leftarrow \max\{ \deg_Y(h_{gp}) \mid g \in G,\ p \in F \}$, where
    $g = h_g(1 - Yf) + \sum_{p \in F} h_{gp}p$ for all $g \in G$
end IDEALDIV2
```

Combining the above algorithm with Corollary 6.18, we can now decide whether or not some power of a polynomial f lies in a given ideal I: this is obviously equivalent to $1 \in I : f^\infty$. What we are looking at is of course a *radical membership test* (see Definition 4.12).

Corollary 6.41 *Assume that K is computable. Let F be a finite subset of $K[\underline{X}]$ and $0 \neq f \in K[\underline{X}]$. Then the algorithm* RADICALMEMTEST *of Table* 6.7 *decides whether or not there exists $s \in \mathbb{N}$ with $f^s \in \mathrm{Id}(F)$, and if so, it computes such an s.* □

TABLE 6.7. Algorithm RADICALMEMTEST

```
Specification: $v \leftarrow$ RADICALMEMTEST$(F, f)$
               Computation of $s \in \mathbb{N}$ with $f^s \in I$ if existent,
               message otherwise
Given: $F$ = a finite subset of $K[\underline{X}]$ and $0 \neq f \in K[\underline{X}]$
Find: $v \in$ {false} $\cup$ ({true} $\times \mathbb{N}$) such that
      $v$ = false implies $f^s \notin I$ for all $s \in \mathbb{N}$, and
      $v$ = (true, $s$) implies $f^s \in I$
begin
$(G, s) \leftarrow$ IDEALDIV2$(F, f)$
if $G \cap K = \emptyset$ then return(false)
else return(true, $s$) end
end RADICALMEMTEST
```

If one wishes to just decide membership in the radical without computing the exponent s, then it clearly suffices to do

$$\mathrm{PROPER}(F \cup \{1 - Yf\}).$$

Exercise 6.42 Let p be a prime number and $K = \mathbb{Z}/p\mathbb{Z}$. Let I be the ideal

$$\mathrm{Id}(X^p - 1, Y^p + 1)$$

of $K[X,Y]$. Use the algorithm RADICALMEMTEST to decide whether or not $X + Y \in \mathrm{rad}(I)$. Use Lemma 1.106 to confirm your answer.

SUBRING MEMBERSHIP

We have now seen how Gröbner bases can be used to decide ideal membership and radical membership. Next, we will show how they can be used to decide membership in the ring $K[f_1,\dots,f_m]$ obtained by adjoining $\{f_1,\dots,f_m\} \subseteq K[\underline{X}]$ to K within $K[\underline{X}]$. It follows immediately from Lemma 1.110 that

$$K[f_1,\dots,f_m] = \{\, h(f_1,\dots,f_m) \mid h \in K[Y_1,\dots,Y_m] \,\}.$$

Lemma 6.43 Let $Y_1,\dots,Y_m$ be new indeterminates, $f_1,\dots,f_m \in K[\underline{X}]$, and I the ideal $\mathrm{Id}(Y_1 - f_1,\dots,Y_m - f_m)$ of $K[\underline{X},\underline{Y}]$. Then the following hold:

(i) $h - h(f_1,\dots,f_m) \in I$ for all $h \in K[\underline{Y}]$.

(ii) $I \cap K[\underline{X}] = \{0\}$.

Proof For the proof of (i), it suffices to note that $Y_i - f_i \in I$ implies $Y_i \equiv f_i \bmod I$ for $1 \le i \le m$, and thus

$$h(Y_1,\dots,Y_m) \equiv h(f_1,\dots,f_m) \mod I.$$

For (ii), we consider a term order $\le$ on $T(\underline{X},\underline{Y})$ with $\underline{X} \ll \underline{Y}$. Then $G = \{Y_1 - f_1,\dots,Y_m - f_m\}$ is a Gröbner basis of I w.r.t. $\le$ by Lemma 5.66. We see that if $0 \ne h \in K[\underline{X}]$, then h is in normal form modulo G and thus $h \notin I$. □

Proposition 6.44 *With the assumptions of Lemma* 6.43, *let $\le$ be a term order on $T(\underline{X},\underline{Y})$ satisfying $\underline{Y} \ll \underline{X}$, G a Gröbner basis of I w.r.t. $\le$, $g \in K[\underline{X}]$. Then $g \in K[f_1,\dots,f_m]$ iff the unique normal form h of g modulo G is in $K[\underline{Y}]$, and in this case, $g = h(f_1,\dots,f_m)$.*

Proof Assume that $g \in K[f_1,\dots,f_m]$, say $g = h(f_1,\dots,f_m)$ where $h \in K[\underline{Y}]$. Then $h-g \in I$ by Lemma 6.43 (i), and so the unique normal form h_0 of h modulo G equals that of g. Lemma 6.14 (iii) tells us that $h_0 \in K[\underline{Y}]$. Conversely, assume that the normal form h of g modulo G lies in $K[\underline{Y}]$. We have $h - h(f_1,\dots,f_m) \in I$ by Lemma 6.43 (i), and this together with $g - h \in I$ implies $g - h(f_1,\dots,f_m) \in I$. It now follows from Lemma 6.43 (ii) that $g = h(f_1,\dots,f_m) \in K[f_1,\dots,f_m]$. □

Corollary 6.45 *Assume that K is computable, and let g, $f_1, \dots, f_m$ be elements of $K[\underline{X}]$. Then the algorithm* SUBRINGMEMTEST *of Table* 6.8 *decides whether $g \in K[f_1, \dots, f_m]$, and if the answer is positive, it computes*

$$h \in K[\underline{Y}] = K[Y_1, \dots, Y_m]$$

with $g = h(f_1, \dots, f_m)$. □

TABLE 6.8. Algorithm SUBRINGMEMTEST

```
Specification: $v \leftarrow$ SUBRINGMEMTEST$(g, f_1, \dots, f_m)$
               Decision whether $g \in K[f_1, \dots, f_m]$, if so,
               computation of $h \in K[\underline{Y}]$ with $g = h(f_1, \dots, f_m)$
Given: $g, f_1, \dots, f_m \in K[\underline{X}]$
Find: $v \in \{$false$\} \cup (\{$true$\} \times K[\underline{Y}])$ such that
      $v = ($true$, h)$ implies $g = h(f_1, \dots, f_m)$,
      $v =$ false implies $g \notin K[f_1, \dots, f_m]$
begin
choose a decidable term order on $T(\underline{X}, \underline{Y})$ with $\underline{Y} \ll \underline{X}$
$G \leftarrow$ a Gröbner basis of $\mathrm{Id}(Y_1 - f_1, \dots, Y_m - f_m)$ w.r.t. $\leq$
$h \leftarrow$ the normal form of $g$ w.r.t. $G$
if $h \in K[\underline{Y}]$ then return((true, $h$))
else return(false) end
end SUBRINGMEMTEST
```

In Section 10.7, we will obtain better results for a particular subring, namely, the one consisting of the *symmetric functions.*

6.3 Dimension of Ideals

The notion of the dimension of an ideal lies at the very heart of ideal theory and its connection with algebraic geometry; it will play an important role in the rest of this book. As before, K will be a field and

$$K[\underline{X}] = K[X_1, \dots, X_n].$$

Recall that for an ideal I of $K[\underline{X}]$ and $\{U_1, \dots, U_r\} \subseteq \{X_1, \dots, X_n\}$, $I_{\underline{U}}$ denotes the elimination ideal $I \cap K[\underline{U}]$, and $\underline{U} \ll \underline{X} \setminus \underline{U}$ means $s < t$ for all $s \in T(\underline{U})$ and $1 \neq t \in T(\underline{X} \setminus \underline{U})$.

Definition 6.46 Let I be a proper ideal of $K[\underline{X}]$ and $\{U_1, \dots, U_r\}$ a subset of $\{X_1, \dots, X_n\}$. Then $\{U_1, \dots, U_r\}$ is called **independent** modulo I if $I_{\underline{U}} = \{0\}$. Moreover, $\{U_1, \dots, U_r\}$ is called **maximally independent**

modulo I if it is independent modulo I and not properly contained in any other independent set modulo I. The **dimension** $\dim(I)$ of I is defined as

$$\dim(I) = \max\{\, |U| \mid U \subseteq \{X_1, \ldots, X_n\} \text{ independent modulo } I \,\}.$$

We will, rather obviously, call an ideal of $K[\underline{X}]$ **zero-dimensional** if it is proper and has dimension zero.

We see that given a finite basis of the ideal I, we can decide whether a given set of variables is independent modulo I by computing the corresponding elimination ideal; the dimension of I can then be computed by testing all subsets of $\{X_1, \ldots, X_n\}$ for independence. This method is of a rather unpleasant combinatorial complexity even when implemented in an intelligent way. We do not discuss details of such an implementation since a much more elegant way to compute dimensions and maximally independent sets will be presented in Section 9.3. However, it is important to note the following consequence of Proposition 6.15. Recall that by our definition, zero is never an element of a Gröbner basis.

Lemma 6.47 Let I be a proper ideal of $K[\underline{X}]$ and $\{U_1, \ldots, U_r\}$ a subset of $\{X_1, \ldots, X_n\}$. Then the following are equivalent:

(i) $\{U_1, \ldots, U_r\}$ is independent modulo the ideal I.

(ii) $G \cap K[\underline{U}] = \emptyset$ for some Gröbner basis G of I w.r.t. a term order satisfying $\underline{U} \ll \underline{X} \setminus \underline{U}$.

(iii) $G \cap K[\underline{U}] = \emptyset$ for every Gröbner basis G of I w.r.t. a term order satisfying $\underline{U} \ll \underline{X} \setminus \underline{U}$. □

It is clear that every independent set modulo an ideal I is contained in a maximally independent set modulo I. We will later see that for prime ideals P of $K[\underline{X}]$, the cardinality of every maximally independent set modulo P equals the dimension of P. The following example shows that this is not true in general.

Example 6.48 Let $K = \mathbb{Q}$, $n = 3$, $G = \{X_1X_3 + X_3, X_2X_3 + X_3\}$. Considering that X_1X_3 and X_2X_3, respectively, are necessarily the head terms of the two polynomials, one immediately verifies that G is a Gröbner basis w.r.t. every term order. Independent sets modulo $\mathrm{Id}(G)$ are thus $\{X_1\}$, $\{X_2\}$, $\{X_3\}$, and $\{X_1, X_2\}$. Among these, $\{X_3\}$ and $\{X_1, X_2\}$ are maximally independent, and $\dim(\mathrm{Id}(G)) = 2$.

The next lemma is immediate from the definitions.

Lemma 6.49 If I and J are proper ideals of $K[\underline{X}]$ with $I \subseteq J$, then $\dim(J) \leq \dim(I)$. □

Recall from Corollary 2.31 that a non-trivial univariate polynomial ideal over a field contains a unique monic element of minimal degree, namely, its monic generator. With the definition of the dimension and Proposition 6.15 in mind, it is now easy to prove the following important lemma.

Lemma 6.50 Let I be a proper ideal of $K[\underline{X}]$. Then the following hold:

(i) I is zero-dimensional iff it contains a non-constant univariate polynomial in each of the variables in $\{X_1, \ldots, X_n\}$. In this case, I actually contains a unique monic univariate polynomial f_i of minimal degree in each variable X_i, namely, the monic generator of the elimination ideal $I \cap K[X_i]$, and f_i is the polynomial of minimal degree in $G \cap K[X_i]$ whenever G is a Gröbner basis of I w.r.t. a term order $\leq$ satisfying $\{X_i\} \ll \{X_1, \ldots, X_n\} \setminus \{X_i\}$.

(ii) If I is zero-dimensional, then so is every proper ideal J of $K[\underline{X}]$ that contains I, and the elimination ideal $I_{\underline{U}}$ is a zero-dimensional ideal of $K[\underline{U}]$ for each subset $\{U_1, \ldots, U_r\}$ of $\{X_1, \ldots, X_n\}$. □

We see that the univariate polynomials of minimal degree in a zero-dimensional ideal can in principle be found by means of n Gröbner basis computations. A much more efficient way to achieve this that uses a single Gröbner basis w.r.t. any term order will be presented in Proposition 9.6.

For the rest of this section, let I be a proper ideal of the ring $K[\underline{X}]$ and $T = T(X_1, \ldots, X_n)$. We will now establish a connection between the dimension of I, the set $\mathrm{HT}(I)$ of head terms of elements of I (w.r.t. a term order), and properties of the K-vector space $K[\underline{X}]/I$ (Example 3.2 (iii)). To avoid confusion with the dimension $\dim(I)$ of the ideal I, we will denote the dimension of the K-vector space $K[\underline{X}]/I$ by $\dim_K(K[\underline{X}]/I)$. One of the main results will be that I is zero-dimensional iff $K[\underline{X}]/I$ is finite-dimensional.

If $\leq$ is a term order on T, then the set of **reduced terms** w.r.t. I and $\leq$ is defined as $T \setminus \mathrm{HT}(I)$ and denoted by $\mathrm{RT}(I)$.

Lemma 6.51 Let $\leq$ be a term order on T and G a Gröbner basis of I w.r.t. $\leq$. Then

$$\begin{aligned} \mathrm{RT}(I) &= \{\, t \in T \mid s \nmid t \text{ for all } s \in \mathrm{HT}(I) \,\} \\ &= \{\, t \in T \mid s \nmid t \text{ for all } s \in \mathrm{HT}(G) \,\}. \end{aligned}$$

Proof The first equation claims that $\mathrm{RT}(I) = T \setminus \mathrm{mult}(\mathrm{HT}(I))$. But if $f \in I$ and $s \in T$, then $sf \in I$ and $\mathrm{HT}(sf) = s \cdot \mathrm{HT}(f)$, and thus $\mathrm{HT}(I) = \mathrm{mult}(\mathrm{HT}(I))$. For the second equation, just recall that $\mathrm{mult}(\mathrm{HT}(G)) = \mathrm{HT}(I)$ by Theorem 5.35. □

The residue class $g + I \in K[\underline{X}]/I$ of an element $g \in K[\underline{X}]$ will from now on be denoted by $\overline{g}$. If $A \subseteq K[\underline{X}]$, then $\overline{A} = \{\, \overline{g} \mid g \in A \,\}$.

Proposition 6.52 *Let $\leq$ be any term order on T, and let $B = \overline{\mathrm{RT}(I)}$. Then B is a basis of the K-vector space $K[\underline{X}]/I$; moreover, the map $\mathrm{RT}(I) \longrightarrow B$ given by $t \longmapsto \overline{t}$ is bijective.*

Proof Recall that scalar multiplication in $K[\underline{X}]/I$ is defined by $a\cdot\overline{f} = \overline{af}$. We begin by showing that B is a generating system for $K[\underline{X}]/I$. Suppose G is a Gröbner basis of I w.r.t. $\leq$. Let $f \in K[\underline{X}]$, and let h be the normal form of f w.r.t. $\xrightarrow[G]{}$. Then $\overline{f} = \overline{h}$, and $T(h) \subseteq \mathrm{RT}(I)$. It follows that

$$\begin{aligned} \overline{f} &= \overline{h} \\ &= \overline{\sum_{t\in T(h)} a_t t} \qquad (a_t \in K) \\ &= \sum_{t\in T(h)} \overline{a_t t} \\ &= \sum_{t\in T(h)} a_t \cdot \overline{t}. \end{aligned}$$

It remains to show that B is linearly independent. Assume that there exists a linear combination

$$0 = \sum_{i=1}^{k} a_i \cdot \overline{t_i} \qquad (a_i \in K,\ t_i \in \mathrm{RT}(I))$$

where not all a_i $(1 \leq i \leq k)$ equal zero. We may assume w.l.o.g. that $a_1 \neq 0$ and $t_1 > t_i$ for $2 \leq i \leq k$. If we set

$$h = \sum_{i=1}^{k} a_i t_i\,,$$

then $h \neq 0$, $\mathrm{HT}(h) = t_1$, and $h \in I$ because $\overline{h} = 0$. By Theorem 5.35, there exists $s \in \mathrm{HT}(G)$ with $s\,|\,\mathrm{HT}(h) = t_1$, contradicting $t_1 \in \mathrm{RT}(I)$. The indicated map from $\mathrm{RT}(I)$ to B is clearly surjective. To see that it is injective, let $s, t \in \mathrm{RT}(I)$ with $s < t$ and $\overline{s} = \overline{t}$. Then $\overline{s - t} = 0$, so $s - t \in I$, and thus $t = \mathrm{HT}(s - t) \in \mathrm{HT}(I)$, a contradiction. □

B as in the proposition above is called the **canonical term basis** of $K[\underline{X}]/I$ w.r.t. $\leq$.

Let us inspect the proof of the last proposition a little more closely. The equation

$$\overline{f} = \overline{\sum_{t\in T(f)} a_t t} = \sum_{t\in T(f)} a_t \cdot \overline{t}$$

holds for any $0 \neq f \in K[\underline{X}]$. Using this and the same arguments as in the proof above, the following facts are straightforward consequences of the definitions of linear independence, the canonical term basis, and scalar multiplication in the K-vector space $K[\underline{X}]/I$. Here, $\leq$ is any term order on T and $\mathrm{RT}(I)$ is the set of reduced terms w.r.t. $\leq$.

Lemma 6.53 (i) If $0 \neq f \in I$, say $f = \sum_{t \in T(f)} a_t t$, then

$$0 = \sum_{t \in T(f)} a_t \cdot \bar{t}$$

and thus $\overline{T(f)}$ is linearly dependent in $K[\underline{X}]/I$.

(ii) If D is a subset of T such that $\overline{D}$ is linearly dependent in $K[\underline{X}]/I$, say

$$0 = \sum_{t \in D'} a_t \cdot \bar{t} \qquad (a_t \neq 0 \text{ for all } t \in D'),$$

where $D' \neq \emptyset$ is a finite subset of D, then $0 \neq f = \sum_{t \in D'} a_t t \in I$.

(iii) If $D \subseteq \mathrm{RT}(I)$, then $|D| = |\overline{D}|$.

(iv) If $f \in K[\underline{X}]$, then a representation of $\bar{f} \in K[\underline{X}]/I$ as a linear combination of elements of $\overline{\mathrm{RT}(I)}$ is given by

$$\bar{h} = \sum_{t \in T(h)} a_t \cdot \bar{t},$$

where h is a normal form of f modulo any Gröbner basis of I w.r.t. $\leq$.

(v) If K is computable, $\leq$ is decidable, and a Gröbner basis G of I w.r.t. $\leq$ has been computed, then for any given $f \in K[\underline{X}]$, one can effectively express $\bar{f} \in K[\underline{X}]/I$ as a linear combination of elements of the canonical term basis w.r.t. $\leq$. □

As an application of the above results, we can now give criteria for I to be zero-dimensional that are of great importance in the theory as well as for computational purposes.

Theorem 6.54 *Let I be a proper ideal of $K[\underline{X}]$. Then the following assertions are equivalent:*

(i) $\dim(I) = 0$.

(ii) $K[\underline{X}]/I$ is finite-dimensional as a K-vector space.

(iii) There exists a term order $\leq$ on $T(\underline{X})$ and a Gröbner basis G of I w.r.t. $\leq$ such that for each $1 \leq i \leq n$, there is $g_i \in G$ with $\mathrm{HT}(g_i) = X_i^{\nu_i}$ *for some $0 < \nu_i \in \mathbb{N}$.*

(iv) For every term order $\leq$ on $T(\underline{X})$ and every Gröbner basis G of I w.r.t. $\leq$ there exists, for each $1 \leq i \leq n$, $g_i \in G$ with $\mathrm{HT}(g_i) = X_i^{\nu_i}$ *for some $0 < \nu_i \in \mathbb{N}$.*

Proof (i)$\Longrightarrow$(iv): Let G be any Gröbner basis of I, and let $1 \leq i \leq n$. Since I is zero-dimensional, there exists $0 \neq f_i \in I \cap K[X_i]$. Moreover, f_i is reducible modulo G and not constant, and so G must contain a polynomial whose head term equals $X_i^{\nu_i}$ for some $0 < \nu_i \in \mathbb{N}$.

(iv)$\Longrightarrow$(iii) is trivial.

(iii)$\Longrightarrow$(ii): Let $B = \overline{\mathrm{RT}(I)}$ be the canonical term basis of $K[\underline{X}]/I$ w.r.t. $\leq$. It is clear that every $t \in \mathrm{RT}(I)$ must satisfy $\deg_{X_i}(t) < \nu_i$ for $1 \leq i \leq n$. But there are at most $\nu_1 \cdot \cdots \cdot \nu_n$ terms with this property, and we see that B is finite.

(ii)$\Longrightarrow$(i): To prove zero-dimensionality of I, we will show that I contains a non-zero univariate polynomial in each variable. Let $1 \leq i \leq n$ and consider the set $C_i = \{\overline{X_i^k} \mid k \in \mathbb{N}\}$. If C_i is finite, then there exist $k \neq l \in \mathbb{N}$ with $\overline{X_i^k} = \overline{X_i^l}$, and so $0 \neq X_i^k - X_i^l \in I$. If C_i is infinite, then (ii) together with Theorem 3.20 tells us that it is linearly dependent, and Lemma 6.53 (ii) guarantees the existence of $0 \neq f \in I \cap K[X_i]$. □

Proposition 8.27 will add another important equivalent condition to the characterization of zero-dimensional ideals. The fact that residue class rings of the type $K[\underline{X}]/I$ are K-vector spaces will be exploited further in Chapter 9.

Using (iii) and (iv) above, we can now decide whether or not I is zero-dimensional simply by computing a single Gröbner basis G w.r.t. a term order of our choice and looking at the head terms of G. The advanced method for computing dimensions which we will present in Section 9.3 basically states that the same is true for arbitrary dimension. The proof of (iii)$\Longrightarrow$(ii) above shows in fact a little more.

Corollary 6.55 *Suppose there exists a term order $\leq$ on $T(\underline{X})$ and a Gröbner basis G of I w.r.t. $\leq$ such that for each $1 \leq i \leq n$, there is $g_i \in G$ with* $\mathrm{HT}(g_i) = X_i^{\nu_i}$ *for some* $0 < \nu_i \in \mathbb{N}$. *Then*

$$\dim_K(K[\underline{X}]/I) \leq \nu_1 \cdot \cdots \cdot \nu_n. \quad \square$$

The following corollary is now immediate from the fact that

$$\mathrm{mult}\big(\mathrm{HT}(I)\big) = \mathrm{mult}\big(\mathrm{HT}(G)\big)$$

for any term order $\leq$ and Gröbner basis G of I w.r.t. $\leq$.

Corollary 6.56 *Let I be a proper ideal of $K[\underline{X}]$. Then the following assertions are equivalent:*

(i) $\dim(I) = 0$.

(ii) There exists a term order $\leq$ on $T(\underline{X})$ such that for each $1 \leq i \leq n$, there is $g_i \in I$ with $\mathrm{HT}(g_i) = X_i^{\nu_i}$ *for some* $0 < \nu_i \in \mathbb{N}$.

(iii) For every term order $\leq$ on $T(\underline{X})$ there exists, for each $1 \leq i \leq n$, $g_i \in I$ with $\mathrm{HT}(g_i) = X_i^{\nu_i}$ *for some* $0 < \nu_i \in \mathbb{N}$. □

Exercise 6.57 Let $K = \mathbb{Q}$, $n = 2$, $F = \{f_1, f_2\}$ with $f_1 = X^2 + Y + 1$ and $f_2 = 2XY + Y$. Show that $\dim(\mathrm{Id}(F)) = 0$.

Exercise 6.58 Write an algorithm that tests an ideal for zero-dimensionality based on Theorem 6.54.

6.4 Uniform Word Problems

In this section we show how Gröbner bases can be employed for the solution of a class of classical decision problems, namely, *word problems*. This section forms an aside within this book; none of the material presented here will be used in the rest of the book.

In order to give a general and rigorous definition of what a word problem is, one needs concepts from mathematical logic and model theory such as formal language, terms, formulas, models, and the like. However, in order to understand the problems presented here and their solutions by means of Gröbner bases, such formal rigor is not at all necessary. The following informal discussion of the first problem that we will discuss should sufficiently explain the idea that is being pursued here. Recall that "ring" always means "commutative ring with unity."

Let K be a field and R an extension ring of K. As usual, we will write $K[\underline{X}]$ for the polynomial ring $K[X_1, \ldots, X_n]$. If $\boldsymbol{c} = (c_1, \ldots, c_n) \in R^n$, then for any $f \in K[\underline{X}]$, Lemma 2.17 (i) defines $f(\boldsymbol{c})$ as an element of R. If f, $g \in K[\underline{X}]$, then it may or may not be true that $f(\boldsymbol{c}) = g(\boldsymbol{c})$ in R. Viewing $f(\boldsymbol{c})$ and $g(\boldsymbol{c})$ as nothing but strings of symbols, i.e., *words* formed according to certain rules from the letters of a certain alphabet, the equation $f(\boldsymbol{c}) = g(\boldsymbol{c})$ in R can be interpreted as saying that the words $f(\boldsymbol{c})$ and $g(\boldsymbol{c})$ have the same meaning in the ring R. Now let $f_0, \ldots, f_m$, $g_0, \ldots, g_m \in K[\underline{X}]$. Then according to the discussion above, the condition

(∗) "whenever R is an extension ring of K, $\boldsymbol{c} = (c_1, \ldots, c_n) \in R^n$, and $f_i(\boldsymbol{c}) = g_i(\boldsymbol{c})$ holds in R for $1 \leq i \leq m$, then $f_0(\boldsymbol{c}) = g_0(\boldsymbol{c})$ holds in R"

can be interpreted as saying, "whenever $X_1, \ldots, X_n$ are substituted for in an extension ring of K in such a way that f_i and g_i receive the same meaning for $1 \leq i \leq m$, then f_0 and g_0 receive the same meaning." It is precisely the condition (∗) whose decidability is proved in the first theorem of this section, and we have just demonstrated why this decision problem deserves to be called the **word problem** for extension rings of K. The qualification *uniform* in the title of the section alludes to the fact that we are looking for a single algorithm that decides (∗) for arbitrary given polynomials.

Two remarks are in order before we can state our first theorem. It is clear that the equation $f_i(\boldsymbol{c}) = g_i(\boldsymbol{c})$ which occurs in (∗) above is equivalent to

$(f_i - g_i)(c) = 0$. This means that we do not give up generality by considering conditions of the form $f(c) = 0$ only. The second remark concerns not just the next theorem. Although we will not assume any familiarity with the formalisms of mathematical logic and model theory, we will try to conform with the notation that is commonly used in connection with word problems. This means that condition $(*)$ above—with $f_i(c) = g_i(c)$ replaced by $f_i(c) = 0$ according to the previous remark—will be written as follows.

$(**)$ The implication

$$\forall x_1 \cdots \forall x_n \left(\bigwedge_{i=1}^{m} f_i(\underline{x}) = 0 \longrightarrow f_0(\underline{x}) = 0 \right)$$

holds in the class of all extension rings of K.

As long as we view $(**)$ as nothing but a different notation for $(*)$, this should need no further explanation, except perhaps that here, as in all formalized mathematics, the symbol "$\forall$" means "for all," "$\wedge$" means "and," "$\longrightarrow$" means "implies," and "$(\underline{x})$" stands for "$(x_1, \ldots, x_n)$."

We are now in a position to formulate the theorem that will allow us to decide condition $(**)$ above in case K is computable. The decision method is obtained by reducing the word problem to an ideal membership test. The notation $K[\underline{X}] = K[X_1, \ldots, X_n]$ where K is any field will be used throughout this section.

Theorem 6.59 *Let K be a field and f_0, f_1, …, $f_m \in K[\underline{X}]$. Then the following are equivalent:*

(i) The implication

$$\forall x_1 \cdots \forall x_n \left(\bigwedge_{i=1}^{m} f_i(\underline{x}) = 0 \longrightarrow f_0(\underline{x}) = 0 \right)$$

holds in the class of all extension rings of K.

(ii) $f_0 \in \mathrm{Id}(f_1, \ldots, f_m)$, where the ideal is taken in $K[\underline{X}]$.

Proof (i)$\Longrightarrow$(ii): We prove the contrapositive. Assume that (ii) does not hold, i.e., that $f_0 \notin I$, where we have set $I = \mathrm{Id}(f_1, \ldots, f_m)$. Then I is proper, and so we may form the residue class ring $K[\underline{X}]/I$. The canonical homomorphism

$$\begin{array}{rccc} \chi: & K[\underline{X}] & \longrightarrow & K[\underline{X}]/I \\ & f & \longmapsto & f + I \end{array}$$

is injective when restricted to K, for otherwise I would contain an invertible element of $K[\underline{X}]$ and would thus not be proper. We see that up to a natural isomorphism which is obtained by replacing the image of K in $K[\underline{X}]/I$ under χ with K itself, $K[\underline{X}]/I$ is an extension ring of K. Let us now consider the element $(\chi(X_1),\dots,\chi(X_n))$ of $(K[\underline{X}]/I)^n$. Then we have, for $1 \leq i \leq m$,

$$f_i\big(\chi(X_1),\dots,\chi(X_n)\big) = \chi\big(f_i(X_1,\dots,X_n)\big) = 0$$

because $f_i \in I$. On the other hand,

$$f_0\big(\chi(X_1),\dots,\chi(X_n)\big) = \chi\big(f_0(X_1,\dots,X_n)\big) \neq 0$$

because $f_0 \notin I$, and we have proved that condition (i) is violated with $R = K[\underline{X}]/I$ and $\mathbf{c} = (\chi(X_1),\dots,\chi(X_n))$.

(ii)$\Longrightarrow$(i): Condition (ii) means that there exist $q_1, \dots, q_m \in K[\underline{X}]$ with

$$f_0 = \sum_{i=1}^{m} q_i f_i \,.$$

Now let R be an extension ring of K and $\mathbf{c} \in R^n$ such that $f_1(\mathbf{c}) = \cdots = f_m(\mathbf{c}) = 0$. Then

$$f_0(\mathbf{c}) = \sum_{i=1}^{m} q_i(\mathbf{c}) f_i(\mathbf{c}) = 0\,. \quad \square$$

If K is computable, then there is an algorithm that decides condition (ii) of the theorem for arbitrary input $f_0, \dots, f_m$: simply use the ideal membership test which computes a Gröbner basis G of $\mathrm{Id}(f_1,\dots,f_m)$ and outputs a positive answer if $f_0 \xrightarrow[G]{*} 0$. Referring to the decidability of all instances of condition (i) by means of a single algorithm as the *word problem for extension rings of K*, we have thus proved the following corollary.

Corollary 6.60 *If K is a computable field, then the word problem for extension rings of K is decidable.* □

The next theorem is the basis for the solution of the word problem for extension fields of the given field K. Here, the word problem will be reduced to a radical membership test. As it turns out, the same solution applies to the word problem for integral domains extending K.

Theorem 6.61 *Let K be a field and $f_0, f_1, \dots, f_m \in K[\underline{X}]$. Then the following are equivalent:*

(i) The implication

$$\forall x_1 \cdots \forall x_n \Big(\bigwedge_{i=1}^{m} f_i(\underline{x}) = 0 \longrightarrow f_0(\underline{x}) = 0 \Big)$$

holds in the class of all extension fields of K.

(ii) The implication

$$\forall x_1 \cdots \forall x_n \left(\bigwedge_{i=1}^{m} f_i(\underline{x}) = 0 \longrightarrow f_0(\underline{x}) = 0 \right)$$

holds in the class of all integral domains extending K.

(iii) There exists $s \in \mathbb{N}$ with $f_0^s \in \mathrm{Id}(f_1, \ldots, f_m)$, where the ideal is taken in $K[\underline{X}]$.

Proof (i)$\Longrightarrow$(ii): For a proof of the contrapositive, assume that (ii) does not hold. Then there exists an integral domain R that extends K and $c \in R^n$ with

$$f_1(c) = \cdots = f_m(c) = 0 \quad \text{and} \quad f_0(c) \neq 0. \tag{$*$}$$

The domain R is naturally embedded in its quotient field Q_R. Replacing the image of R in Q_R under this embedding by R itself, we see that up to a natural isomorhism, Q_R is an extension field of K. Viewing ($*$) as taking place in Q_R, we see that condition (i) is violated.

(ii)$\Longrightarrow$(iii): Again, we will prove the contrapositive. Assume that (iii) does not hold, i.e., that $f_0^s \notin I$ for all $s \in \mathbb{N}$, where we have set $I = \mathrm{Id}(f_1, \ldots, f_m)$. Since the set $\{ f_0^s \mid s \in \mathbb{N} \}$ is closed under multiplication, Proposition 4.11 provides a prime ideal J of $K[\underline{X}]$ with $I \subseteq J$ and $f_0^s \notin J$ for all $s \in \mathbb{N}$. Since J is prime, the residue class ring $K[\underline{X}]/J$ can be formed and is an integral domain. We may now argue as in the proof of (i)$\Longrightarrow$(ii) of the previous theorem. The canonical homomorphism

$$\begin{array}{cccc} \chi: & K[\underline{X}] & \longrightarrow & K[\underline{X}]/J \\ & f & \longmapsto & f + J \end{array}$$

is injective when restricted to K, and so up to a natural isomorphism which is obtained by replacing the image of K in $K[\underline{X}]/J$ under χ with K itself, $K[\underline{X}]/J$ is an integral domain which extends K. We now consider the element $(\chi(X_1), \ldots, \chi(X_n))$ of $(K[\underline{X}]/J)^n$. Then we have, for $1 \leq i \leq m$,

$$f_i(\chi(X_1), \ldots, \chi(X_n)) = \chi(f_i(X_1, \ldots, X_n)) = 0$$

because $f_i \in I \subseteq J$. On the other hand,

$$f_0(\chi(X_1), \ldots, \chi(X_n)) = \chi(f_0(X_1, \ldots, X_n)) \neq 0$$

because $f_0 \notin J$, and we see that (ii) is violated.

(iii)$\Longrightarrow$(i): If (iii) holds, then there exist $s \in \mathbb{N}$ and $q_1, \ldots, q_m \in K[\underline{X}]$ with

$$f_0^s = \sum_{i=1}^{m} q_i f_i \,.$$

Now let L be an extension field of K and $\boldsymbol{c} \in L^n$ such that $f_1(\boldsymbol{c}) = \cdots = f_m(\boldsymbol{c}) = 0$. Then

$$(f_0(\boldsymbol{c}))^s = \sum_{i=1}^{m} q_i(\boldsymbol{c}) f_i(\boldsymbol{c}) = 0,$$

and so necessarily $f_0(\boldsymbol{c}) = 0$. □

It is clear that condition (iii) of the theorem above can be read as

$$f_0 \in \operatorname{rad}(\operatorname{Id}(f_1, \ldots, f_m)).$$

In view of the radical membership test of Corollary 6.41 (see also the remarks preceding and following that corollary), we have thus proved the following.

Corollary 6.62 *If K is a computable field, then the word problem for extension fields of K is decidable, and so is the one for integral domains extending K.* □

In Section 7.2, we will define and investigate a class of fields that are called *algebraically closed.* One may then consider the word problem for the class of algebraically closed extension fields of a given field K. It will turn out that condition (i) of the theorem above with "extension field" replaced by "algebraically closed extension field" is still equivalent to condition (iii) of the theorem. This result, which we will prove in Section 7.4, is known as the *Hilbert Nullstellensatz* (theorem on zeroes). From our present point of view, the Hilbert Nullstellensatz is nothing but the solution to another word problem. However, it will turn out that it is a considerably deeper and more important theorem than the results of this section and does therefore not belong here. We point out that because of its proximity to the Hilbert Nullstellensatz, the previous theorem is also referred to as the *weak Nullstellensatz,* i.e., the *weak theorem on zeroes.*

Before we turn to our next word problem, we prove another corollary to the last theorem which is perhaps another reason for calling the latter the weak theorem on zeroes.

Corollary 6.63 *Let K be a field and $f_1, \ldots, f_m \in K[\underline{X}]$. Then the following are equivalent:*

(i) There exists an extension field L of K and $\boldsymbol{c} \in L^n$ such that $f_i(\boldsymbol{c}) = 0$ for $1 \le i \le m$.

(ii) There exists an integral domain R extending K and $\boldsymbol{c} \in R^n$ such that $f_i(\boldsymbol{c}) = 0$ for $1 \le i \le m$.

(iii) $1 \notin \operatorname{Id}(f_1, \ldots f_m)$.

Proof If we set $f_0 = 1 \in K[\underline{X}]$, then the implication

$$\forall x_1 \cdots \forall x_n \left(\bigwedge_{i=1}^{m} f_i(\underline{x}) = 0 \longrightarrow f_0(\underline{x}) = 0 \right)$$

holds in an extension ring R of K if and only if its premise is always false, i.e., if and only if

$$f_1(\mathbf{c}) = \cdots = f_m(\mathbf{c}) = 0$$

does not hold for any $\mathbf{c} \in R^n$. It is now clear that (i)–(iii) of the corollary are precisely the negations of (i)–(iii), respectively, of the last theorem with $f_0 = 1$. □

Exercise 6.64 Prove the equivalence (i)$\Longleftrightarrow$(iii) of the corollary above directly. (Hint: Use the fact that every proper ideal can be extended to a maximal one.)

In view of the fact that we can decide properness of polynomial ideals over a computable field (cf. Corollary 6.18), we have just proved that we can decide whether or not finitely many given polynomials over a computable field have a common zero in some extension field of K. For those who already have an understanding of algebraically closed fields, we mention that one of the central results of Section 7.4 will be as follows: if $f_1, \ldots, f_m \in K[\underline{X}]$, then $\mathrm{Id}(f_1, \ldots, f_m)$ is proper iff $f_1, \ldots, f_m$ have a common zero in *every algebraically closed* extension field of K.

The word problems that we have considered thus far were all for classes of structures extending a given field K. The words in question were polynomials over K evaluated in certain extensions of K. Next, we consider the word problem for the class of all rings (i.e., commutative rings with unity). At first glance, this does not even seem to be a meaningful question, because it is not clear what the words could be in the absence of a common ground field. To see how we can make sense out of this word problem, recall from Section 1.9 that for any ring R, the map

$$\begin{array}{cccc} \varphi: & \mathbb{Z} & \longrightarrow & R \\ & n & \longmapsto & n \cdot 1_R \end{array}$$

is a homomorphism of rings. If $\mathbf{c} \in R^n$, then according to Proposition 2.15, φ extends uniquely to a homomorphism $\varphi_{\mathbf{c}} : \mathbb{Z}[\underline{X}] \longrightarrow R$ which is obtained by first mapping coefficients by means of φ and then evaluating at $\mathbf{c}$. Now if $f_0, \ldots, f_m \in \mathbb{Z}[\underline{X}]$, then we define our "word implication"

$$\forall x_1 \cdots \forall x_n \left(\bigwedge_{i=1}^{m} f_i(\underline{x}) = 0 \longrightarrow f_0(\underline{x}) = 0 \right)$$

to hold in R if for all $\mathbf{c} \in R^n$,

$$\varphi_{\mathbf{c}}(f_1) = \cdots = \varphi_{\mathbf{c}}(f_m) = 0 \quad \text{implies} \quad \varphi_{\mathbf{c}}(f_0) = 0.$$

With this understanding, we can easily reduce the word problem for rings to a certain ideal membership test. The proof of the theorem below differs from the one of Theorem 6.59 only by some formal subtleties. We will be using the obvious fact that for any homomorphism $\psi : R \longrightarrow S$ of rings,

$$(n \cdot 1_S)\psi(r) = n \cdot \psi(r) = \psi(n \cdot r)$$

for all $n \in \mathbb{Z}$ and $r \in R$.

Theorem 6.65 *Let $f_0, f_1, \ldots, f_m \in \mathbb{Z}[\underline{X}]$. Then the following are equivalent:*

(i) The implication

$$\forall x_1 \cdots \forall x_n \left(\bigwedge_{i=1}^{m} f_i(\underline{x}) = 0 \longrightarrow f_0(\underline{x}) = 0 \right)$$

holds in the class of all commutative rings with unity.

(ii) $f_0 \in \mathrm{Id}(f_1, \ldots, f_m)$, where the ideal is taken in $\in \mathbb{Z}[\underline{X}]$.

Proof (i)$\Longrightarrow$(ii): We prove the contrapositive. Assume that (ii) does not hold, i.e., that $f_0 \notin I$, where $I = \mathrm{Id}(f_1, \ldots, f_m)$. Then I is proper, and so we may form the residue class ring $\mathbb{Z}[\underline{X}]/I$. Consider the canonical homomorphism

$$\begin{array}{rccc} \chi : & \mathbb{Z}[\underline{X}] & \longrightarrow & \mathbb{Z}[\underline{X}]/I \\ & f & \longmapsto & f + I, \end{array}$$

and set $\mathbf{c} = (\chi(X_1), \ldots, \chi(X_n)) \in (\mathbb{Z}[\underline{X}])^n$. Then we have, for arbitrary $g = \sum_{j=1}^{m} a_j X_1^{\nu_{j1}} \cdot \cdots \cdot X_n^{\nu_{jn}} \in \mathbb{Z}[\underline{X}]$,

$$\begin{aligned} \varphi_{\mathbf{c}}(g) &= \varphi_{\mathbf{c}}\left(\sum_{j=1}^{m} a_j X_1^{\nu_{j1}} \cdot \cdots \cdot X_n^{\nu_{jn}} \right) \qquad (a_j \in \mathbb{Z} \text{ for } 1 \le j \le m) \\ &= \sum_{j=1}^{m} a_j \cdot (\chi(X_1))^{\nu_{j1}} \cdot \cdots \cdot (\chi(X_n))^{\nu_{jn}} \\ &= \chi\left(\sum_{j=1}^{m} a_j \cdot X_1^{\nu_{j1}} \cdot \cdots \cdot X_n^{\nu_{jn}} \right) = \chi(g). \end{aligned}$$

It is now easy to see that $\varphi_{\mathbf{c}}(f_i) = 0$ because $f_i \in I$ for $1 \le i \le m$, while $\varphi_{\mathbf{c}}(f_0) \neq 0$ because $f_0 \notin I$.

(ii)$\Longrightarrow$(i): Condition (ii) says that there exist $q_1, \ldots, q_m \in \mathbb{Z}[\underline{X}]$ with

$$f_0 = \sum_{i=1}^{m} q_i f_i \,.$$

Now let R be a ring and $\boldsymbol{c} \in R^n$ such that $\varphi_{\boldsymbol{c}}(f_i) = 0$ for $1 \leq i \leq m$. Then

$$\varphi_{\boldsymbol{c}}(f_0) = \sum_{i=1}^{m} \varphi_{\boldsymbol{c}}(q_i)\varphi_{\boldsymbol{c}}(f_i) = 0 . \quad \square$$

The theorem above reduces the word problem for rings to the ideal membership test in $\mathbb{Z}[\underline{X}]$. Our Gröbner basis theory thus far, however, allows the ideal membership test only for polynomial rings over a *field*. An algorithmic realization of the ideal membership test in $\mathbb{Z}[\underline{X}]$ will be presented in Section 10.1. Since the results of the present section will not be used in the sequel, there is no harm in making a forward quote by stating the following corollary.

Corollary 6.66 *The word problem for commutative rings with unity is decidable.* □

Next, we consider the word problem for a special class of rings. Let R be a ring and $a \in R$. Then a is called **idempotent** if $a^2 = a$. A ring R is called **Boolean** if every $a \in R$ is idempotent. An obvious example is the ring $\mathbb{Z}/2\mathbb{Z}$; moreover, if $R_1, \ldots, R_m$ are Boolean rings, then the direct product $R_1 \times \cdots \times R_m$ of Proposition 1.113 is again a Boolean ring. The following lemma collects some facts about idempotents and Boolean rings that are needed for the next theorem. Recall from Section 1.9 that a ring of characteristic m contains an isomorphic copy of the ring $\mathbb{Z}/m\mathbb{Z}$. In the special case $\operatorname{char}(R) = 2$, the isomorphic copy of $\mathbb{Z}/2\mathbb{Z}$ contained in R is the subring $\{0_R, 1_R\}$ of R.

Lemma 6.67 Let R be a ring. Then the following hold:

(i) If a and b are idempotents of R, then so is ab. If in addition, $\operatorname{char}(R) = 2$, then $-a$ and $a + b$ are idempotents of R.

(ii) Let $\operatorname{char}(R) = 2$, and let F be the subfield $\{0, 1\}$ of R. If $M \subseteq R$ is a set of idempotents of R, then the subring $F[M]$ generated by M in R is a Boolean ring.

(iii) If R is Boolean, then $\operatorname{char}(R) = 2$.

(iv) If R is a Boolean domain, then $R = \{0, 1\}$.

Proof (i) If a and b are idempotents, then $(ab)^2 = a^2b^2 = ab$, and so ab is idempotent. Now assume that in addition, $\operatorname{char}(R) = 2$. Then $a + a = 0$ which means that $-a = a$, and so $-a$ is idempotent. Moreover,

$$(a + b)^2 = a^2 + ab + ab + b^2 = a^2 + b^2 = a + b,$$

and so $a + b$ is idempotent.

(ii) We first note that 1 and 0 are idempotents. This together with (i) above shows that the set S of all idempotents of R is in fact a subring of R. The ring $F[M]$ is the intersection of all those subrings of R that contain F and M. Since S occurs in this intersection, we must have $F[M] \subseteq S$.

(iii) Since R is Boolean, we have

$$1_R + 1_R = (1_R + 1_R)^2 = 1_R^2 + 1_R^2 + 1_R^2 + 1_R^2,$$

and so $1_R + 1_R = 0$.

(iv) If $a \in R$, then $a(a-1) = a^2 - a = 0$, and so $a = 0$ or $a = 1$. □

Exercise 6.68 Let R be a ring and $a \in R$ idempotent. Show that $a^s = a$ for all $s \in \mathbb{N}^+$.

There is a rather obvious way of posing a word problem for the class of rings of characteristic 2. Strictly speaking, the class of extension rings of $\mathbb{Z}/2\mathbb{Z}$ is smaller than the class of all rings of characteristic 2. But every ring R in the larger class contains an isomorphic copy F_R of $\mathbb{Z}/2\mathbb{Z}$ and is thus naturally isomorphic to one in the smaller class. This shows that there is no harm in identifying F_R with $\mathbb{Z}/2\mathbb{Z}$ via the natural isomorphism. Under this point of view, the class of all rings of characteristic 2 equals the class of all extension rings of $\mathbb{Z}/2\mathbb{Z}$. Let R be a ring in this class. If $f_0, \ldots, f_m \in \mathbb{Z}/2\mathbb{Z}[\underline{X}]$, then we may, as in Theorem 6.59, define the word implication

$$\forall x_1 \cdots \forall x_n \left(\bigwedge_{i=1}^{m} f_i(\underline{x}) = 0 \longrightarrow f_0(\underline{x}) = 0 \right)$$

to hold in R if for all $\mathbf{c} \in R^n$,

$$f_1(\mathbf{c}) = \cdots = f_m(\mathbf{c}) = 0 \quad \text{implies} \quad f_0(\mathbf{c}) = 0.$$

The resulting word problem has been solved in Theorem 6.59. In view of (iii) of the last lemma, we may pose a new word problem now by restricting ourselves to the class of all Boolean rings. The next theorem shows how this problem can in fact be reduced to the general one.

Theorem 6.69 *Let $f_0, \ldots, f_m \in \mathbb{Z}/2\mathbb{Z}[\underline{X}]$. Then the following are equivalent:*

(i) The implication

$$\forall x_1 \cdots \forall x_n \left(\bigwedge_{i=1}^{m} f_i(\underline{x}) = 0 \longrightarrow f_0(\underline{x}) = 0 \right)$$

holds in the class of all Boolean rings.

(ii) *The implication*

$$\forall x_1 \cdots \forall x_n \left(\bigwedge_{i=1}^{m} f_i(\underline{x}) = 0 \wedge \bigwedge_{i=1}^{n} x_i^2 - x_i = 0 \longrightarrow f_0(\underline{x}) = 0 \right)$$

holds in the class of all rings of characteristic 2.

Proof (ii)⟹(i): This is immediate from the fact that every Boolean ring R has characteristic 2 and satisfies $a^2 - a = 0$ for all $a \in R$.

(i)⟹(ii): Let R be a ring with $\operatorname{char}(R) = 2$, and let $c = (c_1, \ldots, c_n) \in R^n$ such that

$$f_1(c) = \cdots = f_m(c) = 0 \quad \text{and} \quad c_1^2 - c_1 = \cdots = c_m^2 - c_m = 0.$$

Then c_1, …, c_n are idempotents, and so by (ii) of the previous lemma, $S = \mathbb{Z}/2\mathbb{Z}[c_1, \ldots, c_n]$ is a Boolean subring of R. It now follows from our hypothesis (i) that $f_0(c) = 0$ in S and hence in R. □

The theorem above combined with Theorem 6.59 provides the decidability of the word problem for Boolean rings: if f_0, …, $f_m \in \mathbb{Z}/2\mathbb{Z}[\underline{X}]$, then the implication of (i) of the last theorem holds iff

$$f_0 \in \operatorname{Id}(f_1, \ldots, f_m, X_1^2 - X_1, \ldots, X_n^2 - X_n),$$

where the ideal is taken in $\mathbb{Z}/2\mathbb{Z}[\underline{X}]$. The next theorem provides an entirely different solution to the same word problem.

Theorem 6.70 *Let* f_0, …, $f_m \in \mathbb{Z}/2\mathbb{Z}[\underline{X}]$. *Then the following are equivalent:*

(i) *The implication*

$$\forall x_1 \cdots \forall x_n \left(\bigwedge_{i=1}^{m} f_i(\underline{x}) = 0 \longrightarrow f_0(\underline{x}) = 0 \right)$$

holds in the class of all Boolean rings.

(ii) *The implication*

$$\forall x_1 \cdots \forall x_n \left(\bigwedge_{i=1}^{m} f_i(\underline{x}) = 0 \longrightarrow f_0(\underline{x}) = 0 \right)$$

holds in $\mathbb{Z}/2\mathbb{Z}$.

Proof (i)⟹(ii): This is trivial because $\mathbb{Z}/2\mathbb{Z}$ is a Boolean ring.

(ii)⟹(i): We prove the contrapositive. Suppose there exists a Boolean ring R and $c \in R^n$ such that

$$f_1(c) = \cdots = f_m(c) = 0 \quad \text{and} \quad f_0(c) \neq 0.$$

Then $(f_0(c))^s = f_0(c) \neq 0$ for all $s \in \mathbb{N}$, and so the previous theorem together with the remark following it tells us that

$$f_0^s \notin \mathrm{Id}(f_1, \ldots, f_m, X_1^2 - X_1, \ldots, X_n^2 - X_n)$$

for all $s \in \mathbb{N}$. We may now apply Theorem 6.61 to conclude that the implication

$$\forall x_1 \cdots \forall x_n \left(\bigwedge_{i=1}^{m} f_i(\underline{x}) = 0 \wedge \bigwedge_{i=1}^{n} x_i^2 - x_i = 0 \longrightarrow f_0(\underline{x}) = 0 \right)$$

fails in some integral domain R extending $\mathbb{Z}/2\mathbb{Z}$, i.e., there exists $b = (b_1, \ldots, b_n) \in R^n$ with

$$f_1(b) = \cdots = f_m(b) = 0, \quad b_1^2 - b_1 = \cdots = b_m^2 - b_m = 0,$$

and $f_0(b) \neq 0$. Since R is a domain, the equations $b_i^2 - b_i = b_i(b_i - 1) = 0$ imply that $b_i \in \{0_R, 1_R\} = \mathbb{Z}/2\mathbb{Z}$ for $1 \leq i \leq n$, and so the implication

$$\forall x_1 \cdots \forall x_n \left(\bigwedge_{i=1}^{m} f_i(\underline{x}) = 0 \longrightarrow f_0(\underline{x}) = 0 \right)$$

fails in $\mathbb{Z}/2\mathbb{Z}$. □

The last theorem shows that the validity of a word implication

$$\forall x_1 \cdots \forall x_n \left(\bigwedge_{i=1}^{m} f_i(\underline{x}) = 0 \longrightarrow f_0(\underline{x}) = 0 \right)$$

in the class of Boolean rings can also be decided by testing the 2^n many possible substitutions for $x_1, \ldots, x_n$ in $\mathbb{Z}/2\mathbb{Z}$. What we have proved in two different ways is the following corollary.

Corollary 6.71 *The word problem for the class of Boolean rings is decidable.* □

For the reader who is familiar with propositional logic we digress briefly at this point to describe an interesting application of the last two theorems. If φ is a compound statement in the formal system of propositional logic whose statement variables are $q_1, \ldots, q_n$, then one may assign to φ an element f_φ of $\mathbb{Z}/2\mathbb{Z}[X_1, \ldots, X_n]$ as follows. One replaces q_i by X_i for $1 \leq i \leq n$ and then recursively replaces the logical connectives by ring operations according to the following rules.

$$\begin{aligned}
\varphi \wedge \psi &\longmapsto f_\varphi f_\psi \\
\varphi \vee \psi &\longmapsto f_\varphi + f_\psi + f_\varphi f_\psi \\
\neg\varphi &\longmapsto 1 + f_\varphi \\
\varphi \longrightarrow \psi &\longmapsto 1 + f_\varphi + f_\varphi f_\psi
\end{aligned}$$

(The table is of course redundant.) The definitions of the logical connectives and the operations in $\mathbb{Z}/2\mathbb{Z}$ are such that φ is a tautology iff $f_\varphi(\boldsymbol{c}) = 1$ for all $\boldsymbol{c} \in (\mathbb{Z}/2\mathbb{Z})^n$. We see that a statement φ_0 is a logical consequence of the statements $\varphi_1, \ldots, \varphi_m$ iff the implication

$$\forall x_1 \cdots \forall x_n \left(\bigwedge_{i=1}^{m} f_{\varphi_i}(\underline{x}) + 1 = 0 \longrightarrow f_{\varphi_0}(\underline{x}) + 1 = 0 \right) \qquad (*)$$

holds in $\mathbb{Z}/2\mathbb{Z}$. One way of testing this is by looking at the finitely many possible substitutions for the x_i from $\mathbb{Z}/2\mathbb{Z}$; this is the method of truth tables. From the last two theorems above and the remark between them, however, we may conclude that $(*)$ is also equivalent to

$$f_{\varphi_0} + 1 \in \mathrm{Id}(f_{\varphi_1} + 1, \ldots, f_{\varphi_m} + 1, X_1^2 - X_1, \ldots, X_n^2 - X_n),$$

a condition which can be tested by means of a Gröbner basis computation over $\mathbb{Z}/2\mathbb{Z}$.

Resuming our discussion of word problems, we will now investigate the word problem for Abelian monoids. Here, the monoid operation will be written as multiplication and the neutral element as 1. Let M be an Abelian monoid and T the monoid of terms in n variables $X_1, \ldots, X_n$. Then for each $\boldsymbol{c} = (c_1, \ldots, c_n) \in M^n$, there is a natural homomorphism

$$\begin{array}{cccc} \varphi_{\boldsymbol{c}}: & T & \longrightarrow & M \\ & X_1^{\nu_1} \cdot \cdots \cdot X_n^{\nu_n} & \longmapsto & c_1^{\nu_1} \cdot \cdots \cdot c_n^{\nu_n}. \end{array}$$

In view of the obvious formal analogy to evaluation of polynomials, we will write $\varphi_{\boldsymbol{c}}(t) = t(\boldsymbol{c})$ whenever $t \in T$. These "evaluated terms" will be the words that the word problem for Abelian monoids is about. In other words, if $s_0, \ldots, s_m, t_0, \ldots, t_m \in T$, then we define the word implication

$$\forall x_1 \cdots \forall x_n \left(\bigwedge_{i=1}^{m} s_i(\underline{x}) = t_i(\underline{x}) \longrightarrow s_0(\underline{x}) = t_0(\underline{x}) \right)$$

to hold in M if $s_i(\boldsymbol{c}) = t_i(\boldsymbol{c})$ for $1 \leq i \leq n$ implies $s_0(\boldsymbol{c}) = t_0(\boldsymbol{c})$ whenever $\boldsymbol{c} \in M^n$. The following theorem reduces the word problem for Abelian monoids to an ideal membership problem in $K[\underline{X}]$, where K is a field that can be arbitrarily chosen.

Theorem 6.72 *Let K be field, and let $s_0, \ldots, s_m, t_0, \ldots, t_m \in T$. Then the following are equivalent:*

(i) The implication

$$\forall x_1 \cdots \forall x_n \left(\bigwedge_{i=1}^{m} s_i(\underline{x}) = t_i(\underline{x}) \longrightarrow s_0(\underline{x}) = t_0(\underline{x}) \right)$$

holds in the class of all Abelian monoids.

(ii) $s_0 - t_0 \in \mathrm{Id}(s_1 - t_1, \dots, s_m - t_m)$ *in* $K[\underline{X}]$.

Proof (i)$\Longrightarrow$(ii): Every extension ring R of K is an Abelian monoid under multiplication. Moreover, if we view a term t as an element of $K[\underline{X}]$, then t evaluated as a polynomial at $\boldsymbol{c} \in R^n$ coincides with $t(\boldsymbol{c})$ in the sense that we are using it here, namely, as $\varphi_{\boldsymbol{c}}(t)$. We see that the implication

$$\forall x_1 \cdots \forall x_n \left(\bigwedge_{i=1}^{m} (s_i - t_i)(\underline{x}) = 0 \longrightarrow (s_0 - t_0)(\underline{x}) = 0 \right)$$

holds in the class of all extension rings of K. The claim now follows from Theorem 6.59.

(ii)$\Longrightarrow$(i): Let M be an Abelian monoid and $\boldsymbol{c} \in M^n$ such that $s_i(\boldsymbol{c}) = t_i(\boldsymbol{c})$ for $1 \leq i \leq m$. Let N be the additive monoid $\mathbb{N}^n$. Then $K[\underline{X}]$ is by definition the monoid ring KN over K and N, and both K and N are naturally embedded in KN. Moreover, $\boldsymbol{c}$ gives rise to a homomorphism

$$\begin{array}{rccc} \tau: & N & \longrightarrow & M \\ & (\nu_1, \dots, \nu_n) & \longmapsto & c_1^{\nu_1} \cdot \dots \cdot c_n^{\nu_n}, \end{array}$$

and finally, both K and M are naturally embedded in the monoid ring KM over K and M. We obtain a diagram

$$\begin{tikzcd} K \arrow[r, "\iota"] \arrow[dr, "\varphi"'] & K[\underline{X}] & N \arrow[l, "\eta"'] \arrow[dl, "\sigma"] \\ & KM & \end{tikzcd}$$

where ι, η, and φ are natural embeddings, and σ is τ followed by the natural embedding of M in KM. Proposition 2.13 now provides a ring homomorphism $\overline{\varphi}$ which completes the diagram as follows.

$$\begin{tikzcd} K \arrow[r, "\iota"] \arrow[dr, "\varphi"'] & K[\underline{X}] \arrow[d, "\overline{\varphi}"'] & N \arrow[l, "\eta"'] \arrow[dl, "\sigma"] \\ & KM & \end{tikzcd}$$

Here, $\overline{\varphi}$ maps a term t to the monomial $t(\boldsymbol{c})$ because t corresponds to its exponent tuple under η. Suppose now that

$$s_0 - t_0 = \sum_{i=1}^{m} h_i \cdot (s_i - t_i) \qquad (h_i \in K[\underline{X}]).$$

If we map this equation into KM by means of $\overline{\varphi}$, then the right-hand side becomes zero because $s_i(\boldsymbol{c}) = t_i(\boldsymbol{c})$ for $1 \leq i \leq m$, and we see that

$s_0(\boldsymbol{c}) = t_0(\boldsymbol{c})$ in KM and thus in M, because the monomials in KM correspond to the elements of M under the natural embedding of M in KM. □

For computational purposes, it is of course advantageous to use the simplest possible field $K = \mathbb{Z}/2\mathbb{Z}$.

Corollary 6.73 *The word problem for the class of Abelian monoids is decidable.* □

Finally, we consider the word problem for the class of Abelian groups. Again, we write the group operation as multiplication and the neutral element as 1. Accordingly, the inverse of an element a will be written as a^{-1}. Let G be an Abelian group and $\boldsymbol{c} = (c_1, \ldots, c_n) \in G^n$. Then we may consider the map

$$\begin{array}{cccc} \varphi_{\boldsymbol{c}}: & \mathbb{Z}^n & \longrightarrow & G \\ & (k_1, \ldots, k_n) & \longmapsto & c_1^{k_1} \cdot \cdots \cdot c_n^{k_n}. \end{array}$$

Again, there is a certain formal analogy to evaluation of terms, and so a natural notation is given by $\varphi_{\boldsymbol{c}}(u) = u(\boldsymbol{c})$ for $u \in \mathbb{Z}^n$. These $u(\boldsymbol{c})$ are the words whose equality is to be uniformly decided here. In other words, if u_0, $\ldots$, u_m, v_0, $\ldots$, $v_m \in \mathbb{Z}^n$, then we define the word implication

$$\forall x_1 \cdots \forall x_n \left(\bigwedge_{i=1}^{m} u_i(\underline{x}) = v_i(\underline{x}) \longrightarrow v_0(\underline{x}) = u_0(\underline{x}) \right)$$

to hold in G if $u_i(\boldsymbol{c}) = v_i(\boldsymbol{c})$ for $1 \le i \le m$ implies $u_0(\boldsymbol{c}) = v_0(\boldsymbol{c})$ whenever $\boldsymbol{c} \in G^n$. Let us denote by $T(\underline{X}, \underline{Y})$ the set of all terms in the variables X_1, $\ldots$, X_n, Y_1, $\ldots$, Y_n. To each element $u = (k_1, \ldots, k_n)$ of $\mathbb{Z}^n$, we assign an element $t_u = X_1^{\nu_1} \cdot \cdots \cdot X_n^{\nu_n} \cdot Y_1^{\mu_1} \cdot \cdots \cdot Y_n^{\mu_n}$ of $T(\underline{X}, \underline{Y})$ by setting

$$\nu_i = \begin{cases} k_i & \text{if } k_i > 0 \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad \mu_i = \begin{cases} -k_i & \text{if } k_i < 0 \\ 0 & \text{otherwise.} \end{cases}$$

The following theorem reduces the word problem for Abelian groups to the one for Abelian monoids.

Theorem 6.74 *Let u_0, $\ldots$, u_m, v_0, $\ldots$, $v_m \in \mathbb{Z}^n$. Then the following are equivalent:*

(i) The implication

$$\forall x_1 \cdots \forall x_n \left(\bigwedge_{i=1}^{m} u_i(\underline{x}) = v_i(\underline{x}) \longrightarrow u_0(\underline{x}) = v_0(\underline{x}) \right)$$

holds in the class of all Abelian groups.

(ii) The implication

$$\forall x_1 \cdots \forall x_n \forall y_1 \cdots \forall y_n \Bigl(\bigwedge_{i=1}^{m} t_{u_i}(\underline{x}, \underline{y}) = t_{v_i}(\underline{x}, \underline{y}) \wedge \bigwedge_{i=1}^{n} x_i y_i = 1$$

$$\longrightarrow t_{u_0}(\underline{x}, \underline{y}) = t_{v_0}(\underline{x}, \underline{y}) \Bigr)$$

holds in the class of all Abelian monoids.

Proof (i)$\Longrightarrow$(ii): Let M be an Abelian monoid, and let

$$\boldsymbol{a} = (a_1, \ldots, a_n) \quad \text{and} \quad \boldsymbol{b} = (b_1, \ldots, b_n)$$

be elements of M^n such that

$$t_{u_i}(\boldsymbol{a}, \boldsymbol{b}) = t_{v_i}(\boldsymbol{a}, \boldsymbol{b}) \quad \text{and} \quad a_j b_j = 1$$

for $1 \leq i \leq m$ and $1 \leq j \leq n$. It is not hard to prove that

$$G = \{ a_1^{\nu_1} \cdot \cdots \cdot a_n^{\nu_n} \cdot b_1^{\mu_1} \cdot \cdots \cdot b_n^{\mu_n} \mid \nu_1, \ldots, \nu_n, \mu_1, \ldots, \mu_n \in \mathbb{N} \}$$

is an Abelian group under the operation of M, and that $w(\boldsymbol{a}) = t_w(\boldsymbol{a}, \boldsymbol{b})$ for all $w \in \mathbb{Z}^n$. It now follows easily from (i) that $t_{u_0}(\boldsymbol{a}, \boldsymbol{b}) = t_{v_0}(\boldsymbol{a}, \boldsymbol{b})$.

(ii)$\Longrightarrow$(i): Let G be an Abelian group and $\boldsymbol{a} = (a_1, \ldots, a_n) \in G^n$ such that $u_i(\boldsymbol{a}) = v_i(\boldsymbol{a})$ for $1 \leq i \leq m$. Setting $b_i = a_i^{-1}$ for $1 \leq i \leq n$, it is easy to see that $w(\boldsymbol{a}) = t_w(\boldsymbol{a}, \boldsymbol{b})$ for all $w \in \mathbb{Z}^n$. It follows that

$$t_{u_i}(\boldsymbol{a}, \boldsymbol{b}) = t_{v_i}(\boldsymbol{a}, \boldsymbol{b}) \quad \text{and} \quad a_j b_j = 1$$

for $1 \leq i \leq m$ and $1 \leq j \leq n$. Since G is an Abelian monoid, we may apply (ii) to conclude that

$$u_0(\boldsymbol{a}) = t_{u_0}(\boldsymbol{a}, \boldsymbol{b}) = t_{v_0}(\boldsymbol{a}, \boldsymbol{b}) = v_0(\boldsymbol{a}). \quad \square$$

Together with the corollary to the last theorem, we obtain the following decidability result.

Corollary 6.75 *The word problem for the class of Abelian groups is decidable.* □

We have demonstrated how Gröbner bases can be used to decide a large number of word problems. We mention that in the area of word problems, undecidability tends to be much harder to prove. A famous result asserts that the word problem for groups is undecidable.

Notes

Our approach to the computation of syzygies imitates that of Apel and Lassner (1988), where the non-commutative case is treated. The commutative case is discussed in Zacharias (1978), Furukawa et al. (1986), and Wall (1989). The proposition and theorem on the lifting of syzygies at the end of Section 6.1 go back to Lazard (1983); the idea was further developed and discussed in Möller (1985), Möller and Mora (1986a), Möller (1988), and Gebauer and Möller (1988). In Möller et al. (1992), a version of the Buchberger algorithm is given that fully exploits Buchberger's second criterion in the strong form of Theorem 6.13. Here, an S-polynomial is tested out if the corresponding head term syzygy lies in the module generated by the syzygies corresponding to the S-polynomials that have already been treated. This version is indeed capable of detecting more superfluous critical pairs than any other known implementation; the cost of testing submodule membership, however, has thus far turned out to be too high to translate the deletion of more pairs into a computational gain (cf. Section 10.4).

The standard pre-Gröbner-bases work on algorithmic problems in ideal theory is Hermann (1926); an update with corrections and improvements was given by Seidenberg (1974). In particular, Hermann had an algorithm to test ideal membership, a problem which is sometimes referred to as the "main problem of the theory of polynomial ideals" (see, e.g., van der Waerden, 1966, §131). Hermann's method is based on an effective bound on the degrees of the polynomials that are needed to represent a given polynomial as a sum of multiples of other given polynomials; the problem thus reduces to solving systems of linear equations after comparing coefficients. The bound depends on the number of variables and the degrees of the given polynomials. The method thus makes, in a manner of speaking, the worst case assumption for each instance of the ideal membership test, whereas the Gröbner basis method is, in a sense, a flexible one.

The Gröbner basis solutions that are given in Section 6.2 are substantially different from the classical ones; their relevance lies in the fact that actual implementations and computations have now become feasible. The methods that we describe have become part of the folklore of the theory; many of them were found by Buchberger himself (see Buchberger, 1985a). For more on multivariate interpolation using Gröbner bases, see Becker and Weispfenning (1990).

Our definition of independent sets and dimension is that of Gröbner (1970), Chapter II, §1. Classically, the dimension is first defined for prime ideals as the transcendence degree of the residue class ring over the ground field (cf. Section 7.1). The dimension of an arbitrary ideal is then defined as the maximum of the dimensions of the associated prime ideals of the primary components (cf. Section 8.5). Gröbner's definition is clearly more in the spirit of algorithmic ideal theory; in particular, it makes it easy to see how to determine the dimension via Gröbner bases (see also Section 9.3).

This was already noticed by Buchberger himself (see, e.g., Buchberger, 1985a). The equivalence of Gröbner's definition with the classical one will be proved for prime ideals in Lemma 7.25; the general case will be the subject of Exercise 8.58.

The concept of a word problem was introduced by Thue in 1914. Word problems are a central topic in the theory of decidability and complexity. A standard argument to prove the undecidability of some problem is to argue that if it were decidable, then a word problem whose undecidability has already been established would be decidable too. The decision methods using Gröbner bases can be found in Kandri-Rody et al. (1986).

7

Field Extensions and the Hilbert Nullstellensatz

7.1 Field Extensions

We have now reached a point in the theory of polynomial ideals where some classical results concerning field extensions are needed. Throughout this section, K will be a field, and until further notice K' will be an **extension field** of K, meaning of course that K is a subfield of K'.

We begin by formulating an analogue to ring adjunction for fields. If A is a subset of K', then the intersection of all subfields of K' that contain both K and A is called the field obtained by **adjunction** of A to K, and it is denoted by $K(A)$. The set whose intersection we are taking is not empty since K' itself contains K and A. $K(A)$ is clearly a subfield of K' which extends K. If $A = \{a_1, \ldots a_n\}$, then we will also write $K(a_1, \ldots, a_n)$ for $K(A)$. In this case, K' is called a **finite extension** of K. If $a \in K'$, then $K(a)$ is called a **simple extension** of K with **primitive element** a. The following lemma can easily be proved in the same way as Lemma 1.110.

Lemma 7.1 Let $A \subseteq K'$. Then $K(A)$ consists of all elements of K' that can be written in the form

$$f(a_1, \ldots, a_m) \cdot (g(a_1, \ldots, a_m))^{-1},$$

where $m \in \mathbb{N}$, f, $g \in K[X_1, \ldots, X_m]$, and $a_1, \ldots, a_m \in A$ such that $g(a_1, \ldots, a_m) \neq 0$. □

Exercise 7.2 Let A, $B \subseteq K'$. Show that $(K(A))(B) = K(A \cup B)$.

An element a of K' is called **algebraic over** K if there exists $0 \neq f \in K[X]$ with $f(a) = 0$, **transcendental over** K otherwise. K' is called **algebraic (transcendental) over** K, or an **algebraic (transcendental) extension of** K, if every $a \in K' \setminus K$ is algebraic (transcendental) over K. An element $a \in K'$ is thus either algebraic or transcendental over K, whereas, as we will see, K' may be neither algebraic nor transcendental over K. In the mathematical literature, transcendental extensions are also called *purely transcendental*.

Lemma 7.3 Let $a \in K'$ be algebraic over K. Then there exists a unique monic polynomial $0 \neq f \in K[X]$ of least degree with $f(a) = 0$.

Proof From the fact that $f(a) = 0$ implies $cf(a) = 0$ for all $c \in K$ and from Corollary 0.4, it follows immediately that there exists a monic polynomial $0 \neq f \in K[X]$ of least degree vanishing at a. Now let $g \in K[X]$ be another monic polynomial with $g(a) = 0$ and $\deg(g) = \deg(f)$. Since $K[X]$ with the degree function is a Euclidean ring, there exist q, $r \in K[X]$ with

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

Evaluating at a, we see that $r(a) = 0$, and thus $r = 0$ by the minimality of $\deg(f)$. (Setting the highest coefficient of r to 1 does not change the degree.) Since f and g are both monic and of the same degree, we must have $q = 1$ and hence $f = g$ as desired. □

The polynomial described in the above lemma is called the **minimal polynomial** of a over K and is denoted by $\min_K^a$. The minimal polynomial of an element depends heavily on the chosen ground field; for example, the minimal polymial of the complex number i over $\mathbb{Q}$ is $X^2 + 1$, whereas the one over $\mathbb{C}$ itself is $X - i$. In fact, whenever $a \in K$, then a is algebraic over K with minimal polynomial $X - a$, and thus K is an algebraic extension of itself.

Exercise 7.4 What is the minimal polynomial of the real number $1 - \sqrt[3]{2}$ over $\mathbb{Q}$? (Hint: Set up monic polynomials of increasing degree with unknown coefficients, try to compute rational coefficients to achieve vanishing at $1 - \sqrt[3]{2}$. There is a much more elegant way of doing this particular one which, if you don't see it now, you will be shown in a later example from a higher point of view).

Lemma 7.5 Let $a \in K'$ be algebraic over K, and let $f \in K[X]$. Then $f(a) = 0$ iff $\min_K^a$ divides f in $K[X]$.

Proof The direction "$\Longleftarrow$" is trivial. Now assume that $f(a) = 0$. Dividing f by $\min_K^a$ with remainder, we obtain $f = q \cdot \min_K^a$ with $q \in K[X]$ by the same argument as in the proof of Lemma 7.3. □

Corollary 7.6 *Let $a \in K'$ be algebraic over K. Then $\min_K^a$ is the unique monic irreducible polynomial in $K[X]$ that vanishes at a.*

Proof Assume for a contradiction that $\min_K^a$ has a proper factorization $\min_K^a = fg$ in $K[X]$. Then we have

$$\deg(f) < \deg(\min_K^a) \quad \text{and} \quad \deg(g) < \deg(\min_K^a),$$

and also $f(a) = 0$ or $g(a) = 0$ since $f(a)g(a) = \min_K^a(a) = 0$, contradicting the minimality of $\deg(\min_K^a)$. Any other monic irreducible element of $K[X]$ vanishing at a would have to be a multiple of $\min_K^a$ by Lemma 7.5 and must thus equal $\min_K^a$. □

Next, we wish to obtain a structural description of simple field extensions. To this end, we temporarily change our point of view: let us forget now about the given extension field K'. Given nothing but K, we will *construct* an extension field K' of K in such a way that $K' = K(a)$ for some $a \in K'$, where a is prescribed as being transcendental over K, or as being algebraic with prescribed minimal polynomial over K.

The transcendental case is easy: we take for K' the rational function field over K in the variable X, i.e., the field of fractions of $K[X]$. Then K' is an extension field of K. Now if we look at the field obtained by adjoining $X \in K'$ to K within K', then we see from Lemma 7.1 that this gives all of K'. Moreover, X is transcendental over K since $g = g(X) = 0$ in K' only if g is the zero polynomial. This phenomenon justifies the double notation $K(X)$ for the rational function field in the variable X over K and the simple extension obtained by adjoining the transcendental element X to K within some given extension of K.

Now let g be a monic irreducible polynomial in $K[X]$. $K[X]$ is a PID, and so by Lemma 2.48 (ii) and Proposition 1.94, the residue class ring $K[X]/\mathrm{Id}(g)$ is a field. (See also the remarks following Lemma 2.48.) We will denote the residue class $f + \mathrm{Id}(g)$ of $f \in K[X]$ by $\overline{f}$. We claim that the canonical homomorphism

$$\begin{array}{ccc} K[X] & \longrightarrow & K[X]/\mathrm{Id}(g) \\ f & \longmapsto & \overline{f} \end{array}$$

is injective when restricted to K: if $a \in K$, then $\overline{a} = 0$ implies $a \in \mathrm{Id}(g)$ and so $a = 0$ since g is not a unit and thus not constant by the definition of irreducibility. We may therefore simply identify every $a \in K$ with $\overline{a}$ and thus operate on the assumption that K is a subfield of $K[X]/\mathrm{Id}(g)$. Now let $f = \sum_{i=1}^{m} a_i X^i \in K[X]$. Then we have, by the homomorphism property of the bar,

$$\overline{f} = \overline{\sum_{i=1}^{m} a_i X^i} = \sum_{i=1}^{m} a_i \overline{X}^i = f(\overline{X}).$$

In particular, $g(\overline{X}) = \overline{g} = 0$, and $f(\overline{X}) = \overline{f} \neq 0$ whenever $\deg(f) < \deg(g)$. So if we set $K' = K[X]/\mathrm{Id}(g)$, then $\overline{X} \in K'$ is algebraic over K with minimal polynomial $g \in K[X]$. Moreover, the elements of K' are of the form $\overline{f} = f(\overline{X})$ with $f \in K[X]$ and thus $K' = K(\overline{X})$ by Lemma 7.1. Note that for all $f \in K[X]$, we get $\overline{f} = \overline{r}$ where r is the remainder of f upon division by g, so we even have

$$K' = \{\, f(\overline{X}) \mid f \in K[X],\ \deg(f) < \deg(g)\,\}.$$

We have proved the following proposition.

Proposition 7.7 *Let K be a field. Then the following hold:*

(i) Let K' be the rational function field over K in the indeterminate X. Then K' is a simple extension of K with primitive transcendental element X.

(ii) *Let g be a monic irreducible element of $K[X]$. Then $K' = K[X]/\mathrm{Id}(g)$ is a simple extension of K with primitive algebraic element $\overline{X}$ whose minimal polynomial equals g.* □

The constructions of the above proposition can of course be iterated. We can, for example, extend K by an algebraic primitive element a and then extend $K(a)$ by a primitive element which is transcendental over $K(a)$ and hence over K. The field $K(a)(b) = K(a,b)$ is then an extension of K which is neither algebraic nor transcendental.

Let K' and K'' both be extension fields of K. An isomorphism (embedding) $\varphi : K' \longrightarrow K''$ is called a **K-isomorphism** (**K-embedding**) if it is an isomorphism (embedding) and satisfies $\varphi \restriction K = \mathrm{id}_K$. It is easy to see from Lemma 7.1 that a K-embedding φ from a simple extension of K to some other extension of K is completely determined by the φ-value of the primitive element.

Proposition 7.7 now suggests how the structure of simple extensions of K in some given extension field K' can be described.

Proposition 7.8 *Let K' be an extension field of K, $a \in K'$. If a is transcendental over K, then $K(a)$ is K-isomorphic to the rational function field $K(X)$ where X is mapped to a; else, it is K-isomorphic to $K[X]/\mathrm{Id}(\mathrm{min}_K^a)$ where $\overline{X} = X + \mathrm{Id}(\mathrm{min}_K^a)$ is mapped to a.*

Proof Let $\varphi : K[X] \longrightarrow K'$ be the homomorphism that maps f to $f(a)$ (Lemma 2.17 (i)). If a is transcendental over K, then $\ker(\varphi) = \{0\}$. So φ is an embedding and thus extends to a unique embedding $\psi : K(X) \longrightarrow K'$ by the universal property of the field of fractions. It is clear that ψ is a K-embedding and that $\psi(K(X)) = K(a)$. If a is algebraic over K, then $\ker(\varphi) = \mathrm{Id}(\mathrm{min}_K^a)$ by Lemma 7.5. By Corollary 1.56, $K[X]/\mathrm{Id}(\mathrm{min}_K^a)$ is isomorphic to $\varphi(K[X])$ under the map

$$\begin{array}{rccc}\psi : & K[X]/\mathrm{Id}(\mathrm{min}_K^a) & \longrightarrow & \varphi(K[X]) \\ & \overline{f} & \longmapsto & f(a)\end{array}$$

It is clear that ψ is a K-embedding. $\psi(K[X]/\mathrm{Id}(\mathrm{min}_K^a))$ is a subfield of K' containing a and all of K, and each of its elements is of the form $f(a)$ with $f \in K[X]$. We see that $\psi(K[X]/\mathrm{Id}(\mathrm{min}_K^a))$ must equal $K(a)$. □

The following corollary is now obvious from the fact that the composition of two K-isomorphisms is again a K-isomorphism.

Corollary 7.9 *Let $K(a)$ and $K(b)$ be simple extensions of K such that either a and b are both transcendental over K, or they are both algebraic and have the same minimal polynomial over K. Then there exists a unique K-isomorphism $\varphi : K(a) \longrightarrow K(b)$ with $\varphi(a) = b$.* □

We will later on need the following more general version of the above corollary.

Corollary 7.10 *Let K_1 and K_2 be fields, and $\varphi : K_1 \longrightarrow K_2$ an isomorphism of fields. Let $K_1(a)$ and $K_2(b)$ be simple algebraic extensions of K_1 and K_2, respectively, such that $\min^b_{K_2} = \varphi'(\min^a_{K_1})$, where φ' is the isomorphism*

$$\begin{array}{ccc} K_1[X] & \longrightarrow & K_2[X] \\ \sum_{i=0}^{m} a_i X^i & \longmapsto & \sum_{i=0}^{m} \varphi(a_i) X^i \end{array}$$

of Lemma 2.17 (ii). Then φ extends to an isomorphism $\overline{\varphi} : K_1(a) \longrightarrow K_2(b)$ with $\overline{\varphi}(a) = b$.

Proof By Proposition 7.8 and Lemma 1.30 (ii) there exists a K_1-isomorphism

$$\psi_1 : K_1(a) \longrightarrow K_1[X]/\mathrm{Id}(\min^a_{K_1})$$

with $\psi_1(a) = X + \mathrm{Id}(\min^a_{K_1})$ and a K_2-isomorphism

$$\psi_2 : K_2[X]/\mathrm{Id}(\min^b_{K_2}) \longrightarrow K_2(b)$$

with $\psi_2(X + \mathrm{Id}(\min^b_{K_2})) = b$. From the fact that $\varphi' : K_1[X] \longrightarrow K_2[X]$ is an isomorphism with $\min^b_{K_2} = \varphi'(\min^a_{K_1})$, one easily concludes that

$$\mathrm{Id}(\min^b_{K_2}) = \varphi'(\mathrm{Id}(\min^a_{K_1})).$$

Lemma 1.66 now provides an isomorphism

$$\psi_3 : K_1[X]/\mathrm{Id}(\min^a_{K_1}) \longrightarrow K_2[X]/\mathrm{Id}(\min^b_{K_2})$$

with $\psi_3(f + \mathrm{Id}(\min^a_{K_1})) = \varphi'(f) + \mathrm{Id}(\min^b_{K_2})$. Then $\overline{\varphi} = \psi_2 \circ \psi_3 \circ \psi_1$ is an isomorphism from $K_1(a)$ to $K_2(b)$. Furthermore, we have

$$\begin{aligned} \overline{\varphi}(a) &= \psi_2(\psi_3(\psi_1(a))) \\ &= \psi_2(\psi_3(X + \mathrm{Id}(\min^a_{K_1})) \\ &= \psi_2(\varphi'(X) + \mathrm{Id}(\min^b_{K_2})) \\ &= \psi_2(X + \mathrm{Id}(\min^b_{K_2})) \\ &= b, \end{aligned}$$

and, for all $c \in K_1$,

$$\begin{aligned} \overline{\varphi}(c) &= \psi_2(\psi_3(\psi_1(c))) \\ &= \psi_2(\psi_3(c + \mathrm{Id}(\min^a_{K_1})) \\ &= \psi_2(\varphi'(c) + \mathrm{Id}(\min^b_{K_2})) \\ &= \psi_2(\varphi(c) + \mathrm{Id}(\min^b_{K_2})) \\ &= \varphi(c). \quad \square \end{aligned}$$

Proposition 7.8 together with the reults of Section 4.6 tells us how to compute in any simple extension of a computable field K, provided we know whether the primitive element is algebraic or transcendental over K and, in the former case, we know what the minimal polynomial over K is: just treat a as an indeterminate, and compute with elements of $K(a)$ as with rational functions in the transcendental case, as with polynomials modulo $\mathrm{Id}(\min_K^a)$ otherwise. In particular, in the algebraic case, every element of $K(a)$ equals $f(a)$ for some $f \in K[X]$ with $\deg(f) < \deg(\min_K^a)$ (cf. Example 4.84 and the comments following it). So if, for example, a is a real number which is algebraic over $\mathbb{Q}$ and whose minimal polynomial we know, then we can rewrite every expression $f(a)/g(a)$ ($f, g \in \mathbb{Q}[X]$ with $g(a) \neq 0$) in such a way that only rational denominators occur. This is what is called "rationalizing the denominator" in elementary algebra.

Exercise 7.11 Rationalize the denominator of $1/(a^2+a-1)$ where $a = 1 - \sqrt[3]{2}$ (cf. Exercise 7.4).

Now that we know how to compute in simple algebraic extensions, we can also do polynomial arithmetic and long division of polynomials over these fields. This means that we can compute in iterated algebraic extensions.

Exercise 7.12 Let $z_1, z_2 \in \mathbb{C}$ such that z_1 is a zero of of X^2+X+1 and z_2 is a zero of $Y^3 - z_1$. Find a polynomial $f \in \mathbb{Q}[X,Y]$ such that $f(z_1, z_2) = (z_2^2 - z_1)^{-1}$. If you have a computer algebra system at your disposal, then check your answer by substituting $z_1 = (1/2)(1+i\sqrt{3})$ and $z_2 = z_1^{(1/3)}$.

In view of Propositions 7.7 and 7.8, we may now speak of simple extensions of the field K without specifying whether or not they sit in some previously given extension field K' of K.

Proposition 7.13 *Suppose a is algebraic over K and b is algebraic over $K(a)$. Then b is algebraic over K.*

Proof Let $0 \neq f \in K[X]$ with $f(a) = 0$, e.g., $f = \min_K^a$. Let $0 \neq g \in K(a)[Y]$ monic with $g(b) = 0$, e.g., $g = \min_{K(a)}^b$, say

$$\begin{aligned} g &= Y^m + \sum_{i=0}^{m-1} c_i Y^i \qquad (c_i \in K(a)) \\ &= Y^m + \sum_{i=0}^{m-1} g_i(a) Y^i \qquad (g_i \in K[X]). \end{aligned}$$

We set

$$h = Y^m + \sum_{i=0}^{m-1} g_i(X) Y^i \in K[X,Y]$$

and view f as an element of $K[X,Y]$ too. We see that $f(a,b) = h(a,b) = 0$. f and h thus lie in the kernel I of the homomorphism

$$\begin{array}{ccc} K[X,Y] & \longrightarrow & K(a,b) \\ p & \longmapsto & p(a,b). \end{array}$$

I is proper and satisfies condition (ii) of Corollary 6.56 if we take for $\leq$ the inverse lexicographical order, where $X \ll Y$. I thus contains a non-zero polynomial $q \in K[Y]$, and this has the desired property $q(b) = 0$. □

Inspection of the proof above shows that in the situation of the proposition, we can read off from a certain Gröbner basis a polynomial $q \in K[Y]$ with $q(b) = 0$. The question arises if we can actually find $\min_K^b$ in this way. The following proposition gives a positive answer.

Proposition 7.14 *Suppose a is algebraic over K and b is algebraic over $K(a)$. Let $f = \min_K^a \in K[X]$ and $h \in K[X,Y]$ such that $h(a,Y) = \min_{K(a)}^b$. Furthermore, let G be the reduced Gröbner basis of* Id(f,h) *w.r.t. the lexicographical term order, where $Y \ll X$. Then*

$$G \cap K[Y] = \{\min_K^b\}.$$

In particular, $\min_K^b$ can be computed from f and h if K is computable.

Proof We have already argued in the proof of the last proposition that $G \cap K[Y]$ is not empty, and that each of its members vanishes at b. Moreover, since G is reduced, $G \cap K[Y]$ has no more than one element, which is monic, and it remains to prove that this element is irreducible. We claim that it suffices to show that $\mathrm{Id}(f,h)$ is prime. Indeed, if this is the case, then, as one easily proves, $\mathrm{Id}(f,h) \cap K[Y]$ is prime too, and so its monic generator must be irreducible. By Proposition 6.15, this generator is the element of $G \cap K[Y]$.

To see that $\mathrm{Id}(f,h)$ is prime, suppose p_1, $p_2 \in K[X,Y]$ with $p_1 p_2 \in \mathrm{Id}(f,h)$, say $p_1 p_2 = q_1 h + q_2 f$. From $f(a) = 0$ we conclude that

$$p_1(a,Y) \cdot p_2(a,Y) = q_1(a,Y) \cdot h(a,Y),$$

and thus, since $h(a,Y)$ is an irreducible polynomial in $K(a)[Y]$, at least one of $p_1(a,Y)$ and $p_2(a,Y)$ must be a multiple of $h(a,Y)$, say

$$p_1(a,Y) = q_3(a,Y) \cdot h(a,Y) \quad \text{with} \quad q_3 \in K[X,Y]. \qquad (*)$$

Now consider the homomorphism

$$\begin{array}{rccc} \psi: & K[X,Y] & \longrightarrow & K(a)[Y] \\ & g & \longmapsto & g(a,Y). \end{array}$$

Viewing $g \in K[X,Y]$ as an element of $K[X][Y]$, we see that $g \in \ker(\psi)$ iff every coefficient of g vanishes at a iff every coefficient of g is a multiple of f iff g is a multiple of f. We have proved that $\ker(\psi)$ is the ideal generated by f in $K[X,Y]$. Going back to $(*)$, we see that $p_1 - q_3 h \in \ker(\psi)$ and so

$$p_1 = q_3 h + q_4 f$$

with $q_4 \in K[X,Y]$, i.e., $p_1 \in \mathrm{Id}(f,h)$. □

We mention that primeness of the ideal $\mathrm{Id}(f,h)$ of the proof above will also be an immediate consequence of Proposition 7.44.

Exercise 7.15 Let K be a computable field, and let $K(a)$ be a simple algebraic extension of K, given by the minimal polynomial $\min_K^a \in K[X]$. Show how one may compute the minimal polynomial over K of an element $g(a)$ of $K(a)$ when $g \in K[X]$ is given. (Hint: The minimal polynomial over $K(a)$ of $g(a)$ is $Y - g(a)$.)

Exercise 7.16 Redo Exercise 7.4, and try to see how you could have solved it even then with no effort at all. Now compute the minimal polynomial over $\mathbb{Q}$ of $\sqrt[3]{4} + \sqrt[3]{2} + 1$.

Corollary 7.17 *(i) If b is algebraic over $K(a_1, \dots, a_m)$ with $a_1, \dots, a_m$ algebraic over K, then b is algebraic over K.*

(ii) If K' is an extension field of K, and $A \subseteq K'$ such that $K' = K(A)$ and each $a \in A$ is algebraic over K, then K' is algebraic over K. In particular, if a is algebraic over K, then $K(a)$ is an algebraic extension of K.

(iii) If K' is algebraic over K and K'' is algebraic over K', then K'' is algebraic over K.

(iv) If $K_0 \subseteq K_1 \subseteq \cdots \subseteq K_m$ and K_i is algebraic over K_{i-1} for $1 \le i \le m$, then K_m is algebraic over K_0.

Proof (i) We proceed by induction on m. For $m = 0$ there is nothing to prove. Now let $m > 0$. Since a_m is algebraic over K, it is trivially algebraic over $K(a_1, \dots, a_{m-1})$, and b is algebraic over

$$K(a_1, \dots, a_{m-1})(a_m) = K(a_1, \dots, a_m).$$

By Proposition 7.13 above, b is algebraic over $K(a_1, \dots, a_{m-1})$, and so by induction hypothesis, it is algebraic over K.

(ii) Let $a \in K'$. By Lemma 7.1, there exist $a_1, \dots, a_m \in A$ such that $a \in K(a_1, \dots, a_m)$. In particular, a is algebraic over $K(a_1, \dots, a_m)$ and thus over K by (i) above.

(iii) Let $b \in K''$ and $\min_{K'}^b = \sum_{i=0}^m a_i X^i$. Then rather obviously, b is already algebraic over the subfield $K(a_0, \dots, a_m)$ of K'. Since $a_0, \dots, a_m$ are algebraic over K, so is b by (i) above.

(iv) This is easy to prove from (iii) by induction on m. □

In view of (ii) above, we may now refer to simple extensions with algebraic primitive element as *simple algebraic extensions.*

Exercise 7.18 Let K' be an extension field of K. Show that there exists a field K_{a} such that $K \subseteq K_{\mathrm{a}} \subseteq K'$, K_{a} is algebraic over K, and K' is transcendental over K_{a}.

Corollary 7.17 (iii) and (iv) have obvious analogues for the transcendental case: if K' is transcendental over K and K'' is transcendental over K', then K'' is trivially transcendental over K. We will now provide an analogue to Corollary 7.17 (ii) for the special case of a simple extension.

Proposition 7.19 *Suppose a is transcendental over K and*

$$b = \frac{f(a)}{g(a)} \in K(a) \setminus K \qquad (f, g \in K[X],\ g \neq 0).$$

Then the following hold:

(i) b is transcendental over K and a is algebraic over $K(b)$.

(ii) $K(b) = K(a)$ if and only if, after reduction of f/g to lowest terms, f and g are linear.

Proof (i) The polynomial $h = gb - f \in K(b)[X]$ is not the zero polynomial since otherwise we could conclude $b \in K$ by comparing non-zero coefficients in gb and f. Obviously, $h(a) = 0$, and thus a is algebraic over $K(b)$. Now b must be transcendental over K since otherwise a would be algebraic over K by Proposition 7.13.

(ii) Let now f/g be reduced to lowest terms. Note that we already know that b is transcendental over K and thus "behaves like an indeterminate over K." We claim that the univariate polynomial

$$h = gb - f \in K(b)[X]$$

is irreducible. If this were not so, then by Lemma 2.62 (i), h would have a factorization $h = h_1 h_2$ with $h_1, h_2 \notin K(b)$ and $h_1, h_2 \in K[b][X] = K[X][b]$. Since h is linear in b, one of the factors, say h_1, would have to be in $K[X]$, while the other would be linear in b, say $h_2 = g^* b + f^*$. It would follow that

$$gb - f = h = h_1 h_2 = h_1(g^* b + f^*),$$

and we see that h_1 would have to be a common factor of f and g, a contradiction.

For the direction "$\Longleftarrow$" of (ii), suppose f and g are linear, i.e.,

$$b = \frac{sa + t}{ua + v} \qquad (s, t, u, v \in K,\ u \text{ and } v \text{ not both zero}).$$

We conclude that $(ub - s)a = (-vb + t)$. From the fact that $b \notin K$, it follows easily that $ub - s \neq 0$, and so

$$a = \frac{-vb + t}{ub - s}.$$

This shows that $a \in K(b)$, and together with $K(b) \subseteq K(a)$ we obtain $K(a) = K(b)$. For "$\Longrightarrow$," assume that the latter equality holds. Then the minimal polynomial $\min^a_{K(b)}$ of a over $K(b)$ equals $X - a$. Now

$$h = gb - f \in K(b)[X]$$

is a polynomial that vanishes at a, and so $(X-a) \mid h$ in $K(b)[X]$ by Lemma 7.5. But we have proved that h is irreducible as a polynomial in X, and so it must be linear in X. If we view h as an element of $K[X][b]$ and observe that b behaves like an indeterminate, we see that both f and g must be linear in X. □

In contrast to Corollary 7.17 (ii), it is not true that $K(A)$ is a transcendental extension if every $a \in A$ is transcendental over K: if $\mathbb{R}(T)$ is a simple transcendental extension of $\mathbb{R}$, then T is clearly transcendental over the subfield $\mathbb{Q}$ of $\mathbb{R}$. Furthermore, $T+\sqrt{2} \in \mathbb{R}(T)$ is transcendental over $\mathbb{R}$ and thus over $\mathbb{Q}$ by Proposition 7.19. So each member of $A = \{T, T+\sqrt{2}\}$ is transcendental over $\mathbb{Q}$, but $\mathbb{Q}(A)$ contains the algebraic element $\sqrt{2}$. The following definition describes sets whose adjunction results in a transcendental extension. Let K' be any extension field of K. A subset $\{a_1, \ldots, a_n\}$ of K' is called **algebraically independent** over K if $f(a_1, \ldots, a_n) \neq 0$ for all $0 \neq f \in K[X_1, \ldots, X_n]$. An arbitrary subset A of K' is called algebraically independent (over K) if every finite subset of A is algebraically independent (over K). It is clear that $\emptyset$ is always algebraically independent. We will also refer to algebraically independent sets as simply "independent" if there is no danger of confusion. A dependent set is of course one that is not independent.

Proposition 7.20 *If K' is an extension field of K such that $K' = K(A)$ for some algebraically independent set A over K, then K' is transcendental over K.*

Proof Let $b \in K' \setminus K$, and assume for a contradiction that $f(b) = 0$ for some monic non-zero $f \in K[X]$, say

$$f = \sum_{i=0}^{m} c_i X^i \qquad (c_i \in K \text{ for } 0 \leq i \leq m)$$

where $c_m = 1$. By Lemma 7.1, we can write $b = g(\boldsymbol{a}) \cdot (h(\boldsymbol{a}))^{-1}$ with g, $h \in K[\underline{Y}] = K[Y_1, \ldots, Y_n]$ and $\boldsymbol{a} = (a_1, \ldots, a_n) \in A^n$. If we substitute this into the equation $f(b) = 0$ and multiply by $(h(\boldsymbol{a}))^m$, we see that $p(\boldsymbol{a}) = 0$ where

$$p = \sum_{i=0}^{m} c_i h^{m-i} g^i.$$

To obtain the desired contradiction, we need to show that p is not the zero polynomial. If c_m is the only non-zero coefficient of f, then $p = g^m \neq 0$. If h is constant, then $\mathrm{HT}(g^m)$ (w.r.t. any term order) occurs only in the first summand, and so $p \neq 0$. Finally, assume that $c_i \neq 0$ for some $i < m$, and that h is not constant. Since $K[\underline{Y}]$ is a UFD, we may assume that g and h have no prime factors in common. Now if p were the zero polynomial, then we could write

$$g^m = -\sum_{i=0}^{m-1} c_i h^{m-i} g^i = -h \cdot \sum_{i=0}^{m-1} c_i h^{m-(i+1)} g^i,$$

and we see that $h \mid g^m$ contrary to our assumption that g and h are relatively prime. □

We say that an algebraically independent subset of K' over K is **maximal** if it is not properly contained in any independent subset of K' over K. This is equivalent to saying that $A \cup \{a\}$ is dependent for all $a \in K' \setminus A$ since subsets of independent sets are clearly again independent.

We are now going to show that algebraically independent sets behave much like linearly independent subsets of a vector space do. The table of correspondences below will make the next two lemmas and the theorem following them more plausible.

K' a field extension of K and $A \subseteq K'$	V a K-vector space and $B \subseteq V$
A independent over K	B linearly independent
K' algebraic over $K(A)$	B a generating system for V
A maximally independent over K	B a basis of V

Lemma 7.21 Let K' be an extension field of K, $A \subseteq K'$ algebraically independent over K, and $a \in K'$. Then the following are equivalent:

(i) $A \cup \{a\}$ is no longer independent over K.

(ii) $a \notin A$ and a is algebraic over $K(A)$.

Proof (i)$\Longrightarrow$(ii): There exist $a_1, \ldots, a_n \in A$ such that $f(a_1, \ldots, a_n, a) = 0$ for some polynomial $0 \neq f \in K[X_1, \ldots, X_{n+1}]$. The degree of f in X_{n+1} must be positive since otherwise A would be dependent over K. So if we regard f as a univariate polynomial in X_{n+1} over $K(A)$, we see that a is algebraic over the latter field.

(ii)$\Longrightarrow$(i): a is a zero of a non-constant univariate polynomial f over $K(A)$. If we multiply f by the product of the denominators of all its coefficients in $K(A)$, we obtain a non-zero polynomial $g \in K[A][X]$ with $g(a) = 0$, and this clearly shows the dependence of $A \cup \{a\}$ over K. □

The next lemma is an immediate consequence of the last one.

Lemma 7.22 Let K' be an extension field of K and $A \subseteq K'$ independent over K. Then A is maximally independent over K iff K' is algebraic over $K(A)$. □

Theorem 7.23 *Let K' be an extension field of K and assume that there exists a finite maximally independent subset B of K' over K. Then the following hold:*

(i) *Every algebraically independent subset A of K' over K is finite with $|A| \leq |B|$, and if in addition, A is maximal too, then $|A| = |B|$.*

(ii) *If $A \subseteq K'$ is such that K' is algebraic over $K(A)$, then A has at least $|B|$ many elements, and if it has exactly $|B|$ many elements, then it is maximally independent over K.*

Proof (i) In view of Corollary 3.18, it suffices to prove that the collection of all independent subsets of K' over K satisfies axioms U1 and U2 of Section 3.2. We have already observed that U1 holds trivially. Now let $A \subseteq K'$ and a, b_1, $b_2 \in K'$ be as in the premise of U2, and assume for a contradiction that both $A \cup \{b_1, a\}$ and $A \cup \{a, b_2\}$ are dependent over K. Then by Lemma 7.21, a is algebraic over $K(A)(b_1)$ and b_2 is algebraic over $K(A)(a)$ and thus over $K(A)(b_1)(a)$. It follows that b_2 is algebraic over $K(A)(b_1)$, and thus again by Lemma 7.21, $A \cup \{b_1, b_2\}$ is dependent over K, a contradiction.

(ii) Suppose K' is algebraic over $K(A)$, and consider the set

$$M = \{\, C \subseteq A \mid C \text{ independent over } K \,\}.$$

Let $D \in M$ be maximal w.r.t. inclusion. Then we may conclude from Lemma 7.21 that every $a \in A \setminus D$ is algebraic over $K(D)$. Corollary 7.17 (ii) tells us that $K(D)(A \setminus D) = K(A)$ is algebraic over $K(D)$, and (iii) of the same corollary says that K' is algebraic over $K(D)$. So $D \subseteq K'$ is maximally independent over K by the last lemma and thus has $|B|$ many elements by (i) above. It follows that A had at least that many elements, and if $|A| = |B|$, then $D = A$ and so A is maximally independent over K. □

In view of the results above, a maximal algebraically independent subset of K' over K is called a **transcendence base** of K' over K. If there exists a finite transcendence base of K' over K, then the invariant number of elements of such a base is called the **transcendence degree** of K' over K.

Exercise 7.24 Let K' be an extension field of K. Show the following:

(i) There exists a transcendence base of K' over K (Hint: Zorn's lemma).

(ii) If $B \subseteq K'$ is a transcendence base over K, then K' is algebraic over $K(B)$.

(iii) There exists a field K_{t} such that $K \subseteq K_{\mathrm{t}} \subseteq K'$, K_{t} is transcendental over K, and K' is algebraic over K_{t} (cf. Exercise 7.18).

We are now in a position to prove a property of prime ideals in multivariate polynomial rings over K that we have mentioned before in Section 6.3. With the notation $K[\underline{X}] = K[X_1, \ldots, X_n]$, let I be a prime ideal of $K[\underline{X}]$. Set $R = K[\underline{X}]/I$. Then R is an integral domain, and we may consider its field of fractions Q_R. For $f \in K[\underline{X}]$, we will denote the residue class $f + I \in R$ by $\overline{f}$. The composition of canonical homomorphism and canonical inclusion gives a homomorphism

$$\begin{tikzcd}
K[\underline{X}] \arrow[r] & R \arrow[r] & Q_R \\
f \arrow[r, mapsto] & \overline{f} \arrow[r, mapsto] & \overline{f}
\end{tikzcd}$$

from $K[\underline{X}]$ to Q_R. The first homomorphism becomes injective when restricted to K, for otherwise I would contain a constant and thus not be proper. We may thus identify each $a \in K$ with its image $\overline{a}$ in Q_R and operate on the assumption that Q_R is an extension field of K. With this setup, we get the following connection between algebraic independence over K and independence modulo I (Definition 6.46).

Lemma 7.25 Let $\{U_1, \ldots, U_r\} \subseteq \{X_1, \ldots, X_n\}$. Then the following are equivalent:

(i) $\{U_1, \ldots, U_r\}$ is maximally independent modulo I.

(ii) The residue classes $\overline{U_1}$, ..., $\overline{U_r}$ are pairwise different, and the set

$$B = \{\, \overline{U_i} \mid 1 \leq i \leq r \,\}$$

is a transcendence base of Q_R over K.

Proof Renumbering variables if necessary, we may assume w.l.o.g. that $\{U_1, \ldots, U_r\} = \{X_1, \ldots, X_r\}$.

(i)$\Longrightarrow$(ii): If $\overline{X_i}$ were equal to $\overline{X_j}$ for some $1 \leq i < j \leq r$, then

$$\overline{X_i} - \overline{X_j} = \overline{X_i - X_j} = 0$$

and thus $0 \neq X_i - X_j \in I$, contradicting the independence of $\{X_1, \ldots, X_r\}$ modulo I. If B were not algebraically independent over K, then there would be a polynomial $0 \neq f \in K[Y_1, \ldots, Y_r]$ with

$$0 = f(\overline{X_1}, \ldots, \overline{X_r}) = \overline{f(X_1, \ldots, X_r)},$$

and thus $0 \neq f(X_1, \ldots, X_r) \in I \cap K[X_1, \ldots, X_r]$, again a contradiction. It remains to show that B is in fact a maximal algebraically independent set over K. Since $\{X_1, \ldots, X_r\}$ is maximally independent modulo I, we can find, for each $r < i \leq n$, a non-zero polynomial $f_i \in I \cap K[X_1, \ldots, X_r, X_i]$. Then

$$0 = \overline{f} = f(\overline{X_1}, \ldots, \overline{X_r}, \overline{X_i})$$

in Q_R, and thus $\overline{X_i}$ is algebraic over $K(\overline{X_1}, \ldots, \overline{X_r})$ by Lemma 7.21. It is not hard to see that

$$Q_R = K(\overline{X_1}, \ldots, \overline{X_n}) = K(\overline{X_1}, \ldots, \overline{X_r})(\overline{X_{r+1}}, \ldots, \overline{X_n}).$$

In view of Corollary 7.17 (ii), we may now conclude that Q_R is an algebraic extension of $K(\overline{X_1}, \ldots, \overline{X_r})$, and it follows easily from Lemma 7.22 that B is maximal as an independent subset of Q_R over K.

(ii)$\Longrightarrow$(i): To see that $\{X_1, \ldots, X_r\}$ is independent modulo I, let $f \in I \cap K[X_1, \ldots, X_r]$. Then

$$0 = \overline{f(X_1, \ldots, X_r)} = f(\overline{X_1}, \ldots, \overline{X_r}),$$

and thus $f = 0$ by the algebraic independence of B. If $r < i \leq n$, then $\{\overline{X_1}, \ldots, \overline{X_r}, \overline{X_i}\}$ is no longer algebraically independent over K, and thus

$$0 = f(\overline{X_1}, \ldots, \overline{X_r}, \overline{X_i}) = \overline{f(X_1, \ldots X_r, X_i)}$$

for some $0 \neq f \in K[X_1, \ldots, X_r, X_i]$. This means that

$$f_i \in I \cap K[X_1, \ldots, X_r, X_i],$$

and we have proved that the set $\{X_1, \ldots, X_r\}$ is maximally independent modulo I. □

Together with Theorem 7.23 (i), we immediately obtain the following proposition.

Proposition 7.26 *If I is a prime ideal in a polynomial ring over a field, then every maximally independent set modulo I has* $\dim(I)$ *many elements.*

A second proof of this fact which does not use field theory will surface in Section 7.5.

7.2 The Algebraic Closure of a Field

Definition 7.27 A field K is called **algebraically closed** if every non-constant polynomial $f \in K[X]$ has a zero in K, i.e., there is $a \in K$ with $f(a) = 0$.

The *fundamental theorem of algebra* states that the field $\mathbb{C}$ of complex numbers is algebraically closed. This theorem can of course only be proved on the basis of a rigorous definition of $\mathbb{C}$. Here, we show that assuming the axiom of choice, every field has an algebraic extension which is algebraically closed and has a universal embedding property.

Following are some equivalent characterizations of algebraically closed fields.

Lemma 7.28 Let K be a field. Then the following are equivalent:

(i) K is algebraically closed.

(ii) Every irreducible polynomial in $K[X]$ has a zero in K.

(iii) In $K[X]$, every irreducible polynomial is linear.

(iv) In $K[X]$, every non-constant polynomial has a factorization into linear polynomials.

Proof (i)$\Longrightarrow$(ii): Every irreducible polynomial is a non-unit in $K[X]$ and hence not constant.

(ii)$\Longrightarrow$(iii): Let $f \in K[X]$ be irreducible. Then f has a zero in K, say $f(a) = 0$ with $a \in K$. By Proposition 2.95, there exists $g \in K[X]$ with $f = g \cdot (X - a)$, and g must be a constant by the irreducibility of f.

(iii)$\Longrightarrow$(iv): This is immediate from Theorem 2.51.

(iv)$\Longrightarrow$(i): Let $f \in K[X]$ be non-constant, and let $aX + b$ with $a, b \in K$ be a non-constant linear factor of f, i.e., $a \neq 0$. Then $-b/a \in K$ is a zero of f. □

Theorem 7.29 *Let K be a field. Then there exists an algebraically closed algebraic extension field $\overline{K}$ of K with the the following universal embedding property over K: whenever K' is an algebraically closed extension field of K, then there exists a K-embedding $\varphi : \overline{K} \longrightarrow K'$, and if, in addition, K' is algebraic over K, then φ is a K-isomorphism.*

Proof In preparation of the proof, we will, for arbitrary field L, define a certain set A_L of extension fields of L. Let $L[\boldsymbol{X}]$ denote the polynomial ring in the variables $\{ X_f \mid f \in L[X] \setminus L \}$ over L as defined in the discussion following Lemma 2.22. We thus have one variable for each univariate non-constant polynomial over L. For each such f, there is an embedding $\varphi_f : L[X] \longrightarrow L[\boldsymbol{X}]$ which is determined by the requirement $\varphi_f(X) = X_f$. Then $\varphi_f(f) = f(X_f)$, and we set

$$P = \{ f(X_f) \mid f \in L[X] \setminus L \}.$$

We claim that the ideal $\mathrm{Id}(P)$ of $L[\boldsymbol{X}]$ is proper. Assume for a contradiction that $\mathrm{Id}(P) = L[\boldsymbol{X}]$. Then there exist $f_1, \ldots, f_m \in P$ and $q_1, \ldots, q_m \in L[\boldsymbol{X}]$ with

$$1 = \sum_{i=1}^{m} q_i f_i. \qquad (*)$$

Let $Y_1, \ldots, Y_r$ be the indeterminates occuring in $q_1, \ldots, q_m, f_1, \ldots, f_m$. The set $F = \{f_1, \ldots, f_m\}$ is a Gröbner basis in $L[\underline{Y}]$ because any two elements of F have disjoint head terms. F does not contain a constant, and hence $\mathrm{Id}(F)$ is proper in $L[\underline{Y}]$ by Corollary 6.16, contradicting $(*)$. We now define the set A_L by setting

$$A_L = \{ L[\boldsymbol{X}]/M \mid P \subseteq M \text{ a maximal ideal of } L[\boldsymbol{X}] \}.$$

Using the axiom of choice, we see from Lemma 4.9 that A_L is not empty. Now let $L' \in A_L$, i.e., $L' = L[\boldsymbol{X}]/M$ for some maximal ideal M extending P, and denote by $\overline{f}$ the residue class of $f \in L[\boldsymbol{X}]$ in $L[\boldsymbol{X}]/M$. L' is a field because M is maximal. It is easy to see that the canonical homomorphism $f \longmapsto \overline{f}$ is injective when restricted to L (cf. the discussion on page 295), and so we may regard L' as an extension field of L. Under this point of view, every non-constant polynomial $f \in L[X]$ has a zero in L': from the homomorphism property of the bar, we see that $f(\overline{X_f}) = \overline{f} = 0$. We claim that L' is algebraic over L. Every element of L' is of the form $\overline{f}$ for some $f \in L[\boldsymbol{X}]$, and thus we have

$$L' = L(\{\,\overline{X_f} \mid f \in L[X] \setminus L\,\}).$$

Each $\overline{X_f}$ is algebraic over L, and so L' is algebraic over L according to Corollary 7.17 (ii).

We are now in a position to define the desired extension $\overline{K}$ of the given field K. According to the discussion at the end of Section 4.1, there exists an ascending chain

$$K = K_0 \subseteq K_1 \subseteq K_2 \subseteq \cdots$$

of fields with $K_0 = K$ and $K_{n+1} \in A_{K_n}$ for all $n \in \mathbb{N}$, and we let

$$\overline{K} = \bigcup_{n \in \mathbb{N}} K_n.$$

It is easy to see that $\overline{K}$ is a field: any two elements of $\overline{K}$ lie in some common K_n where they may be added, subtracted or multiplied in a well-defined manner, and one then verifies the field axioms by inspection. $\overline{K}$ obviously extends K. Moreover, it is an algebraic extension of K: if $a \in \overline{K}$, then $a \in K_n$ for some $n \in \mathbb{N}$ and thus a is algebraic over K by Corollary 7.17 (iv). If f is a non-constant polynomial in $\overline{K}[X]$, then, since it has only finitely many coefficients, $f \in K_n[X]$ for some $n \in \mathbb{N}$, and we may conclude that f has a zero in $K_{n+1} \subseteq \overline{K}$ because $K_{n+1} \in A_{K_n}$. We have proved that $\overline{K}$ is an algebraically closed algebraic extension of K.

To verify the universal embedding property of $\overline{K}$, let K' be an algebraically closed extension of K. Let Π be the set of all partial K-embeddings from $\overline{K}$ to K', i.e., the set of all K-embeddings from intermediate fields $\widehat{K}$ with $K \subseteq \widehat{K} \subseteq \overline{K}$ to K'. Then $\Pi \subseteq \mathcal{P}(\overline{K} \times K')$, and Π is not empty since $\mathrm{id}_K \in \Pi$. By Zorn's lemma, Π has a maximal element φ w.r.t. inclusion. We claim that the domain $\widehat{K}$ of φ equals $\overline{K}$. To this end, we view φ as a K-isomorphism from $\widehat{K}$ to $\widehat{K}' = \varphi(\widehat{K})$.

$$\begin{tikzcd}
K \arrow[r, phantom, "\subseteq"] \arrow[d, equal] & \widehat{K} \arrow[r, phantom, "\subseteq"] \arrow[d, "\varphi"'] & \overline{K} \\
K \arrow[r, phantom, "\subseteq"] & \widehat{K}' \arrow[r, phantom, "\subseteq"] & K'
\end{tikzcd}$$

Now assume for a contradiction that $a \in \overline{K} \setminus \widehat{K}$. Since a is algebraic over K, it is trivially algebraic over the extension field $\widehat{K}$ of K. Let

$$f = \min\nolimits_{\widehat{K}}^{a} \in \widehat{K}[X],$$

and consider $g = \varphi'(f) \in \widehat{K}'[X]$ where φ' is the isomorphism of Lemma 2.17 (ii). Then g is monic and $\deg(g) > 1$ because the same is true for f.

Moreover, g is irreducible in $\widehat{K}'[X]$ since a proper factorization of g over $\widehat{K}'$ would give rise to a proper factorization of f over $\widehat{K}$ via the inverse isomorphism $(\varphi')^{-1}$. Since K' is algebraically closed, g has a zero b in K', and $g = \min^b_{\widehat{K}'}$ by Corollary 7.6. Corollary 7.10 provides an extension

$$\overline{\varphi} : \widehat{K}(a) \longrightarrow \widehat{K}'(b)$$

of φ, contradicting the maximality of φ in Π.

Finally, let K' be algebraic over K, and assume for a contradiction that there exists $b \in K' \setminus \varphi(\overline{K})$. Being algebraic over K, b is trivially algebraic over the extension field $\varphi(\overline{K})$ of K, and $g = \min^b_{\varphi(\overline{K})}$ is a non-linear monic irreducible element of $\varphi(\overline{K})[X]$ since $b \notin \varphi(\overline{K})$. But then the preimage of g under the induced isomorphism

$$\varphi' : \overline{K}[X] \longrightarrow \varphi(\overline{K})[X]$$

would be a non-linear irreducible polynomial over the algebraically closed field $\overline{K}$, contradicting Lemma 7.28 (iii). □

An algebraically closed algebraic extension of a field K is called an **algebraic closure** of K. By the theorem above, every field K has an algebraic closure which is unique up to K-isomorphism. It is therefore referred to as *the* algebraic closure of K, and we will denote it by $\overline{K}$.

7.3 Separable Polynomials and Perfect Fields

In this section, we discuss a class of fields that plays an important role in the theory of polynomial ideals. We will make liberal use of the elementary facts on zeroes of polynomials that were proved at the beginning of Section 2.7. Throughout this section, K will be a field. In view of Lemma 7.28, we now have the following result: there exists an extension field $\overline{K}$ of K, namely, the algebraic closure of K, such that every polynomial $f \in K[X]$ has a factorization into linear factors in $\overline{K}[X]$. In this factorization, associated factors correspond to multiple zeroes of f in $\overline{K}$, i.e., zeroes with multiplicity greater than 1.

Definition 7.30 A polynomial $f \in K[X]$ is called **separable** if it is either a non-zero constant, or the factorization into non-constant linear polynomials of f in $\overline{K}[X]$ consists of pairwise non-associated factors.

A separable polynomial is thus a polynomial that does not have multiple zeroes in $\overline{K}$. Using the universal embedding property of the algebraic closure, it is easy to see that in this case, f cannot have multiple zeroes in any algebraically closed extension of K.

Exercise 7.31 Let $f \in K[X]$ be separable, K' any extension field of K. Show that f is still separable when viewed as an element of $K'[X]$, and that every zero of f in K' has multiplicity 1. (Hint: Consider $\overline{K'}$.)

It is clear that a separable polynomial f must be squarefree, because if f has a non-constant factor g^2, then the zeroes of g in $\overline{K}$ will be zeroes of f of multiplicity at least 2. We will later see a large class of fields over which every squarefree polynomial is separable. Let us first demonstrate how a squarefree polynomial can fail to be separable.

Example 7.32 Let p be a prime number, and let K be the rational function field $\mathbb{Z}/p\mathbb{Z}(T)$. Consider the polynomial $f = X^p - T \in K[X]$. Since f is not constant, it has a zero a in $\overline{K}$. Then $a^p = T$, and using Lemma 1.106, we obtain the factorization

$$f = X^p - T = X^p - a^p = (X - a)^p$$

in $\overline{K}[X]$. We claim that f is irreducible (and hence squarefree) in $K[X]$. Assume for a contradiction that this is not the case. Then there are non-constant monic $g, h \in K[X]$ with $f = gh$. Passing to $\overline{K}[X]$, we conclude from the unique prime factor decomposition that there is $0 < i < p$ with

$$g = (X - a)^i \quad \text{and} \quad h = (X - a)^{p-i}.$$

Looking at the constant coefficient of g, we see that $a^i \in K$, i.e., there exist $s, q \in \mathbb{Z}/p\mathbb{Z}[T]$ with $a^i = s/q$. We conclude that

$$s^p = (a^i q)^p = a^{ip} q^p = T^i q^p.$$

The degree in T of the polynomial on the left-hand side is a multiple of p, whereas the one on the right is congruent i modulo p, a contradiction.

Recall from Lemma 2.84 that a polynomial that is prime to its derivative is always squarefree. The next proposition refines this result.

Proposition 7.33 *Let $f \in K[X]$. Then the following are equivalent:*

(i) $\gcd(f, f') = 1$.

(ii) f is squarefree in $K'[X]$ for every extension field K' of K.

(iii) f is separable.

Proof (i)$\Longrightarrow$(ii): Property (i) is invariant under field extensions by Proposition 2.38, and it implies that f is squarefree by Lemma 2.84.

(ii)$\Longrightarrow$(iii): If f had multiple zeroes in $\overline{K}$, then it would no longer be squarefree in $\overline{K}[X]$.

(iii)$\Longrightarrow$(i): If f is a non-zero constant, then the claim is trivial. Else, assume for a contradiction that $q \mid f'$ for some irreducible factor q of f in $K[X]$. Then $f = qg$ for some $g \in K[X]$, and thus $f' = q'g + qg'$. We see that $q \mid q'g$, and so $q \mid q'$ or $q \mid g$. The latter is impossible since the prime factors of f are pairwise non-associated, and the former is possible only if $q' = 0$ for reasons of the degree. Being irreducible, q is not constant, and so by Lemma 2.79, $q' = 0$ implies that $\mathrm{char}(K) = p \neq 0$, and that we can write

$$q = \sum_{i=0}^{m} a_i X^{ip} \qquad (a_0, \ldots, a_m \in K).$$

By the same argument that we have used in Example 7.32 above, a_i has a pth root b_i in $\overline{K}$ for $0 \leq i \leq m$, and we obtain the factorization

$$q = \sum_{i=0}^{m} b_i^p X^{ip} = \left(\sum_{i=0}^{m} b_i X^i\right)^p$$

of q in $\overline{K}[X]$. This contradicts the separability of f. □

Definition 7.34 A field K is called **perfect** if every irreducible polynomial $f \in K[X]$ is separable.

From the fact that every irreducible polynomial over an algebraically closed field is linear, one easily deduces that every algebraically closed field is perfect. Example 7.32 shows that if p is a prime number, then the rational function field $\mathbb{Z}/p\mathbb{Z}(T)$ is not perfect. We will soon see examples of perfect fields that are not algebraically closed.

Lemma 7.35 A field K is perfect iff every squarefree polynomial in $K[X]$ is separable.

Proof The direction "$\Longleftarrow$" is trivial because every irreducible polynomial is squarefree. For "$\Longrightarrow$," suppose K is perfect, and let $f \in K[X]$ be squarefree. If f is a constant, then it must be non-zero and the claim is trivial. Otherwise, there are pairwise non-associated, irreducible polynomials $p_1, \ldots, p_r \in K[X]$ with $f = p_1 \cdot \cdots \cdot p_r$. Assume for a contradiction that K' is an extension field of K and $a \in K'$ with $(X - a)^2 \mid f$ in $K'[X]$. Then $X - a \mid p_i$ in $K'[X]$ for some $1 \leq i \leq r$ because $X - a$ is irreducible and $K'[X]$ is a UFD. We cannot have $X - a \mid p_j$ for any $j \neq i$ because the gcd of p_i and p_j equals 1 in $K[X]$ and thus in $K'[X]$ (Proposition 2.38). It follows easily that in fact $(X - a)^2 \mid p_i$, contradicting the fact that p_i is separable. □

Theorem 7.36 *Let K be a field. Then the following are equivalent:*

(i) K is perfect.

(ii) A non-constant polynomial $f \in K[X]$ is squarefree iff $\gcd(f, f') = 1$.

Proof (i)$\Longrightarrow$(ii): The direction "$\Longleftarrow$" is Lemma 2.84. If f is squarefree, then it is separable by the previous lemma, and so $\gcd(f, f') = 1$ by Proposition 7.33.

(ii)$\Longrightarrow$(i): If K is not perfect, then there exists a squarefree polynomial $f \in K[X]$ which is not separable. Proposition 7.33 tells us that 1 is not a gcd of f and f', and we see that "$\Longrightarrow$" of (ii) is violated. □

Example 7.32 shows that (ii) of the theorem above does not hold for arbitrary field K: here, f is squarefree because it is irreducible, but $f' = 0$ and so $\gcd(f, f') = f \neq 1$. The following corollary is immediate from Lemmas 2.84 and 2.85.

Corollary 7.37 *All fields of characteristic zero and all finite fields are perfect.* □

Exercise 7.38 Show that a field K of characteristic $p \neq 0$ is perfect iff every element of K has a pth root in K.

Corollary 7.39 *A field K is perfect iff every squarefree polynomial $f \in K[X]$ is still squarefree in $K'[X]$ for every extension field K' of K.*

Proof If K is perfect, then by (ii) of the theorem, squarefreeness is preserved under field extensions along with gcd's. If K is not perfect, then there exists an irreducible polynomial in $K[X]$ which is not squarefree in $\overline{K}[X]$ because it has multiple zeroes in $\overline{K}$. □

7.4 The Hilbert Nullstellensatz

The main purpose of this section is to prove the following Nullstellensatz (theorem on zeroes).

Theorem 7.40 (HILBERT NULLSTELLENSATZ) *Let K be a field, L an algebraically closed extension field of K, and $f, g_1, \dots, g_m \in K[X_1, \dots, X_n]$. Then the following are equivalent:*

(i) For all $z \in L^n$, $g_1(z) = \cdots = g_m(z) = 0$ implies $f(z) = 0$.

(ii) There exists $0 < s \in \mathbb{N}$ with $f^s \in \mathrm{Id}(g_1, \dots, g_m)$.

Condition (ii) above can also be read as "$f \in \mathrm{rad}(\mathrm{Id}(g_1, \dots, g_m))$," and the validity of this condition can be decided by means of the algorithm RADICALMEMTEST of Corollary 6.41. The theorem thus allows us to decide a property of $f, g_1, \dots, g_m$ which concerns geometric configurations in L^n by means of a computation that takes place in $K[X_1, \dots, X_n]$. The most natural example to think of throughout this section is of course $K = \mathbb{Q}$ and $L = \mathbb{C}$.

Before we tackle the proof of the Hilbert Nullstellensatz, we give an alternate formulation which uses some terminology that is common in algebraic geometry. Let K' be an extension field of K, $z \in (K')^n$, and $P \subseteq K[X_1, \dots, X_n]$. Then we say that z is a **zero** of P if it is a zero of every $p \in P$. The **variety** of P in $(K')^n$ is the set of all zeroes of P in $(K')^n$. It is clear that every zero of P is a zero of $\mathrm{Id}(P)$ and vice versa, so that the varieties of P and $\mathrm{Id}(P)$ are equal. If $V \subseteq (K')^n$ and $f \in K[X_1, \dots, X_n]$, then we say, rather obviously, that f *vanishes on* V

if $f(z) = 0$ for all $z \in V$. It is now easy to see that the statement of the Hilbert Nullstellensatz can be rephrased as follows.

(HILBERT NULLSTELLENSATZ, ALTERNATE FORMULATION) *Let L be an algebraically closed extension field of K and I an ideal of $K[X_1, \ldots, X_n]$. Then* $\operatorname{rad}(I)$ *consists of precisely those $f \in K[X_1, \ldots, X_n]$ that vanish on the variety of I in L^n.*

There is now an immediate corollary, which will of course not be used until the Hilbert Nullstellensatz has been proved. Recall that a radical ideal is an ideal that equals its radical.

Corollary 7.41 *Let L be an algebraically closed extension field of K. Then a radical ideal of $K[X_1, \ldots, X_n]$ consists of precisely those polynomials that vanish on its variety in L^n. If I_1 and I_2 are radical ideals of $K[X_1, \ldots, X_n]$, then $I_1 = I_2$ iff the varieties of I_1 and I_2 in L^n agree.* □

Our proof of the Hilbert Nullstellensatz relies strongly on Gröbner bases; only elementary ring and field theory are being used otherwise. We begin with several results that are instrumental in the proof and are also interesting in their own right. Throughout this section, K will be a field.

Proposition 7.42 *Let I be a zero-dimensional prime ideal of the polynomial ring $K[X_1, \ldots, X_n]$, and let G be the reduced Gröbner basis of I w.r.t. the inverse lexicographical term order $\leq$, where $X_n \gg \cdots \gg X_1$. Then the following hold:*

(i) I is maximal.

(ii) G has n elements $g_1, \ldots, g_n$, and $\mathrm{HM}(g_i) = X_i^{\nu_i}$ with $\nu_i \geq 1$ for $1 \leq i \leq n$. (In particular, $g_i \in K[X_1, \ldots, X_i]$ for $1 \leq i \leq n$.)

Proof We proceed by induction on n. The case $n = 1$ is trivial: $K[X]$ is a PID, and every non-trivial prime ideal is maximal (Proposition 1.97). Now assume that $n > 1$. Let I_{n-1} be the elimination ideal of I w.r.t. $\{X_1, \ldots, X_{n-1}\}$ and $G_{n-1} = G \cap K[X_1, \ldots, X_{n-1}]$. It is easy to see that I_{n-1} is a prime ideal of $K[X_1, \ldots, X_{n-1}]$, and it is zero-dimensional by Lemma 6.50 (ii). By induction hypothesis and Proposition 6.15, I_{n-1} is a maximal ideal of $K[X_1, \ldots, X_{n-1}]$, and $G_{n-1} = \{g_1, \ldots, g_{n-1}\}$ with $\mathrm{HM}(g_i) = X_i^{\nu_i}$ ($\nu_i \geq 1$) for $1 \leq i \leq n-1$. The set

$$N = \{\, f \in I \mid \mathrm{HM}(f) = X_n^{\nu},\ f \text{ in normal form modulo } G_{n-1} \,\}$$

is not empty because I contains a non-constant univariate polynomial in X_n. Let $g_n \in N$ be of minimal degree in X_n, say $\mathrm{HM}(g_n) = X_n^{\nu_n}$. Then $\nu_n \geq 1$ since I is proper. $\{g_1, \ldots, g_n\}$ is a Gröbner basis w.r.t. $\leq$ by Lemma 5.66, and it is even a reduced Gröbner basis by the choice of g_n. To prove (ii), it remains to show that $I = \mathrm{Id}(g_1, \ldots, g_n)$. The inclusion "$\supseteq$" is trivial. Assume for a contradiction that there exists $0 \neq f \in I \setminus \mathrm{Id}(g_1, \ldots, g_n)$.

We may assume w.l.o.g. that f is in normal form modulo $\{g_1,\dots,g_n\}$. Regarding f as an element of $K[X_1,\dots,X_{n-1}][X_n]$, we may write

$$f=\sum_{i=0}^{r} h_i X_n^i$$

with $h_i \in K[X_1,\dots,X_{n-1}]$ for $0 \leq i \leq r$ and $h_r \neq 0$. Then we must have $0 < r < \nu_n$ because f is in normal form modulo g_n. Furthermore, $h_r \notin I_{n-1}$ since otherwise f would have to be reducible modulo the Gröbner basis $G_{n-1} = \{g_1,\dots,g_{n-1}\}$ of I_{n-1}. Because of the maximality of I_{n-1}, there exist $p \in K[X_1,\dots,X_{n-1}]$ and $q \in I_{n-1}$ with $ph_r + q = 1$. Then $g = pf + qX_n^r \in I$, and

$$\begin{aligned} g &= ph_rX_n^r + p\sum_{i=0}^{r-1} h_iX_n^i + qX_n^r \\ &= X_n^r + p\sum_{i=0}^{r-1} h_iX_n^i. \end{aligned}$$

We see that $\deg_{X_n}(g) = r < \nu_n$, contradicting the choice of g_n.

It remains to show that I is maximal. Assume for a contradiction that it is not. Then by Lemma 4.9, there exists a maximal ideal J of $K[X_1,\dots,X_n]$ with $I \subseteq J$ and $I \neq J$. Let J_{n-1} be the elimination ideal of J w.r.t. $\{X_1,\dots,X_{n-1}\}$. We trivially have $I_{n-1} \subseteq J_{n-1}$, and this together with maximality of I_{n-1} and properness of J_{n-1} implies $I_{n-1} = J_{n-1}$. We see that G_{n-1} is a Gröbner basis of J_{n-1}. J is prime because it is maximal, and it has dimension zero because it contains the zero-dimensional ideal I (Lemma 6.50). We can thus extend G_{n-1} to a Gröbner basis $G^* = \{g_1,\dots,g_{n-1},g_n^*\}$ of J in the exact same way as we did for I. From $g_n \in J$ we may conclude that

$$\nu_n' = \deg_{X_n}(g_n^*) \leq \deg_{X_n}(g_n) = \nu_n.$$

For the same reason, g_n must be reducible to 0 modulo the Gröbner basis G^* of J. Reducing g_n completely modulo g_n^* first, we obtain $h \in J$ with

$$h = g_n - qg_n^*$$

for some $q \in K[X_1,\dots,X_n]$, and $\deg_{X_n}(h) < \deg_{X_n}(g_n^*) = \nu_n'$ or $h = 0$. If h is not yet equal to zero, then it must reduce to zero modulo G^*. Under the current term order, a reduction step does not increase the degree in X_n. It follows easily that

$$h \xrightarrow[\{g_1,\dots,g_{n-1}\}]{*} 0,$$

and thus $h \in I$. We conclude that $qg_n^* = g_n - h \in I$, and thus $q \in I$ or $g_n^* \in I$ since I is prime. $g_n^* \in I$ would imply $I = J$ contrary to our assumption. The fact that

$$h = 0 \quad \text{or} \quad \deg_{X_n}(h) < \nu_n' \leq \nu_n$$

implies that $\mathrm{HT}(g_n - h) = X_n^{\nu_n}$, and we see that $\mathrm{HT}(q) = X_n^{\nu_n - \nu_n'}$. Now if q were in I, we would either be contradicting the properness of I (if $\nu_n' = \nu_n$), or the choice of g_n (if $\nu_n' < \nu_n$). □

If I is a prime ideal, then the reduced Gröbner basis of I w.r.t. the inverse lexicographical term order, of which the proposition above gives a description, is also called the **prime basis** of I.

We will later show that a polynomial ideal is maximal if *and only if* it is zero-dimensional and prime. The direction "$\Longleftarrow$" that we have just proved can also be shown using linear algebra arguments in the K-vector space $K[X_1, \ldots, X_n]/I$. The proof then becomes a little simpler, but it is noteworthy that using Gröbner bases, one can do it within the theory of polynomial rings.

Exercise 7.43 Use linear algebra in the K-vector space $K[X_1, \ldots, X_n]/I$ to show that every zero-dimensional prime ideal is maximal. (Hint: Use Lemma 3.23 (ii) to imitate the proof of Lemma 1.19 (iii) to show that $K[X_1, \ldots, X_n]/I$ is a field.)

A natural question arising at this point is whether an ideal that has a basis that looks like a prime basis is actually a zero-dimensional prime ideal. The next proposition states that this is true under an additional assumption.

Proposition 7.44 *Let I be an ideal of the ring $K[X_1, \ldots, X_n]$, and assume that I has a basis G as described in (ii) of the previous proposition, where the head terms are taken w.r.t. the inverse lexicographical term order. Assume further that for $1 \leq i \leq n$, there does not exist a representation*

$$g_i = f_1 f_2 + \sum_{j=1}^{i-1} q_j g_j$$

with f_1, f_2, q_1, $\ldots$, $q_{i-1} \in K[X_1, \ldots, X_i]$ such that f_1, $f_2 \neq 0$ and

$$\deg_{X_i}(f_1) < \deg_{X_i}(g_i) \quad \text{and} \quad \deg_{X_i}(f_2) < \deg_{X_i}(g_i).$$

Then I is a zero-dimensional prime ideal.

Proof All statements regarding Gröbner bases will be referring to the inverse lexicographical term order. G is a Gröbner basis of I since every two elements of G have disjoint head terms. It now follows from Theorem 6.54 that I is zero-dimensional. Next, we note that if we mutually reduce the elements of G, then no top-reductions will take place, and it is now easy to see that the *reduced* Gröbner basis G' of I still fits the description of (ii) of the previous proposition. Moreover, using Lemma 6.14 and the fact that

$$\{g_1,\dots,g_i\} \subseteq K[X_1,\dots,X_i] \quad \text{for} \quad 1 \leq i \leq n,$$

it is not hard to prove by induction on n that G' still satisfies the additional hypothesis of the present proposition. We may thus assume w.l.o.g. that G is reduced.

To show that I is a prime ideal, we will actually prove that it is maximal. We proceed by induction on n. If $n = 1$, then I is generated by an irreducible polynomial and hence is maximal. Now let $n > 1$, and assume for a contradiction that I is not maximal. Lemma 4.9 provides the existence of a maximal ideal J of $K[X_1,\dots,X_n]$ with $I \subseteq J$ and $I \neq J$. J is prime because it is maximal, and it has dimension zero because it contains the zero-dimensional ideal I. Let $H = \{h_1,\dots,h_n\}$ be the prime basis of J, and let I_{n-1} and J_{n-1} be the elimination ideals of I and J w.r.t. $\{X_1,\dots,X_{n-1}\}$, respectively. Then by Proposition 6.15, the sets

$$H_{n-1} = \{h_1,\dots,h_{n-1}\} \quad \text{and} \quad G_{n-1} = \{g_1,\dots,g_{n-1}\}$$

are the reduced Gröbner bases of J_{n-1} and I_{n-1}, respectively. By the induction hypothesis, I_{n-1} is a zero-dimensional prime ideal of $K[X_1,\dots,X_{n-1}]$, and hence it is maximal by the previous proposition. From the inclusion $I_{n-1} \subseteq J_{n-1}$ and the properness of J_{n-1} it now follows that $I_{n-1} = J_{n-1}$, and so H_{n-1} and G_{n-1} are equal by the uniqueness of the reduced Gröbner basis. It is easy to see that we must in fact have $g_i = h_i$ for $1 \leq i \leq n-1$. From $g_n \in J$ we conclude that

$$g_n \xrightarrow[H]{*} 0.$$

We see that $\deg_{X_n}(g_n) < \deg_{X_n}(h_n)$ is impossible, because then g_n would not be top-reducible modulo H. We may now perform the reduction of g_n modulo H in such a way that we first reduce modulo h_n until the degree in X_n is less than $\deg_{X_n}(h_n)$. Since reduction modulo $\{g_1,\dots,g_{n-1}\}$ does not increase the degree in X_n, the remaining reduction steps modulo H must actually be modulo $\{g_1,\dots,g_{n-1}\}$, and we arrive at a representation

$$g_n = q_n h_n + \sum_{i=1}^{n-1} q_i g_i,$$

where $q_1, \dots, q_n \in K[X_1,\dots,X_n]$, and either

$$q_n = 1 \quad \text{or} \quad 1 < \deg_{X_n}(q_n) = \deg_{X_n}(g_n) - \deg_{X_n}(h_n) < \deg_{X_n}(g_n),$$

depending on whether

$$\deg_{X_n}(h_n) = \deg_{X_n}(g_n) \quad \text{or} \quad \deg_{X_n}(h_n) < \deg_{X_n}(g_n).$$

If $q_n = 1$, then it follows that $h_n \in I$, contradicting the fact that $I \neq J$. Otherwise, we are contradicting the additional assumption made on G. □

The following example shows how the criterion of the previous proposition can often be verified for simple ideal bases.

Example 7.45 Let $K = \mathbb{Q}$, $n = 2$, and $G = \{X_2^2 + X_1^2, X_1^2 - 2\}$. We claim that $\mathrm{Id}(G)$ is a zero-dimensional prime ideal. G meets the description of Proposition 7.42 (ii). The polynomial $X_1^2 - 2$ is clearly irreducible in $\mathbb{Q}[X_1]$. Moreover, if we were given a representation

$$X_2^2 + X_1^2 = h_1 h_2 + q(X_1^2 - 2) \qquad (*)$$

with h_1, h_2, $q \in K[X_1, X_2]$) and $\deg_{X_2}(h_i) < 2$ for $i = 1, 2$, then we could set $X_1 = \sqrt{2}$ and conclude that $X_2^2 + 2$ factors into linear polynomials over $K = \mathbb{Q}$, which is clearly not true. Note that there may well be representations of the type $(*)$ with $\deg_{X_2}(h_i) \leq 2$ for $i = 1, 2$, e.g.,

$$X_2^2 + X_1^2 = (X_1^2/2)(X_2^2 + X_1^2) - ((X_2^2 + X_1^2)/2)(X_1^2 - 2).$$

Exercise 7.46 Let I be an ideal of the ring $K[X_1, \ldots, X_n]$, and assume that I has a basis G as described in (ii) of Proposition 7.42, where the head terms are taken w.r.t. the inverse lexicographical term order. Show the following:

(i) For $1 \leq i \leq n$, we have $g_i \notin \mathrm{Id}(g_1, \ldots, g_{i-1})$. (This means that the condition "f_1, $f_2 \neq 0$" in Proposition 7.44 is not an essential one; it was only made to be able to express the degree condition on f_1 and f_2 without further ado.)

(ii) The additional assumption on G of Proposition 7.44 is in fact equivalent to I being prime. Moreover, the following is a third equivalent condition: for $1 \leq i \leq n$, the residue class $\overline{g_i} = g_i + \mathrm{Id}(g_1, \ldots, g_{i-1})$ of g_i is irreducible in in the ring $K[X_1, \ldots, X_i]/\mathrm{Id}(g_1, \ldots, g_{i-1})$, i.e., $\overline{g_i}$ is not a unit and cannot be written as a product of two non-units in that ring.

The next proposition provides a generalization of Proposition 7.42 to prime ideals whose dimension is greater than zero. To this end, we need to discuss an important technique for reducing proofs in ideal theory to the zero-dimensional case. If $1 \leq d \leq n$, then $M = K[X_1, \ldots, X_d] \setminus \{0\}$ is a multiplicatively closed subset of $K[X_1, \ldots, X_n]$ with $1 \in M$ and $0 \notin M$, and we may form the ring of quotients $K[X_1, \ldots, X_n]_M$ of $K[X_1, \ldots, X_n]$ w.r.t. M. It is easy to see that

$$K[X_1, \ldots, X_n]_M = K(X_1, \ldots, X_d)[X_{d+1}, \ldots, X_n].$$

We may now consider extensions to the ring of quotients $K[X_1, \ldots, X_n]_M$ and contractions to $K[X_1, \ldots, X_n]$ of ideals as defined preceding Lemma 1.122. In addition to the statements of Lemmas 1.122 and 1.123, we will need the following results that are specific to the present situation.

Lemma 7.47 Suppose $1 \le d \le n$, let I be an ideal of $K[X_1, \ldots, X_n]$, and let I^e be its extension to $K[X_1, \ldots, X_n]_M$, where $M = K[X_1, \ldots, X_d] \setminus \{0\}$. Then the following hold:

(i) I^e is proper iff $\{X_1, \ldots, X_d\}$ is independent modulo I.

(ii) If $\{X_1, \ldots, X_d\}$ is maximally independent modulo I, then I^e is a zero-dimensional ideal of $K[X_1, \ldots, X_n]_M$.

(iii) If I is a prime ideal and I^e is zero-dimensional, then $\{X_1, \ldots, X_d\}$ is maximally independent modulo I.

Proof (i) is immediate from Lemma 1.122 (ii) and the definition of independence.

(ii) By (i) above, I^e is a proper ideal. Now I contains a non-zero element f_i of $K[X_1, \ldots, X_d, X_i]$ for each $d+1 \le i \le n$. If we view

$$f_i \in K(X_1, \ldots, X_d)[X_{d+1}, \ldots, X_n],$$

then it is a univariate polynomial in X_i, and we see that $\dim(I^e) = 0$.

(iii) $\{X_1, \ldots, X_d\}$ is independent modulo I by (i) above. From $\dim(I^e) = 0$ we conclude that there is a non-zero element

$$f_i \in I^e \cap K(X_1, \ldots, X_d)[X_i]$$

for each $d+1 \le i \le n$. If we multiply each f_i by the product of the denominators of its coefficients in $K(X_1, \ldots, X_d)$, then we obtain a non-zero element of

$$I^{ec} \cap K[X_1, \ldots, X_d][X_i]$$

for each $d+1 \le i \le n$. Moreover, $I = I^{ec}$ because I is prime according to Lemma 1.122 (iv), and we see that $\{X_1, \ldots, X_d\}$ is in fact maximally independent modulo I. □

Proposition 7.48 *Let I be a prime ideal of $K[X_1, \ldots, X_n]$, and assume that $\{X_1, \ldots, X_d\}$ is maximally independent modulo I, where $1 \le d < n$. Then there exist polynomials f, $g_{d+1}, \ldots, g_n \in K[X_1, \ldots, X_n]$, none of them equal to zero, with the following properties:*

(i) $f \in K[X_1, \ldots, X_d]$.

(ii) $g_i \in K[X_1, \ldots, X_i]$ for $d+1 \le i \le n$.

(iii) If, for $d+1 \le i \le n$, g_i is viewed as an element of

$$K[X_1, \ldots, X_{i-1}][X_i],$$

then its head coefficient lies in $K[X_1, \ldots, X_d]$.

(iv) $I = \mathrm{Id}(g_{d+1}, \ldots, g_n) : f$.

Proof Let $M = K[X_1,\ldots,X_d] \setminus \{0\}$ and consider the extension I^{e} of I to $K[X_1,\ldots,X_n]_M$. Then I^{e} is a zero-dimensional prime ideal of

$$K[X_1,\ldots,X_n]_M = K(X_1,\ldots,X_d)[X_{d+1},\ldots,X_n],$$

and thus, by Proposition 7.42, has a basis $\{f_{d+1},\ldots,f_n\}$ such that

$$f_i \in K(X_1,\ldots,X_d)[X_{d+1},\ldots,X_i]$$

for $d+1 \leq i \leq n$, and if we view

$$f_i \in K(X_1,\ldots,X_d)[X_{d+1},\ldots,X_{i-1}][X_i],$$

then f_i is monic and non-constant. Clearing all denominators of coefficients in $K(X_1,\ldots,X_d)$, we obtain g_{d+1}, $\ldots$, $g_n \in I^{\mathrm{ec}} = I$, none of them zero, with properties (ii) and (iii). The set $\{g_{d+1},\ldots,g_n\}$ is still a basis of I^{e} in the ring

$$K(X_1,\ldots,X_d)[X_{d+1},\ldots,X_n].$$

Now let $\{h_1,\ldots,h_m\}$ be any basis of I in $K[X_1,\ldots,X_n]$. Since $I \subseteq I^{\mathrm{e}}$, there exist

$$q_{ij} \in K(X_1,\ldots,X_d)[X_{d+1},\ldots,X_n] \qquad (1 \leq i \leq m,\ d+1 \leq j \leq n)$$

with

$$h_i = \sum_{j=d+1}^{n} q_{ij} g_j \qquad (1 \leq i \leq m).$$

We now define f to be the product of all denominators of coefficients in $K(X_1,\ldots,X_d)$ of the q_{ij}, where $1 \leq i \leq m$ and $d+1 \leq j \leq n$. Then $f \neq 0$, (i) holds, and it remains to prove (iv). Let $g \in I$. Then there exist q_1, $\ldots$, $q_m \in K[X_1,\ldots,X_n]$ with

$$g = \sum_{i=1}^{m} q_i h_i = \sum_{i=1}^{m} q_i \sum_{j=d+1}^{n} q_{ij} g_j,$$

and we see that $fg \in \mathrm{Id}(g_{d+1},\ldots,g_n)$. Conversely, let $g \in K[X_1,\ldots,X_n]$ with

$$fg \in \mathrm{Id}(g_{d+1},\ldots,g_n) \subseteq I.$$

Then $g \in I$ because I is prime and moreover, $f \notin I$ by the independence of $\{X_1,\ldots,X_d\}$ modulo I. □

Note that the above proof is constructive: f, g_{d+1}, $\ldots$, g_n can be found by a Gröbner basis computation in the ring $K(X_1,\ldots,X_d)[X_{d+1},\ldots,X_n]$.

If an ideal I of $K[X_1,\ldots,X_n]$ has a zero in some extension field K' of K, then clearly $1 \notin I$ and thus I is proper. The proof of the Hilbert Nullstellensatz rests upon the important fact that conversely, every proper ideal of $K[X_1,\ldots,X_n]$ has at least one zero in L^n for every algebraically closed extension field L of K. This can be viewed as another instance of a univariate result carrying over to the multivariate case: for $n = 1$, we are looking at the fact that every non-constant univariate polynomial over K has a zero in L. For the proof we need four lemmas, each of which is interesting in its own right.

Lemma 7.49 Let L be an algebraically closed field. Then L has infinitely many elements.

Proof Assume for a contradiction that L is an algebraically closed field with finitely many elements. Then in particular, $\operatorname{char}(L) = p \neq 0$. Let $n \in \mathbb{Z}$ be such that $|L| < n$ and $p \nmid n$, and set $f = X^n - 1$. Then $f' = nX^{n-1} \neq 0$, and so $\gcd(f, f') = 1$ since the only prime factor occurring in f' is X, which does not divide f. We know from Corollary 7.37 that L is perfect, and so f must be squarefree by Theorem 7.36. Since L is also algebraically closed, this means that f factors into n pairwise non-associated linear factors in $L[X]$ and thus has n different zeroes in L, contradicting the fact that L has fewer than n elements. □

Lemma 7.50 Let R be a domain with infinitely many elements, and let $0 \neq f \in R[X_1, \ldots, X_n]$. Then there are infinitely many different $\boldsymbol{z} \in R^n$ with $f(\boldsymbol{z}) \neq 0$.

Proof We proceed by induction on n. If $n = 1$, then by Corollary 2.97, f has at most finitely many zeroes in Q_R and hence a fortiori in R itself. This leaves infinitely many $z \in R$ with $f(z) \neq 0$. Now let $n > 1$. We view f as an element of $R[X_1, \ldots, X_{n-1}][X_n]$ and let $g \in R[X_1, \ldots, X_{n-1}]$ be any non-zero coefficient of f. By induction hypothesis, there exist z_1, …, $z_{n-1} \in R$ with $g(z_1, \ldots, z_{n-1}) \neq 0$. Then $f(z_1, \ldots, z_{n-1}, X_n)$ is a non-zero polynomial in $R[X_n]$ and thus has only finitely many zeroes in R, and so we can find infinitely many $z_n \in R$ with $f(z_1, \ldots, z_n) \neq 0$. □

Lemma 7.51 Let L be an algebraically closed extension field of K, and let I be a zero-dimensional prime ideal of $K[X_1, \ldots, X_n]$. Then the following hold:

(i) If $G = \{g_1, \ldots, g_n\}$ is the prime basis of I of Proposition 7.42, and $(z_1, \ldots, z_i) \in L^i$ is a zero of $\{g_1, \ldots, g_i\}$, where $1 \leq i < n$, then there exist z_{i+1}, …, $z_n \in L$ such that $(z_1, \ldots, z_n)$ is a zero of I.

(ii) I has a zero in L^n.

Proof (i) We proceed by induction on n. If $n = 1$, then the claim follows from the fact that the polynomial g_1, which cannot be a non-zero constant, has a zero in L. Next, suppose $n > 1$, and let $(z_1, \ldots, z_i) \in L^i$ be a zero of $\{g_1, \ldots, g_i\}$, where $1 \leq i < n$. By induction hypothesis, there exist z_{i+1}, …, $z_{n-1} \in L$ such that $(z_1, \ldots, z_{n-1})$ is a zero of

$$I \cap K[X_1, \ldots, X_{n-1}] = \operatorname{Id}(\{g_1, \ldots, g_{n-1}\}).$$

The polynomial $g_n(z_1, \dots, z_{n-1}, X_n) \in L[X_n]$ is non-constant because the head term of g_n was a power of X_n, and so it has a zero z_n in the algebraically closed field L. It is clear that $\boldsymbol{z} = (z_1, \dots, z_n)$ is a zero of I.

(ii) Let G be as in (i). The non-constant polynomial $g_1 \in K[X_1]$ has a zero $z_1 \in L$, which can be extended to a zero $\boldsymbol{z} \in L^n$ by (i). □

A more hands-on way of looking at the proof of (ii) above is as follows: to obtain a zero of I, start with a zero $z_1 \in L$ of the non-constant polynomial $g_1 \in K[X_1]$, plug it into g_2 to obtain a non-constant polynomial $g_2(z_1, X_2)$, and continue the process in the obvious way.

Lemma 7.52 Let L be an algebraically closed extension field of K, and let I be a prime ideal of $K[X_1, \dots, X_n]$ such that $\{X_1, \dots, X_d\}$ is maximally independent modulo I. Then there exists a non-zero polynomial $p \in K[X_1, \dots, X_d]$ such that every $(z_1, \dots, z_d) \in L^d$ with the possible exception of the zeroes of p extends to a zero $(z_1, \dots, z_n) \in L^n$ of I.

Proof Consider the polynomials f, g_{d+1}, …, g_n of Proposition 7.48. For $d+1 \leq i \leq n$, let $h_i \in K[X_1, \dots, X_d]$ be the head coefficient of g_i when viewed as a polynomial in X_i, and set

$$p = f \cdot \prod_{i=d+1}^{n} h_i.$$

Let z_1, …, $z_d \in L$ with $p(z_1, \dots, z_d) \neq 0$. We consider the polynomials

$$g_i(z_1, \dots, z_d, X_{d+1}, \dots, X_i) \in L[X_{d+1}, \dots, X_i] \qquad (d+1 \leq i \leq n).$$

The head coefficient of $g_i(z_1, \dots, z_d, X_{d+1}, \dots, X_\imath)$ when viewed as a polynomial in X_i is $h_i(z_1, \dots, z_d) \neq 0$. Arguing as in the remark preceding the lemma, we can thus inductively find z_{d+1}, …, $z_n \in L$ such that $g_i(z_1, \dots, z_n) = 0$ for $d+1 \leq i \leq n$. We claim that $\boldsymbol{z} = (z_1, \dots, z_n)$ is a zero of I. Let $g \in I$. Then $fg \in \mathrm{Id}(g_{d+1}, \dots, g_n)$ by Proposition 7.48, and thus

$$0 = (fg)(\boldsymbol{z}) = f(z_1, \dots, z_d)g(\boldsymbol{z}).$$

But $f(z_1, \dots, z_d) \neq 0$ by the choice of z_1, …, z_d, and so $g(\boldsymbol{z}) = 0$. □

Proposition 7.53 *Let L be an algebraically closed extension field of K. Then every proper ideal of $K[X_1, \dots, X_n]$ has a zero in L^n.*

Proof Let I be a proper ideal of $K[X_1, \dots, X_n]$ with $\dim(I) = d$. It clearly suffices to find a zero in L^n of some ideal J with $I \subseteq J$.
Case 1: $d = 0$.
We extend I to a maximal ideal J of $K[X_1, \dots, X_n]$. Then J is a zero-dimensional prime ideal, and thus it has a zero by Lemma 7.51.

Case 2: $d > 0$.
Renumbering variables if necessary, we may assume that $\{X_1, \ldots, X_d\}$ is maximally independent modulo I. Set

$$M = K[X_1, \ldots, X_d] \setminus \{0\}.$$

Then $I \cap M = \emptyset$, and we let J be a prime ideal with $I \subseteq J$ and $J \cap M = \emptyset$ (Proposition 4.11). The set $\{X_1, \ldots, X_d\}$ is still maximally independent modulo J. Now let

$$0 \neq p \in K[X_1, \ldots, X_d]$$

be as in Lemma 7.52. Lemma 7.50 together with Lemma 7.49 provides z_1, $\ldots$, $z_d \in L$ with $p(z_1, \ldots, z_d) \neq 0$, and this extends to a zero of J by Lemma 7.52. □

Lemma 7.52 together with Corollary 2.97 provides a clue to the geometric meaning of the dimension of an ideal (and to the reason for the choice of the terms "dimension" and "independent set"). Suppose I is a two-dimensional prime ideal of $\mathbb{Q}[X, Y, Z]$, and assume $\{X, Y\}$ is maximally independent modulo I. Then every point $(z_1, z_2) \in \mathbb{C}^2$ with the possible exception of the zeroes of $0 \neq p \in \mathbb{Q}[X, Y]$ can be extended to a zero $(z_1, z_2, z_3) \in \mathbb{C}^3$ in at least one but at most finitely many different ways. But the set of zeroes of p forms what one would call a *curve* in $\mathbb{C}^2$ (a fixed choice for X leaves only finitely many possibilities for Y, or else there are only finitely many possibilities for X), and we see that the set of zeroes in $\mathbb{C}^3$ of I is—intuitively speaking—a *surface*, i.e., a two-dimensional configuration.

Proof of the Hilbert Nullstellensatz The implication "(ii)$\Longrightarrow$(i)" is trivial. Now assume that (i) holds, and set $I = \mathrm{Id}(g_1, \ldots, g_m)$. Let Y be a new indeterminate. The ideal

$$J = \mathrm{Id}(I, 1 - Yf)$$

of the ring $K[X_1, \ldots, X_n][Y]$ does not have a zero in L^{n+1} because the vanishing of g_1, $\ldots$, g_m at z implies that $(1 - Yf)(z) = 1$. We see that J is not proper and thus $1 \in J$. Proposition 6.37 now tells us that

$$1 \in I : f^\infty$$

and thus $f^s \in I$ for some $0 < s \in \mathbb{N}$. □

Another important consequence of Proposition 7.53 is the following addition to Proposition 7.42 that we announced earlier. Recall that we denote by $\overline{K}$ the algebraic closure of K.

Proposition 7.54 *The zero-dimensional prime ideals of* $K[X_1, \ldots, X_n]$ *are precisely the maximal ones.*

Proof In view of Proposition 7.42 (i), it remains to show that every maximal ideal is zero-dimensional. Let I be a proper ideal with $\dim(I) > 0$. Then $I \cap K[X_i] = \{0\}$ for some $1 \le i \le n$. Since I is proper, it has a zero

$$z = (z_1, \ldots, z_n) \in \overline{K}^n.$$

Since $\overline{K}$ is algebraic over K, there is $0 \ne f \in K[X_i]$ with $f(z_i) = 0$. Then $J = \mathrm{Id}(I, f)$ is proper because it has the zero z. Moreover, J properly extends I and thus I was not maximal. □

Exercise 7.55 Let I be a zero-dimensional ideal of $K[X_1, \ldots, X_n]$, and let $1 \le i \le n$. Suppose $z \in \overline{K}^i$ is a zero of the elimination ideal $I \cap K[X_1, \ldots, X_i]$. Show that z extends to a zero of I in $\overline{K}^n$. (Hint: Reduce the problem to the case where I is radical, and use the fact that a radical ideal equals the intersection of all prime ideals containing it.) Demonstrate that the claim is false in higher dimensions.

We close this section with an easy lemma which shows that the Hilbert Nullstellensatz can be strengthened considerably if the polynomials g_1, $\ldots$, g_m generate a maximal ideal.

Lemma 7.56 Let L be an extension field of K and I a maximal ideal of $K[X_1, \ldots, X_n]$.

(i) If $f \in K[X_1, \ldots, X_n]$ such that I and f have a common zero in L^n, then $f \in I$.

(ii) If J is another maximal ideal of $K[X_1, \ldots, X_n]$ such that I and J have a common zero in L^n, then $I = J$.

Proof To prove (i), assume for a contradiction that $f \notin I$. Then there exists $g \in K[X_1, \ldots, X_n]$ and $h \in I$ such that $1 = h + gf$. Evaluating at the common zero of f and I, we arrive at the contradiction $1 = 0$. Part (ii) is an easy consequence of (i). □

7.5 Height and Depth of Prime Ideals

The material of this section is not directly connected to the Hilbert Nullstellensatz, but the flavor of the proofs is similar to the ones in the preceding section. Of the material in this section, only Lemma 7.57 will ever be used again in this book.

Throughout this section, K will be a field. We will repeatedly make use of Lemmas 1.122, 1.123, and 7.47. If I_0, $\ldots$, I_m $(m \ge 0)$ are prime ideals of $K[X_1, \ldots, X_n]$ with $I_i \subseteq I_{i+1}$ and $I_i \ne I_{i+1}$ for $0 \le i \le m-1$, then we call $(I_0, \ldots, I_m)$ a **chain of prime ideals** of **length** m. The **depth** $\mathrm{d}(I)$ of a prime ideal of $K[X_1, \ldots, X_n]$ is the maximal length of a chain $(I_0, \ldots, I_m)$ of prime ideals with $I_0 = I$. The **height** $\mathrm{h}(I)$ of a prime ideal of $K[X_1,\ldots,X_n]$ is the maximal length of a chain $(I_0,\ldots,I_m)$ of prime ideals with $I_m = I$. Our aim is now to show that $\mathrm{d}(I) = \dim(I)$ and $\mathrm{h}(I) = n - \dim(I)$ for every prime ideal I of $K[X_1,\ldots,X_n]$.

Lemma 7.57 Let I and J be prime ideals of $K[X_1,\ldots,X_n]$ with $I \subseteq J$ and $I \neq J$. Then $\dim(J) < \dim(J)$.

Proof We have already observed in Lemma 6.49 that $\dim(J) \leq \dim(I)$. Assume for a contradiction that $\dim(I) = \dim(J) = d$. W.l.o.g., we may assume that $\{X_1,\ldots,X_d\}$ is maximally independent modulo J. Then this set is a fortiori independent modulo I, and since $\dim(I) = d$, it must even be maximally independent modulo I. So if we form I^{e} and J^{e} w.r.t. $M = K[X_1,\ldots,X_d] \setminus \{0\}$, then I^{e} and J^{e} are both zero-dimensional prime ideals and hence maximal. This together with $I^{\mathrm{e}} \subseteq J^{\mathrm{e}}$ implies $I^{\mathrm{e}} = J^{\mathrm{e}}$ and thus $I = I^{\mathrm{ec}} = J^{\mathrm{ec}} = J$, a contradiction. □

As an immediate consequence, we obtain the following proposition.

Proposition 7.58 *Let $(I_0,\ldots,I_m)$ be a chain of prime ideals in the ring $K[X_1,\ldots,X_n]$. Then $m \leq \dim(I_0) - \dim(I_m)$.* □

Corollary 7.59 *If I is a prime ideal of $K[X_1,\ldots,X_n]$, then $\mathrm{d}(I) \leq \dim(I)$ and $\mathrm{h}(I) \leq n - \dim(I)$.*

Proof Let $(I_0,\ldots,I_m)$ be a chain of prime ideals. If $I = I_0$, then $m \leq \dim(I) - \dim(I_m) \leq \dim(I)$ and thus the maximal possible length of such a chain is $\dim(I)$. If $I = I_m$, then $m \leq \dim(I_0) - \dim(I) \leq n - \dim(I)$ and thus the maximal length of such a chain is $n - \dim(I)$. □

The next two propositions will provide the reverse inequalities of the corollary above.

Proposition 7.60 *Let I be a prime ideal of $K[X_1,\ldots,X_n]$ and $0 \leq d \leq n$ such that $\{X_1,\ldots,X_d\}$ is maximally independent modulo I. Then there exists a chain $(I_0,\ldots,I_d)$ of prime ideals with $I_0 = I$.*

Proof We proceed by induction on d. If $d = 0$, then (I_0) is the desired chain. Now let $d > 0$, and let $M = K[X_1,\ldots,X_{d-1}] \setminus \{0\}$. Consider the extension I^{e} of I to

$$K[X_1,\ldots,X_n]_M = K(X_1,\ldots,X_{d-1})[X_d,\ldots,X_n].$$

Then

$$I^{\mathrm{e}} \cap K(X_1,\ldots,X_{d-1})[X_d] = \{0\}$$

since otherwise we would obtain a non-zero element of $I \cap K[X_1,\ldots,X_d]$ by clearing denominators of coefficients in $K(X_1,\ldots,X_{d-1})$. We see that $\dim(I^{\mathrm{e}}) > 0$, and thus I^{e} is not maximal by Proposition 7.54. Let J be a maximal ideal of $K(X_1,\ldots,X_{d-1})[X_d,\ldots,X_n]$ with $I^{\mathrm{e}} \subseteq J$. Then J^{c} is a prime ideal of $K[X_1,\ldots,X_n]$ with $I = I^{\mathrm{ec}} \subseteq J^{\mathrm{c}}$. Being maximal, $J = J^{\mathrm{ce}}$

has dimension zero, and thus $\{X_1, \ldots, X_{d-1}\}$ is maximally independent modulo J^c. In particular, $I \neq J^c$. The induction hypothesis provides a chain $(I_0, \ldots, I_{d-1})$ of prime ideals with $I_0 = J^c$, and hence $(I, I_0, \ldots, I_{d-1})$ is a chain with the desired properties. □

If $\dim(I) = d$, then we may assume w.l.o.g. that $\{X_1, \ldots, X_d\}$ is maximally independent modulo I, and so there exists a chain of prime ideals as described in the proposition. Together with Corollary 7.59 we thus obtain the following corollary.

Corollary 7.61 $\mathrm{d}(I) = \dim(I)$ *for every prime ideal of* $K[X_1, \ldots, X_n]$. □

To show that the height of a prime ideal I equals $n - \dim(I)$, we need two technical lemmas.

Lemma 7.62 Let I be a zero-dimensional prime ideal of $K[X_1, \ldots, X_n]$ and $G = \{g_1, \ldots, g_n\}$ the prime basis of I of Proposition 7.42. Let $1 < i \leq n$, and set $G_i = \{g_i, \cdots g_n\}$, $I_i = \mathrm{Id}(G_i)$, and

$$M_i = (K[X_1, \ldots, X_{i-1}]) \cdot (K[X_1, \ldots, X_n] \setminus I).$$

Then $M_i \cap I_i = \{0\}$.

Proof Let $\leq$ be the inverse lexicographical term order, where $X_n \gg \cdots \gg X_1$. The head term of g_j is a power of X_j for $i \leq j \leq n$, and so G_i is a Gröbner basis of I_i w.r.t. $\leq$ by Lemma 5.66. Now assume for a contradiction that $0 \neq f \in M_i \cap I_i$. Then $f = gh$ with $g \in K[X_1, \ldots, X_{i-1}]$ and $h \notin I$. Let h_0 be the unique normal form of h modulo G_i. Then $h_0 \neq 0$ since otherwise $h \in I_i \subseteq I$. Furthermore, $h - h_0 \in I_i$ and thus $gh - gh_0 = g(h - h_0) \in I_i$. It follows that $gh_0 \in I_i$, and so gh_0 must be reducible modulo G_i. But h_0 was in normal form modulo G_i, i.e.,

$$\deg_{X_j}(t) < \deg_{X_j}(g_j)$$

for all $t \in T(h_0)$ and $i \leq j \leq n$. Since $g \in K[X_1, \ldots, X_{i-1}]$, the latter inequality remains true for all $t \in T(gh_0)$, a contradiction. □

Lemma 7.63 Let I be a prime ideal of $K[X_1, \ldots, X_n]$ and $0 \leq d < n$ such that $\{X_1, \ldots, X_d\}$ is maximally independent modulo I. Then there exists a prime ideal J of $K[X_1, \ldots, X_n]$ such that $J \subseteq I$ and $\{X_1, \ldots, X_{d+1}\}$ is maximally independent modulo J.

Proof The univariate case $n = 1$ is trivial: then $d = 0$, $I \neq \{0\}$, and we may take $J = \{0\}$. So let $n > 1$.
Case 1: $d = 0$.
Then $\emptyset$ is maximally independent modulo I and so $\dim(I) = 0$. Let G_2, I_2, and M_2 be as defined in Lemma 7.62. $K[X_1]$ is trivially multiplicative, and so is $K[X_1, \ldots, X_n] \setminus I$ since I is prime. It follows that their product

M_2 is multiplicative too, and by Lemma 7.62, $I_2 \cap (M_2 \setminus \{0\}) = \emptyset$. We can thus extend I_2 to a prime ideal J with $J \cap M_2 = \{0\}$. The inclusion

$$J \cap (K[X_1, \dots, X_n] \setminus I) \subseteq J \cap (M_2 \setminus \{0\}) = \emptyset$$

implies that $J \subseteq I$, and from the inclusion

$$J \cap K[X_1] \subseteq J \cap M_2 = \{0\}$$

we see that $\{X_1\}$ is independent modulo J. It remains to show that $\{X_1\}$ is maximally independent modulo J. This follows from the fact that $G_2 \subseteq J$ and hence the extension J^e of J to $K[X_1, \dots, X_n]_M$ with $M = K[X_1] \setminus \{0\}$ is a zero-dimensional ideal of

$$K[X_1, \dots, X_n]_M = K(X_1)[X_2, \dots, X_n]$$

by Corollary 6.56.
Case 2: $d > 0$.
We consider I^e w.r.t. $K[X_1, \dots, X_d] \setminus \{0\}$. Then I^e is a zero-dimensional prime ideal of the ring $K(X_1, \dots, X_d)[X_{d+1}, \dots, X_n]$, and Case 1 above provides us with a prime ideal J of this ring such that $J \subseteq I^e$ and $\{X_{d+1}\}$ is maximally independent modulo J. Then J^c is a prime ideal of $K[X_1, \dots, X_n]$ with $J^c \subseteq I^{ec} = I$. We have

$$J^c \cap K[X_1, \dots, X_{d+1}] = \{0\}$$

since otherwise

$$J \cap K(X_1, \dots, X_d)[X_{d+1}] \neq \{0\},$$

and thus $\{X_1, \dots, X_{d+1}\}$ is independent modulo J^c. It remains to show that this latter set is maximally independent modulo J^c. Let $d+1 < i \leq n$. Then

$$J \cap K(X_1, \dots, X_d)[X_{d+1}, X_i] \neq \{0\},$$

and since we can clear denominators of coefficients in $K(X_1, \dots, X_d)$, we see that

$$J^c \cap K[X_1, \dots, X_{d+1}, X_i] \neq \{0\}. \quad \square$$

The following proposition can now easily be proved by means of a repeated application of the above lemma and an argument like the one we used to prove Corollary 7.61.

Proposition 7.64 *Let I be a prime ideal of $K[X_1, \dots, X_n]$ and $0 \leq d \leq n$ such that $\{X_1, \dots, X_d\}$ is maximally independent modulo I. Then there exists a chain $(I_0, \dots, I_{n-d})$ of prime ideals with $I_{n-d} = I$. In particular,* $\mathrm{h}(I) = n - \dim(I)$. $\square$

We can now give a second, independent proof of Proposition 7.26.

Corollary 7.65 *Let I be a prime ideal of $K[X_1, \ldots, X_n]$, and suppose $U \subseteq \{X_1, \ldots, X_n\}$ is maximally independent modulo I. Then $|U| = \dim(I)$.*

Proof Renumbering variables if necessary, we may assume w.l.o.g. that $U = \{X_1, \ldots, X_d\}$ for some $0 \leq d \leq n$. Then $d \leq \dim(I)$ by the definition of the dimension. By Proposition 7.64, there exists a chain $(I_0, \ldots, I_{n-d})$ of prime ideals with $I = I_{n-d}$. By Proposition 7.58, we must have $n - d \leq \dim(I_0) - \dim(I) \leq n - \dim(I)$ and thus $d \geq \dim(I)$. □

Exercise 7.66 Use Corollary 7.65 to prove Theorem 7.23 without the use of the abstract theory of independent sets.

7.6 Implicitization of Rational Parametrizations

In this section, we demonstrate how the Hilbert Nullstellensatz can be used together with Gröbner basis techniques to solve a problem that is geometric in nature. This section forms an aside within this book; the material presented here will not be used again.

Throughout, K will be a field, and we will use the notation

$$K[\underline{X}] = K[X_1, \ldots, X_n].$$

Recall that for an ideal I of $K[\underline{X}]$ and an extension field L of K, we have called the set of zeroes of I in L^n the *variety* of I in L^n; in addition, we now introduce the notation $V_L(I)$ for this set. Moreover, we call a subset V of L^n a **K-variety** if $V = V_L(I)$ for some ideal I of $K[\underline{X}]$. We will abbreviate n-tuples $(z_1, \ldots, z_n) \in L^n$ to $\boldsymbol{z}$. Our first lemma is hardly more than another reformulation of the Hilbert Nullstellensatz.

Lemma 7.67 Let I_1 and I_2 be ideals of $K[\underline{X}]$ and L an algebraically closed extension field of K. Then the following are equivalent:

(i) $I_1 \subseteq \operatorname{rad}(I_2)$.

(ii) $V_L(I_2) \subseteq V_L(I_1)$.

Proof (i)$\Longrightarrow$(ii): If $\boldsymbol{z} \in V_L(I_2)$, then every $f \in I_2$ vanishes at $\boldsymbol{z}$. This trivially implies that every element of $\operatorname{rad}(I_2)$ vanishes at $\boldsymbol{z}$, and using (i), we see that $\boldsymbol{z} \in V_L(I_1)$.

(ii)$\Longrightarrow$(i): If $f \in I_1$, then by (ii), it vanishes on the variety of I_2 in L^n and is thus an element of $\operatorname{rad}(I_2)$ by the Hilbert Nullstellensatz. □

The next lemma relates elimination ideals to projections of varieties: it says that the variety of an elimination ideal of an ideal I is the smallest variety containing the projection of the variety of I on the corresponding components.

Lemma 7.68 Let I be an ideal of $K[\underline{X}]$ and L an algebraically closed extension field of K. Let $1 \leq d \leq n$, and set

$$W = \{\, (z_1, \ldots, z_d) \in L^d \mid \text{there exist } z_{d+1}, \ldots, z_n \in L \text{ with } \boldsymbol{z} \in V_L(I) \,\}.$$

Then $V_L(I \cap K[X_1, \ldots, X_d])$ is the smallest K-variety in L^d extending W.

Proof To prove that $W \subseteq V_L(I \cap K[X_1, \ldots, X_d])$, let $(z_1, \ldots, z_d) \in W$. Then there exist $z_{d+1}, \ldots, z_n \in L$ with $\boldsymbol{z} \in V_L(I)$. In particular,

$$f(z_1, \ldots, z_d) = f(\boldsymbol{z}) = 0 \quad \text{for} \quad f \in I \cap K[X_1, \ldots, X_d].$$

It remains to show that $V_L(I \cap K[X_1, \ldots, X_d]) \subseteq V_L(J)$ for all ideals J of $K[X_1, \ldots, X_d]$ with $W \subseteq V_L(J)$. Let J be such an ideal. According to the previous lemma, it suffices to prove that

$$J \subseteq \operatorname{rad}(I \cap K[X_1, \ldots, X_d]).$$

But if $f \in J$, then we may consider f as an element of $K[\underline{X}]$. The assumption $W \subseteq V_L(J)$ implies that f vanishes on the variety of I in L^n, and consequently, $f^s \in I \cap K[X_1, \ldots, X_d]$ for some $s \in \mathbb{N}$ by the Hilbert Nullstellensatz. □

We know from elementary mathematics that subsets of $\mathbb{R}^2$ or $\mathbb{R}^3$ can often be described alternatively as varieties or by parametrizations: the unit circle, for example, has the representations

$$\{\, (x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1 \,\} \quad \text{and} \quad \{\, (\cos t, \sin t) \mid t \in \mathbb{R} \,\}.$$

The following theorem shows how one can, in a certain restricted sense, transform a parametrization by rational functions into a representation as a variety. Here, $T_1, \ldots, T_m$ will be indeterminates, and the notations $K[\underline{T}]$ and $K[\underline{T}, \underline{X}]$ will be used in the obvious sense.

Theorem 7.69 *Let* $f_1, \ldots, f_n, g_1, \ldots, g_n \in K[\underline{T}]$ *with*

$$g = g_1 \cdot \cdots \cdot g_n \neq 0.$$

Let L be an extension field of K with infinitely many elements, and set

$$\begin{array}{cccc} \varphi: & L^m \setminus V_L(g) & \longrightarrow & L^n \\ & \boldsymbol{a} & \longmapsto & \big(f_1(\boldsymbol{a})/g_1(\boldsymbol{a}), \ldots, f_n(\boldsymbol{a})/g_n(\boldsymbol{a})\big). \end{array}$$

Furthermore, let Y be a new indeterminate and set

$$I = \operatorname{Id}(g_1X_1 - f_1, \ldots, g_nX_n - f_n, gY - 1) \subseteq K[\underline{T}, \underline{X}, Y].$$

Then $V_L(I \cap K[\underline{X}])$ *is the smallest* K*-variety in* L^n *containing the image of* φ.

Proof It is not hard to see that the image of φ equals the "projection of $V_L(I)$ on the X-components," i.e., the set W of all those $z \in L^n$ such that the ideal

$$\mathrm{Id}(g_1 z_1 - f_1, \ldots, g_n z_n - f_n, gY - 1) \subseteq L[\underline{T}, Y]$$

has a zero in L^{m+1}. Now let J be an ideal of $K[\underline{X}]$ with $W \subseteq V_L(J)$. We must show that $V_L(I \cap K[\underline{X}]) \subseteq V_L(J)$. To this end, we lift the entire situation of the theorem to the algebraic closure $\overline{L}$ of L, i.e., we set

$$\begin{array}{rccc} \overline{\varphi}: & \overline{L}^m \setminus V_{\overline{L}}(g) & \longrightarrow & \overline{L}^n \\ & a & \longmapsto & \big(f_1(a)/g_1(a), \ldots, f_n(a)/g_n(a)\big) \end{array}$$

and let $\overline{W}$ be the image of $\overline{\varphi}$ in $\overline{L}$.

Claim: $\overline{W} \subseteq V_{\overline{L}}(J)$.

Proof: Let $h \in J$. It is easy to see that there exist $\nu \in \mathbb{N}$ and a polynomial $q \in K[\underline{T}]$ such that

$$gq = g^{\nu} \cdot h(f_1/g_1, \ldots, f_n/g_n). \qquad (*)$$

From our assumption $W \subseteq V_L(J)$ and the fact that gq can be written in the form $(*)$, one easily deduces that gq vanishes on all of L^m and must thus, in view of Lemma 7.50, be the zero polynomial. Going back to the representation $(*)$, it is now easy to see that h vanishes on the image $\overline{W}$ of the map $\overline{\varphi}$.

Recalling that $\overline{W}$ is a projection of the variety of I in $\overline{L}^{m+n+1}$, we find ourselves in a position to apply the previous lemma and conclude that

$$V_{\overline{L}}(I \cap K[\underline{X}]) \subseteq V_{\overline{L}}(J).$$

Intersecting with L^n, we see that indeed

$$V_L(I \cap K[\underline{X}]) \subseteq V_L(J). \quad \square$$

In view of Corollary 6.17, it is clear that a basis of the elimination ideal of the theorem above can be computed from the f_i and g_i whenever K is a computable field.

Exercise 7.70 Show that the unit circle is the smallest $\mathbb{Q}$-variety in $\mathbb{R}^2$ that contains the set

$$\left\{ \left(\frac{1-t^2}{t^2+1}, \frac{2t}{t^2+1} \right) \,\middle|\, t \in \mathbb{R} \right\}.$$

Show that this implicitization adds the point $(-1, 0)$ to the given set. (Hint: This is about the upper limit of what you want to do by hand. A computer algebra system comes in handy.)

7.7 Invertibility of Polynomial Maps

We conclude this chapter with the solution of a decision problem that is closely related to a famous open problem, namely, the *Jacobian conjecture.* Here, Gröbner basis techniques will be combined with the result of Theorem 7.23 which led to the definition of the transcendence degree. This section is another aside within this book; the material presented here will not be used in the sequel. We will once again be using the notation $K[\underline{X}] = K[X_1, \ldots, X_n]$.

According to Lemma 2.17 (i), an n-tuple $\boldsymbol{f} = (f_1, \ldots, f_n) \in (K[\underline{X}])^n$ gives rise to a map

$$\begin{array}{rccc} \varphi_{\boldsymbol{f}}: & K^n & \longrightarrow & K^n \\ & (a_1, \ldots, a_n) & \longmapsto & \big(f_1(\boldsymbol{a}), \ldots, f_n(\boldsymbol{a})\big), \end{array}$$

where $\boldsymbol{a}$ stands for $(a_1, \ldots, a_n)$. Here, $\varphi_{\boldsymbol{f}}$ is called **invertible** if there exist $g_1, \ldots, g_n \in K[\underline{X}]$ such that

$$g_i(f_1, \ldots, f_n) = X_i \quad \text{for} \quad 1 \leq i \leq n,$$

so that $\varphi_{\boldsymbol{g}} \circ \varphi_{\boldsymbol{f}} = \mathrm{id}_{K^n}$. The Jacobian conjecture (which we won't go into here) states that over fields K of characteristic zero, $\varphi_{\boldsymbol{f}}$ is invertible iff the determinant of the Jacobi matrix

$$\left(\frac{\partial f_i}{\partial X_j}\right)_{\substack{i=1,\ldots,n \\ j=1,\ldots,n}}$$

is in $K \setminus \{0\}$. Another way of looking at the same problem is as follows. It is easy to see that every endomorphism ψ of $K[\underline{X}]$ that satisfies $\psi \restriction K = \mathrm{id}_K$ is of the form

$$h \longmapsto h(f_1, \ldots, f_n),$$

where $f_i = \psi(X_i)$ for $1 \leq i \leq n$. The existence of $g_1, \ldots, g_n$ with

$$g_i(f_1, \ldots, f_n) = X_i \quad \text{for} \quad 1 \leq i \leq n$$

is then equivalent to the surjectivity of the endomorphism in question.

The next proposition shows how Gröbner bases provide a means to decide, for given $\boldsymbol{f}$, whether or not $\varphi_{\boldsymbol{f}}$ is invertible.

Proposition 7.71 *Let $\boldsymbol{f} = (f_1, \ldots, f_n) \in (K[\underline{X}])^n$, and let $Y_1, \ldots, Y_n$ be new indeterminates. Set $I = \mathrm{Id}(F)$, where*

$$F = \{Y_1 - f_1, \ldots, Y_n - f_n\}.$$

Furthermore, let $\leq$ be a term order on $T(\underline{X}, \underline{Y})$ that satisfies $\underline{Y} \ll \underline{X}$. Then the following are equivalent:

(i) φ_f is invertible.

(ii) The reduced Gröbner basis G of I w.r.t. $\leq$ is of the form

$$G = \{X_1 - g_1, \ldots, X_n - g_n\} \quad \text{with} \quad g_1, \ldots, g_n \in K[\underline{Y}].$$

Moreover, if (ii) holds, then $g_i(f_1, \ldots, f_n) = X_i$ for $1 \leq i \leq n$.

Proof For the direction "(ii)$\Longrightarrow$(i)," suppose that the reduced Gröbner basis G of I w.r.t. $\leq$ is of the form

$$G = \{X_1 - g_1, \ldots, X_n - g_n\} \quad \text{with} \quad g_1, \ldots, g_n \in K[\underline{Y}].$$

Then there exist $q_{ij} \in K[\underline{X}, \underline{Y}]$ with

$$X_i - g_i = \sum_{j=1}^{n} q_{ij}(Y_j - f_j) \qquad (1 \leq i, j \leq n),$$

and substitution of f_j for Y_j yields $X_i - g_i(f_1, \ldots, f_n) = 0$ for $1 \leq i \leq n$.

Conversely, suppose that φ_f is invertible, and let $g_1, \ldots, g_n \in K[\underline{Y}]$ such that

$$g_i(f_1, \ldots, f_n) = X_i \quad \text{for} \quad 1 \leq i \leq n. \tag{$*$}$$

Lemma 6.43 (i) tells us that

$$g_i(f_1, \ldots, f_n) - g_i = X_i - g_i \in I \quad \text{for} \quad 1 \leq i \leq n.$$

The set $G = \{X_1 - g_1, \ldots, X_n - g_n\}$ is thus a subset of I, and since our term order $\leq$ satisfies $\underline{Y} \ll \underline{X}$, it is clearly reduced. In order to show that G is in fact a Gröbner basis of I w.r.t. $\leq$, it suffices by Proposition 5.38 to prove that every $0 \neq h \in I$ is reducible modulo G. This is obviously true whenever $\deg_{X_i}(h) > 0$ for some $1 \leq i \leq n$. The proof will thus be finished once we have proved the following claim.

Claim: $I \cap K[\underline{Y}] = \{0\}$.

Proof: We first note that from $(*)$, it follows that ring adjunction of f_1, $\ldots$, f_n to K within $K[\underline{X}] = K[X_1, \ldots, X_n]$ yields

$$K[f_1, \ldots, f_n] = K[X_1, \ldots, X_n].$$

Taking fields of quotients, we may apply Theorem 7.23 (ii) with

$$K' = K(\underline{X}), \quad B = \{X_1, \ldots, X_n\}, \quad \text{and} \quad A = \{f_1, \ldots, f_n\}$$

to conclude that $\{f_1, \ldots, f_n\}$ is algebraically independent over K. Whenever $h \in K[\underline{Y}]$, then $h(f_1, \ldots, f_n) - h \in I$ by Lemma 6.43 (i), and so $h \in I \cap K[\underline{Y}]$ implies

$$h(f_1, \ldots, f_n) \in I \cap K[\underline{X}].$$

Now Lemma 6.43 (ii) says that $h(f_1, \ldots, f_n) = 0$, and so $h = 0$ because of the algebraic independence of $\{f_1, \ldots, f_n\}$ over K. □

It is clear now that over a computable field K, one can decide the invertibility of $\varphi_{\boldsymbol{f}}$ for any given $\boldsymbol{f} \in (K[\underline{X}])^n$ and compute an inverse if one exists.

Exercise 7.72 Let $f_1, \ldots, f_n, g_1, \ldots, g_n \in K[\underline{X}]$ such that $g_i(f_1, \ldots, f_n) = X_i$ for $1 \leq i \leq n$. Show that then $f_i(g_1, \ldots, g_n) = X_i$ as well. Conclude that, with the notation for polynomial maps introduced at the beginning of this section, $\varphi_{\boldsymbol{g}} \circ \varphi_{\boldsymbol{f}} = \mathrm{id}_{K^n}$ implies $\varphi_{\boldsymbol{f}} \circ \varphi_{\boldsymbol{g}} = \mathrm{id}_{K^n}$. Moreover, if an endomorphism ψ of $K[\underline{X}]$ satisfying $\psi \restriction K = \mathrm{id}_K$ is surjective, then it is injective, whereas the converse fails in general.

Notes

The first comprehensive treatment of field extensions is Steinitz (1910). Kronecker (1882) introduced the idea of the purely symbolic construction of a simple algebraic field extension in the absence of an existing extension field. A crucial point in the theory of field extensions is the fact that every element b that is algebraic over a simple algebraic extension field $K(a)$ of a field K is algebraic over K as well. Following the method of Steinitz, most algebra textbooks prove this non-constructively. One first shows that a simple algebraic extension field of a field L is finite-dimensional as a vector space over L, and that, conversely, an extension field of L whose vector space dimension over L is finite cannot contain transcendental elements over L. These results are then combined with the fact that the property of a field extension to be finite-dimensional in this sense is transitive. The approach via Gröbner bases has the advantage of providing a means to compute $\min_K^b$ from $\min_K^a$ and $\min_{K(a)}^b$. Another way to achieve this is by using resultants (see Loos, 1982).

The first proof of the fundamental theorem of algebra that stands up to modern mathematical standards was given in 1799 by C.F. Gauss in his doctoral dissertation. The existence of the algebraic closure of a field is proved in Steinitz (1910). In order to fully appreciate the difficulties of the proof, one must first understand why the following set-theoretically naive proof is not legitimate: consider the collection of all algebraic extensions of the given field and choose a maximal one. The catch is of course the fact that the collection of all algebraic extensions of a given field is not a set in the sense of Zermelo and Fraenkel and therefore does not allow an application of Zorn's lemma. The proof that we have given here is that of Lang (1971), except that we have added a pinch of Gröbner basis flavor. Lang's proof is an elegant one because it neatly separates the set-theoretical part from the algebraic construction (see the discussion at the end of Section 4.1).

The notion of separability is again due to Steinitz (1910). Steinitz himself used the term "of the first kind"; the term "separable" was suggested by van der Waerden (1931) and has since come to be universally accepted.

The Hilbert Nullstellensatz is proved in Hilbert (1893), Part II, §3. It is the pivotal point at the juncture of algebra and algebraic geometry, and there is hardly another theorem with stronger repercussions in commutative algebra. The relevance of the theorem in ideal theory will become obvious in Chapter 8 of this book; another important concept that it is related to, namely, *quantifier elimination*, will be briefly touched upon in Section "Gröbner Bases and Automatic Theorem Proving" on p. 518 in the appendix. Our approach to the proof of the Nullstellensatz via prime bases follows Gröbner (1970), Chapter I, §3 and Chapter II, §§2,3, except that we have used Gröbner basis arguments in a number of places. The idea of introducing an additional variable in the actual proof of the Hilbert Nullstellensatz is due to Rabinowitsch (1930).

Chains of prime ideals were first investigated by Krull (1928). We have collected here those results that can be proved in the spirit of the preceding section, using prime bases, Gröbner bases, and extension-contraction arguments. A related result that does not seem to be accessible on that level is the fact that for two prime ideals I_1 and I_2 of $K[\underline{X}]$ with $I_1 \subseteq I_2$, there exists a chain of prime ideals of length $\dim(I_1) - \dim(I_2)$ connecting the two (see, e.g., Zariski and Samuel, 1958/1960, Vol. 2, VII, §7, Corollary 1; cf. also the remarks at the beginning of that paragraph).

The solution of the implicitization problem by means of Gröbner bases was initiated by Buchberger (1987a) and carried out by Kalkbrener (1990a). Kalkbrener considers varieties in the algebraic closure of the ground field; the generalization to infinite fields appears in Cox et al. (1992).

The Jacobian conjecture is stated in Keller (1939); it has thus far resisted all attempts of proof or refutation. Our treatment of the Gröbner basis approach to the corresponding decision problem follows van den Essen (1990); see also Abhyankar and Li (1989) and Audoly et. al (1991).

8

Decomposition, Radical, and Zeroes of Ideals

If R is a PID, $0 \neq a$ is a non-unit of R, and $a = p_1^{\nu_1} \cdot \dots \cdot p_r^{\nu_r}$ is a prime factor decomposition of a, then, according to Proposition 1.89, we have

$$aR = \bigcap_{i=1}^{r} p_i^{\nu_i} R .$$

We see that every ideal of R can be decomposed into an intersection of ideals that are generated by powers of pairwise non-associated irreducible elements of R. In particular, this is true for univariate polynomial rings over fields. In multivariate polynomial rings, we still have the unique prime factor decomposition of non-zero non-units. It is easy to see that one still obtains a corresponding decomposition of *principal* ideals as described above. (This is in fact true in every UFD.) However, we know that multivariate polynomial rings over fields are noetherian but not PID's. The central theme of this chapter is the fact that in a noetherian ring, every ideal can be decomposed into an intersection of *primary* ideals, where a primary ideal is an ideal that "behaves like" an ideal that is generated by a power of an irreducible element. This *primary decomposition* of ideals is thus, in a manner of speaking, a generalization of the unique prime factor decomposition to non-principal ideals.

We will also see how primary decompositions can be computed in polynomial rings over certain kinds of fields including the rational numbers. As it turns out, questions concerning radical and zeroes of polynomial ideals are closely related to the concept of the primary decomposition.

8.1 Preliminaries

In this section, we collect some definitions and results concerning ideals whose relevance will soon become apparent. Recall that by a ring, we always mean a commutative ring with unity.

If $I_1, \dots, I_r$ are ideals of a ring R, then we understand by $I_1 \cdot \dots \cdot I_r$ the set of all products $a_1 \cdot \dots \cdot a_r \in R$ with $a_i \in I_i$ for $1 \leq i \leq r$, and we define the **ideal product** of $I_1, \dots, I_r$ as $\mathrm{Id}(I_1 \cdot \dots \cdot I_r)$. The ideal product of the I_i thus consists of all sums of multiples of elements of $I_1 \cdot \dots \cdot I_r$, and it is easy to see that it actually consists of all sums of elements of $I_1 \cdot \cdots \cdot I_r$. Throughout the mathematical literature, the standard notation for the ideal product of I and J is $I \cdot J$. The authors of this book, however, have been actively involved in too many errors and pointless discussions arising from this truly misleading notation to carry it any further.

For an ideal I of a ring R and $\nu \in \mathbb{N}^+$, we use the obvious notation

$$\mathrm{Id}(I^\nu) = \mathrm{Id}(\underbrace{I \cdot \cdots \cdot I}_{\nu \text{ times}}) \quad \text{and} \quad \mathrm{Id}(I^0) = R,$$

and we refer to $\mathrm{Id}(I^\nu)$ as an **ideal power** of I.

Exercise 8.1 Show the following:

(i) Ideal multiplication is associative, and

$$\mathrm{Id}\big(I_1 \cdot \mathrm{Id}(I_2 \cdot I_3)\big) = \mathrm{Id}(I_1 \cdot I_2 \cdot I_3).$$

(ii) If I is a principal ideal, then $\mathrm{Id}(I \cdot J) = I \cdot J$ for any ideal J.

(iii) If both I and J are principal, say $I = aR$ and $J = bR$, then $\mathrm{Id}(I \cdot J) = abR$. (In particular, this means that $\mathrm{Id}(aR \cdot bR)$ is again principal, and that $\mathrm{Id}((aR)^\nu) = a^\nu R$.)

(iv) If I and J are ideals with finite bases B and C, respectively, then $\mathrm{Id}(I \cdot J)$ is generated by the finite set

$$B \cdot C = \{ ab \in R \mid a \in B,\ b \in C \}.$$

It is clear that the ideal product of finitely many ideals is contained in their intersection. The converse is not true in general: if a is a non-unit in a domain R, then

$$aR \cap a^2R = a^2R \neq a^3R = aR \cdot a^2R = \mathrm{Id}(aR \cdot a^2R).$$

Recall that two ideals I and J are called comaximal if $1 \in I + J$.

Lemma 8.2 If $I_1, \ldots, I_r$ are pairwise comaximal ideals of a ring R, then

$$\bigcap_{i=1}^{r} I_i = \mathrm{Id}\Big(\prod_{i=1}^{r} I_i\Big).$$

Proof In view of the remarks preceding the lemma, it suffices to prove the inclusion "$\subseteq$." We proceed by induction on r. If $r = 1$, then there is nothing to prove. Let $r > 1$, and suppose $a \in I_i$ for $1 \leq i \leq r$. By Lemma 6.26, we have

$$1 \in I_1 + \bigcap_{i=2}^{r} I_i,$$

say $1 = s_1 + s_2$ with s_1 and s_2 in the first and second summand, respectively. From the induction hypothesis we may conclude that a, $s_2 \in \prod_{i=2}^{r} I_i$, and we see that

$$a = a \cdot 1 = as_1 + as_2 \in \prod_{i=1}^{r} I_i \,. \quad \square$$

Note that the statement of the last lemma has already been proved for PID's in Proposition 1.89.

Lemma 8.3 Let R be a ring, $I_1, \ldots, I_r$ ideals of R with radicals $J_i = \operatorname{rad}(I_i)$ for $1 \leq i \leq r$, respectively. Assume further that P is a prime ideal of R which does not contain any one of the J_i for $1 \leq i \leq r$. Then there exists

$$b \in (I_1 \cdot \cdots \cdot I_r) \setminus P.$$

Proof Let $b_i \in J_i \setminus P$ for $1 \leq i \leq r$. Then there are $\nu_i \in \mathbb{N}$ with $b_i^{\nu_i} \in I_i$ for $1 \leq i \leq r$, and so

$$b = (b_1^{\nu_1} \cdot \cdots \cdot b_r^{\nu_r}) \in I_1 \cdot \cdots \cdot I_r.$$

But $b \notin P$ since otherwise at least one of the b_i would have to be in P. □

Lemma 8.4 Let R be a ring, $M, M_1, \ldots, M_r$ pairwise different maximal ideals of R. Then M does not contain the intersection $\bigcap_{i=1}^{r} M_i$, i.e.,

$$M \cap \bigcap_{i=1}^{r} M_i \neq \bigcap_{i=1}^{r} M_i \,.$$

Proof Since M and the M_i are pairwise different and maximal, M does not contain any one of the M_i. We may now apply the previous lemma with $I_i = J_i = M_i$ for $1 \leq i \leq r$ and $P = M$ to obtain

$$b \in (M_1 \cdot \cdots \cdot M_r) \setminus M \subseteq (M_1 \cap \cdots \cap M_r) \setminus P.$$

(The ideal product actually equals the intersection here because, as one easily sees, the M_i are pairwise comaximal.) □

Next, we prove a technical lemma that is a generalization to multivariate polynomial rings of Proposition 1.89. For the rest of this section, let K be a field and $K[\underline{X}] = K[X_1, \ldots, X_n]$.

Lemma 8.5 Let I be an ideal of $K[\underline{X}]$. Assume that $f, g_1, \ldots, g_r \in K[X_1]$ are such that $f = g_1 \cdot \cdots \cdot g_r$ is a factorization of f in $K[X_1]$ into pairwise relatively prime factors. Then

$$\operatorname{Id}(I, f) = \bigcap_{i=1}^{r} \operatorname{Id}(I, g_i).$$

Proof The inclusion "$\subseteq$" is trivial. For the reverse inclusion, let h be an element of the intersection on the right-hand side. Then for $1 \le i \le r$, there exist $q_i \in K[\underline{X}]$ and $s_i \in I$ with $h = q_i g_i + s_i$. Now if we set

$$f_i = \prod_{\substack{j=1 \\ j \neq i}}^{r} g_j \quad \text{for} \quad 1 \le i \le r,$$

then it follows that $hf_i \in \mathrm{Id}(I, f)$ for $1 \le i \le r$. From the fact that g_i and g_j have no prime factor in common for $1 \le i < j \le r$, one easily concludes that the gcd of $f_1, \ldots, f_r$ in $K[X_1]$ equals 1, and so there exist $t_1, \ldots, t_r \in K[X_1]$ with $1 = t_1 f_1 + \cdots + t_r f_r$, from which it follows that

$$h = \sum_{i=1}^{r} t_i h f_i \in \mathrm{Id}(I, f). \quad \square$$

We will now show how the previous lemma can be used to decompose a given zero-dimensional polynomial ideal I into an intersection of pairwise comaximal ideals $I_1, \ldots, I_r$ such that for $1 \le j \le r$ and $1 \le i \le n$, the unique monic generator of $I_j \cap K[X_i]$ is a power of an irreducible polynomial. This decomposition is not itself of particular interest, but it is often used in practice in order to "preprocess" an ideal for certain purposes. Recall that by Lemma 6.50, we can compute the monic univariate polynomial f_i of minimal degree in $\mathrm{Id}(F) \cap K[X_i]$ for $1 \le i \le n$ whenever F is a finite subset of $K[\underline{X}]$, $\mathrm{Id}(F)$ is zero-dimensional, and K is computable. (Recall further that a more efficient method to achieve this will be given in Proposition 9.6.) The algorithm PREDEC of the following lemma computes these univariate polynomials, factorizes them into products of prime powers, and then forms all ideals $\mathrm{Id}(I, p_1^{s_1}, \ldots, p_n^{s_n})$, where for $1 \le i \le n$, the prime power $p_i^{s_i}$ is taken from the factorization of the univariate polynomial in the variable X_i.

Lemma 8.6 Let I be a zero-dimensional ideal of $K[\underline{X}]$. Then there exist pairwise comaximal ideals $I_1, \ldots, I_r$ such that $I = I_1 \cap \cdots \cap I_r$, and for $1 \le j \le r$ and $1 \le i \le n$, the unique monic generator of $I_j \cap K[X_i]$ is of the form p^s with $s \in \mathbb{N}^+$ and p irreducible in $K[X_i]$. If K is computable and allows effective factorization of univariate polynomials, then the algorithm PREDEC of Table 8.1 computes finite bases of $I_1, \ldots, I_r$ from any finite basis of I.

Proof We will prove the correctness of the algorithm; for the general case of an arbitrary field, this amounts to a mathematical existence proof. It clearly suffices to prove that for $0 \le k \le n$, after the run $i = k$ through the **for**-loop, properties (ii) and (iii) as stated under "**Find**" hold, and (i) holds with n replaced by k. This claim is trivial for $k = 0$, i.e., upon initialization of H. Now suppose $k > 0$ and the claim was true after the run $i = k - 1$

TABLE 8.1. Algorithm PREDEC

```
Specification: $H \leftarrow$ PREDEC($F$)
               Decomposition of a zero-dimensional ideal into an
               intersection of ideals each of which contains a power
               of an irreducible polynomial in each variable
Given: a finite subset $F$ of $K[\underline{X}]$ with Id($F$) zero-dimensional
Find: a set $H$ of finite subsets of $K[\underline{X}]$ such that
      (i) for all $G \in H$ and $1 \leq i \leq n$, there exists an irreducible
          $p \in K[X_i]$ and $s \in \mathbb{N}^+$ with $p^s \in \mathrm{Id}(G) \cap K[X_i]$,
      (ii) the ideals generated by the elements of $H$ are pairwise
          comaximal, and
      (iii) $\mathrm{Id}(F) = \bigcap_{G \in H} \mathrm{Id}(G)$.
begin
$H \leftarrow \{F\}$
for i=1 to n do
    $S \leftarrow H$;  $H \leftarrow \emptyset$
    $f \leftarrow$ the monic generator of $\mathrm{Id}(F) \cap K[X_i]$
    while $f$ is not constant do
        $p \leftarrow$ an irreducible factor of $f$
        $s \leftarrow \max\{ r \in \mathbb{N} \mid p^r | f \}$
        $f \leftarrow f/p^s$
        $T \leftarrow S$
        while $T \neq \emptyset$ do
            select $G$ from $T$
            $T \leftarrow T \setminus \{G\}$
            if PROPER($G \cup \{p^s\}$) then
               $H \leftarrow H \cup \{ G \cup \{p^s\} \}$
            end
        end
    end
end
end PREDEC
```

through the **for**-loop, and consider the run $i = k$ through the loop. Each element that is placed into H during this run is of the form $G \cup \{p^s\}$ with G in the value of H after the previous run and $p \in K[X_k]$ irreducible. It is now obvious that property (i) will hold with n replaced by k after the present run.

To see that property (ii) continues to hold, it suffices to prove that every time a new element $G \cup \{p^s\}$ is added to H, the ideals $\mathrm{Id}(G \cup \{p^s\})$ and $\mathrm{Id}(G')$ are comaximal for all G' that are already in H at the time. If G' entered H during the same run through the outer **while**-loop, then

$G' = G'' \cup \{p^s\}$ for some $G'' \neq G$, and G and G'' were both elements of H after the previous run through the **for**-loop. It follows that $\mathrm{Id}(G'')$ and $\mathrm{Id}(G)$ are comaximal, and this property is trivially preserved when the ideals are being enlarged. If G' was added to H during an earlier run through the outer **while**-loop, then, since G' has been processed through the inner **while**-loop, there exists $q^r \in K[X_k]$ with $q^r \in G'$ and $\gcd(p^s, q^r) = 1$. The claim now follows from the fact that this gcd is a sum of multiples of p^s and q^r in $K[X_k]$.

For property (iii), we first note that rather obviously, we have $\mathrm{Id}(F) \subseteq \mathrm{Id}(G)$ for every G that is ever an element of H during the course of the computation. It follows that the polynomial f that is computed during the present run $i = k$ through the **for**-loop is in $\mathrm{Id}(G)$ for all G that are in the value H_k of H as this run is entered. It is not hard to see that at the end of the run, the value of H has turned into

$$H_{k+1} = \left\{ G \cup \{p_j^{s_j}\} \mid G \in H_k,\ 1 \leq j \leq m \right\},$$

where $p_1^{s_1}, \ldots, p_m^{s_m}$ are the powers of irreducible factors of f that the outer **while**-loop provides. Using Lemma 8.5, we thus obtain

$$\bigcap_{G \in H_{k+1}} \mathrm{Id}(G) = \bigcap_{G \in H_k} \bigcap_{j=1}^{m} \mathrm{Id}(G, p_j^{s_j}) = \bigcap_{G \in H_k} \mathrm{Id}(G) = \mathrm{Id}(F). \quad \square$$

Exercise 8.7 The elements of the set H that PREDEC outputs are clearly of the form

$$G = F \cup \{p_1^{s_1}, \ldots, p_n^{s_n}\},$$

where $p_i \in K[X_i]$ is a prime factor with multiplicity s_i of the unique monic generator of $\mathrm{Id}(F) \cap K[X_i]$ for $1 \leq i \leq n$. Explain why $p_i^{s_i}$ may not be the unique monic generator of $\mathrm{Id}(G) \cap K[X_i]$.

Exercise 8.8 Write an alternate version of PREDEC where the univariate polynomial that is used to further "branch" a set $G \in H$ on the level i is not a generator of $\mathrm{Id}(F) \cap K[X_i]$, but a generator of $\mathrm{Id}(G) \cap K[X_i]$.

Exercise 8.9 Apply both the algorithm PREDEC and the version of the previous exercise to the subset

$$F = \{X^2 - XY + X - Y, XY + 2X + Y + 2, Y^3 + 4Y^2 + 4Y\}$$

of $\mathbb{Q}[X, Y]$. Here, PREDEC will actually come across improper ideals. Show that the version of the previous exercise need not worry about improper ideals. (Hint: you may want to look at the first part of the proof of Proposition 8.69.)

8.2 The Radical of a Zero-Dimensional Ideal

Throughout this section, K will be a field and $K[\underline{X}] = K[X_1, \ldots, X_n]$. Recall that the Hilbert Nullstellensatz characterizes the radical of an ideal

I of $K[\underline{X}]$ as the set of all polynomials that vanish at every zero of I in L^n, where L is any given algebraically closed extension of K. We have an algorithm RADICALMEMTEST that tests a given $f \in K[\underline{X}]$ for membership in the radical of a given ideal. What we do not yet have is an algorithm that decides whether a given ideal I is a radical ideal, i.e., whether $I = \mathrm{rad}(I)$, nor one that computes a basis of the radical of I. In this section, we will solve these problems for zero-dimensional ideals. We remind the reader that some elementary results on the radical were given at the end of Section 4.1.

It is immediate from the definition of the radical that a prime ideal I of any ring is a radical ideal. It is also easy to see that the intersection of radical ideals is again a radical ideal. We already know that the radical of any ideal I equals the intersection of all prime ideals containing I. Now if I is a zero-dimensional ideal of $K[\underline{X}]$, then every ideal containing I is zero-dimensional too, and thus, in view of Proposition 7.42, every prime ideal containing I is maximal. These observations combine into the following lemma.

Lemma 8.10 Let I be a zero-dimensional ideal of $K[\underline{X}]$. Then $\mathrm{rad}(I)$ equals the intersection of all maximal ideals containing I. I is itself a radical ideal iff it is an intersection of maximal ideals. □

Following are two easy observations that will be needed below.

Exercise 8.11 Show the following:

(i) If I is an ideal of a UFD and $a \in I$, then the squarefree part of a is in $\mathrm{rad}(I)$.

(ii) If I and J are a ideals of any ring R with $I \subseteq J \subseteq \mathrm{rad}(I)$, then $\mathrm{rad}(I) = \mathrm{rad}(J)$.

Lemma 8.12 A zero-dimensional radical ideal I of $K[\underline{X}]$ contains a univariate squarefree polynomial in each of the n variables.

Proof Being zero-dimensional, I contains a univariate polynomial in each variable, and by (i) of the exercise above, it contains the squarefree part of each of these. □

Lemma 8.13 (SEIDENBERG'S LEMMA 92) Let I be a zero-dimensional ideal of $K[\underline{X}]$, and assume that for $1 \leq i \leq n$, I contains a polynomial $f_i \in K[X_i]$ with $\gcd(f_i, f_i') = 1$. Then I is an intersection of finitely many maximal ideals. In particular, I is then radical.

Proof We first note that by Lemma 2.84, f_i is squarefree for $1 \leq i \leq n$. To prove the lemma, we proceed by induction on n. If $n = 1$, then the generator f of I must be squarefree since every multiple of a polynomial that is not squarefree is not squarefree either. Let $f = g_1 \cdot \cdots \cdot g_r$ with pairwise non-associated, irreducible polynomials $g_1, \ldots, g_r \in K[X_1]$. Then the g_i are pairwise relatively prime, and so

$$I = \mathrm{Id}(f) = \bigcap_{i=1}^{r} \mathrm{Id}(g_i)$$

by Lemma 8.5 (applied with I of that lemma being the zero ideal), and the ideals occurring in the intersection are all maximal since their generators are irreducible. Now let $n > 1$. As before, we may write $f_1 = g_1 \cdot \cdots \cdot g_r$ with pairwise non-associated, irreducible polynomials $g_1, \ldots, g_r \in K[X_1]$, and again by Lemma 8.5, we obtain

$$I = \mathrm{Id}(I, f_1) = \bigcap_{i=1}^{r} \mathrm{Id}(I, g_i).$$

It now suffices to prove that the ideals occuring in the intersection on the right-hand side are intersections of finitely many maximal ideals, and this obviously means that we may assume w.l.o.g. that f_1 is in fact irreducible. Then $K[X_1]/\mathrm{Id}(f_1)$ is a field, and we may consider the canonical homomorphism

$$\psi : K[X_1] \longrightarrow K[X_1]/\mathrm{Id}(f_1)$$

which induces a natural surjective homomorphism

$$\varphi : K[X_1][X_2, \ldots, X_n] \longrightarrow K[X_1]/\mathrm{Id}(f_1)[X_2, \ldots, X_n]$$

according to Lemma 2.17 (ii). The kernel of φ, as one easily proves, equals the ideal generated by f_1 in $K[\underline{X}]$ and is thus contained in I. Now $J = \varphi(I)$ is an ideal of the ring on the right-hand side by Lemma 1.62 (ii), and we claim that the induction hypothesis applies to J. Indeed, according to the remarks preceding Proposition 7.7, we may view $K[X_1]/\mathrm{Id}(f_1)$ as an extension field of K, whence $\varphi \upharpoonright K = \mathrm{id}_K$, and so the polynomials

$$\varphi(f_i) = f_i \in J \qquad (2 \leq i \leq n)$$

satisfy $\gcd(f_i, f_i') = 1$ over the field $K[X_1]/\mathrm{Id}(f_1)$ by Proposition 2.38. We conclude that J equals the intersection of finitely many maximal ideals M_1, $\ldots$, M_s, and so

$$I = \varphi^{-1}(J) = \bigcap_{i=1}^{s} \varphi^{-1}(M_i)$$

by Lemmas 1.62 (iii) and 0.10 (i). Moreover, it is easy to prove from Lemma 1.62 (iii) that the ideals $\varphi^{-1}(M_i)$ on the right-hand side are maximal ideals of $K[\underline{X}]$. □

The following proposition is an immediate consequence of of the last two lemmas together with Theorem 7.36.

Proposition 8.14 *If K is perfect, then a zero-dimensional ideal of $K[\underline{X}]$ is a radical ideal iff it contains a univariate squarefree polynomial in each variable.* □

Together with Corollary 7.39, we obtain the following lemma on the invariance of the radical property under field extensions.

Lemma 8.15 If K is perfect, then for every zero-dimensional radical ideal I of $K[\underline{X}]$ and every extension field K' of K, the ideal generated by I in $K'[\underline{X}]$ is again radical. □

The following example shows that the last proposition is not true for an arbitrary field.

Example 8.16 Let p be a prime number, let K be the rational function field $\mathbb{Z}/p\mathbb{Z}(T)$, and consider $G = \{f, g\} \subseteq K[X, Y]$ with $f = X^p - T$ and $g = Y^p - T$. We have proved in Example 7.32 that f and g are irreducible and hence squarefree. Consider the polynomial $h = X - Y$. G is a Gröbner basis w.r.t. every term order since the head terms of f and g are disjoint, and we see that $h \notin \mathrm{Id}(G)$. But we have

$$h^p = (X - Y)^p = X^p - Y^p = f - g \in \mathrm{Id}(G),$$

and so $\mathrm{Id}(G)$ is not a radical ideal.

Exercise 8.17 Show that for *univariate* polynomial ideals, the equivalence of Proposition 8.14 holds over arbitrary fields.

If we want to use Proposition 8.14 in order to effectively test a zero-dimensional ideal for being radical, then we must know where to look for the univariate squarefree polynomials.

Lemma 8.18 Let I be any proper ideal of $K[\underline{X}]$. If I contains a squarefree polynomial f in some variable X_i, then the unique monic univariate polynomial g of minimal degree in $I \cap K[X_i]$ is squarefree.

Proof Since g is a generator of the ideal $I \cap K[X_i]$, it divides the squarefree polynomial f and is thus itself squarefree. □

For the actual computation of the radical of a zero-dimensional ideal, we need one more lemma.

Lemma 8.19 Assume that K is perfect, and let I be a zero-dimensional ideal of $K[\underline{X}]$. For $1 \leq i \leq n$, let f_i be the unique monic polynomial of minimal degree in $I \cap K[X_i]$ with squarefree part g_i. Then

$$\mathrm{rad}(I) = \mathrm{Id}(I, g_1, \ldots, g_n).$$

Proof If we set $J = \mathrm{Id}(I, g_1, \ldots, g_n)$, then we have

$$I \subseteq J \subseteq \mathrm{rad}(I)$$

according to Exercise 8.11 (i). Moreover, J is a radical ideal by Proposition 8.14, and Exercise 8.11 (ii) tells us that $\mathrm{rad}(I) = \mathrm{rad}(J) = J$. □

If K is a computable field that is either finite or has characteristic zero, then it is perfect by Corollary 7.37, and it allows the computation of squarefree decompositions of univariate polynomials by Proposition 2.86. The next two theorems will therefore apply. We will once again use the fact that one can compute the monic univariate polynomial f_i of minimal degree in $\mathrm{Id}(F) \cap K[X_i]$ for $1 \leq i \leq n$ whenever F is a finite subset of $K[\underline{X}]$, $\mathrm{Id}(F)$ is zero-dimensional, and K is computable. (Let us emphasize again that an efficient method to achieve this will be given in Proposition 9.6.) If K is perfect, then by Proposition 8.14 and Lemma 8.18, $\mathrm{Id}(F)$ is a radical ideal iff each f_i is squarefree. In view of Theorem 7.36, this latter condition is equivalent to $\gcd(f_i, f_i') = 1$. We have proved the following theorem.

Theorem 8.20 *Assume that K is perfect and computable. Then the algorithm* ZRADICALTEST *of Table* 8.2 *decides, for given finite subset F of $K[\underline{X}]$ with* $\mathrm{Id}(F)$ *zero-dimensional, whether* $\mathrm{Id}(F)$ *is radical.* □

TABLE 8.2. Algorithm ZRADICALTEST

```
Specification: $v \leftarrow$ ZRADICALTEST($F$)
               Decision whether a zero-dimensional ideal is radical
Given: a finite subset $F$ of $K[\underline{X}]$ with $\mathrm{Id}(F)$ zero-dimensional
Find: $v \in$ {false, true} such that $v =$ true iff $\mathrm{Id}(F)$ is radical
begin
for i=1 to n do
    $f_i \leftarrow$ the monic generator of $\mathrm{Id}(F) \cap K[X_i]$
    if $\gcd(f_i, f_i') \neq 1$ then return(false) end
end
return(true)
end ZRADICALTEST
```

Exercise 8.21 Show that the ideal $\mathrm{Id}(X^2 + Y, Y^2 + X)$ of $\mathbb{Q}[X, Y]$ is radical.

The correctness of the next algorithm is immediate from Lemma 8.19.

Theorem 8.22 *Assume that K is perfect and computable and allows the computation of squarefree decompositions of univariate polynomials. Then the algorithm* ZRADICAL *of Table* 8.3 *computes a basis of* $\mathrm{rad}(\mathrm{Id}(F))$ *for a given finite subset F of $K[\underline{X}]$ with* $\mathrm{Id}(F)$ *zero-dimensional.* □

TABLE 8.3. Algorithm ZRADICAL

```
Specification: $G \leftarrow$ ZRADICAL($F$)
               Computation of zero-dimensional radical
Given: a finite subset $F$ of $K[\underline{X}]$ with Id($F$) zero-dimensional
Find: a finite basis $G$ of rad(Id($F$))
begin
$G \leftarrow F$
for i=1 to n do
    $f_i \leftarrow$ the monic generator of Id($F$) $\cap K[X_i]$
    $g_i \leftarrow$ the squarefree part of $f_i$
    $G \leftarrow G \cup \{g_i\}$
end
end ZRADICAL
```

Exercise 8.23 The algorithm ZRADICAL only calls for the computation of univariate squarefree *parts* rather than squarefree decompositions. Explain how the algorithms of Proposition 2.86 can be streamlined for this particular purpose.

Another remarkable consequence of Lemma 8.19 is that given a zero-dimensional polynomial ideal I, it is possible to compute *one* natural number μ such that for *all* $f \in \mathrm{rad}(I)$, the power f^μ is in I.

Definition 8.24 Let I be a zero-dimensional ideal of $K[\underline{X}]$. For $1 \leq i \leq n$, let f_i be the unique monic polynomial of minimal degree in $I \cap K[X_i]$, and set

$$\mu_i = \max\{\, \nu \in \mathbb{N} \mid p^\nu \mid f_i \text{ with } p \in K[X_i] \text{ irreducible} \,\},$$

i.e., μ_i is the highest exponent that occurs non-trivially in the squarefree decomposition of f_i. Then we call the natural number

$$\mu = 1 + \sum_{i=1}^{n} (\mu_i - 1)$$

the **univariate exponent** of I.

It is clear that the univariate exponent of a zero-dimensional ideal of $K[\underline{X}]$ can be computed as soon as K is computable and allows effective squarefree decompositions of univariate polynomials.

Exercise 8.25 Explain how the algorithms of Proposition 2.86 can be trimmed down to provide univariate exponents without actually computing the squarefree decomposition.

Proposition 8.26 *Assume that K is perfect, and let I be a zero-dimensional ideal of $K[\underline{X}]$ with univariate exponent μ and radical $J = \mathrm{rad}(I)$. Then $\mathrm{Id}(J^\mu) \subseteq I$. In particular, $f \in \mathrm{rad}(I)$ implies $f^\mu \in I$.*

Proof It clearly suffices to prove that $J^\mu \subseteq I$. For $1 \leq i \leq n$, let f_i be the unique monic polynomial of minimal degree in $I \cap K[X_i]$ with squarefree part g_i. Then

$$J = \mathrm{Id}(I, g_1, \ldots, g_n)$$

by Lemma 8.19, and thus every element $f \in J^\mu$ is of the form

$$f = \prod_{i=1}^{\mu}\left(s_i + \sum_{j=1}^{n} q_{ij}g_j\right)$$

with $s_i \in I$ and $q_{ij} \in K[\underline{X}]$ for $1 \leq i \leq \mu$ and $1 \leq j \leq n$. Expanding the product in a suitable way, we see that there exists $s \in I$ with

$$f = s + \prod_{i=1}^{\mu}\left(\sum_{j=1}^{n} q_{ij}g_j\right). \tag{$*$}$$

Now if we fully expand the product on the right-hand side of $(*)$, then it is clear that every summand is of the form

$$q \cdot \prod_{j=1}^{n} g_j^{\nu_j}$$

with $q \in K[\underline{X}]$ and $\nu_1 + \cdots + \nu_n = \mu$. It follows that there must exist an index $1 \leq j \leq n$ such that $\nu_j \geq \mu_j$, where μ_j is as in the definition of the univariate exponent, and one easily concludes that $f_j \mid g_j^{\nu_j}$. We have proved that every summand in the full expansion of the product in $(*)$ is in I, and thus $f \in I$. □

8.3 The Number of Zeroes of an Ideal

Throughout this section, let K be a field, $K[\underline{X}] = K[X_1, \ldots, X_n]$, and I a proper ideal of $K[\underline{X}]$. Recall from Section 6.3 that each of the following conditions is equivalent to I being zero-dimensional.

(i) I contains a non-zero univariate polynomial in each of the n variables.

(ii) Every Gröbner basis of I contains n polynomials $g_1, \ldots, g_n$ such that the head term of g_i is a power of X_i.

(iii) There exists a term order $\leq$ and a Gröbner basis of I w.r.t. $\leq$ that contains n polynomials $g_1, \ldots, g_n$ such that the head term of g_i is a power of X_i.

(iv) The vector space dimension $\dim_K(K[\underline{X}]/I)$ is finite.

Moreover, we have an upper bound for $\dim_K(K[\underline{X}]/I)$ in this case (Corollary 6.55). We will now establish a connection between all this and the number of different zeroes of I in extension fields of K.

Proposition 8.27 *The following are equivalent:*

(i) $\dim(I) = 0$.

(ii) There exists an algebraically closed extension L of K such that I has only finitely many different zeroes in L^n.

(iii) For every algebraically closed extension L of K, I has only finitely many different zeroes in L^n.

Proof (i)$\Longrightarrow$(iii): Assume that $\dim(I) = 0$ and L is an algebraically closed extension of K. Then I contains an element $0 \neq f_i \in K[X_i]$ for $1 \leq i \leq n$. If $z = (z_1, \ldots, z_n) \in L^n$ is a zero of I, then $f_i(z_i) = 0$ for $1 \leq i \leq n$. By Corollary 2.97, this means that there are only finitely many possibilities for each z_i.

(iii)$\Longrightarrow$(ii): This is trivial in view of the existence of the algebraic closure.

(ii)$\Longrightarrow$(i): Suppose $\dim(I) = d > 0$. W.l.o.g., we may assume that $\{X_1, \ldots, X_d\}$ is maximally independent modulo I. Then I extends to a prime ideal J which is disjoint from the multiplicative set $K[X_1, \ldots, X_d]$. By Lemma 7.52, every $(z_1, \ldots, z_d) \in L^d$ with the possible exception of the zeroes of a polynomial $0 \neq p \in K[X_1, \ldots, X_d]$ can be extended to a zero of J and thus of I. By Lemma 7.50, there are infinitely many possibilities. □

The argument of (i)$\Longrightarrow$(iii) in the proof above can actually be refined to obtain an estimate on the number of zeroes of a zero-dimensional ideal. If, for $1 \leq i \leq n$, we find that $0 \neq f_i \in I \cap K[X_i]$, then by Corollary 2.97, f_i can have at most $m_i = \deg(f_i)$ many different zeroes in any extension field K' of K. So there are at most $m = m_1 \cdot \cdots \cdot m_n$ possibilities for simultaneous zeroes of $f_1, \ldots, f_n$ in $(K')^n$, and we have proved the following corollary.

Corollary 8.28 *Let K' be an extension field of K. If* $\dim(I) = 0$ *and $m = m_1 \cdot \cdots \cdot m_n$, where for $1 \leq i \leq n$,*

$$m_i = \min\{ \deg(f_i) \mid 0 \neq f_i \in I \cap K[X_i] \},$$

then I has at most m different zeroes in $(K')^n$. □

Let us now compare this result to the upper bound for $\dim_K(K[\underline{X}]/I)$ of Corollary 6.55. We fix a term order $\leq$ and consider a Gröbner basis G of I w.r.t. $\leq$. Corollary 6.55 tells us that $\dim_K(K[\underline{X}]/I) \leq \nu$ where $\nu = \nu_1 \cdot \cdots \cdot \nu_n$ with

$$\nu_i = \min\{ \mu_i \mid X_i^{\mu_i} \in \mathrm{HT}(G) \}$$

for $1 \leq i \leq n$. Now every $0 \neq f \in I \cap K[X_i]$ is reducible modulo the Gröbner basis G of I, and so we must have $\nu_i \leq m_i$ for $1 \leq i \leq n$, where m_i is defined as in the corollary above. It follows that $\nu \leq m$, i.e., our bound m for the number of zeroes of I is worse in general than the bound ν for the vector space dimension of $K[\underline{X}]/I$.

Our goal in this section is to prove that the number of zeroes of a zero-dimensional ideal I is actually less than or equal to the better bound $\dim_K(K[\underline{X}]/I)$, and that equality holds in case K is perfect and I is a radical ideal. Let us first illustrate the situation with two simple examples.

Example 8.29 Let I be the ideal $\mathrm{Id}(G)$ of $\mathbb{Q}[X,Y]$, where

$$G = \{X^2 + Y, Y^2 + X\}.$$

Then G is a Gröbner basis w.r.t. every total degree order because the head terms X^2 and Y^2 are disjoint. We see that $\nu = 4$, and the dimension of $\mathbb{Q}[X,Y]/I$ actually equals 4 in this case because the univariate head terms are the only ones, and thus the canonical term basis consists of the residue classes of 1, X, Y, and XY. It is easy to compute by hand Gröbner bases of I w.r.t. the two lexicographical orders, and this produces the univariate polynomials $f_X = X^4 + X$ and $f_Y = Y^4 + Y$ of minimal degree. These are easily factored, and we see that each of them has the zeroes

$$\{z_1, z_2, z_3, z_4\} = \left\{0, -1, \frac{1+i\sqrt{3}}{2}, \frac{1-i\sqrt{3}}{2}\right\}.$$

This leaves us with $m = 16$ possible combinations for simultaneous zeroes of f_X and f_Y. But substitution into G shows that only (z_1, z_1), (z_2, z_2), (z_3, z_4), and (z_4, z_3) are zeroes of I, and we see that the number of zeroes equals $\dim_{\mathbb{Q}}(\mathbb{Q}[X,Y]/I)$.

Exercise 8.30 Let I be the ideal $\mathrm{Id}(G)$ of $\mathbb{Q}[X,Y]$, where

$$G = \{X^2 + Y^2 + 1, Y^2 + 2X\}.$$

Explain why G is a Gröbner basis w.r.t. the total degree–lexicographical order. Show that here, the number of zeroes of I is strictly less than $\dim_{\mathbb{Q}}(\mathbb{Q}[X,Y]/I)$.

Exercise 8.31 Show that the number of zeroes of a zero-dimensional ideal, the bound ν, and the bound m as defined above may all coincide. (Hint: Consider a set of n irreducible univariate polynomials, one in each variable.)

In the proof of the following theorem, we will once again make use of the fact that for any proper ideal I of $K[\underline{X}]$, the canonical homomorphism from $K[\underline{X}]$ to $K[\underline{X}]/I$ becomes injective when restricted to K, and that we may thus identify $a \in K$ with $a+I$ and view K as a subfield of $K[\underline{X}]/I$.

Theorem 8.32 *Assume that* $\dim(I) = 0$, *and let L be an algebraically closed extension field of K. Then the number of zeroes of I in L^n is less than or equal to the vector space dimension* $\dim_K(K[\underline{X}]/I)$. *If K is perfect and I is a radical ideal, then equality holds.*

Proof Let G be a Gröbner basis of I w.r.t. any term order, and let J be the ideal generated by G in $L[\underline{X}]$. Since the set of zeroes of an ideal obviously equals the set of zeroes of any basis of that ideal, I and J have exactly the same zeroes in L^n. Moreover, G is a Gröbner basis of J by Corollary 5.51 (i). It follows that the set of reduced terms is the same w.r.t. I and J, and we see from Proposition 6.52 that the canonical term bases of $K[\underline{X}]/I$ and $L[\underline{X}]/J$ have the same number of elements, i.e.,

$$\dim_K(K[\underline{X}]/I) = \dim_L(L[\underline{X}]/J).$$

Moreover, if I is radical and K is perfect, then J is again radical by Lemma 8.15. All this together shows that we may assume w.l.o.g. that $K = L$ (and thus $I = J$).

Let now $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$ be the different zeroes of I in K^n, where

$$\boldsymbol{a}_i = (a_{i1}, \ldots, a_{in}) \quad \text{for} \quad 1 \le i \le k,$$

and set

$$I_{\boldsymbol{a}_i} = \mathrm{Id}(X_1 - a_{i1}, \ldots, X_n - a_{in}) \quad \text{for} \quad 1 \le i \le k.$$

By Lemma 1.116, the map

$$\begin{array}{rccc} \varphi: & K[\underline{X}] & \longrightarrow & \prod_{i=1}^k K[\underline{X}]/I_{\boldsymbol{a}_i} \\ & f & \longmapsto & (f + I_{\boldsymbol{a}_1}, \ldots, f + I_{\boldsymbol{a}_k}) \end{array}$$

is a homomorphism of rings with kernel $\bigcap_{i=1}^k I_{\boldsymbol{a}_i}$. We claim that φ is surjective. If

$$(f_1 + I_{\boldsymbol{a}_1}, \ldots, f_k + I_{\boldsymbol{a}_k}) \in \prod_{i=1}^{k} K[\underline{X}]/I_{\boldsymbol{a}_i},$$

then Lemma 6.27 together with Lemma 6.28 (iv) provides the existence of a polynomial $f \in \bigcap_{i=1}^k f_i + I_{\boldsymbol{a}_i}$. It is clear that then

$$\varphi(f) = (f + I_{\boldsymbol{a}_1}, \ldots, f + I_{\boldsymbol{a}_k}) = (f_1 + I_{\boldsymbol{a}_1}, \ldots, f_k + I_{\boldsymbol{a}_k}),$$

and we have proved that φ is indeed surjective.

Since $f \in I$ implies that f vanishes at $\boldsymbol{a}_i$ for all $1 \le i \le k$, we see with Lemma 6.28 (iii) that

$$I \subseteq \bigcap_{i=1}^{k} I_{\boldsymbol{a}_i} = \ker(\varphi).$$

The homomorphism theorem for rings now tells us that the map

$$\begin{array}{rccc} \psi: & K[\underline{X}]/I & \longrightarrow & \prod_{i=1}^k K[\underline{X}]/I_{\boldsymbol{a}_i} \\ & f + I & \longmapsto & (f + I_{\boldsymbol{a}_1}, \ldots, f + I_{\boldsymbol{a}_k}) \end{array}$$

is a surjective homomorphism of rings. Lemma 6.28 (ii) tells us that every $f \in K[\underline{X}]$ is congruent to some constant modulo I_{a_i}. With our understanding of viewing K as a subfield of $K[\underline{X}]/I_{a_i}$, we thus have

$$K[\underline{X}]/I_{a_i} = K \quad \text{for} \quad 1 \leq i \leq k,$$

and so we may view $\prod_{i=1}^{k} K[\underline{X}]/I_{a_i}$ as the vector space K^k of Example 3.2 (ii). It is now straightforward to prove from the definitions of the ring and vector space operations in K^k and the definitions of homomorphisms of rings and vector spaces that ψ is in fact a homomorphism of K-vector spaces. Since ψ is surjective and $\dim_K(K^k) = k$, we conclude with Lemma 3.22 (ii) that $k \leq \dim_K(K[\underline{X}]/I)$.

Finally, assume that I is a radical ideal. To prove that

$$k = \dim_K(K[\underline{X}]/I),$$

it suffices to show that ψ is bijective, which, in view of the homomorphism theorem, can be inferred from $I = \bigcap_{i=1}^{k} I_{a_i}$. We have already proved the inclusion $\subseteq$. If f is an element of the intersection on the right-hand side, then f vanishes at a_i for $1 \leq i \leq k$. Since these are all zeroes of I in K^n and K is algebraically closed, the Hilbert Nullstellensatz (applied to any basis of I) tells us that $f \in \text{rad}(I) = I$. □

Corollary 8.33 *Let K be perfect. If* $\dim(I) = 0$ *and L is an algebraically closed extension field of K, then the number of zeroes of I in L^n equals* $\dim_K(K[\underline{X}]/\text{rad}(I))$.

Proof It is immediate from the definition of the radical that I and $\text{rad}(I)$ have the same zeroes in any extension field of K. Moreover, we have already noted (Exercise 4.14) that $\text{rad}(I)$ is a radical ideal. The claim is now an easy consequence of the theorem. □

The condition that K is perfect is in fact equivalent to the statement concerning the number of zeroes of the radical: if K is not perfect, then there exists a univariate irreducible polynomial f over K with multiple zeroes in $\overline{K}$, and we see that $\text{Id}(f)$ is a maximal ideal with less than $\deg(f)$ many different zeroes in $\overline{K}$, while $\dim_K(K[X]/\text{Id}(f)) = \deg(f)$.

Exercise 8.34 Explain why the ideal I of Example 8.29 had exactly four different zeroes, and why it was possible for the ideal I of Exercise 8.30 to have less than four zeroes.

8.4 Primary Ideals

In order to explain what this section is all about, let us once again consider a proper, non-trivial ideal I in a PID R, e.g., a univariate polynomial ring over a field. Then I is generated by a non-zero non-unit $a \in R$ which has a unique prime factor decomposition

$$a = up_1^{\nu_1} \cdot \dots \cdot p_r^{\nu_r},$$

where $p_1, \dots, p_r \in R$ are irreducible and pairwise non-associated, and $u \in R$ is a unit. By Proposition 1.89,

$$I = aR = \bigcap_{i=1}^{r} p_i^{\nu_i} R,$$

and one easily proves that each ideal occurring in the intersection on the right-hand side has the following characteristic properties: whenever

$$bc \in p_i^{\nu_i} R \quad \text{and} \quad b \notin p_i^{\nu_i} R,$$

then $c^\mu \in p_i^{\nu_i} R$ for some $\mu \in \mathbb{N}$, and whenever b is an element of the prime ideal $p_i R$, then $b^{\nu_i} \in p_i^{\nu_i} R$. These considerations do not apply to multivariate polynomial ideals because multivariate polynomial rings are not PID's. Our aim in this section and the next is to prove that nevertheless, the possibility of *decomposing* an ideal in the above manner belongs to those properties of univariate polynomial rings that carry over to the multivariate case, and, moreover, even to any noetherian ring.

Recall that by our understanding, a ring is always a commutative ring with unity.

Definition 8.35 An ideal Q of a ring R is called **primary** if it is proper and satisfies the following condition: whenever $ab \in Q$ and $a \notin Q$, then $b^\nu \in Q$ for some $\nu \in \mathbb{N}$.

It is clear that every prime ideal is primary, but the following examples show that the converse is not true.

Example 8.36 If R is a UFD and $Q = p^\nu R$ for some irreducible $p \in R$ and some $\nu \in \mathbb{N}^+$, then, using the unique prime factor decomposition, it is easy to see that Q is primary. Here, Q is not prime unless $\nu = 1$.

If R is a PID, then $\{0\}$ and the proper ideals of the type $p^\nu R$ with irreducible p are the only primary ideals: as soon as the generator a of a proper non-trivial ideal has two non-associated prime factors, we can write $a = bc$ with non-units b and c that have no prime factor in common. Then b is not in aR because it is a proper divisor of a, and no power of c is a multiple of a because taking powers cannot make the missing prime factor(s) show up.

Example 8.37 For an example of a non-principal primary ideal, let K be a field, $R = K[X, Y]$, and $Q = \mathrm{Id}(X^2, XY, Y^2)$. An easy way to figure out what such an ideal looks like is as follows. Firstly, a set M of monomials is always a Gröbner basis (Corollary 5.49). Moreover, a reduction step modulo M does not change any terms other than the one it is eliminating, and so a polynomial is in $\mathrm{Id}(M)$ if and only if each of its terms is reducible modulo M. We see that here, Q consists of all those polynomials whose constant and linear coefficients are 0. Now if $fg \in Q$ and $f \notin Q$, then f has a non-zero linear or constant coefficient. For fg to be in Q, the constant coefficient of g must be 0, and so $g^2 \in Q$. Q is not prime since $XY \in Q$ but X, $Y \notin Q$. It is easy to make up similar examples using monomials of higher degree.

Note that in the examples above, there is a uniform bound for the exponent ν occurring in the definition of a primary ideal. We will see shortly that this is in fact true whenever R is noetherian.

Lemma 8.38 Let Q be a primary ideal of a ring R, and let $P = \mathrm{rad}(Q)$, i.e.,

$$P = \{\, a \in R \mid a^\nu \in Q \text{ for some } \nu \in \mathbb{N} \,\}.$$

Then P is a prime ideal of R with $Q \subseteq P$.

Proof We already know that $P = \mathrm{rad}(Q)$ is an ideal with $Q \subseteq P$. To see that P is prime in this case, we first note that P is proper since $1 \in P$ would imply $1 \in Q$. Now assume that $ab \in P$, and let $\nu \in \mathbb{N}$ with $(ab)^\nu = a^\nu b^\nu \in Q$. Then either $a^\nu \in Q$, in which case $a \in P$, or $(b^\nu)^\mu = b^{\nu+\mu} \in Q$ for some $\mu \in \mathbb{N}$, which implies that $b \in P$. □

If Q is a primary ideal of a ring R, then the ideal P of the lemma above is called the **associated prime ideal** of Q. Using the unique prime factor decomposition, one easily sees that in Example 8.36, the associated prime ideal of $Q = p^\nu R$ is $P = pR$. We see that for fixed irreducible $p \in R$, the primary ideals $p^\nu R$, which obviously form a chain, all share pR as their associated prime ideal. If R is a PID, then, in view of the remarks following Example 8.36, this is a complete description of primary ideals and associated primes. If R is not a PID, worse things can happen: two primary ideals may share the same associated prime although they are not contained one in the other either way.

Example 8.39 Consider the ideal $Q_1 = \mathrm{Id}(X^2, XY, Y^2)$ of Example 8.37, and let P be the associated prime. If $f \in R$ has a non-zero constant coefficient, then the same is true for every power of f, and so $f \notin P$; on the other hand, as soon as f has constant coefficient 0, then $f^2 \in Q_1$ and thus $f \in P$. We have proved that $P = \mathrm{Id}(X, Y)$. Now consider the ideals

$$Q_2 = \mathrm{Id}(X^2, Y) \quad \text{and} \quad Q_3 = \mathrm{Id}(X, Y^2).$$

It is not hard to prove by the same reasoning as for Q_1 that both Q_2 and Q_3 are primary with associated prime P, and we have the situation

$$Q_2 \supseteq Q_1 \subseteq Q_3$$

with both inclusions being proper, neither $Q_2 \subseteq Q_3$ nor $Q_3 \subseteq Q_2$, and everything properly contained in the common associated prime ideal P. However, if we consider $Q_4 = Q_2 \cap Q_3$, then it is easy to see that

$$Q_4 = \mathrm{Id}(X^2, Y^2),$$

and that this is once again a primary ideal with associated prime P. We thus have the situation

$$\begin{array}{ccccc} Q_2 & \supseteq & Q_1 & \subseteq & Q_3 \\ & & \rotatebox{90}{$\subseteq$} & & \\ & & Q_4 & & \end{array}$$

with all inclusions being proper, $Q_4 = Q_2 \cap Q_3$, and everything properly contained in the common associated prime P.

The next lemma shows that what happened in the example above was no coincidence.

Lemma 8.40 If Q_1 and Q_2 are primary ideals with the same associated prime ideal P, then $Q = Q_1 \cap Q_2$ is again primary with associated prime P.

Proof If $ab \in Q$ and $a \notin Q$, then $ab \in Q_i$ for $i = 1$, 2, and at least one of $a \notin Q_1$ and $a \notin Q_2$ holds, say $a \notin Q_1$. It follows that $b^\nu \in Q_1$ for some $\nu \in \mathbb{N}$ and so $b \in P$ because P is the associated prime of Q_1. Since P is also the associated prime of Q_2, we must have $b^\mu \in Q_2$ for some $\mu \in \mathbb{N}$, and we see that $b^{\max(\nu,\mu)} \in Q$. Now let P' be the associated prime of Q. If $a \in P$, then $a^\nu \in Q_1$ and $a^\mu \in Q_2$ for certain $\mu, \nu \in \mathbb{N}$, and thus $a^{\max(\nu,\mu)} \in Q$. This shows that $a \in P'$. Conversely, if $a \in P'$, then some power of a lies in Q and thus in both Q_1 and Q_2, which means that $a \in P$. We have proved that $P = P'$. □

It is clear from the definitions that if Q_1 and Q_2 are primary ideals with associated primes P_1 and P_2, respectively, then $Q_1 \subseteq Q_2$ implies that $P_1 \subseteq P_2$. Example 8.39 provides ample evidence that properness of the first inclusion does not imply properness of the second. Example 8.39 also shows that $P_1 \subseteq P_2$ does not in general imply $Q_1 \subseteq Q_2$: we had the situation $P_1 = P_2$ with no inclusion between Q_1 and Q_2. The next example shows how we can have a *proper* inclusion $P_1 \subseteq P_2$ between the associated primes without having $Q_1 \subseteq Q_2$. The example also shows that the intersection of Q_1 and Q_2 is then not necessarily primary.

Example 8.41 Let R again be a polynomial ring in the variables X and Y over a field. Consider the following primary ideals Q_1 and Q_2 with associated primes P_1 and P_2:

$$\begin{aligned} Q_1 &= \mathrm{Id}(X), & Q_2 &= \mathrm{Id}(X^2, Y), \\ P_1 &= \mathrm{Id}(X), & P_2 &= \mathrm{Id}(X, Y). \end{aligned}$$

Then clearly $P_1 \subseteq P_2$ but not $Q_1 \subseteq Q_2$. Q_1 consists of all polynomials that have a zero coefficient on 1 and all powers of Y, while Q_2 consists of all those that have a zero coefficient on 1 and X. The intersection of Q_1 and Q_2 thus consists of all polynomials where the coefficients of 1, X, and all powers of Y are 0, and we see that

$$Q_1 \cap Q_2 = \mathrm{Id}(X) \cap \mathrm{Id}(X^2, Y) = \mathrm{Id}(X^2, XY).$$

This latter ideal is not itself primary because it contains XY but neither X nor any power of Y.

Exercise 8.42 Make up an example of two non-prime primary ideals that are not contained one in the other either way but have associated primes one of which is contained in the other. (Hint: Raise the powers of X in the previous example.)

If K is a perfect field and I is a zero-dimensional ideal of a polynomial ring over K, then Proposition 8.26 states that there is a natural number μ such that the μth ideal power of $\mathrm{rad}(I)$ is contained in I. The next proposition shows that this is in fact true for every ideal in a noetherian ring. (The point of Proposition 8.26 was that such a natural number could be obtained by means of Gröbner basis computations and squarefree decompositions.)

Proposition 8.43 *Let R be a ring and I an ideal of R with radical J. If J has a finite basis, then there exists $\nu \in \mathbb{N}$ with* $\mathrm{Id}(J^\nu) \subseteq I$.

Proof Let $B = \{b_1, \ldots, b_m\}$ be a finite basis of J. Then there exist ν_1, $\ldots$, $\nu_m \in \mathbb{N}$ with $b_i^{\nu_i} \in I$ for $1 \leq i \leq m$. Set

$$\nu = 1 + \sum_{i=1}^{m} (\nu_i - 1).$$

According to Exercise 8.1, the ideal J^ν is generated by the set

$$B^\nu = \{\, a_1 \cdot \cdots \cdot a_\nu \mid a_j \in B \text{ for } 1 \leq j \leq \nu \,\}.$$

If we look at any element of this set, then by the choice of ν, there must be $1 \leq i \leq m$ such that b_i occurs more than $\nu_i - 1$ times among the a_j. We see that $B^\nu \subseteq I$ and hence $\mathrm{Id}(J^\nu) \subseteq I$. □

If I is an ideal with radical J and there exists $\nu \in \mathbb{N}$ with $\mathrm{Id}(J^\nu) \subseteq I$, then the least $\nu \in \mathbb{N}$ with this property is called the **exponent** of the ideal I. It is clear that $\mathrm{Id}(J^\mu) \subseteq I$ then holds for all $\mu \geq \nu$. Although its definition pertains to arbitrary ideals, the concept of the exponent is relevant mostly for primary ideals. Note that the exponent of a primary ideal—if it exists at all—is necessarily positive because primary ideals are proper by definition. If R is a noetherian ring, then by the proposition above, every primary ideal Q has an exponent ν. We then have the two inclusions

$$Q \subseteq P \quad \text{and} \quad \mathrm{Id}(P^\nu) \subseteq Q,$$

where P is the associated prime.

Corollary 8.44 *Let Q be a primary ideal of a ring R, and assume that Q has an exponent $\nu \in \mathbb{N}$. Then $ab \in Q$ and $a \notin Q$ implies $b^\nu \in Q$.*

Proof Let P be the associated prime ideal of Q. If $ab \in Q$ and $a \notin Q$, then $b^\mu \in Q$ for some $\mu \in \mathbb{N}$, hence $b \in P$ and so $b^\nu \in \mathrm{Id}(P^\nu) \subseteq Q$. □

The following corollary is immediate from Proposition 8.26 and the definition of the exponent.

Corollary 8.45 *If K is a perfect field and I is a zero-dimensional ideal of a polynomial ring over K, then the exponent of I is less than or equal to the univariate exponent of I.* □

We will demonstrate below that the inequality of the corollary may be strict.

If Q is an ideal of the form $p^\nu R$ with p irreducible in the UFD R, then we already know that the associated prime P is pR, and we see that here, the exponent of Q is ν, and we even have $\mathrm{Id}(P^\nu) = Q$. If R is even a PID, then every prime ideal is of the form pR, and the primary ideals are precisely the powers of prime ideals.

The general situation for primary ideals is much more unpleasant, even if we restrict ourselves to polynomial rings over fields. First of all, Example 8.39 shows that if ν is the exponent of Q, then the inclusion $\mathrm{Id}(P^\nu) \subseteq Q$ will in general be proper: here, the ideals Q_1–Q_3 all have exponent 2, and Q_4 has exponent 3. We have $\mathrm{Id}(P^2) = Q_1$, while the inclusions $\mathrm{Id}(P^2) \subseteq Q_i$ are proper for $i = 2$, 3, and so is the inclusion $\mathrm{Id}(P^3) \subseteq Q_4$. Note that Corollary 8.45 applies to Q_1–Q_4. We see that the univariate exponent is not in general equal to the exponent; it may come out greater than necessary for the inclusion $\mathrm{Id}(P^\nu) \subseteq Q$.

	Q_1	Q_2	Q_3	Q_4
univariate exponent	3	2	2	3
exponent	2	2	2	3

We will now show that in polynomial rings over a field the following holds: whenever the inclusion $\mathrm{Id}(P^\nu) \subseteq Q$ is proper for the exponent ν and associated prime P of Q (which may happen as we just saw), then Q is not equal to an ideal power of a prime ideal at all. To this end, we need the following lemma.

Lemma 8.46 Let $K[\underline{X}]$ be a polynomial ring over a field K. Then the following hold:

(i) If I is a proper ideal of $K[\underline{X}]$, then $\dim(I) = \dim(\mathrm{Id}(I^\nu))$ for all $\nu \in \mathbb{N}^+$.

(ii) If Q is a primary ideal of $K[\underline{X}]$ with associated prime P, then $\dim(Q) = \dim(P)$.

Proof (i) The inequality "$\le$" follows from the inclusion $\mathrm{Id}(I^\nu) \subseteq \mathrm{Id}(I)$. For the reverse inequality, we prove that every subset $\{U_1, \dots, U_r\}$ of $\{X_1, \dots, X_n\}$ that is independent modulo $\mathrm{Id}(I^\nu)$ is independent modulo I. If $\{U_1, \dots, U_r\}$ is dependent modulo I, then there exists a non-zero polynomial $f \in I \cap K[\underline{U}]$, which implies

$$f^\nu \in \mathrm{Id}(I^\nu) \cap K[\underline{U}],$$

and so $\{U_1, \dots, U_r\}$ is dependent modulo $\mathrm{Id}(I^\nu)$. Statement (ii) is proved in a very similar manner, using $Q \subseteq P$ and the definition of P as the associated prime. □

To prove the claim preceding the lemma, let Q be a primary ideal of $K[\underline{X}]$ and assume that the inclusion $\mathrm{Id}(P^\nu) \subseteq Q$ is proper for the exponent ν and associated prime P of Q. Suppose $Q = \mathrm{Id}(P_1^\mu)$ for some prime ideal P_1 and $\mu \in \mathbb{N}$. Then

$$\mathrm{Id}(P_1^\mu) = Q \subseteq P,$$

from which it is easy to conclude that $P_1 \subseteq P$. By the lemma above, the dimensions of all ideals in question agree, and so $P_1 = P$ by Lemma 7.57. But from the fact that $\mathrm{Id}(P^\nu)$ was properly contained in Q, it is easy to conclude that no power of P equals Q.

Finally, we demonstrate that it is not even true in general in multivariate polynomial rings that an ideal power of a prime ideal is primary.

Example 8.47 Let K be a field, $L = K(T)$ the rational function field in the variable T over K, and let G be the subset $\{f_1, f_2, f_3, f_4, f_5\}$ of $K[X, Y, Z]$, where

$$\begin{aligned} f_1 &= XZ - Y^2, \quad f_2 = X^3 - YZ, \quad f_3 = X^2Y - Z^2, \\ f_4 &= Y^5 - Z^4, \quad f_5 = XY^3 - Z^3. \end{aligned}$$

Let $\boldsymbol{a} = (T^3, T^4, T^5) \in L^3$. We claim that

$$\mathrm{Id}(G) = \{\, p \in K[X, Y, Z] \mid p(\boldsymbol{a}) = 0 \,\}. \tag{$*$}$$

The inclusion "$\subseteq$" is easily verified. Now let $p \in K[X, Y, Z]$ with $p(\boldsymbol{a}) = 0$. Let $\le$ be the lexicographical term order on $T(X, Y, Z)$, where $X \gg Y \gg Z$. It is true that G is a Gröbner basis w.r.t. $\le$, but even without knowing that, we may consider a normal form r of p modulo G. Then the terms of r must be among 1, X, X^2, Y, Y^2, Y^3, Y^4, XY, XY^2, and Y^iZ^j for $0 \le i \le 4$ and $j \in \mathbb{N}$. If we substitute $\boldsymbol{a}$ into each of these terms, then we obtain a set of powers of T with pairwise different exponents. Because of the inclusion "$\subseteq$" of $(*)$ and the fact that

$$r \equiv p \mod \mathrm{Id}(G),$$

we have $r(\boldsymbol{a}) = 0$, and we may conclude that $r = 0$. From the equality $(*)$ it now follows that $P = \mathrm{Id}(G)$ is prime: if a product of two polynomials vanishes at some point, then at least one of the factors must vanish there. The product

$$(X^5 - 3X^2YZ + XY^3 + Z^3)X = f_2^2 + f_1 f_3$$

lies in $\mathrm{Id}(P^2)$. The first factor is not in $\mathrm{Id}(P^2)$ because a polynomial in $\mathrm{Id}(P^2)$ cannot have a term of total degree less than 4, and no power of the second factor lies in $\mathrm{Id}(P^2)$ because X does not vanish at $\boldsymbol{a}$. We have proved that the ideal square $\mathrm{Id}(P^2)$ of the prime ideal P is not primary.

Note that the ideal P of the example above was not zero-dimensional. We will soon see that in polynomial rings over a field, ideal powers of *zero-dimensional* prime ideals are indeed always primary. This will come out of the following discussion of some properties that are shared by those primary ideals of a noetherian ring whose associated prime is maximal. Examples are all zero-dimensional primary ideals of a polynomial ring over a field: their associated primes are maximal because they are prime and zero-dimensional (Lemma 6.49 and Proposition 7.42). Note that we have actually encountered this situation: we have seen several primary ideals in $K[X, Y]$ whose associated prime is the zero-dimensional ideal $\mathrm{Id}(X, Y)$.

Lemma 8.48 Let R be a noetherian ring, I a proper ideal of R, and P a prime ideal of R. Then the following are equivalent:

(i) $I \subseteq P$, and P is the only prime ideal that contains I.

(ii) P is maximal, and I is primary with associated prime P.

(iii) P is maximal, and $\mathrm{Id}(P^\nu) \subseteq I$ for some $\nu \in \mathbb{N}$.

Proof (i)$\Longrightarrow$(ii): I is contained in some maximal ideal (Lemma 4.9) which must equal P because it is prime. Now let $ab \in I$ with $a \notin I$. Then the ideal $\mathrm{Id}(I, b)$ is proper, because otherwise we would have $1 = s + rb$ with $s \in I$ and $r \in R$ and thus

$$a = as + rab \in I.$$

The proper ideal $\mathrm{Id}(I, b)$ extends to a maximal ideal P' which must equal P because it contains I, and so $b \in P$. On the other hand, P equals $\mathrm{rad}(I)$ because the latter is the intersection of all prime ideals containing I. We see that $b^\nu \in I$ for some $\nu \in \mathbb{N}$.

(ii)$\Longrightarrow$(iii): Take for ν the exponent of I.

(iii)$\Longrightarrow$(i): I is contained in some maximal ideal, and hence in some prime ideal. Let P' be any such ideal, i.e., P' is a prime ideal with $I \subseteq P'$. We then have

$$P^\nu \subseteq I \subseteq P'.$$

One easily concludes that $P \subseteq P'$, and so $P = P'$ by the maximality of P. □

A primary ideal whose associated prime is maximal is called **monadic.** A monadic primary ideal Q thus satisfies the equivalent conditions of the lemma above with its associated prime taken for P. In the discussion preceding the lemma, we have proved the following.

Lemma 8.49 Every zero-dimensional primary ideal of a polynomial ring over a field is monadic. □

The following proposition is an immediate consequence of the implication "(iii)$\Longrightarrow$(ii)" of Lemma 8.48, applied with P maximal and $I = \mathrm{Id}(P^\nu)$.

Proposition 8.50 *Let R be a noetherian ring, P a maximal ideal of R, and $\nu \in \mathbb{N}^+$. Then $\mathrm{Id}(P^\nu)$ is a monadic primary ideal with associated prime P. In particular, all proper ideal powers of a zero-dimensional prime ideal P of a polynomial ring over a field are monadic primary ideals with associated prime P.* □

We can now summarize our knowledge about primary ideals and associated primes in polynomial rings over fields as follows. In the univariate case, the primary ideals come in descending chains $\{\mathrm{Id}(p^\nu)\}_{\nu\in\mathbb{N}^+}$, with irreducible p. The top element $\mathrm{Id}(p)$ is the associated prime of each element of the chain. Every primary ideal is thus an ideal power of a prime ideal, and every ideal power of a prime ideal is primary. In the multivariate case, the following pathologies are possible.

(i) Two primary ideals that are not contained one in the other either way may share the same associated prime. Their intersection is then again primary with the same associated prime.

(ii) Two primary ideals with no inclusion between them may also have associated primes one of which is properly contained in the other. We will see in the next section that in this case, the intersection of the two primary ideals is not primary.

(iii) Not every primary ideal is an ideal power of a prime ideal, and for dimensions greater than zero, not every ideal power of a prime ideal is primary.

8.5 Primary Decomposition in Noetherian Rings

An ideal I of a ring R is called **reducible** if there exist ideals I_1 and I_2 of R such that

$$I = I_1 \cap I_2 \quad \text{and} \quad I_1, I_2 \neq I,$$

irreducible otherwise. Obvious examples of irreducible ideals are all maximal ideals and R itself; for examples of reducible ones, read the first paragraph of the previous section.

Lemma 8.51 Let R be a noetherian ring. Then every ideal I of R is an intersection of finitely many irreducible ideals, i.e., there exist irreducible ideals $I_1, \ldots, I_r$ of R with

$$I = \bigcap_{i=1}^{r} I_i .$$

Proof Let N be the set of all those ideals of R that can not be written as an intersection of finitely many irreducible ideals, and assume for a contradiction that $N \neq \emptyset$. Every $I \in N$ is reducible, for otherwise I itself would be the desired intersection. This means that there exist ideals I_1 and I_2, both different from I, with $I = I_1 \cap I_2$. Now at least one of I_1 and I_2 must be in N, because otherwise we could easily arrive at a representation of I as an intersection of irreducible ideals by concatenating the representations of I_1 and I_2. What we have proved is that for each $I \in N$, there exists $J \in N$ such that I is properly contained in J. Now if we define, for $I \in N$,

$$N_I = \{ J \in N \mid I \subseteq J,\ I \neq J \},$$

then the axiom of choice provides a function

$$F : N \longrightarrow \bigcup_{I \in N} N_I \subseteq N$$

with $F(I) \in N_I$ for all $I \in N$. The sequence $\{I_n\}_{n \in \mathbb{N}}$ defined by $I_0 \in N$ arbitrary and $I_{n+1} = F(I_n)$ is now a strictly ascending chain of ideals in R, which, in view of Lemma 4.5, contradicts the fact that R is noetherian. □

Lemma 8.52 Every proper irreducible ideal of a noetherian ring R is primary.

Proof Suppose the proper ideal I of R is not primary. We show that I is reducible. Let $a, b \in R$ with $ab \in I$, $a \notin I$, and $b^\nu \notin I$ for all $\nu \in \mathbb{N}$. By Lemma 6.36, there exists $\mu \in \mathbb{N}$ with

$$I : b^\mu = I : b^{\mu+1}.$$

We claim that

$$I = \mathrm{Id}(I, a) \cap \mathrm{Id}(I, b^\mu),$$

and that I is properly contained in each of the two ideals on the right-hand side. This latter claim is immediate from the fact that a, $b^\mu \notin I$. To see that the equality above holds, we first note that the inclusion "$\subseteq$" is trivial. Now let c be an element of the intersection on the right. Then there exist s_1, $s_2 \in I$ and t_1, $t_2 \in R$ with

$$c = s_1 + t_1 a = s_2 + t_2 b^\mu .$$

It follows that

$$s_2 b + t_2 b^{\mu+1} = s_1 b + t_1 ab \in I,$$

and so $t_2 b^{\mu+1} \in I$, meaning that $t_2 \in I : b^{\mu+1}$. By the choice of μ, we have $t_2 \in I : b^{\mu}$, so that actually $t_2 b^{\mu} \in I$, and we see that $c \in I$. □

From the last two lemmas, we conclude immediately that every proper ideal of a noetherian ring R has a representation as a finite intersection of primary ideals. In order to have some sort of uniqueness property, however, we must try to obtain a representation that is in some sense minimal. So assume that I is an ideal of the noetherian ring R, and let

$$I = \bigcap_{i=1}^{r} Q_i \qquad (*)$$

be any representation of I as an intersection of primary ideals. First of all, it could be that there is an ideal Q_i occurring in the intersection that contains the intersection of the rest and may thus be dropped. To see how this can happen in a non-trivial way, consider the intersection

$$\mathrm{Id}(X^2, XY) = \mathrm{Id}(X) \cap \mathrm{Id}(X^2, Y)$$

in $K[X,Y]$ of Example 8.41. If we throw in the primary ideal $\mathrm{Id}(Y^2)$, then we obtain

$$\mathrm{Id}(XY^2) = \mathrm{Id}(Y^2) \cap \mathrm{Id}(X) \cap \mathrm{Id}(X^2, Y),$$

and we see that $\mathrm{Id}(X^2, Y)$ is now superfluous in the intersection. In order to "minimize" the representation $(*)$, we may thus, in a first step, do the following as long as possible: pick an ideal occurring in the intersection that contains the intersection of the rest, and drop it.

This done, we now look for pairs Q_j, Q_k with $j \neq k$ of ideals occurring in the intersection that have the same associated prime ideal. To see that this may still be the case, simply look at the intersection

$$\mathrm{Id}(X^2, Y^2) = \mathrm{Id}(X^2, Y) \cap \mathrm{Id}(X, Y^2)$$

of Example 8.39. According to Lemma 8.40, the intersection $Q_i \cap Q_j$ is now again primary with the same associated prime ideal. We may thus, as long as it is possible, find such pairs in the intersection $(*)$ and replace them by their intersection. What we obtain is an intersection in which all primary ideals and all associated primes are pairwise different.

Two questions arise naturally at this point. Firstly, is it possible that after performing the second step, we can go back to the first one and drop a redundant primary ideal from the intersection? The answer is no for a rather trivial set-theoretic reason. Assume for a contradiction that after replacing a pair Q_j, Q_k of primary ideals with associated prime P by the intersection $Q = Q_j \cap Q_k$, there is an ideal occurring in the intersection

$$Q \cap \bigcap_{\substack{i=1 \\ i \neq j,k}}^{r} Q_i$$

that is redundant, i.e., contains the intersection of the rest. If this ideal is Q_i for some $i \neq j$, k, then it was redundant before the replacement, and if it is Q, then both Q_j and Q_k were redundant before the replacement.

The second, more interesting question is whether after performing the second step, it is possible to have a finite set of ideals occurring in the intersection $(*)$ whose intersection is again primary, so that a replacement similar to the ones of step two could be made. The following lemma says that the answer is no.

Lemma 8.53 Let $2 \leq r \in \mathbb{N}$, and let $Q_1, \ldots, Q_r$ be primary ideals of a ring R such that the associated prime ideals of the Q_i are pairwise different, and none of the Q_i contains the intersection of the rest. Then the intersection

$$I = \bigcap_{i=1}^{r} Q_i$$

is not primary.

Proof For $1 \leq i \leq r$, we let P_i be the associated prime of Q_i. Assume w.l.o.g. that P_1 is minimal w.r.t. inclusion among the P_i. Then there exist $a_2, \ldots, a_r \in R$ with

$$a_i \in P_i \setminus P_1 \quad \text{for} \quad 2 \leq i \leq r.$$

Some power of a_i lies in Q_i for $2 \leq i \leq r$, meaning there are $\mu_i \in \mathbb{N}$ with $a_i^{\mu_i} \in Q_i$ for $2 \leq i \leq r$. The inclusion $I \subseteq Q_1$ must be proper since otherwise every Q_i other than Q_1 would contain the intersection of the rest. Let $a \in Q_1 \setminus I$, and consider the product

$$a \cdot (a_2^{\mu_2} \cdot \cdots \cdot a_r^{\mu_r}) \in \bigcap_{i=1}^{r} Q_i = I.$$

The first factor a is not in I, and no power of the second can be in I either, for otherwise there would be $\nu \in \mathbb{N}$ with

$$(a_2^{\mu_2} \cdot \cdots \cdot a_r^{\mu_r})^{\nu} \in I \subseteq Q_1 \subseteq P_1,$$

and so $a_i \in P_1$ for some $2 \leq i \leq r$, a contradiction. □

The lemma above is actually a step towards a uniqueness theorem. For the moment, we note that we have proved the following existence theorem.

Theorem 8.54 (Primary Decomposition—Existence) *If R is a noetherian ring and I is a proper ideal of R, then there exist primary ideals $Q_1, \ldots, Q_r$ of R such that*

(i) $I = \bigcap_{i=1}^{r} Q_i$,

(ii) none of the Q_i contains the intersection of the rest, and

(iii) the associated prime ideals of the Q_i are pairwise different. □

Representations as described in the theorem are called **primary decompositions**. Any primary ideal that occurs in a primary decomposition of an ideal I is called a **primary component** of I. A trivial but important observation that should be kept in mind is that whenever Q is a primary component of I, then I is contained in Q, which in turn is contained in its associated prime P:

$$I = \bigcap_{i=1}^{r} Q_i \subseteq Q_j \subseteq P_j \qquad (1 \leq j \leq r).$$

If R is a PID, then it is clear from the uniqueness of the prime factor decomposition that the primary decomposition of a proper non-trivial ideal aR is—up to order—uniquely determined to be the decomposition

$$aR = \bigcap_{i=1}^{r} p_i^{\nu_i} R$$

as described at the beginning of the previous section. In a multivariate polynomial ring, things are once again less smooth. To see how, consider the decomposition

$$\mathrm{Id}(X^2, XY) = \mathrm{Id}(X) \cap \mathrm{Id}(X^2, Y)$$

of Example 8.41. The reader should have no trouble by now verifying that

$$\mathrm{Id}(X^2, XY) = \mathrm{Id}(X) \cap \mathrm{Id}(X^2, XY, Y^2)$$

is another primary decomposition of $\mathrm{Id}(X^2, XY)$. Note, however, that the number of primary ideals is 2 and the associated primes are $\mathrm{Id}(X)$ and $\mathrm{Id}(X, Y)$ for both representations. The following first uniqueness theorem says that this was no coincidence. Note that the definition of primary decompositions was not tied to R being noetherian.

Theorem 8.55 (Primary Decomposition—Uniqueness 1) *Any two primary decompositions of an ideal I of a ring R have the same number of components and the same set of associated primes.*

Proof Let $Q_1, \ldots, Q_r, Q'_1, \ldots, Q'_s$ be primary ideals of R with associated primes $P_1, \ldots, P_r, P'_1, \ldots, P'_s$, respectively, and suppose

$$I = \bigcap_{i=1}^{r} Q_i = \bigcap_{j=1}^{s} Q'_j$$

are primary decompositions. We proceed by induction on r. If $r = 1$, then Q_1 is primary, and it follows that $s = 1$ and thus $Q_1 = Q'_1$ and $P_1 = P'_1$, because otherwise we would be contradicting Lemma 8.53.

Now let $r > 1$. Among the finitely many ideals $P_1, \ldots, P_r, P_1', \ldots, P_s'$, there must be one that is maximal w.r.t. inclusion. We claim that this ideal must occur on both sides, i.e., it must be among the P_i as well as among the P_j'. Assume w.l.o.g. that P_1 is the ideal in question. We are thus claiming that P_1 must equal one of the P_j'. Assume for a contradiction that this were not so. What we we are going to show is that then Q_1 can be dropped from the first intersection, contradicting one of the properties of a primary decomposition. In other words, we are claiming that now

$$\bigcap_{i=2}^{r} Q_i = \bigcap_{j=1}^{s} Q_j' .$$

The inclusion "$\supseteq$" being trivial, let a be an element of the left-hand side, and let $1 \leq j \leq s$. By our assumption on P_1 we can find b_j such that $b_j \in P_1 \setminus P_j'$. Because of $b_j \in P_1$, there exists $\nu \in \mathbb{N}$ with $b_j^\nu \in Q_1$, whence

$$ab_j^\nu \in \bigcap_{i=1}^{r} Q_i \subseteq Q_j'.$$

We may now conclude that $a \in Q_j'$, for otherwise some power of b_j^ν and thus of b_j would have to lie in Q_j', which is not the case because of $b_j \notin P_j'$.

We may now assume w.l.o.g. that $P_1 = P_1'$, and we claim that

$$\bigcap_{i=2}^{r} Q_i = \bigcap_{j=2}^{s} Q_j' ,$$

which finishes the proof in view of the induction hypothesis. Because of the symmetry of the problem, it suffices to prove one inclusion. Let a be an element of the left-hand side, and let $2 \leq j \leq s$. We know that $P_1' \neq P_j'$ for $2 \leq j \leq s$, and that $P_1 = P_1'$ is maximal w.r.t. inclusion among all prime ideals occurring. It follows that there exists $b_j \in P_1 \setminus P_j'$, and we may now argue literally as before to conclude that $a \in Q_j'$. □

On the basis of the first uniqueness theorem, we may now state the following definition. A primary component of an ideal I is called **isolated** if its associated prime does not properly contain the associated prime of some other primary component of I, **embedded** otherwise. The associated prime ideal of an embedded primary is sometimes also called isolated, and the same goes for the embedded case. The choice of the terminology reflects a geometric point of view: if R is a polynomial ring over a field, then the inclusion $P_1 \subseteq P_2$ between prime ideals is equivalent to the variety of P_2 being contained in the variety of P_1 (cf. the alternate version of the Hilbert Nullstellensatz on p. 313).

We see that in the decomposition

$$\mathrm{Id}(X^2, XY) = \mathrm{Id}(X) \cap \mathrm{Id}(X^2, Y)$$

which we have mentioned several times before, the primary component $\mathrm{Id}(X)$ is isolated, while $\mathrm{Id}(X^2, Y)$ is embedded: the associated primes are $\mathrm{Id}(X)$ and $\mathrm{Id}(X, Y)$, respectively. To see that there may, at the same time, also be isolated primary components whose associated primes do not have an embedded one above them (as is the case with $\mathrm{Id}(X)$ here), consider the example

$$\mathrm{Id}(X^2Z^2, XYZ^2) = \mathrm{Id}(X) \cap \mathrm{Id}(X^2, Y) \cap \mathrm{Id}(Z^2),$$

where the associated primes are $\mathrm{Id}(X)$, $\mathrm{Id}(X, Y)$, and $\mathrm{Id}(Z)$.

In the example

$$\mathrm{Id}(X^2, XY) = \mathrm{Id}(X) \cap \mathrm{Id}(X^2, Y) = \mathrm{Id}(X) \cap \mathrm{Id}(X^2, XY, Y^2)$$

of a non-unique primary decomposition that we gave preceding the last theorem, the primary component that was modified was embedded. The second uniqueness theorem states that just as the example suggests, only embedded components can be modified in a primary decomposition.

Theorem 8.56 (PRIMARY DECOMPOSITION—UNIQUENESS 2) *Let R be a ring, I an ideal of R which posesses a primary decomposition, and P the associated prime of some isolated primary component of I. Then there exists a primary ideal Q_P of R such that in every primary decomposition of I, the primary ideal whose associated prime is P equals Q_P.*

Proof Let $I = Q \cap \bigcap_{i=1}^r Q_i$ be a primary decomposition, and assume that Q is isolated with associated prime P. The claim is trivial if $r = 0$. Otherwise, we set $M = R \setminus P$, and

$$I_M = \{\, a \in R \mid am \in I \text{ for some } m \in M \,\}.$$

We claim that necessarily $Q = I_M$, which means that Q is uniquely determined by P and I. Let $a \in Q$. Since P and the P_i are pairwise different and P is isolated, we may apply Lemma 8.3 with $I_i = Q_i$ and $J_i = P_i$ to obtain an element

$$m \in (Q_1 \cdot \cdots \cdot Q_r) \setminus P = (Q_1 \cdot \cdots \cdot Q_r) \cap M.$$

It follows that

$$am \in Q \cap \bigcap_{i=1}^{r} Q_i = I,$$

and we see that $a \in I_M$. Conversely, suppose $a \in I_M$. Let $m \in M$ with $am \in I$. Then $am \in Q$, but no power of m can be in Q because m would then have to be in P. It follows that $a \in Q$. □

Exercise 8.57 Use the primary decomposition in PID's as described at the beginning of the previous section to visualize the proof of the proposition above.

Exercise 8.58 Let K be a field and I a proper ideal of $K[X_1, \ldots, X_n]$. Show that the dimension of I equals the maximum of the dimensions of the associated primes of the primary components of I.

We close this section with an important proposition concerning monadic primary components. Note that a monadic primary component of an ideal need not be isolated: the component $\mathrm{Id}(X^2, Y)$ of

$$\mathrm{Id}(X^2, XY) = \mathrm{Id}(X) \cap \mathrm{Id}(X^2, Y)$$

is monadic because its associated prime is the maximal ideal $\mathrm{Id}(X, Y)$, and it is also embedded because its associated prime $\mathrm{Id}(X, Y)$ contains the associated prime $\mathrm{Id}(X)$ of the primary component $\mathrm{Id}(X)$. All we can say is that if a primary component is monadic, then its associated prime cannot have an embedded one above it, as is the case with $\mathrm{Id}(X)$ in the above example.

Proposition 8.59 *Let R be a noetherian ring, I a proper ideal of R, and Q an isolated monadic primary component of I with associated prime P and exponent $\nu \in \mathbb{N}$. Then*

$$Q = I + \mathrm{Id}(P^\mu)$$

for all $\mu \geq \nu$, and the exponent ν is in fact the least natural number with this property.

Proof The inclusion "$\supseteq$" is immediate from the inclusions

$$I \subseteq Q \quad \text{and} \quad \mathrm{Id}(P^\mu) \subseteq \mathrm{Id}(P^\nu) \subseteq Q.$$

For the reverse inclusion, we let

$$I = Q \cap \bigcap_{i=1}^{r} Q_i$$

be a primary decomposition of I. (Q must in fact occur in *every* primary decomposition of I because it is isolated.) If $r = 0$, then $I = Q$ and the claim is trivial. Otherwise, we first note that by "(iii)$\Longrightarrow$(ii)" of Lemma 8.48, $I + \mathrm{Id}(P^\mu)$ is primary with associated prime P. Since P and the P_i are pairwise different and P is isolated, Lemma 8.3 applies with $I_i = Q_i$ and $J_i = P_i$ and provides an element

$$b \in (Q_1 \cdot \cdots \cdot Q_r) \setminus P.$$

Now let $a \in Q$. Then

$$ab \in Q \cap \bigcap_{i=1}^{r} Q_i = I \subseteq I + \mathrm{Id}(P^\mu),$$

but no power of b is in $I + \mathrm{Id}(P^\mu)$ because of $b \notin P$, and so $a \in I + \mathrm{Id}(P^\mu)$. To see that ν is minimal with the property that we have just proved, let $\mu < \nu$. By the definition of ν as the exponent of Q, there exists

$$a \in \mathrm{Id}(P^\mu) \setminus Q \subseteq \big(I + \mathrm{Id}(P^\mu)\big) \setminus Q. \quad \square$$

A good example to visualize the statement of the proposition is once again given by the decomposition

$$I = aR = \bigcap_{i=1}^{r} p_i^{\nu_i} R$$

of a proper non-trivial ideal in a PID as described at the beginning of the previous section. Here, every primary component is monadic and isolated because of the equivalence of primeness and maximality for ideals of a PID. According to the proposition, we must have

$$p_j^{\nu_j} R = aR + p_j^{\mu} R$$

whenever $1 \leq j \leq r$ and $\mu \geq \nu_j$. It is easy to see from an elementary point of view that this is true: the ideal on the right-hand side is generated by the gcd of a and p_j^{μ}, which obviously equals $p_j^{\nu_j}$.

For more instances of the proposition, turn to the next section.

8.6 Primary Decomposition of Zero-Dimensional Ideals

Throughout this section, K will be a field with algebraic closure $\overline{K}$, and $K[\underline{X}] = K[X_1, \ldots, X_n]$. Our ultimate goal in this section is to show how one may compute the primary decomposition of a zero-dimensional ideal of $K[\underline{X}]$ for certain K.

If I is a zero-dimensional ideal of $K[\underline{X}]$, then every prime ideal containing I is zero-dimensional too and thus maximal. As we have mentioned before, it follows that the associated primes of the primary components of I are all maximal. We see that here, every primary component of I is monadic. For the same reason, I cannot have any embedded primary components, so that Proposition 8.59 does in fact apply to every primary component of I. By the second uniqueness theorem on the primary decomposition, we may also conclude that primary decompositions of I are uniquely determined up to the order of the components. We may thus speak of *the* primary decomposition of I. The next lemma collects this and some related information.

Lemma 8.60 Let I be a zero-dimensional ideal of $K[\underline{X}]$. Then the following hold:

(i) Any two primary decompositions of I differ only by the order of the components.

(ii) Every primary component of I is isolated and monadic.

(iii) If Q is a primary component of I with associated prime P, then

$$Q = I + \mathrm{Id}(P^{\mu}),$$

where μ is the exponent of Q. If in addition, K is a perfect field, then

$$Q = I + \mathrm{Id}(P^{\sigma}),$$

where σ is the univariate exponent of I.

(iv) Every prime ideal with $I \subseteq P$ is the associated prime of some primary component of I.

(v) The primary components of $\mathrm{rad}(I)$ are precisely the associated primes of the primary components of I.

(vi) If I is itself radical, then its primary components are precisely the prime ideals which it is contained in.

Proof We have proved (i) and (ii) in the discussion preceding the lemma.

(iii) The first statement is immediate from (ii) and Proposition 8.59. Now assume that K is perfect. Let μ be the exponent of Q and ρ its univariate exponent. From $I \subseteq Q$ it follows that for $1 \leq i \leq n$, the monic generator of $Q \cap K[X_i]$ divides the monic generator of $I \cap K[X_i]$, and one easily concludes that $\rho \leq \sigma$. Corollary 8.45 says that $\mu \leq \rho$, and the claim is now obvious from Proposition 8.59 together with Corollary 8.45.

(iv) Let $I = \bigcap_{i=1}^{r} Q_i$ be the primary decomposition of I, and let P_1, ..., P_r be the associated primes of Q_1, ..., Q_r, respectively. Assume for a contradiciton that there exists a prime ideal P with $I \subseteq P$ that is not among the P_i. Since P and the P_i are maximal ideals, it follows that P does not contain any one of the P_i, and Lemma 8.3 applied with $I_i = Q_i$ and $J_i = P_i$ provides an element

$$a \in (Q_1 \cdot \cdots \cdot Q_r) \setminus P \subseteq (Q_1 \cap \cdots \cap Q_r) \setminus P = I \setminus P,$$

a contradiction.

(v) We first recall that the radical of an ideal is always the intersection of all prime ideals that contain the ideal. We may conclude that here,

$$\mathrm{rad}(I) = P_1 \cap \cdots \cap P_r \qquad (*)$$

where P_1, ..., P_r are the different prime ideals that contain I. By (iv) above, these are the associated primes of the primary components of I. We claim that $(*)$ is the primary decomposition of $\mathrm{rad}(I)$. The ideals occurring in the intersection are pairwise different and primary. Moreover, they are identical with their associated primes, and Lemma 8.4 now tells us that none of them contains the intersection of the rest.

(vi) This immediate from (iv) and (v). □

Exercise 8.61 (i) Prove (vi) of the last lemma directly from Lemma 8.4.

(ii) Show that the primary components of a zero-dimensional ideal of $K[\underline{X}]$ are pairwise comaximal.

We begin our investigation of how to compute primary decompositions of zero-dimensional ideals by discussing a special case, namely, the one where the zeroes of the ideal are in the ground field. For a univariate ideal I with generator f, this means that the irreducible factors of f are all linear, so that there is a one-to-one correspondence between zeroes and primary components of I. We are going to show that this remains true for multivariate ideals. Recall that for $\boldsymbol{a} = (a_1, \ldots, a_n) \in K^n$, the corresponding *vanishing ideal* was defined as

$$I_{\boldsymbol{a}} = \mathrm{Id}(X_1 - a_1, \ldots, X_n - a_n).$$

Lemma 8.62 (i) If $\boldsymbol{a} \in K^n$, then $I_{\boldsymbol{a}}$ is a maximal ideal of $K[\underline{X}]$ which contains every ideal of which $\boldsymbol{a}$ is a zero.

(ii) If K is algebraically closed, then every maximal ideal of $K[\underline{X}]$ is of the form $I_{\boldsymbol{a}}$.

Proof (i) We have proved in Lemma 6.28 (iii) that for $\boldsymbol{a} \in K^n$, the ideal $I_{\boldsymbol{a}}$ consists of all $f \in K[\underline{X}]$ that vanish at $\boldsymbol{a}$. Now a product of two polynomials in $K[\underline{X}]$ vanishes at a given point iff at least one of the factors does, and we see that $I_{\boldsymbol{a}}$ is prime. Clearly, it is also zero-dimensional, and thus it is maximal. The second part of the claim is obvious from the fact that $I_{\boldsymbol{a}}$ consists of all those polynomials that vanish at $\boldsymbol{a}$.

(ii) Assume that K is algebraically closed, and let M be a maximal ideal of $K[\underline{X}]$. Being proper, M has a zero $\boldsymbol{a} \in K^n$, and so $M \subseteq I_{\boldsymbol{a}}$. The claim now follows from the maximality of M. □

Proposition 8.63 *Let I be a zero-dimensional ideal of $K[\underline{X}]$, and assume that $\boldsymbol{a}$ is a zero of I in K^n. Then $I_{\boldsymbol{a}}$ is the associated prime of some primary component Q of I. If the different zeroes $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r$ of I in $\overline{K}^n$ are all in K^n, then the associated primes of the primary components of I are precisely $I_{\boldsymbol{a}_1}, \ldots, I_{\boldsymbol{a}_r}$.*

Proof To prove the first statement, it suffices by Lemma 8.60 (iv) to show that $I_{\boldsymbol{a}}$ is a prime ideal of $K[\underline{X}]$ that contains I, and this is stated in (i) of the previous lemma. Now suppose the zeroes of I are all in K^n, and let Q be any primary component of I with associated prime P. Then P has a zero $\boldsymbol{a}$ in $\overline{K}^n$, which must also be a zero of I. We may conclude that

$P \subseteq I_{\boldsymbol{a}}$, and maximality of P implies that we must actually have equality. □

Combining the last proposition with Lemma 8.60 (iii), we see that the primary component Q of I that corresponds to the zero $\boldsymbol{a} \in K^n$ is given by $I + \mathrm{Id}(I_{\boldsymbol{a}}^{\mu})$, where μ is the exponent of Q. Moreover, if K is perfect, then we may replace μ by the univariate exponent of I, which can be effectively found in case K is computable and allows squarefree decomposition of univariate polynomials. We will now show that in this special case, where Q corresponds to a zero of I in K^n, we may use a possibly lower exponent whose computation requires no more than univariate polynomial division.

Lemma 8.64 Let I and $\boldsymbol{a} = (a_1, \ldots, a_n) \in K^n$ be as in the proposition, and assume that K is perfect. For $1 \leq i \leq n$, let f_i be the unique monic generator of $I \cap K[X_i]$. Set

$$\nu = 1 + \sum_{i=1}^{n} (\nu_i - 1),$$

where for $1 \leq i \leq n$, the number ν_i is the multiplicity of a_i as a zero of f_i. Then the primary component Q of I whose associated prime is $I_{\boldsymbol{a}}$ equals $I + \mathrm{Id}(I_{\boldsymbol{a}}^{\nu})$.

Proof In view of Proposition 8.59, it suffices to prove that the exponent μ of Q is less than or equal to ν. We claim that the unique monic generator of $Q \cap K[X_i]$ is of the form $(X_i - a_i)^{\rho_i}$. Indeed, $I \cap K[X_i]$ contains the polynomial f_i that is divided by $(X_i - a_i)^{\nu_i}$, and $I_{\boldsymbol{a}}^{\mu}$ contains the polynomial $(X_i - a_i)^{\mu}$. Since $Q = I + \mathrm{Id}(I_{\boldsymbol{a}}^{\mu})$ contains the gcd of f_i and $(X_i - a_i)^{\mu}$, the unique monic generator of $Q \cap K[X_i]$ must be of the desired form. From $I \subseteq Q$ we conclude that $(X_i - a_i)^{\rho_i} \mid f_i$ and thus $\rho_i \leq \nu_i$. But the exponent μ of Q is less than or equal to its univariate exponent ρ, and we get

$$\mu \leq \rho = 1 + \sum_{i=1}^{n} (\rho_i - 1) \leq 1 + \sum_{i=1}^{n} (\nu_i - 1) = \nu. \quad \square$$

Evaluating the lemma for the case of a univariate ideal, we see that the number ν of the lemma is in a sense a measure for the multiplicity of $\boldsymbol{a}$ as a zero of I. The lemma and the proposition preceding it show that over perfect computable fields, the primary decomposition of zero-dimensional ideals can be computed in the (admittedly unlikely) event that the zeroes of I are known and are all in K^n.

Exercise 8.65 Find the primary decomposition in $\mathbb{R}[X, Y]$ of the ideal

$$\mathrm{Id}\big(\{Y^2 - 2, YX^2 + X^3, X^4 - 2X^2\}\big).$$

It should be noted that here, in contrast to the univariate case, specifying a zero and its "multiplicity" does not determine the corresponding primary component. As an easy example, let us once again consider the two ideals

$$I_1 = \mathrm{Id}(\{X^2, Y^2\}) \quad \text{and} \quad I_2 = (\{X^2, XY, Y^2\})$$

of $\mathbb{Q}[X, Y]$. Each of these has but one zero which lies in $\mathbb{Q}$. It follows that there can be only one primary component, and we have confirmed a result of Example 8.39: both I_1 and I_2 are primary. In both cases, $(0, 0)$ is a zero with "multiplicity" 3, but the corresponding primary components are I_1 and I_2 themselves, which are clearly different from each other. The associated prime is $I_{(0,0)} = \mathrm{Id}(X, Y)$, and we see that the "multipicity" ν that the previous lemma uses as the exponent may still be greater than necessary:

$$I_1 = I_1 + I_{(0,0)}^3 \neq I_1 + I_{(0,0)}^2,$$

while

$$I_2 = I_2 + I_{(0,0)}^3 = I_2 + I_{(0,0)}^2 = I_{(0,0)}^2.$$

We now turn to the general problem of computing primary decompositions of zero-dimensional ideals. The results of the last four sections suggest that zero-dimensional ideals tend to be nicer than arbitrary polynomial ideals: they often behave very much like univariate polynomial ideals. As an example, recall that if K is perfect, then the radical of a zero-dimensional ideal is computed by throwing in the squarefree parts of the univariate polynomials of minimal degree, just like the radical of a univariate polynomial ideal is found by adding in the squarefree part of its generator. This suggests that the primary decomposition of a zero-dimensional ideal can be found by doing something with the univariate polynomials that resembles the construction of the primary decomposition of a univariate ideal as described at the beginning of Section 8.4. The following example shows that there cannot be anything obvious along those lines in general.

Example 8.66 Let $K = \mathbb{Q}$ and $n = 2$. Consider the ideal I that is generated by $G = \{g_1, g_2\}$, where

$$g_1 = X_1^2 - 2 \quad \text{and} \quad g_2 = X_2^2 - 2.$$

G is a Gröbner basis w.r.t. every term order because the head terms are disjoint. We see that the univariate polynomials in I are g_1 and g_2, both of which are irreducible over $\mathbb{Q}$, and so there is nothing that could be done in the way of factoring univariate polynomials. According to the conjecture above, we would have to conclude that I is primary. We claim that this is not true. We first note that

$$(X_1 + X_2)(X_1 - X_2) = X_1^2 - X_2^2 = g_1 - g_2 \in I.$$

The first factor is not in I because it is in normal form w.r.t. G. If any power of $X_1 - X_2$ were in I, then every zero (z_1, z_2) of I in an extension field of $\mathbb{Q}$ would have to satisfy $z_1 = z_2$. This is not true because the zeroes of I are precisely

$$(\sqrt{2}, \sqrt{2}),\ (\sqrt{2}, -\sqrt{2}),\ (-\sqrt{2}, \sqrt{2}),\ \text{and}\ (-\sqrt{2}, -\sqrt{2}).$$

We are actually in a position to give the primary decomposition of I. In Exercise 6.22, you proved that

$$I = \mathrm{Id}(X_1^2 - 2, X_1 + X_2) \cap \mathrm{Id}(X_1^2 - 2, X_1 - X_2).$$

Using the argument of Example 7.45, one easily verifies that the ideals ocurring on the right-hand side are both prime. The two are clearly different, and we see that I has exactly two primary components, each of which is actually prime.

If $\boldsymbol{a} = (a_1, \ldots, a_n) \in \overline{K}^n$ and $1 \leq i \leq n$, then we will, rather obviously, call a_i the *X_i-component* of $\boldsymbol{a}$. Note that in the example above, neither the X_1- nor the X_2-components of the zeroes of I in $\overline{K}$ are pairwise different.

Definition 8.67 An ideal I of $K[\underline{X}]$ is said to be in **normal position** w.r.t. X_i if the X_i-components of the zeroes of I in $\overline{K}^n$ are pairwise different.

We have already encountered an example of an ideal in normal position in Example 8.29, where I is in normal position w.r.t. each variable. An easy observation which will be used repeatedly is as follows. If I and J are ideals of $K[\underline{X}]$ with $I \subseteq J$, then all zeroes of J are zeroes of I; so if I is in normal position w.r.t. some X_i, then so is J.

The next lemma and proposition show that for ideals that are in normal position w.r.t. some variable, the computation of the primary decomposition bears a strong analogy to the univariate case.

Lemma 8.68 Let I be a zero-dimensional ideal of $K[\underline{X}]$ which is in normal position w.r.t. X_1. Assume further that $I \cap K[X_1]$ contains a polynomial of the form p^ν with p irreducible and $\nu \in \mathbb{N}$. Then I is primary.

Proof We may of course assume that p is monic. Let P_1 be a prime ideal of $K[\underline{X}]$ that contains I. We verify condition (i) of Lemma 8.48. Suppose P_2 is another prime ideal with $I \subseteq P_2$. Then p is in both P_1 and P_2. Being irreducible, p must actually be the unique monic generator of

$$P_1 \cap K[X_1] \quad \text{and} \quad P_2 \cap K[X_1],$$

and hence it is the univariate polynomial in the prime bases of P_1 and P_2, respectively, whenever X_1 is the lexicographically least variable. We claim that P_1 and P_2 have the same zeroes in $\overline{K}^n$. Assume that

$$\boldsymbol{z} = (z_1, \ldots, z_n) \in \overline{K}^n$$

is a zero of P_1. Then $p(z_1) = 0$, and so (z_1) extends to a zero

$$z' = (z_1, z_2' \dots, z_n') \in \overline{K}^n$$

of P_2 by Lemma 7.51 (i). Both z and z' are zeroes of I, and so they must be equal because I is in normal position w.r.t. X_1. We have proved that z is a zero of P_2. Symmetry of the problem implies that every zero of P_2 in $\overline{K}^n$ is a zero of P_1. We may now conclude from Corollary 7.41 or Lemma 7.56 that $P_1 = P_2$. □

Proposition 8.69 *Let I be a zero-dimensional ideal of $K[\underline{X}]$ which is in normal position w.r.t. X_1. Let f be the unique monic univariate polynomial in $I \cap K[X_1]$, and let*

$$f = p_1^{\nu_1} \cdot \dots \cdot p_r^{\nu_r}$$

with $p_1, \dots, p_r \in K[X_1]$ irreducible and pairwise non-associated. Then the primary decomposition of I is given by

$$I = \bigcap_{i=1}^{r} \mathrm{Id}(I, p_i^{\nu_i}).$$

Proof It is immediate from Lemma 8.5 that I is equal to the indicated intersection. We claim that the ideals occurring in the intersection are all proper. Assume for a contradiction that $1 \in \mathrm{Id}(I, p_j^{\nu_j})$ for some $1 \le j \le r$. It is easy to see that then

$$\prod_{\substack{i=1 \\ i \ne j}}^{r} p_i^{\nu_i} \in \mathrm{Id}(I, f) = I,$$

contradicting the choice of f. (Note that this part of the proof does not depend on I being in normal position.) Containing I, the ideals occurring in the intersection are clearly zero-dimensional and in normal position w.r.t. X_1. The previous lemma now tells us that they are all primary. Assume for a contradiction that one of them, say $\mathrm{Id}(I, p_j^{\nu_j})$, contains the intersection of the rest. Then we have

$$p_j^{\nu_j} \in \mathrm{Id}(I, p_j^{\nu_j}) \quad \text{and} \quad \prod_{\substack{i=1 \\ i \ne j}}^{r} p_i^{\nu_i} \in \mathrm{Id}(I, p_j^{\nu_j}).$$

The ideal $\mathrm{Id}(I, p_j^{\nu_j})$ thus contains two univariate polynomials that are relatively prime, and so $1 \in \mathrm{Id}(I, p_j^{\nu_j})$, contradicting an earlier conclusion.

It remains to show that the associated prime ideals P_i of the ideals $\mathrm{Id}(I, p_i^{\nu_i})$ are pairwise different for $1 \le i \le r$. From $P_i = P_j$ with $1 \le i < j \le n$ it would follow that $p_i, p_j \in P_i$, and thus $1 \in P_i$ because p_i and p_j are univariate and relatively prime. Again, we obtain the contradiction $1 \in \mathrm{Id}(I, p_i^{\nu_i})$. □

Note how the statement of the proposition specializes to the now familiar univariate primary decomposition when applied with $I = \mathrm{Id}(f)$. As a nontrivial application, we can write down the primary decomposition of the ideal $I = \mathrm{Id}(X^2+Y, Y^2+X)$ of Example 8.29, which, as we have mentioned before, is in normal position w.r.t. both X and Y:

$$\begin{aligned} I &= \mathrm{Id}(I,Y) \cap \mathrm{Id}(I,Y+1) \cap \mathrm{Id}(I,Y^2-Y+1) \\ &= \mathrm{Id}(X,Y) \cap \mathrm{Id}(X+1,Y+1) \cap \mathrm{Id}(X+Y-1,Y^2-Y+1) \end{aligned}$$

Here, the primary components happen to be all prime, as one easily verifies using the criterion of Proposition 7.44 as in Example 7.45. This is of course a coincidence. (More precisely, it is because I is a radical ideal.) As a matter of fact, when it comes to associated primes, the analogy with the univariate case must be taken with a grain of salt: the associated primes P_i of the primary components $Q_i = \mathrm{Id}(I, p_i^{\nu_i})$ are not found by simply setting ν_i to 1. Rather, P_i has to be computed as what it is, namely, the radical of Q_i. This is exemplified by the trivial example of $\mathrm{Id}(X^2, Y^2)$, which is itself primary and has $\mathrm{Id}(X,Y)$ as its associated prime.

The proposition above provides the correctness proof for the algorithm of the following theorem. Termination of the algorithm is trivial. Let us emphasize once again that an efficient method to compute univariate polynomials in zero-dimensional ideals will be given in Proposition 9.6.

Theorem 8.70 *Assume that K is computable and allows effective factorization of univariate polynomials. Then the algorithm* NORMPRIMDEC *of Table* 8.4 *computes, for a given finite subset F of $K[\underline{X}]$ with* $\mathrm{Id}(F)$ *zero-dimensional, a set P of finite subsets of $K[\underline{X}]$ with*

$$I = \bigcap_{G \in P} \mathrm{Id}(G)$$

in such a way that if $\mathrm{Id}(F)$ *is in normal position w.r.t. X_i, then this is the primary decomposition of* $\mathrm{Id}(F)$. □

Exercise 8.71 Let I be the ideal $\mathrm{Id}(G)$ of $\mathbb{Q}[X,Y]$, where

$$G = \{X^2 + Y + 1, 2XY + Y\}.$$

Show that I is in normal position w.r.t. X but not w.r.t. Y. Compute the primary decomposition of I. What happens when you try to apply NORMPRIMDEC w.r.t. the "bad variable" Y?

The results of this section thus far are of course of extremely limited practical value, simply because they were proved under the hypothesis of I being in normal position. The idea behind our general strategy for finding the primary decomposition of a zero-dimensional ideal I is as follows. We introduce a new variable Z and try to extend I in such a way that the extended ideal is in normal position w.r.t. Z. We can then perform the

TABLE 8.4. Algorithm NORMPRIMDEC

```
Specification: $G \leftarrow$ NORMPRIMDEC($F, X_i$)
               Computation of the primary components of
               a zero-dimensional ideal in normal position
Given: a finite subset $F$ of $K[\underline{X}]$ with Id($F$) zero-dimensional
Find: a set $P$ of finite subsets of $K[\underline{X}]$ such that $\mathrm{Id}(F) = \bigcap_{G \in P} \mathrm{Id}(G)$,
      and this is the primary decomposition of Id($F$) if Id($F$) is in
      normal position w.r.t. $X_i$
begin
$P \leftarrow \emptyset$
$f \leftarrow$ the monic generator of $\mathrm{Id}(F) \cap K[X_i]$
while $f$ is not constant do
    $p \leftarrow$ an irreducible factor of $f$
    $\mu \leftarrow \max\{\nu \in \mathbb{N} \mid p^\nu | f\}$
    $f \leftarrow f/p^\mu$
    $P \leftarrow P \cup \{F \cup \{p^\mu\}\}$
end
end NORMPRIMDEC
```

primary decomposition and retrieve the one of I by means of an elimination process. This idea will encounter difficulties that will call for certain refinements.

Throughout, Z will be a new indeterminate. We will use the obvious convention that $(a_1, \dots, a_n) \in K^n$ is abbreviated by $\boldsymbol{a}$, and the same goes for $\overline{K}^n$ and other letters of the alphabet. The starting point of the strategy that we have just described is the follwoing easy lemma.

Lemma 8.72 Let I be an ideal of $K[\underline{X}]$. Let $\boldsymbol{c} \in K^n$, set

$$g = Z - c_1X_1 - \cdots - c_nX_n \in K[\underline{X}, Z],$$

and let J be the ideal $\mathrm{Id}(I, g)$ of $K[\underline{X}, Z]$. Then the following are equivalent:

(i) Whenever $\boldsymbol{z}_1$, $\boldsymbol{z}_2 \in \overline{K}^n$ are two different zeroes of I, then

$$\sum_{i=1}^{n} c_i z_{1i} \neq \sum_{i=1}^{n} c_i z_{2i} .$$

(ii) The ideal J of $K[\underline{X}, Z]$ is in normal position w.r.t. Z.

Proof The equivalence of (i) and (ii) is an easy consequence of the fact that the set of zeroes in $\overline{K}^{n+1}$ of J is given by

$$\{ (z_1, \dots, z_n, c_1z_1 + \cdots + c_nz_n) \in \overline{K}^{n+1} \mid \boldsymbol{z} \in \overline{K}^n \text{ a zero of } I \}. \quad \square$$

Before we discuss how the lemma above can be exploited for the computation of the primary decomposition, we collect some technical results concerning the connection between I and J.

Lemma 8.73 Let I be an ideal of $K[\underline{X}]$. Let $c \in K^n$, set

$$g = Z - c_1X_1 - \cdots - c_nX_n \in K[\underline{X}, Z],$$

and let J be the ideal $\mathrm{Id}(I, g)$ of $K[\underline{X}, Z]$. Then the following hold:

(i) If I is zero-dimensional, then so is J.

(ii) $J \cap K[\underline{X}] = I$.

(iii) If $h \in K[\underline{X}, Z]$, then

$$h \equiv h(X_1, \ldots, X_n, c_1X_1 + \cdots + c_nX_n) \mod \mathrm{Id}(g).$$

(iv) If I is a radical ideal, then so is J.

(v) Assume that I is a zero-dimensional radical ideal, and let

$$J = P_1 \cap \cdots \cap P_r$$

be the primary decomposition of J in $K[\underline{X}, Z]$. If we set $P_i' = P_i \cap K[\underline{X}]$ for $1 \leq i \leq r$, then

$$I = P_1' \cap \cdots \cap P_r'$$

is the primary decomposition of I in $K[\underline{X}]$.

Proof (i) This is immediate from the fact that by the proof of the previous lemma, the number of zeroes of I in $\overline{K}^n$ is the same as the number of zeroes of J in $\overline{K}^{n+1}$.

(ii) The inclusion "$\supseteq$" is trivial. Now let $f \in J \cap K[\underline{X}]$. Then there exist $h \in I$ and q_1, $q_2 \in K[\underline{X}, Z]$ with $f = q_1h + q_2g$. If we set

$$Z = c_1X_1 + \cdots + c_nX_n,$$

then the equation turns into $f = p_1h$ with $p_1 \in K[\underline{X}]$, and we see that $f \in I$.

(iii) This follows immediately from

$$Z \equiv c_1X_1 + \cdots + c_nX_n \mod \mathrm{Id}(g).$$

(iv) Let $f \in K[\underline{X}, Z]$ and $\nu \in \mathbb{N}$ with $f^\nu \in J$. If we set

$$h = f(X_1, \ldots X_n, c_1X_1 + \cdots + c_nX_n) \in K[\underline{X}] \subseteq K[\underline{X}, Z],$$

then we have $h \equiv f \bmod \mathrm{Id}(g)$ by (iii). It follows that $h^\nu \equiv f^\nu \bmod \mathrm{Id}(g)$, which means that

$$h^\nu \in \left(f^\nu + \mathrm{Id}(g)\right) \cap K[\underline{X}] \subseteq (f^\nu + J) \cap K[\underline{X}] = J \cap K[\underline{X}] = I,$$

and thus $h \in I$ because I was assumed to be radical. From $f \in h + \mathrm{Id}(g)$ it now follows that $f \in J$.

(v) In view of (ii), it is clear that

$$I = J \cap K[\underline{X}] = K[\underline{X}] \cap \bigcap_{i=1}^{r} P_i = \bigcap_{i=1}^{r} (P_i \cap K[\underline{X}]) = \bigcap_{i=1}^{r} P_i' .$$

We claim that the P_i' are pairwise different. By (iv) and Lemma 8.60 (vi), the P_i are pairwise different zero-dimensional prime ideals and thus pairwise different maximal ideals. This means that if $1 \le i < j \le r$, then there exist $f_1 \in P_i$ and $f_2 \in P_j$ with $1 = f_1 + f_2$. If we set

$$Z = c_1 X_1 + \cdots + c_n X_n$$

in the equation, then we obtain $1 = g_1 + g_2$ with

$$g_1 \in \left(f_1 + \mathrm{Id}(g)\right) \cap K[\underline{X}] \subseteq (f_1 + P_i) \cap K[\underline{X}] = P_i \cap K[\underline{X}] = P_i'.$$

The same argument shows that $g_2 \in P_j'$, and we see that indeed P_i' and P_j' are different. Lemma 8.4 now tells us that none of the P_i contains the intersection of the rest, and we have proved that the P_i' are the primary components of I. □

Exercise 8.74 Formulate and prove a statement that generalizes (v) of the lemma above to the case where I is not radical.

Our next goal is to find an n-tuple $\boldsymbol{c} \in K^n$ with the property of Lemma 8.72 for a given zero-dimensional ideal. The following combinatorial lemma will enable us to determine a finite subset C of K^n which must contain a winner.

Lemma 8.75 Let $A_1, \ldots, A_n$ be finite subsets of K with $|A_i| \le m_i$ for $1 \le i \le n$. For $2 \le i \le n$, set

$$m_i' = m_1 \cdot \cdots \cdot m_i, \quad \text{and} \quad k_i = \binom{m_i'}{2}.$$

Now whenever $C_2, \ldots, C_n$ are subsets of K with $|C_i| > k_i$ for $2 \le i \le n$, then there exists $(c_2, \ldots, c_n) \in C_2 \times \cdots \times C_n$ with the following property: for all $\boldsymbol{a}$, $\boldsymbol{b} \in A_1 \times \cdots \times A_n$,

$$\boldsymbol{a} \ne \boldsymbol{b} \quad \text{implies} \quad a_1 + \sum_{i=2}^{n} c_i a_i \ne b_1 + \sum_{i=2}^{n} c_i b_i .$$

Proof The proof is by induction on n. If $n = 1$, then the claim is trivial. Let $n > 1$, and suppose $C_2, \ldots, C_n$ satisfy the requirement on the cardinalities. By induction hypothesis, there exists

$$(c_2, \ldots, c_{n-1}) \in C_2 \times \cdots \times C_{n-1}$$

such that for all $\boldsymbol{a}$, $\boldsymbol{b} \in A_1 \times \cdots \times A_{n-1}$,

$$\boldsymbol{a} \neq \boldsymbol{b} \quad \text{implies} \quad a_1 + \sum_{i=2}^{n-1} c_i a_i \neq b_1 + \sum_{i=2}^{n-1} c_i b_i \,.$$

If $\boldsymbol{a}$, $\boldsymbol{b} \in A_1 \times \cdots \times A_n$ with $\boldsymbol{a} \neq \boldsymbol{b}$, then the linear equation

$$a_1 + \sum_{i=2}^{n-1} c_i a_i + Y a_n = b_1 + \sum_{i=2}^{n-1} c_i b_i + Y b_n \,.$$

in the unknown Y has at most one solution, namely,

$$Y = \frac{a_1 + \sum_{i=2}^{n-1} c_i a_i - b_1 - \sum_{i=2}^{n-1} c_i b_i}{b_n - a_n}$$

in case $a_n \neq b_n$. It is easy to see that there are at most k_n such equations. The set C_n must thus contain an element c_n that is not a solution of any one of these equations, and we see that $(c_2, \ldots, c_n)$ has the required property. □

If $\boldsymbol{z} \in \overline{K}^n$ is a zero of the zero-dimensional ideal I, then for $1 \leq i \leq n$, z_i is a zero of the unique monic generator f_i of $I \cap K[X_i]$, which in turn has at most $\deg(f_i)$ many different zeroes in $\overline{K}$. This observation together with the lemma above yields the following result.

Lemma 8.76 Let I be a zero-dimensional ideal of $K[\underline{X}]$, and for $1 \leq i \leq m$, let $m_i = \deg(f_i)$, where f_i is the unique monic polynomial of minimal degree in $I \cap K[X_i]$. For $2 \leq i \leq n$, set

$$m_i' = m_1 \cdot \cdots \cdot m_i, \quad \text{and} \quad k_i = \binom{m_i'}{2}.$$

Then whenever $C_2, \ldots, C_n$ are subsets of K with $|C_i| > k_i$ for $2 \leq i \leq n$, there exists $(c_2, \ldots, c_n) \in C_2 \times \cdots \times C_n$ such that $(1, c_2, \ldots, c_n)$ satisfies the equivalent conditions of Lemma 8.72. □.

If K is computable, then given the ideal I, we can clearly determine the m_i of the lemma. If in addition, K is infinite, then we can of course find suitable finite $C_i \subset K$. The lemma would thus appear to be constructive. The problem is that we can not in general determine which one of the finitely many $(n-1)$-tuples in the Cartesian product of the C_i is good. The next proposition and lemma state that the problem can be solved if K is perfect and I is a radical ideal. For our proposed strategy to compute the primary decomposition, this means that we will have to pass to the radical and then find a way to get back to the given ideal.

Proposition 8.77 *Assume that K is perfect. Let I be a zero-dimensional radical ideal of $K[\underline{X}]$, and suppose I is in normal position w.r.t. X_1. Then the reduced Gröbner basis G of I w.r.t. any term order satisfying $\{X_1\} \ll \{X_2, \ldots, X_n\}$ is of the form*

$$G = \{g_1, X_2 - g_2, \ldots, X_n - g_n\}$$

with $g_1, \ldots, g_n \in K[X_1]$.

Proof By Proposition 6.15 and the fact that G is reduced, $G \cap K[X_1]$ has exactly one element f_1. Set $d = \deg(f_1)$, and let m_1 be the number of different zeroes of f_1 in $\overline{K}$ and m the number of different zeroes of I in $\overline{K}^n$. Then $m \leq m_1$ because I is in normal position w.r.t. X_1. Being reduced, G cannot contain another element whose head term is a power of X_1, and so the terms $1, X_1, \ldots, X_1^{d-1}$ are reduced. Since the number of elements in the canonical term basis of the K-vector space $K[\underline{X}]/I$ equals the number of reduced terms by Proposition 6.52, it follows that $d \leq \dim_K(K[\underline{X}]/I)$. Theorem 8.32 tells us that $\dim_K(K[\underline{X}]/I) = m$, and we have proved that

$$d \leq \dim_K(K[\underline{X}]/I) = m \leq m_1.$$

On the other hand, f_1 can have no more than d different zeroes in $\overline{K}$, i.e., $m_1 \leq d$, and we see that actually

$$d = \dim_K(K[\underline{X}]/I) = m = m_1.$$

Again using Proposition 6.52, we conclude that there are no reduced terms besides $1, X_1, \ldots, X_1^{d-1}$, and so

$$X_i \in \mathrm{HT}(I) = \mathrm{HT}(G) \quad \text{for} \quad 2 \leq i \leq n.$$

Since G was assumed to be reduced, it follows that besides g_1, it contains nothing but exactly one element f_i with head term X_i for each index $2 \leq i \leq n$, and

$$T(f_i) \subseteq \{X_i, 1, X_1, \ldots X_1^{d-1}\} \quad \text{for} \quad 2 \leq i \leq n. \quad \square$$

Exercise 8.78 Use the results of Sections 7.3 and 8.2 to show that at the very beginning of the proof of the above proposition, one may already conclude that $d = m_1$, thus simplifying the rest of the argument slightly.

The converse to the last proposition is in fact true in a more general situation.

Lemma 8.79 Let I be any ideal of $K[\underline{X}]$ where K is again an arbitrary field, and assume that I has a basis G of the form

$$G = \{g_1, X_2 - g_2, \ldots, X_n - g_n\}$$

with $g_1, \ldots, g_n \in K[X_1]$. Then I is zero-dimensional and in normal position w.r.t. X_1.

Proof I is zero-dimensional because for $1 \leq i \leq n$, it contains a polynomial whose head term w.r.t. any term order satisfying $\{X_1\} \ll \{X_2, \ldots, X_n\}$ is a power of X_i. Moreover, if $z \in \overline{K}^n$ is a zero of I, then z_1 uniquely determines z_i to be $g_i(z_1)$ for $2 \leq i \leq n$. □

Exercise 8.80 Use the last lemma and proposition to show once again that the ideal $I = \mathrm{Id}(X^2+Y+1, 2XY+Y)$ of Exercise 8.71 is in in normal position w.r.t. X but not w.r.t. Y. Why is your new argument more elegant than your original one?

From the last lemma, we conclude that if we chance upon an ideal basis of the indicated form and K is computable and allows factorization of univariate polynomials, then we can go right ahead and compute the primary decomposition by means of the algorithm NORMPRIMDEC. The lemma together with the proposition preceding it states that over a computable perfect field, we can decide whether a given radical ideal is in normal position w.r.t. any variable by inspecting a suitable Gröbner basis. Together with Lemma 8.76 and Lemma 8.73 (iv), this proves the correctness of the following algorithm whose termination is trivial. Note that every computable field of characteristic zero satisfies the assumptions of the theorem.

Theorem 8.81 *Assume that K is computable, infinite, and perfect. Then the algorithm* NORMPOS *of Table* 8.5 *computes, for given finite subset F of $K[\underline{X}]$ with* $\mathrm{Id}(F)$ *zero-dimensional, a Gröbner basis G of an extended ideal*

$$J = \mathrm{Id}(F, Z - X_1 - c_2X_2 - \cdots - c_nX_n) \qquad (c_2, \ldots, c_n \in K)$$

which is of the form $G = \{g, X_1-g_1, \ldots, X_n-g_n\}$ with $g, g_1, \ldots, g_n \in K[Z]$ whenever $\mathrm{Id}(F)$ *is radical.* □

Exercise 8.82 Recall that the ideal $I = \mathrm{Id}(X^2 - 2, Y^2 - 2)$ of Example 8.66 is in normal position neither w.r.t. X nor w.r.t. Y. Find $c \in \mathbb{Q}$ such that the ideal

$$\mathrm{Id}(X^2 - 2, Y^2 - 2, Z - X - cY)$$

of $\mathbb{Q}[X, Y, Z]$ is in normal position w.r.t. Z.

We have in fact reached our goal of being able to compute zero-dimensional primary decompositions. Let K be a computable field and suppose a finite basis of the zero-dimensional ideal I of $K[\underline{X}]$ is given. If K is perfect and allows squarefree decompositions of univariate polynomials, then we can apply the algorithm ZRADICAL to compute $\mathrm{rad}(I)$. If, in addition, K is infinite, then the algorithm NORMPOS computes for us the extended ideal J of $K[\underline{X}, Z]$ which is radical too and in normal position w.r.t. Z. If K also allows factorization of univariate polynomials, then NORMPRIMDEC provides the primary components $P_1, \ldots, P_r$ of J in $K[\underline{X}, Z]$. As a matter

TABLE 8.5. Algorithm NORMPOS

```
Specification: $G \leftarrow$ NORMPOS($F$)
               Extending a zero-dimensional radical ideal
               to a radical ideal in normal position
Given: a finite subset $F$ of $K[\underline{X}]$ with Id($F$) zero-dimensional
Find: a basis $G$ of $J = \mathrm{Id}(F, Z - X_1 - c_2X_2 - \cdots - c_nX_n)$, where
      $c_2, \ldots, c_n \in K$, such that $G = \{g, X_1 - g_1, \ldots, X_n - g_n\}$ with
      $g, g_1, \ldots, g_n \in K[Z]$ whenever Id($F$) is radical
begin
$f \leftarrow$ the monic generator of $\mathrm{Id}(F) \cap K[X_1]$
$m \leftarrow \deg(f_1)$
for $i = 2$ to $n$ do
    $f_i \leftarrow$ the monic generator of $\mathrm{Id}(F) \cap K[X_i]$
    $m_i \leftarrow \deg(f_i)$
    $m \leftarrow m \cdot m_i$
    $C_i \leftarrow$ a finite subset of $K$ with $|C_i| = \binom{m}{2} + 1$
end
$C \leftarrow \{1\} \times C_2 \times \cdots \times C_n$
repeat select $c$ from $C$
       $C \leftarrow C \setminus \{c\}$
       $G \leftarrow$ a reduced Gröbner basis of
           $\mathrm{Id}(\{F, Z - X_1 - c_2X_2 - \cdots - c_nX_n\})$ w.r.t.
           a term order with $\{Z\} \ll \{X_1, \ldots, X_n\}$
until $C = \emptyset$ or $G$ is of the form $\{g, X_1 - g_1, \ldots, X_n - g_n\}$
      with $g, g_1, \ldots, g_n \in K[Z]$
end NORMPOS
```

of fact, NORMPOS has already provided a Gröbner basis of J which is of the form

$$\{g, X_1 - g_1, \ldots, X_n - g_n\}$$

with g, g_1, …, $g_n \in K[Z]$, so that the monic generator g of $J \cap K[Z]$ is at hand and need not be computed by NORMPRIMDEC. Moreover, we know that g is squarefree because J is radical. By Lemma 8.73 (v), the elimination ideals $P_i' = P_i \cap K[\underline{X}]$ are the primary components of rad(I), and these in turn are the associated primes of the primary components of I by Lemma 8.60 (iv). All that remains to be done is to recover the primary components of I from their associated primes. Since we can compute the univariate exponent of I by means of squarefree decompositions, Lemma 8.60 (iii) shows how this can be done.

The discussion above provides the correctness proof for the algorithm of the following theorem. Note that for all practical purposes, the requirement that K be infinite and perfect narrows it down to fields of characteristic zero. An example of a field to which the theorem applies is given by the rationals. Recall that the univariate exponent of a zero-dimensional ideal can be computed whenever effective squarefree decomposition of univariate polynomials is available.

Theorem 8.83 *Assume that K is computable, infinite, and perfect and allows effective factorization of univariate polynomials. Then the algorithm* ZPRIMDEC *of Table* 8.6 *computes the primary components and associated primes of* $\mathrm{Id}(F)$ *whenever F is a finite subset of $K[\underline{X}]$ with* $\mathrm{Id}(F)$ *zero-dimensional.* □

TABLE 8.6. Algorithm ZPRIMDEC

```
Specification: $P \leftarrow$ ZPRIMDEC($F$)
               Computation of primary components and associated
               primes of a zero-dimensional ideal
Given: a finite subset $F$ of $K[\underline{X}]$ with Id($F$) zero-dimensional
Find: a set $P$ of pairs $(G, H)$ of finite subsets of $K[\underline{X}]$ such that
      $\{ \mathrm{Id}(G) \mid (G,H) \in P$ for some $H \}$ is the set of all primary
      components of Id($F$), and Id($H$) is the associated prime of
      Id($G$) for all $(G,H) \in P$
begin
$R \leftarrow$ ZRADICAL($F$)
$G \leftarrow$ NORMPOS($R$)
$Q \leftarrow \emptyset$
$g \leftarrow G \cap K[Z]$
while $g$ is not constant do
      $p \leftarrow$ an irreducible factor of $g$
      $g \leftarrow g/p$
      $Q \leftarrow Q \cup \{ G \cup \{p\} \}$
end
$P \leftarrow \emptyset$
$m \leftarrow$ the univariate exponent of Id($F$)
while $Q \neq \emptyset$ do
      select $A$ from $Q$
      $Q \leftarrow Q \setminus \{A\}$
      $H \leftarrow$ ELIMINATION($A, \{X_1, \ldots X_n\}$)
      $G \leftarrow F \cup H^m$
      $P \leftarrow P \cup \{(G,H)\}$
end
end ZPRIMDEC
```

Exercise 8.84 Let I be the ideal $\mathrm{Id}(X^2 - 2, Y^2 - 2)$ of Example 8.66. Use the algorithm ZPRIMDEC to confirm the result of Example 8.66. (Hint: You have done part of the work in Exercise 8.82.)

We mention that in the presence of effective primary decomposition and radical test, one also obtains, at least in principle, a *primality test* for ideals: it is easy to see that an ideal is prime if and only if it has one primary component that is radical. In a situation where the algorithm ZPRIMDEC is applicable, $\mathrm{Id}(F)$ is thus prime iff $\mathrm{Id}(F)$ is radical and the univariate polynomial in Z in $\mathrm{NORMPOS}(F)$ is irreducible.

The most expensive part of the algorithm ZPRIMDEC in terms of time and space is the application of NORMPOS. The fact that a large number of Gröbner bases may have to be computed until a normal position is found often makes inputs of no more than moderate size impossible to handle. For those with an interest in perfomance, we will now discuss a variant of the algorithm that has turned out to generally improve the running time in practice. The underlying principle here is that it seems to be more advantageous in general to do some univariate factoring and then compute many Gröbner bases of ideals with "small" univariate polynomials than it is to compute fewer Gröbner bases of ideals with "large" univariate polynomials. What this means in this case is that one should first decompose a given ideal I by means of the algorithm PREDEC of Lemma 8.6, then compute the primary decompositions of the non-trivial constituents of that decomposition, and finally collect all primary ideals thus obtained. It is clear that the passage to the radical that ZPRIMDEC calls for can be combined with the application of PREDEC: if PREDEC uses the squarefree parts of those univariate polynomials which it would otherwise use, then it has automatically passed to the radical of I.

The idea that we have just described can actually be exploited further. To this end, we now show how the action of NORMPOS of getting the extended ideal J of Theorem 8.81 into normal position w.r.t. the new variable Z can be broken up into smaller steps so that one can try to apply PREDEC in between the steps. We begin with two observations that are almost trivial.

Lemma 8.85 Let $\leq$ be a term order on $T(\underline{X})$ and I an ideal of $K[\underline{X}]$ such that there exists $f \in I$ with $\mathrm{HT}(f) = X_i$ for some $1 \leq i \leq n$. Let G be a Gröbner basis w.r.t. $\leq$ of I. Then the following hold:

(i) There exists $g \in G$ with $\mathrm{HT}(g) = X_i$.

(ii) If G is reduced, then there exists exactly one $g \in G$ with $\mathrm{HT}(g) = X_i$, and $\deg_{X_i}(t) = 0$ for every t occurring anywhere in G with the exception of $\mathrm{HT}(g)$ itself.

Proof Statement (i) is immediate from the fact that f must be top-reducible modulo G. For (ii), it is clear that a reduced Gröbner basis cannot have two different polynomials with the same head term. Furthermore, every term $t \in T(\underline{X})$ with $t < X_i$ must satisfy $\deg_{X_i}(t) = 0$. This proves the second claim of (ii) for all $t \in T(g) \setminus \{X_i\}$. For all other terms occurring in G, it follows from the fact that G is reduced. □

The following lemma will be used to perform a single step in our proposed "step-by-step" version of NORMPOS.

Lemma 8.86 Let I be a zero-dimensional ideal of $K[\underline{X}]$ and $1 \leq i_1 < i_2 \leq n$, and assume that the elimination ideal $I' = I \cap K[X_{i_1}, X_{i_2}]$ is radical. Suppose $\leq$ is a term order on $T(\underline{X}, Z)$ with

$$\{Z\} \ll \{X_{i_1}, X_{i_2}\},$$

where Z is a new variable. Set $m_1 = \deg(f_1)$ and $m_2 = \deg(f_2)$, where f_1 and f_2 are the unique monic polynomials of minimal degree in $I \cap K[X_{i_1}]$ and $I \cap K[X_{i_2}]$, respectively. Let

$$m = m_1 m_2 \quad \text{and} \quad k = \binom{m}{2}.$$

Then whenever C is a subset of K with $|C| > k$, there exists $c \in C$ such that every Gröbner basis G w.r.t. $\leq$ of the extended ideal

$$J = \operatorname{Id}(I, Z + X_{i_1} + cX_{i_2})$$

contains polynomials g_1 and g_2 with $\operatorname{HT}(g_1) = X_{i_1}$ and $\operatorname{HT}(g_2) = X_{i_2}$.

Proof Consider the elimination ideal $I' = I \cap K[X_{i_1}, X_{i_2}]$ and the extended ideal

$$J' = \operatorname{Id}(I', Z + X_{i_1} + cX_{i_2}).$$

Lemma 8.76 tells us that there exists $c \in C$ such that J' is in normal position w.r.t. Z. We may now conclude from Proposition 8.77 that J' contains polynomials h_1 and h_2 of the form

$$h_1 = X_{i_1} - p_1 \quad \text{and} \quad h_2 = X_{i_2} - p_2$$

with $p_1, p_2 \in K[Z]$. We see that the head terms of h_1 and h_2 w.r.t. $\leq$ are X_{i_1} and X_{i_2}, respectively. The claim now follows from $J' \subseteq J$ together with (i) of the previous lemma. □

It is clear that when actually looking for the c of the lemma, we do not need to know what the bound k for the maximal number of unsuccessful tries is. It suffices to try different c until a hit is scored.

We will now give an informal description of the modified algorithm ZPRIMDEC, proving correctness and termination as we go along. Suppose the field K satisfies the requirements of the algorithm ZPRIMDEC. Let F be a finite subset of $K[\underline{X}]$ such that $\operatorname{Id}(F)$ is zero-dimensional. Assume that F has already been preprocessed by PREDEC combined with a computation of the radical, so that all generators of univariate elimination ideals are irreducible. (It should be clear by now how the primary decomposition of the original ideal can be recovered.) Then one may proceed as follows:

(1) Compute a reduced Gröbner basis of $\mathrm{Id}(F)$ w.r.t. some term order. If, possibly after renumbering variables, G is of the form

$$G = \{g_1, X_2 - g_2, \dots, X_n - g_n\}$$

with $g_1, \dots, g_n \in K[X_1]$, then by Lemma 8.79, I in normal position w.r.t. X_1, and no further action is necessary. If G is not of this form, then, in view of (ii) of the lemma before the last one, there must be two variables X_{i_1} and X_{i_2} that do not occur as linear head terms in G. By the previous lemma, there exists $c \in K$ such that the ideal

$$J = \mathrm{Id}(I, Z_1 + X_{i_1} + cX_{i_2})$$

contains polynomials g_1 and g_2 with $\mathrm{HT}(g_1) = X_{i_1}$ and $\mathrm{HT}(g_2) = X_{i_2}$. Furthermore, such a c can be found by trial and error, varying c and inspecting Gröbner bases of J w.r.t. a term order that satisfies $\{Z_1\} \ll \{X_{i_1}, X_{i_2}\}$. Inspection of the proof of Lemma 8.73 (ii) shows that we again have $J \cap K[\underline{X}] = I$.

(2) When step (1) has been performed, the generator f of $J \cap K[Z_1]$ is visible. This generator should now be factored and J should be decomposed according to Lemma 8.5. There is no point in applying PREDEC to its full extent because the generators of all other univariate elimination ideals were already irreducible, so they cannot have changed. Note also that the prime factors of f all have multiplicity 1 because J is radical.

(3) Now if we look at any one constituent I' of the intersection obtained in step (2), then I' is an ideal of $K[\underline{X}, Z_1]$ that has the same properties as the input ideal $\mathrm{Id}(F)$ had as an ideal of $K[\underline{X}]$. We have one more variable, but since $g_1, g_2 \in I'$, we know that there are at least two more variables that occur as linear head terms. It follows that the number of variables that do not occur as linear head terms has decreased by at least one. At most n repetitions of the process will therefore take us to ideal bases H in $K[Z_1, \dots, Z_r, \underline{X}]$ of the form

$$H = \{g_r, Z_{r-1} - h_{r-1}, \dots, Z_1 - h_1, X_1 - g_1, \dots, X_n - g_n\}$$

with $h_1, \dots, h_{r-1}, g_1, \dots, g_n \in K[Z_r]$. The set H is clearly a Gröbner basis w.r.t. any term order with

$$\{Z_r\} \ll \{Z_1, \dots Z_{r-1}, X_1, \dots X_n\},$$

because that way, the head terms remain pairwise disjoint. Placing $Z_1, \dots, Z_{r-1}$ lexicographically high, we may therefore conclude that

$$\mathrm{Id}(H) \cap K[Z_r, \underline{X}] = \{g_r, X_1 - g_1, \dots, X_n - g_n\}.$$

We see that this ideal is in normal position w.r.t. Z_r, and since in the last run through step (2), we have already decomposed so as to make g_r irreducible, we see that this last ideal is already primary, i.e., it is prime in this case because everything was done on the level of the radical. It is now not hard to see that the intersection of the elimination ideals w.r.t. $\{X_1, \ldots, X_n\}$ of the ideals that we have computed equals $\mathrm{Id}(F)$. This means that these elimination ideals are the primary components of $\mathrm{Id}(F)$.

We mention that the choice between NORMPOS and the above step-by-step version involves yet another trade-off. On the one hand, the step-by-step version requires Gröbner basis computations with up to $2n$ variables as opposed to $n+1$ in NORMPOS. On the other hand, the maximal number of unsuccsessful tries in each of the at most n steps of the step-by-step version is

$$k = \binom{m_1 m_2}{2},$$

where m_1 and m_2 are the degrees of two univariate polynomials in the respective ideal. Comparison with NORMPOS shows that there, the number of tries is, in the worst case, by orders of magnitude larger. This is a kind of situation that occurs quite frequently when Gröbner basis computations are involved: even the most sophisticated complexity theory is—at least at present—not strong enough to allow a clear decision between the two possible versions of the algorithm. One has therefore to rely on practical experience, and it is not impossible for different people to arrive at different conclusions.

We will now show how at least in principle, we can compute zero-dimensional primary decompositions over finite fields. Let K be a computable finite field and I a zero-dimensional ideal of $K[\underline{X}]$. We first note that all finite fields allow, at least in principle, effective univariate factorization because there are only finitely many polynomials of each degree. Furthermore, all finite fields are perfect, and so we may, as before, pass to the radical of I, find its primary decomposition, and then recover the one of I using the univariate exponent of I. So let us assume that I is a zero-dimensional radical ideal.

It is easy to see from Lemma 8.60 (vi) that the primary components of I have been found as soon as we have found pairwise different prime ideals $P_1, \ldots, P_r$ with

$$I = P_1 \cap \cdots \cap P_r.$$

The strategy that we will describe arises from a careful analysis of the proof of Lemma 8.13. We begin by factoring the monic generator of $I \cap K[X_1]$ into its pairwise relatively prime factors and apply Lemma 8.5 to obtain a decomposition

$$I = \bigcap_{i=1}^{r} (I, p_i)$$

with $p_i \in K[X_1]$ irreducible for $1 \leq i \leq r$. It is clear that it now suffices to decompose each one of the radical ideals that occur in the intersection into prime ideals. We are left with the task of finding the primary decomposition of a radical ideal I such that the monic generator f_1 of $I \cap K[X_1]$ is irreducible.

In the remarks following Corollary 7.10, we have proved that the field $K[X_1]/\mathrm{Id}(f_1)$ is again computable. It is clear that it is also finite, because a system of unique representatives for its elements is given by

$$\{\, h \in K[X_1] \mid \deg(h) < \deg(f_1) \,\},$$

and this is clearly a finite set. We may thus call the entire procedure recursively on the image of I under the map

$$\varphi : K[X_1][X_2, \ldots, X_n] \longrightarrow K[X_1]/\mathrm{Id}(f_1)[X_2, \ldots, X_n]$$

of the proof of Lemma 8.13. (It was proved there that this image is again a zero-dimensional radical ideal.) If $M_1, \ldots, M_r$ are the prime ideals that this recursive call returns, then again by the proof of Lemma 8.13, their inverse images $\varphi^{-1}(M_1), \ldots, \varphi^{-1}(M_r)$ are the primary components of I that we are looking for. The M_i are given to us by finite bases G_i, and each element of G_i comes as a representative in $K[\underline{X}]$ of its residue class modulo $\ker(\varphi)$. So in practice, we have $G_i \subseteq K[\underline{X}]$, and it is now easy to see that bases H_i of $\varphi^{-1}(M_i)$ are given by $H_i = G_i \cup \{f_1\}$ for $1 \leq i \leq r$. Correctness of the entire procedure now follows from the fact that it does the right thing in the univariate case.

To conclude this section, we point out how a special case of a classical result on field extensions, namely, the *theorem on the primitive element*, can be deduced from the theory behind the algorithm PRIMDEC. (In its full strength, the theorem is not limited to infinite fields, and a weaker assumption than perfectness of K is used.)

Theorem 8.87 (THEOREM ON THE PRIMITIVE ELEMENT) *Let K be a perfect infinite field and K' a finite algebraic extension of K, say $K' = K(a_1, \ldots, a_n)$ with $a_1, \ldots, a_n \in K'$ algebraic over K. Then K' is a simple extension of K, i.e., there exists $b \in K'$ with $K' = K(b)$. Moreover, b can be chosen as*

$$b = a_1 + c_2 a_2 + \cdots + c_n a_n$$

with $c_2, \ldots, c_n \in K$. Finally, if K is also computable, then $c_2, \ldots, c_n$ can be computed from the minimal polynomials of $a_1, \ldots, a_n$ over K.

Proof For $1 \leq i \leq n$, let $f_i \in K[X_i]$ be the minimal polynomial of a_i over K, and set

$$I = \mathrm{Id}(f_1, \ldots, f_n) \subseteq K[X_1, \ldots, X_n] = K[\underline{X}].$$

Then I is a zero-dimensional radical ideal because it contains the irreducible (and hence squarefree) polynomial f_i in the variable X_i for $1 \leq i \leq n$. Lemma 8.76 provides an $(n-1)$-tuple $(c_2, \ldots, c_n) \in K^{n-1}$ such that the ideal $\mathrm{Id}(I, g)$ is in normal position w.r.t. the new variable Z, where

$$g = Z - X_1 - c_2X_2 - \cdots - c_nX_n.$$

Proposition 8.77 tells us that the reduced Gröbner basis of $\mathrm{Id}(I, g)$ w.r.t. a suitable term order is of the form

$$G = \{h, X_1 - h_1, \ldots, X_n - h_n\}$$

with h, h_1, ..., $h_n \in K[Z]$, and in the computable case, G can be computed from f_1, ..., f_n by means of the algorithm NORMPOS. Now set $b = a_1 + c_2a_2 + \cdots + c_na_n$, and consider the substitution homomorphism

$$\begin{array}{cccc} \varphi : & K[X_1, \ldots, X_n, Z] & \longrightarrow & K' \\ & f & \longmapsto & f(a_1, \ldots, a_n, b) \end{array}$$

Then clearly g, f_1, ..., $f_n \in \ker(\varphi)$, so $\mathrm{Id}(I, g) \subseteq \ker(\varphi)$, and so in particular, $G \subseteq \ker(\varphi)$. It follows that $a_i = h_i(b)$ for $1 \leq i \leq n$. This shows that $K' \subseteq K(b)$, and the obvious fact that $b \in K'$ provides the reverse inclusion $K(b) \subseteq K'$. □

It is perhaps noteworthy that in the computable case of the theorem, the algorithm NORMPOS provides the Gröbner basis G which contains the polynomial $h \in K[Z]$ that satisfies $h(b) = 0$ as well as $h_i \in K[Z]$ with $h_i(b) = a_i$ for $1 \leq i \leq n$.

Exercise 8.88 Show that $\mathbb{Q}(\sqrt{2}, \sqrt{3}) = \mathbb{Q}(\sqrt{2} + \sqrt{3})$. Find the minimal polynomial of $\sqrt{2} + \sqrt{3}$ over $\mathbb{Q}$ as well as h_1, $h_2 \in \mathbb{Q}[Z]$ with $h_1(\sqrt{2} + \sqrt{3}) = \sqrt{2}$ and $h_2(\sqrt{2} + \sqrt{3}) = \sqrt{3}$.

Exercise 8.89 It is clear that the ideal I in the proof of the theorem on the primitive element can be replaced by any zero-dimensional radical ideal J that satisfies $J \subseteq \ker(\varphi)$. Now suppose that we are in the situation of the theorem, but instead of the minimal polynomials f_i of the a_i over K, we are given f_1 and the minimal polynomials h_i of a_i over $K(a_1, \ldots, a_{i-1})$ for $2 \leq i \leq n$, say

$$h_i = X^{m_i} + \sum_{j=0}^{m_i - 1} q_{ij}(a_1, \ldots, a_{i-1}) \cdot X^j \qquad (q_{ij} \in K[X_1, \ldots, X_{i-1}]).$$

For $2 \leq i \leq n$, we now set

$$h_i^* = X_i^{m_i} + \sum_{j=0}^{m_i - 1} q_{ij} X_i^j.$$

Show that $J = \mathrm{Id}(f_1, h_2^*, \ldots, h_n^*)$ has the properties stated above. (Hint: Use Proposition 7.44.)

Exercise 8.90 (i) Let K be a field, and let $K(a)$ and $K(b)$ be simple algebraic extensions of K with $K(a) \subseteq K(b)$. Show that $K(a) = K(b)$ is equivalent to $\deg(\min_K^a) = \deg(\min_K^b)$. (Hint: Use Proposition 7.8, and argue with the vector space dimension over K.)

(ii) Assume that K satisfies the hypotheses of the theorem on the primitive element, and suppose two simple algebraic extensions $K(a)$ and $K(b)$ of K are given by the minimal polynomials of a and b over K. Show how it can be effectively decided whether $K(a) \subseteq K(b)$.

8.7 Radical and Decomposition in Higher Dimensions

As before, we let K be a field. The basic strategy for the computation of radicals and primary decompositions of arbitrary polynomial ideals over K is to reduce the problem to the zero-dimensional case by means of the extension/contraction method that was introduced in Lemmas 1.122, 1.123, and 7.47. We begin by discussing some more general aspects of this method and its connection with Gröbner bases. We let $\{X_1, \ldots, X_n\}$ be a set of indeterminates, and $\{U_1, \ldots, U_r\} \subseteq \{X_1, \ldots, X_n\}$. We will use the now familiar notation

$$K[\underline{X}] = K[X_1, \ldots, X_n],$$

and we denote by $K(\underline{U})$ the rational function field over K in the variables that are in the set $\{U_1, \ldots, U_r\}$. Moreover, we set

$$K(\underline{U})[\underline{X} \setminus \underline{U}] = K(\underline{U})[V_1, \ldots, V_{n-r}]$$

where $\{V_1, \ldots, V_{n-r}\} = \{X_1, \ldots, X_n\} \setminus \{U_1, \ldots, U_r\}$. Extensions of ideals of $K[\underline{X}]$ will be understood to be to $K(\underline{U})[\underline{X} \setminus \underline{U}]$, and contractions are always to $K[\underline{X}]$. The notations $T(\underline{U})$ and $T(\underline{X} \setminus \underline{U})$ have the obvious meanings (cf. the beginning of Section 6.2).

We begin by showing how one can compute contractions of ideals of $K(\underline{U})[\underline{X} \setminus \underline{U}]$. Recall that we have introduced and discussed the notation $I : f^\infty$ following Lemma 6.36.

Lemma 8.91 Let $\leq$ be a term order on $T(\underline{X} \setminus \underline{U})$. Suppose J is an ideal of $K(\underline{U})[\underline{X} \setminus \underline{U}]$, and G is a Gröbner basis w.r.t. $\leq$ of J such that $G \subseteq K[\underline{X}]$. Let I be the ideal generated by G in $K[\underline{X}]$, and set

$$f = \operatorname{lcm}\{\operatorname{HC}(g) \mid g \in G\},$$

where $\operatorname{HC}(g) \in K[\underline{U}]$ is taken of g as an element of $K(\underline{U})[\underline{X} \setminus \underline{U}]$. Then $J^c = I : f^\infty$.

Proof The inclusion "$\supseteq$" would in fact be true for arbitrary $f \in K[\underline{U}]$. To see this, let $g \in I : f^\infty$. Then $f^s g \in I$ for some $s \in \mathbb{N}$. It follows that

$$g = \frac{1}{f^s} \cdot f^s g \in J \cap K[\underline{X}] = J^c.$$

For the inclusion "$\subseteq$," let $g \in J^c$. Then $g \in J$, and so $g \xrightarrow[G]{*} 0$. We use induction on the minimal length n of such a reduction chain to prove that $g \in I : f^\infty$. If $n = 0$, then $g = 0$ and the claim is trivial. Now suppose $n > 0$ and $g_1 \in K(\underline{U})[\underline{X} \setminus \underline{U}]$ is such that

$$g \xrightarrow[G]{} g_1 \xrightarrow[G]{n-1} 0.$$

Then there exist $p \in G$, $h \in K[\underline{U}]$, and $s \in T(\underline{X} \setminus \underline{U})$ such that

$$g_1 = g - \frac{h}{\mathrm{HC}(p)} \cdot s \cdot p. \qquad (*)$$

Since f is a constant in the polynomial ring $K(\underline{U})[\underline{X} \setminus \underline{U}]$, we may conclude that $fg_1 \xrightarrow[G]{n-1} 0$. Moreover, if we multiply the equation $(*)$ by f and observe that $\mathrm{HC}(p) \mid f$ in $K[\underline{U}]$, then we see that $fg_1 \in J^c$. The induction hypothesis tells us that $fg_1 \in I : f^\infty$. Again by looking at $(*)$ multiplied by f, we conclude that $fg \in I : f^\infty$ and so $g \in I : f^\infty$. □

Now suppose G is any Gröbner basis in $K(\underline{U})[\underline{X} \setminus \underline{U}]$. If we multiply each element of G with the lcm of the denominators of all its coefficients in $K(\underline{U})$, then it is clear that we obtain a Gröbner basis H of the same ideal of $K(\underline{U})[\underline{X} \setminus \underline{U}]$ with the additional property that $H \subseteq K[\underline{X}]$. This observation together with the lemma above proves the correctness of the following algorithm.

Proposition 8.92 *The algorithm* CONT *of Table* 8.7 *computes, for any subset* $\{U_1, \dots, U_r\}$ *of* $\{X_1, \dots, X_n\}$ *and finite subset* F *of* $K(\underline{U})[\underline{X} \setminus \underline{U}]$, *a Gröbner basis of the ideal* $(\mathrm{Id}(F))^c$ *of* $K[\underline{X}]$. □

TABLE 8.7. Algorithm CONT

```
Specification: $G \leftarrow \mathrm{CONT}(F, \{U_1, \dots, U_r\})$
                   Computation of contraction ideal
Given: a subset $\{U_1, \dots, U_r\}$ of $\{X_1, \dots, X_n\}$, and
       a finite subset $F$ of $K(\underline{U})[\underline{X} \setminus \underline{U}]$
Find: a Gröbner basis $G \subseteq K[\underline{X}]$ of $(\mathrm{Id}(F))^c$
begin
$H \leftarrow$ a Gröbner basis of $\mathrm{Id}(F)$ w.r.t. any term order on $T(\underline{X} \setminus \underline{U})$
for all $h \in H$ do
      $q \leftarrow$ the lcm of all denominators of coefficients in $K(\underline{U})$ of $h$
      $h \leftarrow qh$
end
$f \leftarrow \mathrm{lcm}\{\mathrm{HC}(h) \mid h \in H\}$
$G \leftarrow$ a basis of $\mathrm{Id}(H) : f^\infty$ (by means of IDEALDIV2)
end CONT
```

Note that we needed no more than the first component of the pair that IDEALDIV2 outputs; the exponent $s \in \mathbb{N}$ that satisfies $\mathrm{Id}(H) : f^s = \mathrm{Id}(H) : f^\infty$ is irrelevant here. We will later on also write $\mathrm{CONT}(F, \underline{U})$ for $\mathrm{CONT}(F, \{U_1, \ldots, U_r\})$.

Recall from Lemma 1.122 that for an ideal I of $K[\underline{X}]$, it is true that $I \subseteq I^{\mathrm{ec}}$. To see that the reverse inclusion will fail in general in the present situation, consider the ideal $I = \mathrm{Id}(XY)$ of $K[X,Y]$. Then, taking the extension of I to $K(X)[Y]$,

$$\frac{1}{X} \cdot XY = Y \in I^{\mathrm{e}} \cap K[X,Y] = I^{\mathrm{ec}},$$

while clearly $Y \notin I$. The next proposition, which is preceded by a lemma, will show how I^{ec} relates to I. To this end, suppose $\leq_1$ and $\leq_2$ are term orders on $T(\underline{U})$ and $T(\underline{X} \setminus \underline{U})$, respectively, and let $\leq$ be the term order where for s_1, $t_1 \in T(\underline{U})$ and s_2, $t_2 \in T(\underline{X} \setminus \underline{U})$, we have $s_1 s_2 \leq t_1 t_2$ iff

$$\begin{array}{ccccc} s_2 & <_2 & t_2, & \text{or} & \\ s_2 & = & t_2 & \text{and} & s_1 \;\leq_1\; t_1. \end{array}$$

In the sequel, we will refer to this type of order as an **inverse block order** w.r.t. $\{U_1, \ldots, U_r\}$, or w.r.t. $\underline{U}$ for short. It is clear that here, with s_1, s_2, t_1, t_2 as above,

$$s_1 s_2 \leq t_1 t_2 \quad \text{implies} \quad s_2 \leq t_2.$$

Lemma 8.93 Let $\leq$ be an inverse block order on $T(\underline{X})$ w.r.t. $\underline{U}$, and suppose $G \subseteq K[\underline{X}]$ is a Gröbner basis w.r.t. $\leq$. Then G is a also a Gröbner basis in $K(\underline{U})[\underline{X} \setminus \underline{U}]$ w.r.t. the restriction $\leq'$ of $\leq$ to $T(\underline{X} \setminus \underline{U})$.

Proof Let I be the ideal generated by G in $K[\underline{X}]$ and J the one generated in $K(\underline{U})[\underline{X} \setminus \underline{U}]$. We must show that for every $f \in J$, there exists $g \in G$ with

$$\mathrm{HT}(g) \mid \mathrm{HT}(f),$$

where g too is viewed as an element of $K(\underline{U})[\underline{X} \setminus \underline{U}]$, and head terms are taken w.r.t. $\leq'$. So let $f \in J$. Then there is a representation of f as a sum of multiples in $K(\underline{U})[\underline{X} \setminus \underline{U}]$ of elements of G. All denominators occurring are in $K[\underline{U}]$, and we see that there is $q \in K[\underline{U}]$ with $qf \in I$. Since G is a Gröbner basis w.r.t. $\leq$ of the ideal I of $K[\underline{X}]$, there exists $g \in G$ with $\mathrm{HT}(g) \mid \mathrm{HT}(qf)$, where head terms are taken w.r.t. $\leq$. While we are viewing g and qf as elements of $K[\underline{X}]$, every term in $T(g)$ and $T(qf)$ can be written uniquely as st with

$$s \in T(\underline{U}) \quad \text{and} \quad t \in T(\underline{X} \setminus \underline{U}).$$

Now we view g and qf as elements of $K(\underline{U})[\underline{X} \setminus \underline{U}]$. This amounts to declaring the $T(\underline{U})$-part of each term to be part of the coefficient. From the remark on inverse block orders preceding the lemma, it is easy to see that now the head terms of g and qf w.r.t. $\leq'$ are precisely the $T(\underline{X} \setminus \underline{U})$-parts of what were the head terms before, in $K[\underline{X}]$ and w.r.t. $\leq$. It follows that the divisibility $\mathrm{HT}(g) \,|\, \mathrm{HT}(qf)$ continues to hold under the new point of view, in $K(\underline{U})[\underline{X} \setminus \underline{U}]$ and w.r.t. $\leq'$. Moreover, still under this point of view, q is a constant, and so $\mathrm{HT}(qf) = \mathrm{HT}(f)$, and we are done. □

Proposition 8.94 *Let $\leq$ be an inverse block order on $T(\underline{X})$ w.r.t. $\underline{U}$, and suppose I is an ideal of $K[\underline{X}]$ and G is a Gröbner basis of I w.r.t. $\leq$. Set*

$$f = \mathrm{lcm}\{\, \mathrm{HC}(g) \mid g \in G \,\},$$

where $\mathrm{HC}(g) \in K[\underline{U}]$ is taken of g as an element of $K(\underline{U})[\underline{X} \setminus \underline{U}]$ and w.r.t. the restriction $\leq'$ of $\leq$ to $T(\underline{X} \setminus \underline{U})$. Then $I^{\mathrm{ec}} = I : f^{\infty}$.

Proof It is an easy consequence of Lemma 1.122 (i) that G, when viewed as a subset of $K(\underline{U})[\underline{X} \setminus \underline{U}]$, generates the ideal I^{e} of $K(\underline{U})[\underline{X} \setminus \underline{U}]$. By the previous lemma, G is even a Gröbner basis of I^{e} in $K(\underline{U})[\underline{X} \setminus \underline{U}]$ w.r.t. the restriction $\leq'$ of $\leq$ to $T(\underline{X} \setminus \underline{U})$, and the first lemma of this section now asserts that $I^{\mathrm{ec}} = I : f^{\infty}$. □

The last proposition would allow us to compute I^{ec} from a Gröbner basis of I, but this is not what we will be interested in. Rather, we will need an ideal I' such that $I = I' \cap I^{\mathrm{ec}}$. This will be provided by the following lemma, which holds true in every commutative ring.

Lemma 8.95 Let R be a ring and I an ideal of R. Suppose $q \in R$ and $s \in \mathbb{N}$ are such that

$$I : q^s = I : q^{\infty}.$$

Then $I = \mathrm{Id}(I, q^s) \cap (I : q^s)$.

Proof The inclusion "$\subseteq$" is easily seen to be trivial. For the reverse inclusion, let $a \in \mathrm{Id}(I, q^s) \cap (I : q^s)$. Then $q^s a \in I$, and there exist $b \in I$ and $r \in R$ with $a = b + q^s r$. We conclude that

$$q^{2s} r = q^s a - q^s b \in I,$$

and thus $r \in I : q^{2s} \subseteq I : q^{\infty} = I : q^s$. It follows that $q^s r \in I$, and hence $a \in I$ because of the equation $a = b + q^s r$. □

Combining the last lemma and proposition, we see that in the situation of the proposition, we have

$$I = (I, f^s) \cap I^{\mathrm{ec}}$$

for any $s \in \mathbb{N}$ with $I : f^s = I : f^{\infty}$. Recall that the algorithm IDEALDIV2 computes such an s from f and any basis of I. We have proved the correctness of the following algorithm.

Proposition 8.96 *If $\{U_1, \ldots, U_r\}$ is a subset of $\{X_1, \ldots, X_n\}$ and F is a finite subset of $K[\underline{X}]$, then the algorithm* EXTCONT *of Table* 8.8 *computes $f \in K[\underline{U}]$ and $s \in \mathbb{N}$ such that*

$$\mathrm{Id}(F) = \mathrm{Id}(F, f^s) \cap \left(\mathrm{Id}(F)\right)^{\mathrm{ec}}. \quad \square$$

TABLE 8.8. Algorithm EXTCONT

```
Specification: $(f, s) \leftarrow$ EXTCONT$(F, \{U_1, \ldots, U_r\})$
               Cut $\mathrm{Id}(F)^{\mathrm{ec}}$ down to $\mathrm{Id}(F)$
Given: $F$ = a finite subset of $K[\underline{X}]$, and
       $\{U_1, \ldots, U_r\}$ = a subset of $\{X_1, \ldots, X_n\}$
Find: $f \in K[\underline{U}]$ and $s \in \mathbb{N}$ with $\mathrm{Id}(F) = \mathrm{Id}(F, f^s) \cap (\mathrm{Id}(F))^{\mathrm{ec}}$
begin
$\leq \leftarrow$ a decidable inverse block order on $T(\underline{X})$ w.r.t. $\underline{U}$
$\leq' \leftarrow$ the restriction of $\leq$ to $T(\underline{X} \setminus \underline{U})$
$G \leftarrow$ a Gröbner basis of $\mathrm{Id}(F)$ w.r.t. $\leq$
$f \leftarrow \mathrm{lcm}\{\, \mathrm{HC}(g) \mid g \in G \,\}$, where $\mathrm{HC}(g) \in K[\underline{U}]$ is taken of $g$ as an
    element of $K(\underline{U})[\underline{X} \setminus \underline{U}]$ and w.r.t. $\leq'$
$s \leftarrow$ a natural number with $\mathrm{Id}(F) : f^s = \mathrm{Id}(F) : f^\infty$
    (by means of IDEALDIV2)
return($(f, s)$)
end EXTCONT
```

Again, we will allow ourselves to write EXTCONT$(F, \underline{U})$ instead of EXTCONT$(F, \{U_1, \ldots, U_r\})$.

We can now describe more precisely the strategy for the computation of the radical and the primary decomposition of a polynomial ideal I. We are going to find a subset $\{U_1, \ldots, U_r\}$ of $\{X_1, \ldots, X_n\}$ which is maximally independent modulo I. We may then apply the methods of Sections 8.2 and 8.6 to the zero-dimensional ideal I^{e} of $K(\underline{U})[\underline{X} \setminus \underline{U}]$. We will then contract this radical or decomposition to $K[\underline{X}]$ and finally repeat the procedure with the ideal I' that satisfies $I = I' \cap I^{\mathrm{ec}}$. The only thing we still need to prove is that the concepts of radical, primary decomposition, contraction, and intersection are sufficiently compatible with one another.

Lemma 8.97 Let R be a ring and M a multiplicative subset of R. Taking extensions to R_M and contractions to R, the following hold:

(i) If J is an ideal of R_M, then $(\mathrm{rad}_{R_M}(J))^{\mathrm{c}} = \mathrm{rad}_R(J^{\mathrm{c}})$, where the subscripts mean that the first radical is taken in R_M and the second in R.

(ii) If I_1 and I_2 are ideals of R, then $\mathrm{rad}(I_1 \cap I_2) = \mathrm{rad}(I_1) \cap \mathrm{rad}(I_2)$.

(iii) If J is a primary ideal of R_M, then J^{c} is a primary ideal of R.

(iv) If J_1 and J_2 are ideals of R_M, then $(J_1 \cap J_2)^c = J_1^c \cap J_2^c$.

Proof (i) If $a \in (\mathrm{rad}_{R_M}(J))^c$, then $a^s \in J$ for some $s \in \mathbb{N}$, and $a \in R$. We see that a^s is actually in J^c, and thus $a \in \mathrm{rad}_R(J^c)$. Conversely, if $a \in \mathrm{rad}_R(J^c)$, then some power of a lies in $J^c \subseteq J$, which shows that $a \in \mathrm{rad}_{R_M}(J)$. Together with $a \in R$, this shows that $a \in (\mathrm{rad}_{R_M}(J))^c$.

(ii) If $a \in \mathrm{rad}(I_1 \cap I_2)$, then some power of a lies in both I_1 and I_2, and so $a \in \mathrm{rad}(I_1) \cap \mathrm{rad}(I_2)$. Conversely, if $a \in \mathrm{rad}(I_1) \cap \mathrm{rad}(I_2)$, then some power of a lies in I_1 and another one in I_2. The higher one then lies in $I_1 \cap I_2$, and we see that $a \in \mathrm{rad}(I_1 \cap I_2)$.

(iii) Suppose J is a primary ideal of R_M, and a and b are elements of R with $ab \in J^c$ and $a \notin J^c$. It follows that $ab \in J$ and $a \notin J$, and thus $a^s \in J$ for some $s \in \mathbb{N}$. But $a \in R$ implies $a^s \in R$, and so $a^s \in J^c$.

(iv) This is immediate from the definition of J^c as $J \cap R$. □

Exercise 8.98 Show that under the hypothesis of the previous lemma, the following hold:

(i) If I is an ideal of R, then $(\mathrm{rad}_R(I))^e = \mathrm{rad}_{R_M}(I^e)$.

(ii) If I is a primary ideal of R, then $I = I^{ec}$.

We are now in a position to give the algorithms that we were looking for. Recall that in Section 2.6, we proved that squarefree decompositions of univariate polynomials can be computed over computable fields that either have characteristic zero or are finite. Furthermore, we know from Corollary 7.37 that all fields of characteristic zero are perfect. The next theorem will require the hypothesis that squarefree decompositions can be computed over any rational function field $K(\underline{U})$, and also that the latter is perfect. This is thus true whenever K is computable and has characteristic zero. Recall further that we can determine maximally independent sets modulo an ideal by means of lexicographical Gröbner bases; a much more elegant way to achieve this will be presented in Section 9.3. Finally, we point out that the following algorithm uses the obvious convention that $K(\emptyset) = K$.

Theorem 8.99 *Assume that K is computable, and suppose that for any finite set $\{U_1, \ldots, U_r\}$ of indeterminates, the rational function field $K(\underline{U})$ is perfect and allows the computation of squarefree decompositions of univariate polynomials. Then the algorithm* RADICAL *of Table* 8.9 *computes a basis of* $\mathrm{rad}(\mathrm{Id}(F))$ *for any given finite subset F of* $K[\underline{X}]$.

Proof *Termination*: If $1 \in \mathrm{Id}(F)$, then the algorithm terminates trivially. Else, we note that by the choice of $\{U_1, \ldots, U_r\}$, we have $\mathrm{Id}(F) \cap K[\underline{U}] = \{0\}$. It follows that the inclusion $\mathrm{Id}(F) \subseteq \mathrm{Id}(F, f)$ is proper. We see that the recursive calls of RADICAL give rise to a strictly ascending chain of ideals, which can not be infinite since $K[\underline{X}]$ is noetherian.

Correctness: Like every correctness proof of an algorithm that calls itself recursively, the proof is by (noetherian) induction: we show that the

TABLE 8.9. Algorithm RADICAL

```
Specification: $G \leftarrow$ RADICAL($F$)
               Computation of radical
Given: a finite subset $F$ of $K[\underline{X}]$
Find: a finite basis $G$ of rad(Id($F$))
begin
$G \leftarrow \{1\}$
if $1 \notin$ Id($F$) then
   $\{U_1, \dots, U_r\} \leftarrow$ a maximally independent set modulo Id($F$)
   $Z \leftarrow$ ZRADICAL($F$), computed in $K(\underline{U})[\underline{X} \setminus \underline{U}]$
   $C \leftarrow$ CONT($Z, \underline{U}$)
   $f \leftarrow$ an element of $K[\underline{U}]$ with $\mathrm{Id}(F) = \mathrm{Id}(F, f^s) \cap (\mathrm{Id}(F))^{\mathrm{ec}}$
        for some $s \in \mathbb{N}$ (by means of EXTCONT)
   $G \leftarrow$ INTERSECTION(RADICAL($F \cup \{f\}$), $C$)
end
end RADICAL
```

algorithm runs correctly if $1 \in \mathrm{Id}(F)$ and then prove correctness for the case that $\mathrm{Id}(F)$ is proper under the assumption of correctness for all larger ideals. The algorithm is trivially correct if $1 \in \mathrm{Id}(F)$. Else, we must show that

$$\mathrm{rad}(\mathrm{Id}(F)) = \mathrm{rad}(\mathrm{Id}(F, f)) \cap \mathrm{Id}(C)$$

As we have mentioned before, Lemma 1.122 (i) easily implies that F, when viewed as a subset of $K(\underline{U})[\underline{X} \setminus \underline{U}]$, generates the ideal

$$(\mathrm{Id}(F))^{\mathrm{e}} \quad \text{of} \quad K(\underline{U})[\underline{X} \setminus \underline{U}].$$

We may now conclude from Lemma 7.47 (ii) and Theorem 8.22 that Z is a basis of

$$\mathrm{rad}\Big((\mathrm{Id}(F))^{\mathrm{e}}\Big).$$

It follows that

$$\mathrm{Id}(C) = \left(\mathrm{rad}\Big((\mathrm{Id}(F))^{\mathrm{e}}\Big)\right)^{\mathrm{c}} = \mathrm{rad}\Big((\mathrm{Id}(F))^{\mathrm{ec}}\Big),$$

the latter equation being true by (i) of the last lemma. Now f is computed such that $\mathrm{Id}(F) = \mathrm{Id}(F, f^s) \cap (\mathrm{Id}(F))^{\mathrm{ec}}$ for some $s \in \mathbb{N}$, and we obtain

$$\begin{aligned} \mathrm{rad}(\mathrm{Id}(F)) &= \mathrm{rad}\Big(\mathrm{Id}(F, f^s) \cap (\mathrm{Id}(F))^{\mathrm{ec}}\Big) \\ &= \mathrm{rad}(\mathrm{Id}(F, f)) \cap \mathrm{rad}\Big((\mathrm{Id}(F))^{\mathrm{ec}}\Big) \\ &= \mathrm{rad}(\mathrm{Id}(F, f)) \cap \mathrm{Id}(C). \end{aligned}$$

Note how in the passage from the first to the second line above, we have used the obvious fact that $\mathrm{rad}(\mathrm{Id}(F, f^s)) = \mathrm{rad}(\mathrm{Id}(F, f))$; this is the reason why we did not have to let EXTCONT provide the exponent s. □

It is easy to see that in case the maximally independent set that RADICAL chooses at the beginning of the **if**-block is empty, the last three lines of that block have no effect on $\mathrm{Id}(Z)$; we may therefore let the algorithm return Z in that case. It is perhaps noteworthy that the algorithm ZRADICAL for the computation of zero-dimensional ideals applies to all computable finite fields, while RADICAL does not: we saw in Example 7.32 that the rational function field $\mathbb{Z}/p\mathbb{Z}(T)$ is not perfect.

Exercise 8.100 Compute the radical of the ideal $\mathrm{Id}(X^2+2XYZ+Z^4, YZ-Z^2)$ of $\mathbb{Q}[X, Y, Z]$.

The following algorithm PRIMDEC pursues much the same strategy as RADICAL, except that on the extended level, it computes zero-dimensional primary decompositions rather than radicals. This will of course require that the hypotheses that ZPRIMDEC needs are satisfied for every rational function field $K(\underline{U})$ over K. We note that this is the case for $K = \mathbb{Q}$: the rational function field $\mathbb{Q}(\underline{U})$ is clearly computable and infinite, it is perfect because it has characteristic zero, and Corollary 2.104 applied with $R = \mathbb{Z}$ and $i = n - 1$ tells us that it allows effective factorization of univariate polynomials.

Theorem 8.101 *Assume that K is computable, and suppose that for any finite set $\{U_1, \ldots, U_r\}$ of indeterminates, the rational function field $K(\underline{U})$ is infinite and perfect and allows effective factorization of univariate polynomials. Suppose F is a finite subset of $K[\underline{X}]$ which generates a proper ideal. Then the algorithm* PRIMDEC *of Table* 8.10 *computes a finite set of primary ideals with intersection* $\mathrm{Id}(F)$ *as well as the corresponding associated primes.*

Proof *Termination*: The proof is essentially the same as in the case of RADICAL. We have $\mathrm{Id}(F) \cap K[\underline{U}] = \{0\}$ while $f \in K[\underline{U}]$, and so the inclusion $\mathrm{Id}(F) \subsetneq \mathrm{Id}(F, f^s)$ is proper. An infinite sequence of recursive calls would thus contradict the fact that $K[\underline{X}]$ is noetherian.

Correctness: The structure of the proof is again similar to that of the correctness proof for RADICAL. Correctness is trivial if $1 \in \mathrm{Id}(F)$. Now let $\mathrm{Id}(F)$ be proper, and assume that correctness is guaranteed for all sets that generate a strictly larger ideal. We begin by proving properties (i) and (iii) as stated in the algorithm. We must show that after the algorithm has performed its tasks, the following hold:

(1) For all $(G, H) \in C$, the ideal $\mathrm{Id}(G)$ is primary with associated prime $\mathrm{Id}(H)$.

(2) $\mathrm{Id}(F) = \mathrm{Id}(F, f^s) \cap \bigcap_{(G,H)\in C} \mathrm{Id}(G)$.

TABLE 8.10. Algorithm PRIMDEC

```
Specification: $P \leftarrow$ PRIMDEC($F$)
                    Computation of primary decomposition
Given: a finite subset $F$ of $K[\underline{X}]$
Find: a set $P$ of pairs $(G, H)$ of finite subsets of $K[\underline{X}]$ such that $P = \emptyset$
      if $1 \in \mathrm{Id}(F)$, while otherwise
        (i) for all $(G, H) \in P$, the ideal Id($G$) is primary with associated
            prime Id($H$),
       (ii) $\mathrm{Id}(G_1) \neq \mathrm{Id}(G_2)$ and $\mathrm{Id}(H_1) \neq \mathrm{Id}(H_2)$ whenever
            $(G_1, H_1), (G_2, H_2) \in P$ with $(G_1, H_1) \neq (G_2, H_2)$, and
      (iii) $\mathrm{Id}(F) = \bigcap_{(G,H) \in P} \mathrm{Id}(G)$.
begin
P $\leftarrow \emptyset$
if $1 \notin \mathrm{Id}(F)$ then
  $\{U_1, \ldots, U_r\} \leftarrow$ a maximally independent set modulo Id($F$)
  $Q \leftarrow$ ZPRIMDEC($F$), computed in $K(\underline{U})[\underline{X} \setminus \underline{U}]$
  $C \leftarrow \emptyset$
  while $Q \neq \emptyset$ do
        select $(A, B)$ from $Q$
        $Q \leftarrow Q \setminus \{(A, B)\}$
        $G \leftarrow$ CONT($A, \underline{U}$)
        $H \leftarrow$ CONT($B, \underline{U}$)
        $C \leftarrow C \cup \{(G, H)\}$
  end
  $(f, s) \leftarrow$ EXTCONT($F$)
  $P \leftarrow C \cup$ PRIMDEC($F \cup \{f^s\}$)
end
end PRIMDEC
```

Recall that Lemma 1.122 (i) implies that F, when viewed as a subset of $K(\underline{U})[\underline{X}\setminus\underline{U}]$, generates the ideal $(\mathrm{Id}(F))^e$ of $K(\underline{U})[\underline{X}\setminus\underline{U}]$. The application of ZPRIMDEC yields a set Q of pairs of finite subsets of $K(\underline{U})[\underline{X}\setminus\underline{U}]$ such that

$$\{ A \mid (A, B) \in Q \text{ for some } B \}$$

is the set of primary components of the zero-dimensional ideal $(\mathrm{Id}(F))^e$, and $(A, B) \in Q$ means that

$$\mathrm{Id}(B) = \mathrm{rad}(\mathrm{Id}(A)).$$

The **while**-loop contracts the elements of the pairs in Q to $K[\underline{X}]$ and assembles them into pairs again. These are then collected in the set C. Property (1) above is now an immediate consequence of Lemma 8.97 (i) and (iii). Furthermore, Lemma 8.97 (iv) tells us that

$$\begin{aligned}(\mathrm{Id}(F))^{\mathrm{ec}} &= \Big(\bigcap_{(A,B)\in Q} \mathrm{Id}(A)\Big)^{\mathrm{c}}\\ &= \bigcap_{(A,B)\in Q} (\mathrm{Id}(A))^{\mathrm{c}}\\ &= \bigcap_{(G,H)\in C} \mathrm{Id}(G).\end{aligned}$$

Finally, f and s are computed such that $\mathrm{Id}(F) = \mathrm{Id}(F, f^s) \cap (\mathrm{Id}(F))^{\mathrm{ec}}$. Together, we obtain

$$\begin{aligned}\mathrm{Id}(F) &= \mathrm{Id}(F, f^s) \cap (\mathrm{Id}(F))^{\mathrm{ec}}\\ &= \mathrm{Id}(F, f^s) \cap \bigcap_{(G,H)\in C} \mathrm{Id}(G).\end{aligned}$$

It remains to prove that property (ii) as stated under "**Find**" holds for the output P. To this end, we first prove that it is true for any two different pairs in C. Let (G_1, H_1), $(G_2, H_2) \in C$ be different pairs. We first note that since ZPRIMDEC computes primary decompositions, (A_1, B_1), $(A_2, B_2) \in Q$ with $(A_1, B_1) \neq (A_2, B_2)$ implies that $\mathrm{Id}(A_1) \neq \mathrm{Id}(A_2)$ since otherwise the decomposition would be redundant, and thus $\mathrm{Id}(B_1) \neq \mathrm{Id}(B_2)$ because otherwise the associated primes of two different primary ideals of the decomposition would coincide. From the construction of C as the "contraction of Q" we see that there must exist two different pairs (A_1, B_1), $(A_2, B_2) \in Q$ with

$$\mathrm{Id}(G_i) = (\mathrm{Id}(A_i))^{\mathrm{c}} \quad \text{and} \quad \mathrm{Id}(H_i) = (\mathrm{Id}(B_i))^{\mathrm{c}} \quad \text{for } i = 1, 2.$$

Now $\mathrm{Id}(A_1)$ and $\mathrm{Id}(A_2)$ are two different ideals of $K(\underline{U})[\underline{X}\backslash\underline{U}]$, and so their contractions G_1 and G_2 must be different by Lemma 1.124 (ii). Similarly, we must have $H_1 \neq H_2$.

Finally, suppose $(G_1, H_1) \in C$ and $(G_2, H_2) \in \mathrm{PRIMDEC}(F, f^s)$. Then $f^s \notin \mathrm{Id}(G_1)$ and $f \notin \mathrm{Id}(H_1)$ because both ideals are contractions of proper ideals of $K(\underline{U})[\underline{X}\backslash\underline{U}]$, while f^s, $f \in K[\underline{U}]$. On the other hand, $f^s \in \mathrm{Id}(G_2)$ because G_2 is an ideal that contains $\mathrm{Id}(F, f^s)$, and $f \in \mathrm{Id}(H_2)$ because the latter ideal is the radical of $\mathrm{Id}(G_1)$. We see that once again, $\mathrm{Id}(G_1) \neq \mathrm{Id}(G_2)$ and $\mathrm{Id}(H_1) \neq \mathrm{Id}(H_2)$. □

If the maximally independent set that PRIMDEC chooses at the beginning of the **if**-block is empty, then we may, in analogy to the same situation in the algorithm RADICAL, let PRIMDEC return ZPRIMDEC(F).

Exercise 8.102 Compute primary ideals whose intersection equals the ideal

$$\mathrm{Id}(X^2 - Z^2 - 6Z - 9, YZ - 2Y - Z^2 + 2Z)$$

of $\mathbb{Q}[X,Y,Z]$. Check your answer by doing it again with a different choice for the first maximally independent set.

For the output of PRIMDEC to be a primary decomposition, we would also have to know that none of the primary ideals in the output is redundant, i.e., contains the intersection of the rest. The following example shows that this is not the case in general, and that there does not seem to be a natural way to force it to happen.

Example 8.103 Let $F = \{XY, XZ\} \subseteq \mathbb{Q}[X,Y,Z]$, and suppose PRIMDEC is applied to F. One possible choice for $\{U_1,\dots,U_r\}$ would be $\{X\}$. The algorithm would then find the primary ideal $\mathrm{Id}(Y,Z)$ on the extended level, and it would naturally take $f = X$ and $s = 1$, which would give $\mathrm{Id}(F \cup \{f^s\}) = \mathrm{Id}(X)$. The resulting decomposition

$$\mathrm{Id}(XY,XZ) = \mathrm{Id}(Y,Z) \cap \mathrm{Id}(X)$$

is clearly a primary decomposition. On the other hand, there is no way to keep the algorithm from taking the course that is exhibited in the following table. (Here, the first line describes the original call of PRIMDEC with $I = \mathrm{Id}(F)$, while each subsequent line represents the recursive call on the ideal $\mathrm{Id}(I, f^s)$ of the previous line.)

$\{U_1,\dots,U_r\}$	I^{ec}	f	s	$\mathrm{Id}(I,f^s)$
$\{Y,Z\}$	$\mathrm{Id}(X)$	YZ	1	$\mathrm{Id}(XY,XZ,YZ)$
$\{Z\}$	$\mathrm{Id}(X,Y)$	Z	1	$\mathrm{Id}(XY,Z)$
$\{Y\}$	$\mathrm{Id}(X,Z)$	Y	1	$\mathrm{Id}(Y,Z)$
$\{X\}$	$\mathrm{Id}(Y,Z)$	1	0	$\mathbb{Q}[X,Y,Z]$

The primary ideals of the output are now to be found in the column "I^{ec}," and we see that the two ideals $\mathrm{Id}(X,Y)$ and $\mathrm{Id}(X,Z)$ have been added gratuitously.

The algorithm PRIMDEC also sheds some light on why and how primary decompositions are not unique in general. An example that we have mentioned several times in connection with the uniqueness theorems in Section 8.5 is

$$\mathrm{Id}(X^2,XY) = \mathrm{Id}(X) \cap \mathrm{Id}(X^2,Y) = \mathrm{Id}(X) \cap \mathrm{Id}(X^2,XY,Y^2)$$

in $K[X,Y]$. The algorithm PRIMDEC applied to $\mathrm{Id}(X^2,XY)$ would have to start with $\{U_1,\dots,U_r\} = \{Y\}$. It would find the ideal $\mathrm{Id}(X)$ on the extended level, and it would naturally take $f = Y$. The least $s \in \mathbb{N}$ with

$$\mathrm{Id}(X^2,XY) : Y^s = \mathrm{Id}(X^2,XY) : Y^\infty$$

is $s = 1$, and this choice would result in the primary decomposition

$$\mathrm{Id}(X^2,XY) = \mathrm{Id}(X) \cap \mathrm{Id}(X^2,Y).$$

However, any $s > 1$ is just as good, and we obtain the infinitely many decompositions

$$\mathrm{Id}(X^2, XY) = \mathrm{Id}(X) \cap \mathrm{Id}(X^2, XY, Y^s) \qquad (s \in \mathbb{N}^+).$$

Exercise 8.104 Write an algorithm that decomposes a given polynomial ideal I into an intersection of finitely many proper ideals $D_1, \ldots, D_r$ such that

(i) for each $1 \leq i \leq r$, all primary components of D_i have the same dimension, and

(ii) $\dim(D_i) \neq \dim(D_j)$ for all $1 \leq i < j \leq r$.

Such a decomposition is called an *unmixed decomposition.*

8.8 Computing Real Zeroes of Polynomial Systems

In this section, we will exclusively consider polynomials over the field $\mathbb{Q}$ of rational numbers. As usual, we write $\mathbb{Q}[\underline{X}]$ for $\mathbb{Q}[X_1, \ldots, X_n]$. Perhaps the most important problem to which the theory of Gröbner bases makes a contribution is the computation of the common zeroes in $\mathbb{R}^n$ of a finite set of polynomials $F \subseteq \mathbb{Q}[\underline{X}]$, i.e., the computation of the real solutions of a system of possibly non-linear equations

$$\begin{aligned} f_1(X_1, \ldots, X_n) &= 0 \\ f_2(X_1, \ldots, X_n) &= 0 \\ &\vdots \\ f_m(X_1, \ldots, X_n) &= 0 \end{aligned}$$

with $f_i(X_1, \ldots, X_n) \in \mathbb{Q}[\underline{X}]$ for $1 \leq i \leq m$. What we are looking for is of course the set of real zeroes of the ideal $\mathrm{Id}(F)$. If $\dim(\mathrm{Id}(F)) > 0$, then by Proposition 8.27, $\mathrm{Id}(F)$ has infinitely many zeroes in $\mathbb{C}^n$ and thus possibly in $\mathbb{R}^n$. The problem of computing the real zeroes is thus not really a meaningful one unless it is further qualified. One could, for example, specify a rational value for each one in a set of maximally independent variables in such a way that substitution of these values leads to a zero-dimensional ideal. (In view of Lemma 7.50, it is an easy exercise to prove that there are always infinitely many such substitutions.) This takes us to the case which we will be discussing here, namely, the case $\dim(\mathrm{Id}(F)) = 0$. Proposition 8.27 tells us that there can then be at most finitely many real zeroes of $\mathrm{Id}(F)$.

If one is interested in no more than the *rational* zeroes of $\mathrm{Id}(F)$, then the problem can be solved as follows. First, compute a Gröbner basis G of $\mathrm{Id}(F)$ w.r.t. a lexicographical term order, say the one where

$$X_1 \ll \cdots \ll X_n.$$

Then compute the rational zeroes of the univariate polynomial f in X_1, using the well-known method of clearing denominators and then trying all linear factors $pX - q$ where p and q are divisors of the head coefficient and the constant coefficient, respectively, of f. Substitute each of these into each element g of $G \cap \mathbb{Q}[X_1, X_2]$, and compute the rational zeroes of the resulting univariate polynomials in X_2. The rational zeroes of $G \cap \mathbb{Q}[X_1, X_2]$ are then precisely the pairs (q_1, q_2), where q_1 was a zero of f and q_2 was a common zero of

$$\{ g(q_1, X_2) \mid g \in G \cap \mathbb{Q}[X_1, X_2] \}.$$

It is easy to see that the rational zeroes of $\mathrm{Id}(G) = \mathrm{Id}(F)$ can be found by continuing this process of iterated substitutions in the obvious manner.

Exercise 8.105 Write an algorithm that computes the set of all rational zeroes of $\mathrm{Id}(F)$ for any given finite subset F of $\mathbb{Q}[\underline{X}]$. Prove correctness and termination.

In order to understand the problem of finding the *real* zeroes of a polynomial ideal over $\mathbb{Q}$, we must first clarify what happens in the univariate case. So let $f \in \mathbb{Q}[X]$, and suppose we want to know for which $\alpha \in \mathbb{R}$ we have $f(\alpha) = 0$. There is an algorithm which for those f with $\deg(f) \leq 4$ computes all real zeroes of f in terms of radicals depending on the coefficients of f. (This was essentially known in the sixteenth century.) One of the major achievements of modern mathematics is the proof—given by N.H. Abel (1802–1829)—that there cannot exist such an algorithm for any degree greater than or equal to 5. There is, however, a host of numerical methods for the *approximation* of real zeroes of univariate polynomials. One might therefore consider to simply imitate the procedure for the computation of rational zeroes as described above, working with approximations instead of precise solutions. To see where the catch lies, suppose we have computed an approximation $q \in \mathbb{Q}$ of a zero $\alpha \in \mathbb{R}$ of a univariate polynomial in X_1 and are now substituting q into a bivariate polynomial g in X_1 and X_2 with the intention of computing approximations of the zeroes of $g(q, X_2)$. The fact that q was no more than an approximation will of course lead to an error propagation. Much worse, however, the variation of the coefficients of $g(q, X_2)$ which is caused by the shift from α to q may radically and uncontrollably change the behavior of $g(q, X_2)$ as far as real zeroes are concerned. It could for example happen that the number of such zeroes changes from many to none, thus rendering any result of the procedure meaningless. (See also the Notes to this chapter on p. 419.)

The aim of this section is to demonstrate that there is a method to determine the real zeroes of a zero-dimensional ideal over $\mathbb{Q}$ to a degree of "certainty" which, in view of Abel's result on unsolvability, cannot be further improved. We will first exhibit an algebraic method of approximating the real zeroes of a univariate polynomial f in the following sense: we will show how one can compute the rational endpoints of pairwise disjoint intervals on the real line such that each interval contains exactly one real zero of f, and every real zero of f lies in one of the intervals. Moreover, the maximal length of these intervals can be prescribed to be arbitrarily small. Such a set of intervals will be called a set of *isolating intervals* for the real zeroes of f. For a multivariate zero-dimensional ideal I, we must first recall that rather obviously, the set of real zeroes of I is a subset of the set

$$N = \{\, (\alpha_1, \dots \alpha_n) \in \mathbb{R}^n \mid f_i(\alpha_i) = 0 \text{ for } 1 \leq i \leq n \,\},$$

where f_i is a generator of $I \cap \mathbb{Q}[X_i]$ for $1 \leq i \leq n$. We will first compute sets $R_1, \dots, R_n$ of isolating intervals for the real zeroes of $f_1, \dots, f_n$, respectively. Then there is an obvious bijection between N and the set

$$M = \{\, ([q_1, r_1], \dots, [q_n, r_n]) \mid [q_i, r_i] \in R_i \text{ for } 1 \leq i \leq n \,\}.$$

We will solve the problem of finding the real zeroes of f by giving an algorithm that selects those n-tuples from M whose corresponding element of N is in fact a zero of I. We mention that due to its practical importance, the problem of solving polynomial systems efficiently is the object of vigorous mathematical research; here, we propose to prove no more than the non-trivial fact that the problem is solvable at all.

We begin by discussing the univariate case, i.e., the computation of the real zeroes of a univariate polynomial over $\mathbb{Q}$ in terms of isolating intervals as described above. The algorithm that achieves this rests upon a classical result known as *Sturm's theorem*, which we will now discuss. Sturm's theorem will allow us to compute the number of real zeroes of a polynomial in a given interval. We use the notation for intervals that distinguishes open intervals from pairs:

$$[\alpha, \beta] = \{\, \gamma \in \mathbb{R} \mid \alpha \leq \gamma \leq \beta \,\} \quad \text{and} \quad]\alpha, \beta[= \{\, \gamma \in \mathbb{R} \mid \alpha < \gamma < \beta \,\}.$$

We will use several results from calculus, namely, the continuity of polynomials as functions from $\mathbb{R}$ to $\mathbb{R}$, the intermediate value theorem, and, later on, the mean value theorem.

Definition 8.106 Let $f \in \mathbb{R}[X]$ and α, $\beta \in \mathbb{R}$ with $\alpha \leq \beta$. A **Sturm sequence** for f and $[\alpha, \beta]$ is an $(r+1)$-tuple $(f_0, \dots, f_r) \in (\mathbb{R}[X])^{r+1}$ such that the following hold:

S0 $\{\, \gamma \in \mathbb{R} \mid f(\gamma) = 0 \,\} = \{\, \gamma \in \mathbb{R} \mid f_0(\gamma) = 0 \,\}$.

S1 $f_r(\gamma) \neq 0$ for all $\gamma \in [\alpha, \beta]$.

S2 $f_0(\alpha) \cdot f_0(\beta) \neq 0$.

S3 If $0 < i < r$ and $\gamma \in [\alpha, \beta]$ with $f_i(\gamma) = 0$, then $f_{i-1}(\gamma) \cdot f_{i+1}(\gamma) < 0$.

S4 Whenever $\gamma \in]a, b[$ with $f_0(\gamma) = 0$, then there exist γ_1, $\gamma_2 \in \mathbb{R}$ with $\gamma_1 < \gamma$ and $\gamma < \gamma_2$ such that $f_0(\delta) \cdot f_1(\delta) < 0$ for all $\delta \in]\gamma_1, \gamma[$, and $f_0(\delta) \cdot f_1(\delta) > 0$ for all $\delta \in]\gamma, \gamma_2[$.

The first and last element of a Sturm sequence cannot be the zero polynomial by properties S1 and S2. The next lemma states that if an intermediate polynomial is the zero polynomial, then it may be dropped from the sequence.

Lemma 8.107 Let $f \in \mathbb{R}[X]$ and $\alpha, \beta \in \mathbb{R}$ with $\alpha \leq \beta$, and let $(f_0, \ldots, f_r)$ be a Sturm sequence for f and $[\alpha, \beta]$. Assume further that $0 < i < r$ is an index with $f_i = 0$. Then the sequence

$$(f_0, \ldots, f_{i-1}, f_{i+1}, \ldots, f_r)$$

is still a Sturm sequence for f and $[\alpha, \beta]$.

Proof It is clear that conditions S0, S1, and S2 continue to hold for the shorter sequence. From S3 applied to the original sequence, it follows that f_{i-1} and f_{i+1} have no zeroes in $[\alpha, \beta]$. This means that if we test the shortened sequence for compliance with S3, then for f_{i-1} and f_{i+1}, its hypothesis will never be satisfied. It now follows easily that S3 continues to hold. Condition S4 is critical only if $i = 1$. But we have already argued that then f_0 has no zeroes in $[\alpha, \beta]$, and so S4 is rendered moot for the shortened sequence. □

It will be shown below how a Sturm sequence can be computed if the coefficients of f are given rational numbers. Let us first show how a Sturm sequence for a polynomial f and an interval $[\alpha, \beta]$ codes the information how many zeroes of f there are in $[\alpha, \beta]$. To this end, we define the *number of sign changes*, or *variations in sign*, of an $(r+1)$-tuple $(\alpha_0, \ldots, \alpha_r)$ of real numbers as the output of the "algorithm" VARSIGN of Table 8.11 which becomes an actual algorithm when applied to an $(r+1)$-tuple of rational numbers. For real numbers, it amounts to a definition by recursion on r. Loosely speaking, the idea is to drop all zeroes and then to count the sign changes as one passes through the tuple.

Exercise 8.108 Show that VARSIGN$((1, 0, 0, -1, 2, 3, 0, 1, 0, 0, 0, -1)) = 3$.

Proposition 8.109 *Let $f \in \mathbb{R}[X]$ and α, $\beta \in \mathbb{R}$ with $\alpha \leq \beta$. Assume further that $(f_0, \ldots, f_r)$ is a Sturm sequence for f and $[\alpha, \beta]$. Then the number of distinct zeroes of f in the interval $[\alpha, \beta]$ equals*

$$\mathrm{VARSIGN}\big((f_0(\alpha), \ldots, f_r(\alpha))\big) - \mathrm{VARSIGN}\big((f_0(\beta), \ldots, f_r(\beta))\big).$$

Proof For simplicity, we will write

$$V_\rho = \mathrm{VARSIGN}\big((f_0(\rho), \ldots, f_r(\rho))\big)$$

whenever $\rho \in \mathbb{R}$. We first note that if we drop all zero polynomials from the sequence, then the result is still a Sturm sequence for f and $[\alpha, \beta]$ by Lemma 8.107. Furthermore, the numbers of sign changes of the theorem are unaffected by dropping zero entries, and so we may assume w.l.o.g. that

TABLE 8.11. "Algorithm" VARSIGN

```
Specification: $v \leftarrow$ VARSIGN($(\alpha_0, \dots, \alpha_r)$)
               Definition and computation of number of variations
               in sign
Given: $(\alpha_0, \dots, \alpha_r) \in \mathbb{R}^{r+1}$ with $\alpha_0 \neq 0$
Define and find: the number $v$ of variations in sign of $(\alpha_0, \dots, \alpha_r)$
begin
$v \leftarrow 0$;  $A \leftarrow \alpha_0$;  $i \leftarrow 1$
while $i \leq r$ do
    repeat $B \leftarrow \alpha_i$
           $i \leftarrow i + 1$
    until $B \neq 0$ or $i > r$
    if $A \cdot B < 0$ then $v \leftarrow v + 1$ end
    $A \leftarrow B$
end
end VARSIGN
```

our Sturm sequence contains no zero polynomials. It is clear from S0 that it suffices to discuss the number of zeroes of f_0. Corollary 2.97 tells us that the set

$$N = \{\, \rho \in [\alpha, \beta] \mid f_i(\rho) = 0 \text{ for some } 0 \leq i \leq r \,\}$$

has finitely many elements. We prove the theorem by induction on $|N|$. If $|N| = 0$, then the number of zeroes that we are looking for is 0. But for $0 \leq i \leq r$, the two values $f_i(\alpha)$ and $f_i(\beta)$ are not zero and, by the intermediate value theorem, have the same sign, so that $V_\alpha = V_\beta$ as desired. Next, let $|N| = 1$, say $N = \{\gamma\}$.

Case 1: $f_0(\gamma) \neq 0$, i.e., f_0 has no zero in $[\alpha, \beta]$.

Set

$$J = \{\, i \in \{0, \dots, r\} \mid f_i(\gamma) = 0 \,\} \subseteq \{1, \dots, r-1\}.$$

Now whenever $i \notin J$, then from the intermediate value theorem and the fact that $|N| = 1$, we see that f_i cannot change its sign anywhere on $[\alpha, \beta]$. Moreover, if $i \in J$, then by S3 of the definition of a Sturm sequence, $i - 1$ and $i + 1$ are not in J, and $f_{i-1}(\gamma)$ and $f_{i+1}(\gamma)$ have opposite signs. But we just saw that f_{i-1} and f_{i+1} cannot change their sign anywhere on $[\alpha, \beta]$, and so their signs at α and β are the same as at γ. We can now describe the difference in passing from

$$(f_0(\alpha), \dots, f_r(\alpha)) \quad \text{to} \quad (f_0(\beta), \dots, f_r(\beta))$$

as follows. The only entries that can possibly change their sign, or change from zero to non-zero or vice versa, have indices i with $1 \leq i \leq r - 1$. Each such entry is flanked by two non-zero entries with opposite signs, none of which changes its sign. A simple case distinction, as indicated below, shows that the number of variations in sign has remained the same, so that $V_\alpha = V_\beta$ as desired.

$$\begin{array}{ccccccc}
++- & \longrightarrow & +-- & \qquad & -++ & \longrightarrow & --+ \\
++- & \longrightarrow & +0- & & -++ & \longrightarrow & -0+ \\
+-- & \longrightarrow & +0- & & --+ & \longrightarrow & -0+ \\
+-- & \longrightarrow & ++- & & --+ & \longrightarrow & -++ \\
+0- & \longrightarrow & ++- & & -0+ & \longrightarrow & -++ \\
+0- & \longrightarrow & +-- & & -0+ & \longrightarrow & --+
\end{array}$$

Case 2: $f_0(\gamma) = 0$.
Here, we first note that S3 implies that $f_1(\gamma) \neq 0$, and thus, once again from the intermediate value theorem and the fact that $|N| = 1$, it follows that $f_1(\alpha)$ and $f_1(\beta)$ have the same sign. Now if we consider the sequences

$$\bigl(f_1(\alpha), \dots, f_r(\alpha)\bigr) \quad \text{and} \quad \bigl(f_1(\beta), \dots, f_r(\beta)\bigr),$$

then we can use property S3 in the exact same way as in Case 1 to argue that these two sequences have the same number of sign changes. So in order to prove the desired result $V_\alpha - V_\beta = 1$, we must prove that $f_0(\alpha)$ and $f_1(\alpha)$ have opposite signs, while $f_0(\beta)$ and $f_1(\beta)$ have the same sign. To this end, we first note that by S2, we must have $\gamma \in \,]\alpha, \beta[$. Condition S4 tells us that in a sufficiently small interval $]\gamma_1, \gamma[$ to the left, f_0 and f_1 have opposite signs, while in a sufficiently small interval $]\gamma, \gamma_2[$ to the right, they have the same sign. But neither one of the two polynomials has a zero in $[\alpha, \beta] \setminus \{\gamma\}$, and so by the intermediate value theorem, these relationships between the signs continue to hold at α to the left and at β to the right, respectively.

Finally, let $|N| > 1$. If $\alpha_1, \dots, \alpha_{|N|}$ are the elements of N in ascending order, then we have

$$\alpha \leq \alpha_1 < \cdots < \alpha_{|N|} \leq \beta,$$

and we may choose $\gamma \in \,]\alpha_1, \alpha_2[$. One easily proves by inspection of the definition that $(f_0, \dots, f_r)$ is still a Sturm sequence for f and $[\alpha, \gamma]$, and also for f and $[\gamma, \beta]$. The sets

$$N_1 = \bigl\{ \rho \in [\alpha, \gamma] \bigm| f_i(\rho) = 0 \text{ for some } 0 \leq i \leq r \bigr\}$$

and

$$N_2 = \bigl\{ \rho \in [\gamma, \beta] \bigm| f_i(\rho) = 0 \text{ for some } 0 \leq i \leq r \bigr\}$$

have 1 and $|N| - 1$ elements, respectively. The induction hypothesis tells us that the number of zeroes of f in $[\alpha, \gamma]$ equals $V_\alpha - V_\gamma$, and the number of zeroes of f in $[\gamma, \beta]$ equals $V_\gamma - V_\beta$. It follows that the number of zeroes of f in $[\alpha, \beta]$ equals

$$(V_\alpha - V_\gamma) + (V_\gamma - V_\beta) = V_\alpha - V_\beta. \quad \square$$

Next, we show that Sturm sequences always exist and how they may be computed.

Proposition 8.110 *Let* $0 \neq f \in \mathbb{R}[X]$ *with* $\deg(f) > 0$. *Then there exists an* $(r+1)$*-tuple*

$$(f_0, \ldots, f_r) \in (\mathbb{R}[X])^{r+1}$$

with f_0 *squarefree and* $f_r = 1$ *which is a Sturm sequence for* f *and every interval* $[\alpha, \beta]$ *with* α, $\beta \in \mathbb{R}$ *satisfying* $\alpha \leq \beta$ *and* $f(\alpha)$, $f(\beta) \neq 0$. *Moreover, if the coefficients of* f *are given rational numbers, then this Sturm sequence is in* $(\mathbb{Q}[X])^{r+1}$ *and can be computed from* f.

Proof We give an algorithm STURMSEQ (Table 8.12) for the computation of the Sturm sequence for the rational case; in the general case, one easily infers a mathematical existence proof. The general idea is to perform the same successive long divisions that the Euclidean algorithm uses to find $\gcd(f, f')$, with the exception that remainders are taken with opposite signs. One then divides $\gcd(f, f')$ out of the sequence of remainders thus obtained. Formally, we view the $(r+1)$-tuple $S = (f_0, \ldots, f_r)$ as a function from $\{0, \ldots, r\}$ to $\mathbb{Q}[X]$; enlarging S by an additional entry g is thus achieved by the assignment $S \leftarrow S \cup \{(r+1, g)\}$.

TABLE 8.12. Algorithm STURMSEQ

```
Specification: $S \leftarrow$ STURMSEQ($f$)
               Computation of a Sturm sequence
Given: $0 \neq f \in \mathbb{Q}[X]$ with $\deg(f) > 0$
Find: $S = (f_0, \ldots, f_r) \in (\mathbb{Q}[X])^{r+1}$ which is a Sturm sequence for $f$
      and $[\alpha, \beta]$ whenever $\alpha, \beta \in \mathbb{R}$ with $\alpha \leq \beta$ and $f(\alpha)$, $f(\beta) \neq 0$
begin
$F \leftarrow f$;  $G \leftarrow f'$
$i \leftarrow 0$;  $T \leftarrow \{(i, F)\}$
while $G \neq 0$ do
      (QUOT, REM) $\leftarrow$ DIV($F, G$)
      $F \leftarrow G$;  $G \leftarrow -$REM
      $i \leftarrow i + 1$;  $T \leftarrow T \cup \{(i, F)\}$
end
$S \leftarrow \emptyset$
while $T \neq \emptyset$ do
      select $(j, h)$ from $T$
      $T \leftarrow T \setminus \{(j, h)\}$
      $S \leftarrow S \cup \{(j, h/F)\}$
end
end STURMSEQ
```

We claim that the first **while**-loop terminates, and that the value of F after the last run is $\gcd(f, f')$. A rigorous proof would, up to a few minus signs, look exactly like the correcrtness proof of the algorithm EXTEUC of Theorem 2.32. An easier way of understanding the claim is to observe that because of the equations

$$qg + r = (-q)(-g) + r \quad \text{and} \quad -(qg + r) = (-q)g - r,$$

the first **while**-loop makes, up to a possible factor of -1, the same assignments to F and G as EXTEUC(f, f') would make to its variables A and B. Furthermore, we claim that the final value $\gcd(f, f')$ of F divides not only f and f', but every value of F during the first **while**-loop, thus making the divisions of the second **while**-loop possible. This is immediate from the fact that the ideal $\mathrm{Id}(F, G)$ is a loop invariant (look up the proof of Theorem 2.32 if you don't see why), and that $G = 0$ at the end of the loop.

It remains to be shown that the final value of S begins with a squarefree polynomial, ends with 1, and is a Sturm sequence for f and all intervals $[\alpha, \beta]$ as indicated. Let this final value be $(f_0, \ldots, f_r)$. For S0, it suffices to note that $f_0 = f/\gcd(f, f')$, and so the irreducible factors of f and f_0 in $\mathbb{Q}[X]$ are the same by Lemma 2.82. This also shows that f_0 is squarefree. S1 holds because rather obviously, $f_r = 1$, and S2 holds by the assumption on $[\alpha, \beta]$. For S3, let $1 < i < r$. It is easy to see from the assignments of the two **while**-loops that

$$f_{i-1} = qf_i - f_{i+1}$$

for some $q \in \mathbb{Q}[X]$, and we see that $f_i(\gamma) = 0$ implies that $f_{i-1}(\gamma)\cdot f_{i+1}(\gamma) \leq 0$, and that $f_{i-1}(\gamma) = 0$ if and only if $f_{i+1}(\gamma) = 0$. But the latter is easily seen to be impossible: from $f_i(\gamma) = 0$ together with $f_{i+1}(\gamma) = 0$ and the equation

$$f_i = qf_{i+1} - f_{i+2},$$

we could conclude that $f_{i+2}(\gamma) = 0$, and, continuing in this way, eventually $f_r(\gamma) = 0$, a contradiction.

Finally, for S4, suppose that $f_0(\gamma) = 0$ for some $\gamma \in \,]\alpha, \beta[$. Then we may choose $\gamma_1, \gamma_2 \in \mathbb{R}$ with $\gamma_1 < \gamma < \gamma_2$ close enough to γ so that neither f nor f' has a zero in $]\gamma_1, \gamma_2[\setminus \{\gamma\}$. Since

$$ff' = \left(f_0 \cdot \gcd(f, f')\right) \cdot \left(f_1 \cdot \gcd(f, f')\right) = f_0 f_1 \cdot \left(\gcd(f, f')\right)^2$$

and

$$\left(\gcd(f, f')\right)^2 > 0 \quad \text{on} \quad]\gamma_1, \gamma_2[\setminus \{\gamma\},$$

it suffices to prove that $ff' < 0$ on $]\gamma_1, \gamma[$ and $ff' > 0$ on $]\gamma, \gamma_2[$. To see this, we write

$$f = (X - \gamma)^e \cdot h \qquad (h \in \mathbb{Q}[X],\ h(\gamma) \neq 0).$$

Here, we must have $e > 0$ because γ is a zero of f by S0. We obtain

$$2ff' = (f^2)' = 2e(X - \gamma)^{2e-1} \cdot h^2 + 2(X - \gamma)^{2e} \cdot hh',$$

and so

$$\frac{ff'}{(X-\gamma)^{2e-1}} = eh^2 + (X-\gamma)\cdot hh' \quad \text{for} \quad X \neq \gamma.$$

The value of the right-hand side at γ is $eh^2(\gamma) > 0$ because of $h(\gamma) \neq 0$, and so it is positive on all of $]\gamma_1, \gamma_2[$. Looking at the left-hand side, we see that ff' must be negative to the left of γ and positive to the right. □

It is clear that given a polynomial $f \in \mathbb{Q}[X]$, we can now compute STURMSEQ(f) and then, using the same sequence over again, count the zeroes of f in intervals $[a, b]$ with a, $b \in \mathbb{Q}$ and $f(a)$, $f(b) \neq 0$ by means of Proposition 8.109.

In order to relate all this to the original version of Sturm's theorem, let us look at what happens if we try to use the sequence T that the first **while**-loop of STURMSEQ computes, thus refraining from dividing $\gcd(f, f')$ out of the sequence. It is clear that T will not in general be a Sturm sequence for f: as soon as the interval in question contains a common zero of f and f', property S3 will fail. But it is easy to see that Proposition 8.109 continues to hold with T used instead of a Sturm sequence: T is obtained from the actual output S of STURMSEQ by multiplying $\gcd(f, f')$ back into the sequence, and for the evaluated sequences of Proposition 8.109, this amounts to multiplication by a non-zero constant. (Recall that we are assuming that f, and thus $\gcd(f, f')$, vanishes neither at α nor at β.) Such a multiplication will clearly not affect the number of variations in sign. This proves Sturm's theorem as stated below. It is important to note though that in practice, the number of variations in sign will usually have to be computed at a large number of points; from a computational point of view, one should therefore by all means divide the common factor $\gcd(f, f')$ out of the sequence before embarking on these evaluations.

Corollary 8.111 (STURM'S THEOREM) *Let* $0 \neq f \in \mathbb{R}[X]$ *with* $\deg(f) > 0$, *and* α, $\beta \in \mathbb{R}$ *with* $\alpha \leq \beta$ *and* $f(\alpha)$, $f(\beta) \neq 0$. *Define the sequence* $(f_0, \dots, f_r)$ *recursively by setting*

$$\begin{aligned} f_0 &= f, \\ f_1 &= f', && \text{and for } i \geq 1, \\ f_{i+1} &= -R, && \text{where } f_{i-1} = Q \cdot f_i + R \text{ with } Q, R \in \mathbb{Q}[X] \text{ such that} \\ & && R \neq 0 \text{ and } \deg(R) < \deg(f_i), \end{aligned}$$

with the understanding that f_r *is the last non-zero remainder thus obtained. Then the number of distinct real zeroes of* f *in the interval* $[\alpha, \beta]$ *is equal to*

$$\text{VARSIGN}\big((f_0(\alpha), \dots, f_r(\alpha))\big) - \text{VARSIGN}\big((f_0(\beta), \dots, f_r(\beta))\big). \quad □$$

We are now going to use Sturm's method to isolate the zeroes of non-zero polynomials over $\mathbb{Q}$.

Definition 8.112 Let $0 \neq f \in \mathbb{R}$ and $\alpha \in \mathbb{R}$ a zero of f. Then an interval $[a, b]$ is called an **isolating interval** for the zero α of f if $a, b \in \mathbb{Q}$, $\alpha \in [a, b]$, and $f(\beta) \neq 0$ for all $\beta \in [a, b] \setminus \{\alpha\}$. A **set of isolating intervals** for the real zeroes of f is a set R of pairwise disjoint intervals such that each element of R is an isolating interval for some real zero of f, and R contains an isolating interval for each real zero α of f.

Note that as part of the definition of an isolating interval, we have required both endpoints to be rational. The strategy for computing a set of isolating intervals for the real zeroes of a non-zero polynomial over $\mathbb{Q}$ is now as follows. We start with an interval which we know will contain all real zeroes of f, if any. Such an interval will be provided by the next lemma. We then subdivide this interval further and further, counting zeroes as we go along, and drop those subintervals that contain no zero of f until every interval that is left contains exactly one zero of f.

Lemma 8.113 Let $0 \neq f \in \mathbb{R}[X]$, say $f = \sum_{i=0}^{m} a_i X^i$. Suppose $m > 0$ and $a_m > 0$, and set

$$M = \max\{1, |a_0|/a_m + \cdots + |a_{m-1}|/a_m\}.$$

Then

$$\begin{aligned} f(\gamma) > 0 &\quad \text{for} \quad \gamma > M, \quad \text{and} \\ (-1)^m \cdot f(\gamma) > 0 &\quad \text{for} \quad \gamma < -M. \end{aligned}$$

Proof First, let $\gamma > M$. Using the facts that $\gamma > M \geq 1$ and $x \geq -|x|$ for all $x \in \mathbb{R}$, we obtain

$$\begin{aligned} f(\gamma) &= \sum_{i=0}^{m} a_i\gamma^i \geq a_m\gamma^m - \left|\sum_{i=0}^{m-1} a_i\gamma^i\right| \geq a_m\gamma^m - \sum_{i=0}^{m-1} |a_i| \cdot \gamma^i \\ &> a_m M \cdot \gamma^{m-1} - \sum_{i=0}^{m-1} |a_i| \cdot \gamma^i \geq \sum_{i=0}^{m-1} |a_i| \cdot \gamma^{m-1} - \sum_{i=0}^{m-1} |a_i| \cdot \gamma^i \\ &\geq \sum_{i=0}^{m-1} |a_i| \cdot \gamma^i - \sum_{i=0}^{m-1} |a_i| \cdot \gamma^i = 0. \end{aligned}$$

If $\gamma < -M$, then we get

$$\begin{aligned} (-1)^m f(\gamma) &= (-1)^m a_m\gamma^m + (-1)^m \sum_{i=0}^{m-1} a_i\gamma^i \\ &\geq a_m(-\gamma)^m - \left|\sum_{i=0}^{m-1} a_i\gamma^i\right| \end{aligned}$$

$$\begin{aligned} &\geq \ a_m(-\gamma)^m - \sum_{i=0}^{m-1} |a_i|\,|\gamma|^i \\ &= \ a_m(-\gamma)^m - \sum_{i=0}^{m-1} |a_i|\,(-\gamma)^i \ > \ 0, \end{aligned}$$

the last inequality being true because $-\gamma > M$ and the bound M is the same for the polynomial $g = a_mX^m + \sum_{i=0}^{m-1} |a_i|X^i$ as for f. □

Exercise 8.114 Let $0 \neq f \in \mathbb{R}[X]$, say $f = \sum_{i=0}^{m} a_iX^i$. Suppose $m > 0$ and $a_m \neq 0$, and set

$$M = \max\{1, |a_0|/|a_m| + \cdots + |a_{m-1}|/|a_m|\}.$$

Show that $f(\alpha) = 0$ implies $-M \leq \alpha \leq M$.

We are now in a position to give an algorithm ISOLATE for the computation of isolating intervals according to the strategy that was described preceding the last lemma. There are two points that need attention. One is the fact that in the process of subdividing the interval $[-M, M]$ and counting zeroes, we may encounter a rational zero at some endpoint $a \in \mathbb{Q}$. We must then divide the linear factor $(X - a)$ out of the first element of our Sturm sequence in order to be able to apply Proposition 8.109. (Recall that the first element of the output of STURMSEQ is always squarefree.) One could of course also compute the rational zeroes separately before even beginning to look for real ones.

The other difficulty occurs when the algorithm encounters an interval $[a, b]$ with two zeroes in it, then divides that interval in the middle at $c = (a+b)/2$ and finds that each of $[a, c]$ and $[c, b]$ contains one zero. It would be a mistake to end the process here, because the two intervals are not disjoint. One could of course fix this by working with half-open intervals $[a, c[$ and $[c, b[$. This is not good enough, however, because for a later application we must have the stronger separation of the zeroes by means of disjoint closed intervals. We will therefore have to make the algorithm continue its process of subdividing intervals in the situation described above. Another way to handle the problem would be to compute a lower bound for the minimal distance between any two distinct zeroes and then make the algorithm run until each interval is less than half as wide as this bound.

The provisions described above will make a formal correctness proof of the mathematically plausible algorithm quite tedious; we therefore leave it to the reader to give such a proof if it is desired. If one disregards effectivity and elegance altogether, then the algorithm ISOLATE below is too complicated; we give a version that does not do anything outrageously awkward.

Theorem 8.115 *The algorithm* ISOLATE *of Table* 8.13 *computes a set of isolating intervals for every non-zero polynomial* $f \in \mathbb{Q}[X]$. □

TABLE 8.13. Algorithm ISOLATE

```
Specification: $R \leftarrow$ ISOLATE($f$)
               Isolation of real zeroes (Uses Subalgorithms ISOREC
               and ISOREFINE of Tables 8.14 and 8.15)
Given: $0 \neq f \in \mathbb{Q}[X]$
Find: a set $R$ of ordered pairs of rational numbers such that
      $\{ [a,b] \mid (a,b) \in R \}$ is a set of isolating intervals for the
      real zeroes of $f$
begin
$F \leftarrow f$;  $R \leftarrow \emptyset$
$M \leftarrow \max\{1, |a_0|/|a_m| + \cdots + |a_{m-1}|/|a_m|\}$, where $F = \sum_{i=0}^{m} a_i X^i$
if $(X - M) \mid F$ then
  $F \leftarrow F/(X - M)^{\mu}$, where $\mu$ is the multiplicity of $M$ as a zero of $F$
  $R \leftarrow R \cup \{(M, M)\}$
end
if $(X + M) \mid F$ then
  $F \leftarrow F/(X + M)^{\nu}$, where $\nu$ is the multiplicity of $-M$ as a zero of $F$
  $R \leftarrow R \cup \{(-M, -M)\}$
end
if $F$ is constant then return($R$) end
$S \leftarrow$ STURMSEQ($F$)
$F \leftarrow$ the first entry of $S$
$R \leftarrow R \cup$ ISOREC($-M, M, F, S$)
while there exist pairs $(u,c), (c,v) \in R$ do
      select pairs $(u,c), (c,v) \in R$
      $R \leftarrow R \setminus \{(u,c), (c,v)\}$
      $R \leftarrow R \cup$ ISOREFINE($u, c, v, F, S$)
end
end ISOLATE
```

Exercise 8.116 Modify the algorithm ISOLATE in such a way that it computes real zeroes with arbitrary prescribed precision, i.e., it computes isolating intervals of prescribed maximal length.

Exercise 8.117 Isolate the zeroes of $3X^3 - X^2 - 6X + 2$ by means of the algorithm ISOLATE. Check your answer.

We have now solved the problem of isolating real zeroes of univariate polynomials. The next two lemmas prepare the ground for the treatment of the multivariate case. First, we need an estimate for the absolute value of the value of a polynomial on an interval.

Lemma 8.118 Let $a, b \in \mathbb{Q}$ with $a \leq b$ and $f \in \mathbb{Q}[X]$, say $f = \sum_{i=0}^{m} a_i X^i$.

TABLE 8.14. Subalgorithm ISOREC

```
Specification: $R \leftarrow$ ISOREC($a, b, F, S$)
                Isolation of real zeroes on an interval
Given: $a, b \in \mathbb{Q}$ with $a \leq b$, $F \in \mathbb{Q}[X]$ squarefree, and
       a Sturm sequence $S$ for $F$ and $[a, b]$
Find: a set $R$ of pairs of rational numbers from the interval $[a, b]$
      such that the intervals corresponding to the pairs in $R$ can
      be elements of a set of isolating intervals for the zeroes of $F$,
      covering the zeroes lying in $]a, b[$, except that two intervals
      $[u_1, v_1]$ and $[u_2, v_2]$ with $(u_1, v_1)$ and $(u_2, v_2)$ two different
      elements of $R$ may still be overlapping by a single point
Comment: The arguments $F$ and $S$ are understood to be called by
         reference. This means that any change that ISOREC
         makes to $F$ and $S$ affects the value of $F$ and $S$ in the
         procedure that is making the present call of ISOREC
begin
$R \leftarrow \emptyset$
$v \leftarrow$VARSIGN($(f_0(a), \ldots, f_r(a))$) $-$ VARSIGN($(f_0(b), \ldots, f_r(b))$), where
    $S = (f_0, \ldots, f_r)$
if $v = 0$ then return($R$) end
if $v = 1$ then return($R \cup \{(a, b)\}$) end
$c \leftarrow (a + b)/2$
if $(X - c) \mid F$ then
  $F \leftarrow F/(X - c)$
  $R \leftarrow R \cup \{(c, c)\}$
  if $F$ is constant then return($R$) end
  $S \leftarrow$ STURMSEQ($F$)
  $F \leftarrow$ the first entry of $S$
end
$R \leftarrow R \cup$ ISOREC($a, c, F, S$) $\cup$ ISOREC($c, b, F, S$)
end ISOREC
```

If we set $M = \max\{|a|, |b|\}$, then

$$|f(\alpha)| \leq \sum_{i=0}^{m} |a_i| \cdot M^i$$

for all $\alpha \in [a, b]$.

Proof The statement of the lemma is immediate from the inequality

$$\left|\sum_{i=0}^{m} a_i \alpha^i\right| \leq \sum_{i=0}^{m} |a_i \alpha^i| = \sum_{i=0}^{m} |a_i|\,|\alpha|^i \leq \sum_{i=0}^{m} |a_i| \cdot M^i$$

which holds for all $\alpha \in [a, b]$. □

TABLE 8.15. Subalgorithm ISOREFINE

```
Specification: $P \leftarrow$ ISOREFINE$(u, c, v, F, S)$
               One step in the refinement of the output of ISOREC to
               a set of isolating intervals
begin
if $c = v$ then
  $d \leftarrow (u + c)/2$
  if $(X - d) \mid F$ then return$(\{(d, d), (c, c)\})$ end
  return$(\{(c, c)\} \cup$ ISOREC$(u, d, F, S) \cup$ ISOREC$(d, c, F, S))$
end
if $u = c$ then
  $d \leftarrow (v + c)/2$
  if $(X - d) \mid F$ then return$(\{(c, c), (d, d)\})$ end
  return$(\{(c, c)\} \cup$ ISOREC$(c, d, F, S) \cup$ ISOREC$(d, v, F, S))$
end
$d_1 \leftarrow (u + c)/2$;   $d_2 \leftarrow (c + v)/2$
if $(X - d_1) \mid F$ then return$(\{(d_1, d_1), (c, v)\})$ end
if $(X - d_2) \mid F$ then return$(\{(u, c), (d_2, d_2)\})$ end
$P \leftarrow$ ISOREC$(u, d_1, F, S) \cup$ ISOREC$(d_1, c, F, S) \cup$
      ISOREC$(c, d_2, F, S) \cup$ ISOREC$(d_2, v, F, S)$
end ISOREFINE
```

Exercise 8.119 Discuss how the estimate of the lemma above can be improved.

The next lemma shows how we can, by means of refinement, control the variation of a polynomial g on an isolating interval for a real zero of some other polynomial f.

Lemma 8.120 Let $[a, b]$ be an isolating interval for a real zero α of a polynomial $f \in \mathbb{Q}[X]$. Furthermore, let $g \in \mathbb{Q}[X]$ and $0 < \varepsilon \in \mathbb{Q}$. Then the algorithm SQUEEZE of Table 8.16 refines $[a, b]$ to an isolating interval $[c, d]$ for α such that

$$\max\{\, |g(\alpha_2) - g(\alpha_1)| \mid \alpha_1, \alpha_2 \in [c, d] \,\} < \varepsilon.$$

Proof *Termination*: The first five lines of the **while**-loop have the effect of leaving the loop or cutting the value of $d - c$ in half. Furthermore, the interval $[c, d]$ is replaced with a subinterval, and one easily sees that the value of B cannot increase during a run through the **while**-loop. It is now clear that the condition $B \cdot (d - c) < \varepsilon$ must eventually be reached.

Correctness: It is clear from the if-conditions that an invariant of the **while**-loop is given by: $[c, d]$ is an isolating interval for the zero α of f with $[c, d] \subseteq [a, b]$. In view of the last lemma, it is easy to see that upon

TABLE 8.16. Algorithm SQUEEZE

```
Specification: $(c,d) \leftarrow$ SQUEEZE$(a,b,f,g,\varepsilon)$
               Limiting the variation of $g$ on an isolating interval for a
               zero of $f$
Given: $a,b,\varepsilon \in \mathbb{Q}$ and $f,g \in \mathbb{Q}[X]$ such that $[a,b]$ is an isolating interval
       for a real zero $\alpha$ of $f$, and $\varepsilon > 0$
Find: $c,d \in \mathbb{Q}$ such that $[c,d]$ is an isolating interval for
      the real zero $\alpha$ of $f$ with $[c,d] \subseteq [a,b]$, and
      $|g(\alpha_2) - g(\alpha_1)| < \varepsilon$ for all $\alpha_1, \alpha_2 \in [c,d]$
begin
$(c,d) \leftarrow (a,b)$
if $f(c) = 0$ then return$((c,c))$ end
if $f(d) = 0$ then return$((d,d))$ end
$(f_0,\ldots,f_r) \leftarrow$ STURMSEQ$(f)$
$M \leftarrow \max\{|c|,|d|\}$
$B \leftarrow |a_0| + |a_1| \cdot M + \cdots + |a_m| \cdot M^m$, where $g' = \sum_{i=0}^m a_i X^i$
while $B \cdot (d-c) \geq \varepsilon$ do
    $s \leftarrow (c+d)/2$
    if $f(s) = 0$ then return$((s,s))$ end
    if VARSIGN$((f_0(c),\ldots,f_r(c)))$ − VARSIGN$((f_0(s),\ldots,f_r(s))) = 1$
      then $(c,d) \leftarrow (c,s)$
    else $(c,d) \leftarrow (s,d)$ end
    $M \leftarrow \max\{|c|,|d|\}$
    $B \leftarrow |a_0| + |a_1| \cdot M + \cdots + |a_m| \cdot M^m$, where $g' = \sum_{i=0}^m a_i X^i$
end
end SQUEEZE
```

termination, we also have $|g'(\beta) \cdot (\alpha_1 - \alpha_2)| < \varepsilon$ for all α_1, α_2, $\beta \in [c,d]$. We claim that this implies

$$\max\{\, |g(\alpha_2) - g(\alpha_1)| \mid \alpha_2, \alpha_1 \in [c,d] \,\} < \varepsilon.$$

To see this, let α_1, $\alpha_2 \in [c,d]$, and assume w.l.o.g. that $\alpha_1 < \alpha_2$. By the mean value theorem, there exists $\beta \in [\alpha_1, \alpha_2]$ such that

$$\frac{g(\alpha_2) - g(\alpha_1)}{\alpha_2 - \alpha_1} = g'(\beta).$$

It follows that

$$|g(\alpha_2) - g(\alpha_1)| = |g'(\beta) \cdot (\alpha_2 - \alpha_1)| < \varepsilon. \quad \square$$

Exercise 8.121 Show that the last two lines inside the while-loop of the algorithm SQUEEZE may be dropped.

Exercise 8.122 Write an algorithm with input $f \in \mathbb{Q}[X]$ and $a, b \in \mathbb{Q}$ with $a \leq b$ and output $0 < C \in \mathbb{Q}$ such that C has the following property: whenever $0 < \varepsilon \in \mathbb{R}$ is given, then $|\alpha_2 - \alpha_1| < C \cdot \varepsilon$ implies $|f(\alpha_2) - f(\alpha_1)| < \varepsilon$ for all α_1, $\alpha_2 \in [a, b]$. (This shows that polynomials are "computably uniformly continuous" on bounded sets.)

The following natural terminology and notation will be used in the next theorem and its proof. If $\boldsymbol{I}$ and $\boldsymbol{J}$ are intervals on the real line, then the **distance between $\boldsymbol{I}$ and $\boldsymbol{J}$** is defined as

$$\operatorname{dist}(\boldsymbol{I}, \boldsymbol{J}) = \inf\{\, |\alpha - \beta| \mid \alpha \in \boldsymbol{I},\ \beta \in \boldsymbol{J} \,\}.$$

If $\boldsymbol{I}$ is an interval and $\alpha \in \mathbb{R}$, then the **distance between α and $\boldsymbol{I}$** is defined as

$$\operatorname{dist}(\alpha, \boldsymbol{I}) = \inf\{\, |\alpha - \beta| \mid \beta \in \boldsymbol{I} \,\}.$$

Since we are working with closed intervals only, it is easy to see that the infima of the definitions of distances will always be assumed, and for non-zero distances, they will be assumed at certain endpoints. Moreover, if points and endpoints are given explicitly as rational numbers, then distances can always be computed.

We are finally in a position to discuss the problem of computing the real zeroes of $\operatorname{Id}(F)$ for a finite subset F of $\mathbb{Q}[\underline{X}]$. We know how to compute sets $R_1, \ldots, R_n$ of isolating intervals for the real zeroes of $f_1, \ldots, f_n$, respectively, where f_i is the unique monic generator of $\operatorname{Id}(F) \cap \mathbb{Q}[X_i]$ for $1 \leq i \leq n$. What remains to be done is to select from the set

$$M = \{\, ([a_1, b_1], \ldots, [a_n, b_n]) \mid [a_i, b_i] \in R_i \text{ for } 1 \leq i \leq n \,\}$$

those n-tuples $([a_1, b_1], \ldots, [a_n, b_n])$ that have the property that the unique n-tuple $(\alpha_1, \ldots, \alpha_n)$ with

$$\alpha_i \in [a_i, b_i] \quad \text{and} \quad f_i(\alpha_i) = 0 \qquad (1 \leq i \leq n)$$

is a zero of $\operatorname{Id}(F)$. We are thus faced with the problem to decide whether or not the polynomials in F, or, for that matter, the polynomials in any other basis of $\operatorname{Id}(F)$, vanish at $(\alpha_1, \ldots, \alpha_n)$. It may not be immediately obvious what the algorithm REALZEROES of the next theorem does and why; the reader is advised to use the correctness proof as a comment.

Theorem 8.123 *The algorithm* REALZEROES *of Table* 8.17 *computes, for any finite subset F of* $\mathbb{Q}[\underline{X}]$ *such that* $\operatorname{Id}(F)$ *is zero-dimensional, a set M of n-tuples*

$$((a_1, b_1), \ldots, (a_n, b_n))$$

of pairs of rational numbers such that the following hold:

(i) *For each real zero $(\alpha_1, \ldots \alpha_n) \in \mathbb{R}^n$ of* $\operatorname{Id}(F)$, *there exists exactly one*

$$((a_1, b_1), \ldots, (a_n, b_n)) \in M$$

with $\alpha_i \in [a_i, b_i]$ for $1 \leq i \leq n$.

(ii) *For each* $((a_1,b_1),\ldots,(a_n,b_n)) \in M$, *there exists exactly one zero*

$$(\alpha_1,\ldots,\alpha_n) \in \mathbb{R}^n$$

of $\mathrm{Id}(F)$ *with* $\alpha_i \in [a_i,b_i]$ *for* $1 \leq i \leq n$.

(iii) *Whenever* $((a_1,b_1),\ldots,(a_n,b_n))$ *and* $((c_1,d_1),\ldots,(c_n,d_n))$ *are in* M *and* $1 \leq i \leq n$, *then* $[a_i,b_i]$ *and* $[c_i,d_i]$ *are either equal or disjoint. Moreover,* $[a_i,b_i] = [c_i,d_i]$ *implies that the ith components* α_i *and* β_i *of the corresponding zeroes of* $\mathrm{Id}(F)$ *are equal.*

TABLE 8.17. Algorithm REALZEROES

```
Algorithm REALZEROES
Specification: $M \leftarrow$ REALZEROES($F$)
                    Computation of real zeroes of Id($F$)
Given: $F$ = a finite subset of $\mathbb{Q}[\underline{X}]$ with Id($F$) zero-dimensional
Find: a set $M$ of $n$-tuples of pairs of rational numbers $(a,b)$ such that
        $M$ has properties (i)–(iii) as stated in Theorem 8.123
begin
$H \leftarrow$ ZRADICAL($F$)
for $i = 1$ to $n$ do
    $f_i \leftarrow$ the monic generator of $\mathrm{Id}(H) \cap K[X_i]$
        (provided by ZRADICAL)
    $R_i \leftarrow$ ISOLATE($f_i$)
    $d_i \leftarrow \min\{\, \mathrm{dist}([a,b],[c,d]) \mid (a,b),(c,d) \in R_i$ with $(a,b) \neq (c,d) \,\}$
end
$G \leftarrow$ NORMPOS($H$), say $G = \{g, X_1 - g_1, \ldots, X_n - g_n\}$
$R \leftarrow$ ISOLATE($g$)
$M \leftarrow \emptyset$
while $R \neq \emptyset$ do
        select $(a,b)$ from $R$
        $R \leftarrow R \setminus \{(a,b)\}$
        for $i = 1$ to $n$ do
            $(a,b) \leftarrow$ SQUEEZE($a,b,g,g_i,d_i/2$)
        end
        $M \leftarrow M \cup \{\, ((a_1,b_1),\ldots,(a_n,b_n)) \,\}$, where $((a_1,b_1),\ldots,(a_n,b_n))$
            is the element of $R_1 \times \cdots \times R_n$ that satisfies

                $\mathrm{dist}\big(g_i(a),[a_i,b_i]\big) = \min\{\, \mathrm{dist}\big(g_i(a),[c,d]\big) \mid (c,d) \in R_i \,\}$

            for $1 \leq i \leq n$
end
end REALZEROES
```

Proof It is clear that $\mathrm{Id}(F)$ and $\mathrm{rad}(\mathrm{Id}(F))$ have the same zeroes in $\mathbb{R}^n$, and so passage to the radical is legitimate. Furthermore, our ground field $\mathbb{Q}$ satisfies the hypothesis of Theorem 8.22, so that the application of ZRADICAL does indeed yield a basis H of the radical of $\mathrm{Id}(F)$. Next, we observe that $\mathbb{Q}$ also satisfies the hypothesis of Theorem 8.81. The algorithm NORMPOS will therefore compute a Gröbner basis of the extended ideal

$$J = \mathrm{Id}(H, Z - X_1 - c_2X_2 - \cdots - c_nX_n) \qquad (c_2, \ldots, c_n \in \mathbb{Q})$$

which is of the form

$$G = \{g, X_1 - g_1, \ldots, X_n - g_n\}$$

with $g, g_1, \ldots, g_n \in \mathbb{Q}[Z]$. Lemma 8.73 (ii) tells us that $J \cap \mathbb{Q}[\underline{X}] = \mathrm{Id}(H)$, so that at this point, f_i is the unique monic generator of $J \cap \mathbb{Q}[X_i]$ for $1 \leq i \leq n$. From the fact that

$$\{g, X_1 - g_1, \ldots, X_n - g_n\} \quad \text{and} \quad H \cup \{Z - X_1 - c_2X_2 - \cdots - c_nX_n\}$$

are bases of J and that $J \cap \mathbb{Q}[\underline{X}] = \mathrm{Id}(H)$, it is easy to see that the zeroes of $\mathrm{Id}(H)$—which are the same as those of $\mathrm{Id}(F)$—are given by

$$\{\, (g_1(\alpha), \ldots, g_n(\alpha)) \mid \alpha \in \mathbb{R},\ g(\alpha) = 0 \,\}.$$

The zeroes of g are given to us by their isolating intervals which are collected in R. Moreover, whenever a zero α of g has been chosen, then $g_i(\alpha)$ is a zero of f_i and thus lies in exactly one interval $[a_i, b_i]$ with $(a_i, b_i) \in R_i$ for $1 \leq i \leq n$. The task of one run through the **while**-loop is to determine the n-tuple

$$((a_1, b_1), \ldots, (a_n, b_n)) \in R_1 \times \cdots \times R_n$$

that satisfies $g_i(\alpha) \in [a_i, b_i]$ for $1 \leq i \leq n$, and then to place this n-tuple in the output set M. Here, the zero α of g is given by an isolating interval $[a, b]$ which is selected from the set R.

To see that the **while**-loop performs its task correctly, let $(a, b) \in R$, and let α be the zero of g that is isolated by the interval $[a, b]$. The application of SQUEEZE shrinks $[a, b]$ in such a way that it is still an isolating interval for the zero α of g, and that

$$\max\{\, |g_i(\alpha_2) - g_i(\alpha_1)| \mid \alpha_1, \alpha_2 \in [a, b] \,\} < \frac{d_i}{2} \quad \text{for} \quad 1 \leq i \leq n, \qquad (*)$$

where d_i is the minimal distance between any two different intervals in R_i. For $1 \leq i \leq n$, let now (a_i, b_i) be the element of R_i that the algorithm is supposed to determine, i.e., the one with $g_i(\alpha) \in [a_i, b_i]$. We must show that $[a_i, b_i]$ is the interval in R_i that $g_i(a)$ is closest to, i.e., that

$$\mathrm{dist}(g_i(a), [a_i, b_i]) < \mathrm{dist}(g_i(a), [c, d])$$

for all $(c,d) \in R_i$ with $(c,d) \neq (a_i,b_i)$. Assume for a contradiction that there exists $(c,d) \in R_i$ with $(c,d) \neq (a_i,b_i)$ and

$$\operatorname{dist}\big(g_i(a),[c,d]\big) \leq \operatorname{dist}\big(g_i(a),[a_i,b_i]\big).$$

It follows that there exists $\beta \in [c,d]$ with

$$|\beta - g_i(a)| \leq \operatorname{dist}\big(g_i(a),[a_i,b_i]\big).$$

From the inequality $(*)$ and the fact that a, $\alpha \in [a,b]$ and $g_i(\alpha) \in [a_i,b_i]$, we conclude that

$$\operatorname{dist}\big(g_i(a),[a_i,b_i]\big) \leq |g_i(a) - g_i(\alpha)| < \frac{d_i}{2}.$$

Using the last two inequalities, we see that

$$\begin{aligned}
|\beta - g_i(\alpha)| &= \big|(\beta - g_i(a)) + (g_i(a) - g_i(\alpha))\big| \\
&\leq |\beta - g_i(a)| + |g_i(a) - g_i(\alpha)| \\
&< \frac{d_i}{2} + \frac{d_i}{2} \;=\; d_i,
\end{aligned}$$

from which it follows that $\operatorname{dist}([c,d],[a_i,b_i]) < d_i$, a contradiction. We have proved that the **while**-loop makes all the right choices for elements of M. Property (iii) of the theorem and the uniqueness properties in (i) and (ii) are easy consequences of the fact that M is a subset of $R_1 \times \cdots \times R_n$. □

Exercise 8.124 Show that the algorithm REALZEROES will never try to place an n-tuple in the set M that is already in there.

Exercise 8.125 Let f, $g \in \mathbb{Q}[X]$, let $[a,b]$ be an isolating interval for a real zero α of f, and let $c \in \mathbb{Q}$ with $c \neq g(\alpha)$. Give an algorithm that decides whether $g(\alpha) > c$ or $g(\alpha) < c$.

Exercise 8.126 Modify the algorithm REALZEROES in such a way that it uses the previous exercise rather than the algorithm SQUEEZE in order to determine, for given zero α of g, the corresponding n-tuple of isolating intervals of the zeroes $g_1(\alpha),\dots,g_n(\alpha)$ of $f_1,\dots,f_n$. Discuss on an intuitive level the difference between the two versions as far as computational expense is concerned.

Exercise 8.127 Write an algorithm that computes the complex zeroes of a polynomial ideal over $\mathbb{Q}$. Here, a complex zero should be given as a pair of isolating intervals, one for the real and one for the imaginary part.

Just as with the algorithm ZPRIMDEC, the algorithm REALZEROES is extremely time and space consuming in general because it calls for an application of NORMPOS. As with ZPRIMDEC, one should therefore try to decompose the given ideal as much as possible into an intersection of ideals with "smaller" univariate polynomials. The techniques described on pages 382–385 apply verbatim. The following lemma shows how the real zeroes of the original ideal can be obtained from those of the ideals occurring in the decomposition.

Lemma 8.128 Let K be any field, L an extension field of K, and let I and J be ideals of $K[\underline{X}]$. If we denote by $N(I)$, $N(J)$, and $N(I \cap J)$ the set of zeroes in L^n of I, J, and $I \cap J$, respectively, then $N(I \cap J) = N(I) \cup N(J)$.

Proof The inclusion "$\supseteq$" is an immediate consequence of the fact that $I \cap J \subseteq I$ and $I \cap J \subseteq J$. Now let $z \in L^n$ be a zero of $I \cap J$, and assume for a contradiction that z is neither a zero of I nor a zero of J. Then there exist polynomials $f \in I$ and $g \in J$ with $f(z) \neq 0$ and $g(z) \neq 0$, and we see that $h = fg$ is an element of $I \cap J$ with $h(z) \neq 0$, a contradiction. □

Exercise 8.129 Let K, L, I, and J be as in the last lemma. Prove the dual statement $N(I \cup J) = N(I) \cap N(J)$.

A further potential improvement of the algorithm REALZEROES can be achieved if a more intricate procedure for the selection of the n-tuples of isolating intervals is available. Suppose we have found a Gröbner basis G of some ideal I of $\mathbb{Q}[\underline{X}]$ which is of the form

$$G = \{X_1^{\nu_1} - g_1, \ldots, X_n^{\nu_n} - g_n\} \qquad (*)$$

with g_i bivariate in X_i and some other variable for $1 \leq i \leq n$, say $g_i \in \mathbb{Q}[X_i, X_{j_i}]$. Assume further that we have computed the generators f_i of $I \cap \mathbb{Q}[X_i]$ for $1 \leq i \leq n$ as well as sets R_i of isolating intervals for their real zeroes. Then an n-tuple

$$([a_1, b_1], \ldots, [a_n, b_n]) \in R_1 \times \cdots \times R_n$$

isolates a zero of I in $\mathbb{R}^n$ if and only if $\alpha_i^{\nu_i} - g_i(\alpha_i, \alpha_{j_i}) = 0$ for $1 \leq i \leq n$, where α_i and α_{j_i} are the zeroes of f_i and f_{j_i} isolated by $[a_i, b_i]$ and $[a_{j_i}, b_{j_i}]$, respectively. The decision whether or not this holds cannot be achieved with the procedures that we have discussed; it is, however, covered by an algorithm that should be available wherever real algebraic decision methods are implemented.

Unfortunately, there does not seem to be a systematic way to compute bases of the form $(*)$. But it may always be that a basis of this form is given, or that the algorithm NORMPOS which is called by REALZEROES finds one before it has found the normal position. The following lemma says that this expectation is not absurdly unfounded: whenever the given ideal has been preprocessed by PREDEC and its reduced Gröbner basis G w.r.t. the lexicographical term order consists of bivariate polynomials only, then G will be of the form $(*)$. The main argument of the proof has already been used in the proof of Proposition 7.42.

Lemma 8.130 Let K be a field, and suppose I is a zero-dimensional ideal of $K[\underline{X}]$ such that for $1 \leq i \leq n$, the ideals $I \cap K[X_i]$ are generated by irreducible polynomials. Assume further that the reduced Gröbner basis G

of I w.r.t. the lexicographical term order consists of bivariate polynomials only. Then G is of the form

$$G = \{X_1^{\nu_1} - g_1, \ldots, X_n^{\nu_n} - g_n\}$$

with g_i bivariate in X_i and some other, lexicographically lesser variable for $1 \leq i \leq n$.

Proof Since I is zero-dimensional and G contains only bivariate polynomials, G has a subset G' of the indicated form. Assume for a contradiction that there exists $f \in G \setminus G'$, say

$$f = \sum_{k=0}^{m} h_k X_i^k \qquad (h_0, \ldots, h_m \in K[X_j],\ 1 \leq i < j \leq n).$$

Let g be the element of G' whose head term is univariate in X_i. From the fact that G is reduced, we may conclude that $m < \deg_{X_i}(g)$, and that $h_m \notin I$. Since $I \cap K[X_j]$ is generated by an irreducible polynomial and thus is maximal, there exists $p \in K[X_j]$ and $q \in I \cap K[X_j]$ with $ph_m + q = 1$. Then $h = pf + qX_i^m \in I$, and

$$\begin{aligned} h &= ph_m X_i^m + p \sum_{k=0}^{m-1} h_k X_i^k + qX_i^m \\ &= X_i^m + p \sum_{k=0}^{m-1} h_k X_i^k. \end{aligned}$$

We see that $\deg_{X_i}(h) = m < \deg_{X_i}(g)$. But h must be top-reducible modulo G, and so G must contain a polynomial with head term X_i^ν with $\nu < \deg_{X_i}(g)$, contradicting the fact that G was reduced. □

Notes

The results of the first three sections of Chapter 8 are more or less ideal theoretic folklore; the algorithm for the computation of the radical in the zero-dimensional case is based on Lemma 92 of Seidenberg (1974).

The concept of the primary decomposition of an ideal originates with the so-called *fundamental theorem* of M. Noether (1873). Noether's theorem concerns ideals generated by two polynomials in two variables. It gives a sufficient condition for a polynomial to be in the ideal which, in a sense, amounts to saying that the polynomial must be in every primary component of the ideal. Subsequent improvements and generalizations of the fundamental theorem are due to Bertini (1889), Lasker (1905), and Macaulay (1916). It was E. Noether (1921) who finally proved that the primary decomposition of ideals is not an intrinsically geometric phenomenon, but is in fact possible in every ring that satisfies the ascending chain condition for ideals.

A method for the computation of primary decompositions of polynomial ideals was given in Hermann (1926) (cf. the Notes to Chapter 6 on p. 291); again, it is important to observe the corrections of Seidenberg (1974). The method of Section 8.6 for the zero-dimensional case is based on Kredel (1987); see also Lazard (1985) and Gianni et al. (1988).

The algorithms of Section 8.7 for the computation of radical and primary decomposition in higher dimensions closely resemble those given in a preliminary version of Gianni et al. (1988); in the final version of the paper, this part is modified. Recent improvements are due to Eisenbud et al. (1992).

Sturm published his theorem on the number of real zeroes of a polynomial in 1835. It can be viewed as an improvement on Descartes' rule of signs. Actual implementations tend to favor the Uspensky algorithm over Sturm chains for the computation of isolating intervals of real zeroes; see, e.g., Collins and Loos (1982). The investigation of real zeroes of polynomials and—more generally—real algebraic geometry is one of the focal points of contemporary computational algebra; we refer the reader to Collins and Loos (1982) and Bochnak et al. (1987) for more information and references.

The problem of solving systems of non-linear polynomial equations, i.e., of computing the zeroes of a given ideal, is treated systematically for the first time in Kronecker (1892) under the heading of *elimination theory.* Elimination theory continued to be a major topic in commutative algebra for several decades to come; see, e.g., Netto (1900) and Macaulay (1916). The classical point of view is that the zeroes of a univariate polynomial, i.e., the solutions of a univariate polynomial equation, have been found as soon as one has constructed an extension of the ground field over which the given polynomial decomposes into linear factors. The general problem is solved by successive elimination of variables by means of resultants, followed by a backsubstitution process that requires extending a field which has itself been obtained as an algebraic extension of the original ground field. For a lucid presentation of the classical theory, we refer the reader to van der Waerden (1931), Chapter 11. He remarks in a footnote that these methods are of little practical relevance due to their enormous complexity. It is interesting that the chapter on elimination theory was dropped altogether from the fourth edition of van der Waerden's book in 1959.

In short, the relevance of Gröbner bases in this context is that a single Gröbner basis computation w.r.t. a lexicographical term order can replace the classical elimination procedure. Suppose such a Gröbner basis of the given set of polynomials is at hand. Assuming that the system generates a zero-dimensional ideal (cf. the remarks at the beginning of Section 8.8), one may then adjoin to the ground field the zeroes of the univariate polynomial of minimal degree in the lexicographically least variable, substitute these into the bivariate ones, and continue the process in the obvious manner, computing with polynomials over successively larger algebraic extensions of the ground field. (It is perhaps noteworthy that in contrast to the classical elimination method, the Gröbner basis approach is not a recursive one by nature: if one substitutes a zero of the univariate polynomial into the remaining elements of the Gröbner basis, then the result will not in general be a Gröbner basis, and neither does it have to be for the method to be successful.)

The point of view that we have adopted is somewhat more specific; we are interested in polynomial systems over the rationals and their real zeroes, which we wish to obtain in terms of isolating intervals. It is shown in Loos (1982) that one can indeed compute over finite algebraic extensions of $\mathbb{Q}$ with this understanding of what a zero is. These ideas were developed for the purpose of real quantifier elimination via cylindrical algebraic decomposition; see also Collins (1975).

The method that we have presented here, which was suggested by Kredel (1988a), could perhaps be called the "primitive element version" of the straightforward one that uses a single Gröbner basis. By first computing a Gröbner basis of an extended ideal of the form $\{g, X_1 - g_1, \ldots, X_n - g_n\}$ with g and the g_i univariate in a new indeterminate, we have shifted most of the expenditure of computing in algebraic extensions to the Gröbner bases part of the computation. The remaining selection of the n-tuple of isolating intervals corresponding to $(g_1(\alpha), \ldots, g_n(\alpha))$, where α is a zero of g given by an isolating interval, is then a simple example of a computation in $\mathbb{Q}(\alpha)$ that can be handled on an elementary level. It is for this reason that we have presented this method; our choice does not reflect any judgement concerning the efficiency of various algorithms.

Using the Collins–Loos methods that we have mentioned above, one may also compute the real zeroes of a polynomial system in the spirit of real quantifier elimination without using Gröbner bases at all; see also Grigor'ev and Vorobjov (1988), Renegar (1992), and the references given there. For another interesting approach that does not use Gröbner bases, see Morgan (1987).

Further references concerning the computation of zeroes of polynomial systems include Trinks (1978), Lazard (1979, 1981, 1983, 1992), Pohst and Yun (1981), Grigor'ev and Chistov (1984), Buchberger (1985a), Bronstein (1986), Czapor (1987, 1989), Canny et al. (1989), Böge et al. (1986), Winkler (1986), Gianni (1987), Kalkbrener (1987), Auzinger and Stetter (1988), Kobayashi, Fujise, and Furukawa (1988), Gianni and Mora (1989), Kobayashi, Moritsugu, and Hogan (1988), Melenk et al. (1989), Gerdt et al. (1990), Kalkbrener (1990), Melenk (1990), and Yokoyama et al. (1992).

9

Linear Algebra in Residue Class Rings

The K-vector space structure on residue class rings of polynomial rings has already been used in Section 6.3 in connection with zero-dimensional ideals. An important result was that an ideal I is zero-dimensional if and only if the residue class ring modulo I is finite-dimensional as a K-vector space. In this chapter we discuss a number of important algorithms that use linear algebra in connection with Gröbner bases. The focus is on zero-dimensional ideals.

9.1 Gröbner Bases and Reduced Terms

Throughout this section, K will be a field, I a proper ideal in $K[\underline{X}] = K[X_1, \dots, X_n]$, and $T = T(X_1, \dots, X_n)$. If $f \in K[\underline{X}]$, then the residue class $f + I \in K[\underline{X}]/I$ of f modulo I will be denoted by $\overline{f}$. Recall that $\mathrm{RT}(I)$ stands for the set $T \setminus \mathrm{HT}(I)$ of reduced terms modulo I. We begin with an easy algorithm that computes the set of reduced terms up to a degree bound in each variable.

Proposition 9.1 *Let G be a Gröbner basis in $K[\underline{X}]$. Assume that K is computable and the term order $\leq$ is decidable, and let $k_1, \dots, k_n \in \mathbb{N}$. Then the algorithm* REDTERMS *of Table* 9.1 *computes the set of all reduced terms $t \in \mathrm{RT}(I(G))$ that satisfy $\deg_{X_i}(t) \leq k_i$.*

Exercise 9.2 Prove correctness and termination of the above algorithm.

Suppose now we have established zero-dimensionality of I by computing a Gröbner basis of I w.r.t. some term order $\leq$ and verifying criterion (iii) of Theorem 6.54 It is then clear that $\mathrm{RT}(I)$ w.r.t. $\leq$ can be computed by applying REDTERMS to G and the minimal values ν_i satisfying the criterion. Recall from Proposition 6.52 that $\overline{\mathrm{RT}(I)}$ is a basis of the K-vector space $K[\underline{X}]/I$. The fact that we can, for given zero-dimensional ideal I, actually compute a natural basis of $K[\underline{X}]/I$ from a Gröbner basis of I will be of great importance in this chapter. The following exercise helps to visualize the situation.

Exercise 9.3 Draw the first quadrant of a Cartesian coordinate system, and label the points 0–10 on each axis. Draw horizontal and vertical lines through

TABLE 9.1. Algorithm REDTERMS

```
Specification: $R \leftarrow$ REDTERMS($G, (k_1, \dots, k_n)$)
               Construction of the set $R$ of reduced terms
               $t \in \mathrm{RT}(I(G))$ such that $\deg_{X_i}(t) \le k_i$ for $1 \le i \le n$
Given: a Gröbner basis $G$ for a proper ideal $I(G)$ and $k_1, \dots, k_n \in \mathbb{N}$
Find: the set $R$ of reduced terms $t \in \mathrm{RT}(I(G))$
      such that $\deg_{X_i}(t) \le k_i$ for $1 \le i \le n$
begin
$R \leftarrow \{1\}$
for $i = 1$ to $n$ do
  $S \leftarrow R$
  while $S \neq \emptyset$ do
      $t \leftarrow$ some element of $S$
      $S \leftarrow S \setminus \{t\}$
      for $l = 1$ to $k_i$ do
          $t \leftarrow t \cdot X_i$
          if $t$ is in normal form modulo $G$ then
            $R \leftarrow R \cup \{t\}$ end
      end
  end
end
end REDTERMS
```

these points. Consider the set

$$G = \{X^5, X^4Y, X^3Y^2, XY^3, Y^4\} \subseteq \mathbb{Q}[X,Y],$$

and let $I = \mathrm{Id}(G)$. Then G is the reduced Gröbner basis of I w.r.t. every term order because it consists of monomials with no divisibilities between them. We see that $G = \mathrm{HT}(G)$. Identifying the term X^iY^j with the point (i, j), draw G into your coordinate system. Following the algorithm REDTERMS, determine $\mathrm{RT}(I)$ and add it into the picture. Now shade every square in your grid that has a point drawn at each of its corners. Point out where the set $\mathrm{mult}(\mathrm{HT}(G)) = \mathrm{HT}(I)$ is. If you had to make up a terminology for the set of head terms of the reduced Gröbner basis of an ideal w.r.t. a term order, what would you choose?

The **stairs** $\mathrm{st}(I)$ of I is the unique minimal finite basis of the set $\mathrm{HT}(I)$ w.r.t. the divisibility relation on T. Recall that the unique reduced Gröbner basis G of I (w.r.t. $\le$) satisfies $\mathrm{HT}(G) = \mathrm{st}(I)$. Moreover, each $g \in G$ is monic by our definition of the reduced Gröbner basis, so that in fact $\mathrm{HM}(G) = \mathrm{st}(I)$.

If I is not zero-dimensional, i.e., the dimension of $K[\underline{X}]/I$ is infinite, then the algorithm REDTERMS computes a basis of the subspace

$$\{\,\overline{f} \in K[\underline{X}]/I \mid f \text{ in normal form mod } G \text{ and } \deg_{X_i}(f) \le k_i\,\}$$

of $K[\underline{X}]/I$. We will now look at a dual problem: it is easy to see that the ideal I itself is a subspace of the K-vector space $K[\underline{X}]$, and so is

$$I_k = \{\, 0 \neq f \in I \mid \deg(f) \leq k \,\} \cup \{0\}$$

for each $k \in \mathbb{N}$.

Proposition 9.4 *Let $G = \{g_1, \ldots, g_m\}$ be a Gröbner basis of I w.r.t. a total degree order, and let $k \in \mathbb{N}$. For $1 \leq i \leq m$ set*

$$B_i = \left\{\, tg_i \;\middle|\; t \in T,\ \deg\bigl(\mathrm{HT}(tg_i)\bigr) \leq k,\ \mathrm{HT}(g_j) \nmid \mathrm{HT}(tg_i) \text{ for all } j < i \,\right\}.$$

Then $B = \bigcup_{i=1}^{m} B_i$ is a basis of the K-vector space I_k.

Proof We have $B \subseteq I_k$ by the choice of the term order. It is clear that the head terms of elements of B_i are pairwise different for fixed $1 \leq i \leq m$. If there were tg_i, $sg_j \in B$ with $i < j$ and $\mathrm{HT}(tg_i) = \mathrm{HT}(sg_j)$, then we would have $\mathrm{HT}(g_i) \mid \mathrm{HT}(sg_j)$ contrary to the construction of B_j. To prove the linear independence of B, let

$$p = \sum_{q \in B} \lambda_q \cdot q \qquad (\lambda_q \in K),$$

where not all λ_q equal zero. Then $\max\{\, \mathrm{HT}(q) \mid \lambda_q \neq 0 \,\} = \mathrm{HT}(h)$ for exactly one $h \in B$, and we see that $\mathrm{HT}(h)$ is a term in p. In particular, $p \neq 0$. It remains to show that B is a generating system for I_k. Let $f \in I_k$. Then $f \xrightarrow[G]{*} 0$. Among all possible reduction chains, consider the one where each reduction step $f_k \xrightarrow[g_i]{} f_{k+1}$ is a top reduction and has the property $\mathrm{HT}(g_j) \nmid \mathrm{HT}(f_k)$ for all $j < i$. It is then easy to see that the resulting representation of f as a sum of monomial multiples of elements of G (Lemma 5.60) is in fact a linear combination of elements of B. □

If $\dim(I) = 0$, then we know from Lemma 6.50 that I contains a unique non-zero monic univariate polynomial in X_i for $1 \leq i \leq n$, namely, the monic generator of the ideal $I \cap K[X_i]$. It is often necessary to compute these univariate polynomials from a given finite ideal basis of I. We already know that in principle this can be done by computing the elimination ideals $I \cap K[X_i]$ by means of n Gröbner basis computations for term orders satisfying

$$\{X_i\} \ll \{X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n\}.$$

We are now going to show how the same result can be achieved using a single Gröbner basis together with some linear algebra computations in $K[\underline{X}]/I$.

As a matter of fact, the necessary theory and algorithms are already at our disposal. Suppose we have computed a Gröbner basis of I w.r.t. any term order $\leq$. To compute the univariate polynomial in any one variable

X_i, we consider the sets $C_m = \{1, \overline{X_i}, \overline{X_i^2}, \ldots, \overline{X_i^m}\}$ for $m \in \mathbb{N}$. By Lemma 6.53 (i) and (ii), the polynomials

$$\{\, 0 \neq f \in I \cap K[X_i] \mid \deg(f) \leq m \,\}$$

correspond to the non-trivial linear combinations of elements of C_m that equal zero. By Lemma 6.53 (iv) and (v), we can express $\overline{X_i^k}$ as a linear combination of elements of the canonical term basis $\overline{\mathrm{RT}(I)}$ w.r.t. $\leq$ for each $k \in \mathbb{N}$. Using the algorithm LINDEP of Proposition 3.15, we can now, for each $m \in \mathbb{N}$, decide whether C_m is linearly dependent, and if so, produce a non-trivial linear combination of its elements that equals zero. So all we have to do is apply LINDEP to C_m for increasing m until we find it to be linearly dependent. The coefficients produced by LINDEP are then the coefficients of the desired polynomial. We will now show how one can actually avoid an explicit call of LINDEP by coding much of its action into a rather elegant computation with polynomials.

Let G be any Gröbner basis of I, let t, s_1, ..., $s_m \in T$, and let h, h_1, ..., h_m, respectively, be normal forms of these modulo G. Let $Y_1, \ldots, Y_m$ be new indeterminates, and set

$$p = h + \sum_{i=1}^{m} Y_i h_i \in K[\underline{Y}, \underline{X}],$$

where $K[\underline{Y}, \underline{X}]$ stands for $K[Y_1, \ldots, Y_m, X_1, \ldots, X_n]$. Finally, let C be the set of all coefficients in $K[\underline{Y}]$ of $p \in K[\underline{Y}][\underline{X}]$. It is clear that each $f \in C$ is linear in each Y_i, and we may thus consider the system

$$S = \{\, f = 0 \mid f \in C \,\}$$

of linear equations in $Y_1, \ldots, Y_m$.

Lemma 9.5 In the situation explained above, an m-tuple $(a_1, \ldots, a_m) \in K^m$ is a solution of S if and only if the polynomial

$$g = t + \sum_{i=1}^{m} a_i s_i$$

lies in the ideal I.

Proof We have $t - h \in I$ and $s_i - h_i \in I$ for $1 \leq i \leq m$. It follows that

$$t + \sum_{i=1}^{m} a_i s_i - \left(h + \sum_{i=1}^{m} a_i h_i \right) \in I.$$

for every m-tuple $\boldsymbol{a} = (a_1, \ldots, a_m) \in K^m$. Since the substitution of a_i for Y_i is a homomorphism, the expression in parenthesis equals $p(\boldsymbol{a}, \underline{X})$. We may now conclude that

$$t + \sum_{i=1}^{m} a_i s_i \in I \quad \text{iff} \quad p(a, \underline{X}) \in I.$$

It is not hard to see from the definition of $p(\underline{Y}, \underline{X})$ that the terms of $p(a, \underline{X})$ in $T(\underline{X})$ are all reduced, so it is in I if and only if its coefficients

$$\{\, f(a) \mid f \in C \,\}$$

all equal zero. □

With the lemma and the discussion preceding it it is now easy to prove correctness and termination of the following algorithm.

Proposition 9.6 *Assume* $\dim(I) = 0$, *K is computable, and a Gröbner basis G of I w.r.t. some decidable term order has been computed. Let* $1 \leq i \leq n$. *Then the algorithm* UNIVPOL *of Table* 9.2 *computes the monic polynomial $f \in I \cap K[X_i]$ of minimal degree.* □

Since the algorithm stops when it has found the first solvable system of linear equations, it is obviously advantageous to choose among the various algorithms for the treatment of such systems one that tends to detect unsolvability fast.

Next, we show how a similar but more sophisticated algorithm uses linear algebra to convert a given Gröbner basis of a zero-dimensional ideal to another one w.r.t. a different term order. This conversion process can be considerably less expensive than a new Gröbner basis computation. Since Gröbner bases w.r.t. total degree orders tend to be much easier to compute than those w.r.t. lexicographical orders, it is often advantageous to compute the former by means of a Buchberger algorithm and then convert to the latter.

The conversion algorithm uses much the same ideas as UNIVPOL. It picks terms t by increasing new term order $\preceq$ and looks for polynomials in the ideal with head term t to be placed in the new Gröbner basis. If it does not find one, it puts t in a set R which builds up to the set of reduced terms w.r.t. $\preceq$. Termination is assured by the facts that the ideal has dimension zero and that the algorithm skips all terms that are multiples of a head term that is already in the new Gröbner basis. It should be obvious by now that the new term order will have to satisfy the following condition.

(D) Whenever $t \in T$ and a finite subset S of T are given, then one can decide whether the set

$$N = \left\{\, u \in T \;\middle|\; t \prec u, \text{ and } s \nmid u \text{ for all } s \in S \,\right\}$$

is empty and compute its $\preceq$-minimal element if it is not.

TABLE 9.2. Algorithm UNIVPOL

```
Specification: $f \leftarrow$ UNIVPOL$(G, i)$
                    Computation of univariate polynomial
Given: a Gröbner basis $G$ w.r.t. any term order such that
          Id$(G)$ is proper with dim(Id$(G)) = 0$, and $1 \leq i \leq n$
Find: the unique monic polynomial of minimal
          degree in $I \cap K[X_i]$
begin
$N \leftarrow \min\{\nu - 1 \mid X_i^\nu \in \mathrm{HT}(G)\}$
create new indeterminates $Y_0, \ldots, Y_N$
$q \leftarrow Y_N X_i^N + \cdots + Y_1 X_i + Y_0$
$t \leftarrow X_i^N$
loop $t \leftarrow t \cdot X_i$
    $h \leftarrow$ a normal form of $t$ modulo $G$
    $p \leftarrow h + q$
    $C \leftarrow$ the set of coefficients in $K[Y_0, \ldots, Y_N]$ of
        $p \in K[Y_0, \ldots, Y_N][X_1, \ldots, X_n]$
    if the system $S = \{f = 0 \mid f \in C\}$ of linear equations in
      the indeterminates $Y_0, \ldots, Y_N$ has a solution then
      $\{c_j \in K \mid 0 \leq j \leq N\} \leftarrow$ a solution of $S$
      $f \leftarrow X_i^{N+1} + \sum_{j=0}^{N} c_j X_i^j$
      return$(f)$
    else $N \leftarrow N + 1$
        create a new indeterminate $Y_N$
        $q \leftarrow Y_N h + q$
    end
end
end UNIVPOL
```

Whenever $\leq$ satisfies (**D**), it will be understood that MINTERM(S, t) is an algorithm that performs the computation described in (**D**): it outputs **false** if N is empty, (**true**, u) where u is the $\leq$-least element of N otherwise. For the lexicographical term order, MINTERM will be given explicitly below.

Proposition 9.7 *Assume K is computable, and G is a Gröbner basis of a zero-dimensional ideal in $K[\underline{X}]$ w.r.t. some decidable term order. Let $\leq$ be a decidable term order on T that satisfies condition* (**D**). *Then the algorithm* CONVGRÖBNER *of Table* 9.3 *computes the reduced Gröbner basis of* Id(G) *w.r.t.* $\leq$.

Proof *Termination*: Assume for a contradiction that the algorithm does not terminate for some input. Suppose first that the **if**-condition is satisfied infinitely many times, and let $\{t_k\}_{k \in \mathbb{N}}$ be the terms in H indexed in the

TABLE 9.3. Algorithm CONVGRÖBNER

```
Specification: $(F, R) \leftarrow$CONVGRÖBNER$(G)$
               Conversion of an arbitrary Gröbner basis of a
               zero-dimensional ideal to one w.r.t. $\leq$
Given: a Gröbner basis $G$ w.r.t. some decidable term order $\leq'$ such
       that Id$(G)$ is zero-dimensional, and a new term order $\leq$ that
       satisfies (D)
Find: the reduced Gröbner basis $F$ of Id$(G)$ w.r.t. $\leq$ and
      the set $R$ of reduced terms w.r.t. Id$(G)$ and $\leq$
begin
$F \leftarrow \emptyset$;  $H \leftarrow \emptyset$
$t \leftarrow 1$;  $R \leftarrow \{t\}$
create a new indeterminate $Y_1$
$\underline{Y} \leftarrow \{Y_1\}$;  $q \leftarrow Y_1$
while   MINTERM$(H, t) \neq$ false  do
    $t \leftarrow u$ where MINTERM$(H, t) =$ (true, $u$)
    $h \leftarrow$ a normal form of $t$ modulo $G$ w.r.t. the old term order $\leq'$
    $p \leftarrow h + q$
    $C \leftarrow$ the set of coefficients in $K[\underline{Y}]$ of
        $p \in K[\underline{Y}][X_1, \ldots, X_n]$
    if the system $S = \{ f = 0 \mid f \in C \}$ of linear equations in
       the indeterminates $\underline{Y}$ has a solution then
       $\{ c_s \in K \mid s \in R \} \leftarrow$ a solution of $S$
       $g \leftarrow t + \sum_{s \in R} c_s s$
       $H \leftarrow H \cup \{t\}$
       $F \leftarrow F \cup \{g\}$
    else $R \leftarrow R \cup \{t\}$
         create a new indeterminate $Y_t$
         $\underline{Y} \leftarrow \underline{Y} \cup \{Y_t\}$
         $q \leftarrow Y_t h + q$
    end
end
return$(F, R)$
end CONVGRÖBNER
```

order that they are placed into H. Then by the way MINTERM chooses the t_k, the infinite sequence $\{t_k\}_{k \in \mathbb{N}}$ would have the property that $t_i \nmid t_j$ for all $i < j$, contradicting Dickson's lemma. Now assume the **else**-condition is satisfied infinitely many times. Note that the following invariant holds at the end of each run through the **while**-loop:

$$q = \sum_{s \in R} Y_s h_s \qquad (h_s \text{ the normal form of } s \text{ modulo } G).$$

Now let $\{s_k\}_{k \in \mathbb{N}}$ be the terms in R indexed in the order that they are placed into R. Then by Lemma 9.5 and the fact that the **if**-condition was false when s_k was placed into R, the ideal $\mathrm{Id}(G)$ contains no polynomial of the form

$$f_k = s_k + \sum_{i=1}^{k-1} a_i s_i \qquad (k \in \mathbb{N}).$$

From this we easily conclude that $\overline{s_i} \neq \overline{s_j}$ for $i \neq j$, and that $\{ \overline{s_k} \mid k \in \mathbb{N} \}$ is linearly independent in $K[\underline{X}]/\mathrm{Id}(G)$, contradicting the zero-dimensionality of $\mathrm{Id}(G)$.

Correctness: We will repeatedly make use of the fact that the sequence of terms chosen by MINTERM in the successive runs through the **while**-loop is strictly ascending. A moment's thought shows that the following are invariants that hold at the end of each run through the **while**-loop:

(i) $F \subseteq \mathrm{Id}(G)$, and

(ii) $R = \{ v \in T \mid v \leq t, \text{ and } s \nmid v \text{ for all } s \in H \}$.

Now let the output set F consist of the polynomials $\{g_1, \ldots, g_m\}$ indexed in the order of their placement into F, let s_i be the value of t at the time when g_i was placed into F, and let H_i and R_i be the values of H and R, respectively, at that point in time. Then clearly

$$H_i = \{s_1, \ldots, s_i\}.$$

Also, $s_1 < \cdots < s_m$, and $s_j \nmid s_k$ for $j < k$. It follows that there are no divisibilities among the s_i at all. Furthermore,

$$T(g_i) \setminus \{s_i\} \subseteq R_i \qquad (1 \leq i \leq m)$$

by the construction of g_i, and so $s_i = \mathrm{HT}(g_i)$ w.r.t. $\leq$ because the elements of R_i are earlier choices made by MINTERM. Moreover, the invariant (ii) above implies that g_i is in normal form modulo $\{g_1, \ldots, g_{i-1}\}$ and thus modulo $F \setminus \{g_i\}$.

We have proved that F is reduced. By (i) above, we also have $F \subseteq I$, where $I = \mathrm{Id}(G)$. To see that it is a Gröbner basis of I, we show that for every $s \in \mathrm{HT}(I)$, there exists $1 \leq i \leq m$ with $s_i \mid s$. Assume for a contradiction that this is not the case for some $s = \mathrm{HT}(f)$ with $f \in I$. We may assume w.l.o.g. that f is in normal form modulo F. Now s cannot be above the value of t in the last run through the loop because then the algorithm would have continued to run. Hence s must satisfy

$$t_1 < s \leq t_2$$

for two successive values t_1 and t_2 of t. Since it is not divisible by any s_j $(1 \leq j \leq m)$ at all, s is not divisible by any element of the value of H at the time when t_2 was chosen by MINTERM. It follows that $t_2 = s$. Let $s' \in T(f) \setminus \{s\}$. Then $s' < s$, and s' is not divisible by anything in $\mathrm{HT}(F)$, hence not by anything in the current value of H. Everything that is strictly between t_1 and s is divisible by something in that value of H by the choices that MINTERM makes. We see that $s' \leq t_1$, and thus s' is in the current value of R by (ii) above. It now follows that the **if**-condition must detect f and place s into H, a contradiction.

It remains to prove that the final value of R equals $\mathrm{RT}(I)$ w.r.t. $\leq$. This is immediate from (ii) above together with the fact that all terms above the final value of t are multiples of $H_m = \mathrm{HT}(F)$. □

We will now prove that the lexicographical term order satisfies (**D**) and give the algorithm MINTERM for this special case. We need a combinatorial lemma.

Lemma 9.8 Let $\leq$ be the lexicographical term order on T (where $X_1 \gg X_2 \gg \cdots \gg X_n$). Suppose $t = X_1^{\nu_1} \cdot \cdots \cdot X_{j+1}^{\nu_{j+1}}$ for some $1 \leq j < n$ with $\nu_{j+1} \neq 0$. Then

$$\min\{\, s \in T \mid t < s \text{ and } t \nmid s \,\} = X_1^{\nu_1} \cdot \cdots \cdot X_{j-1}^{\nu_{j-1}} \cdot X_j^{\nu_j+1}.$$

Proof From the fact that $\nu_{j+1} \neq 0$ we conclude that the term

$$s_{\min} = X_1^{\nu_1} \cdot \cdots \cdot X_{j-1}^{\nu_{j-1}} \cdot X_j^{\nu_j+1}$$

satisfies $t < s_{\min}$ and $t \nmid s_{\min}$. Now assume for a contradiction that $s \in T$ with $s < s_{\min}$ also has these properties. Then we must have

$$\deg_{X_k}(s) < \nu_k \qquad (*)$$

for at least one k with $1 \leq k \leq j+1$ since $t \nmid s$. Furthermore, $t < s < s_{\min}$ implies that

$$\deg_{X_k}(s) = \nu_k \quad \text{for} \quad 1 \leq k \leq j-1,$$

and

$$\nu_j \leq \deg_{X_j}(s) \leq \nu_j + 1.$$

If $\deg_{X_j}(s) = \nu_j$, then $(*)$ requires $\deg_{X_{j+1}}(s) < \nu_{j+1}$ and thus $s < t$ which is a contradiction. If $\deg_{X_j}(s) = \nu_j + 1$, then $s < s_{\min}$ means $\deg_{X_k}(s) < 0$ for some $j+1 \leq k \leq n$, which is absurd. □

Lemma 9.9 Let $\leq$ be the lexicographical term order on T, S a finite subset of T, and $t \in T$. Then the algorithm LMINTERM of Table 9.4 decides whether the set

$$M = \{\, u \in T \mid t < u, \text{ and } s \nmid u \text{ for all } s \in S \,\}$$

is empty and computes its $\leq$-minimal element if it is not.

TABLE 9.4. Algorithm LMINTERM

```
**Specification:** $w \leftarrow$ LMINTERM$(S, t)$
**Given:** a finite subset $S$ of $T$ and $t \in T$
**Find:** $w \in \{\textbf{false}\} \cup (\{\textbf{true}\} \times T)$ such that
        $v = \begin{cases} \textbf{false} & \text{if} \quad M = \emptyset \\ (\textbf{true}, u) & \text{otherwise,} \end{cases}$
        where $u$ is the $\leq$-minimal element of $M$
begin
$u \leftarrow t$
for $i = n$ to 1 do
    $u \leftarrow u \cdot X_i$
    if $s \nmid u$ for all $s \in S$ then    return((true,u))    end
    $u \leftarrow u/X_i^\nu$ where $\nu = \deg_{X_i}(u)$
end
return(false)   end
end LMINTERM
```

Proof Termination of the algorithm is trivial. It is also trivial that the division of u by X_i^ν with $\nu = \deg_{X_i}(u)$ yields an element of T. For correctness, let us now consider a call of the algorithm on a particular pair (S, t) of arguments. Let $n \geq k \geq 1$ be the least value assigned to i by the **for**-command, and for $n \geq j \geq k$, denote by u_j the value of u when testing the **if**-condition in the run $i = j$ through the **for**-loop. Then we have, for $n > j \geq k$,

$$u_j = \frac{u_{j+1}}{X_{j+1}^\nu} \cdot X_j, \tag{$*$}$$

where $\nu = \deg_{X_{j+1}}(u_{j+1})$. We first show that the following three statements are true for $n \geq j \geq k$:

(i) $u_j \in T(X_1, \ldots, X_j)$ with $\deg_{X_j}(u_j) > 0$,

(ii) $t < u_j$, and

(iii) for each term v with $t < v < u_j$, there exists $s \in S$ with $s \,|\, v$.

When $j = n$, then (i) is trivial, and (ii) and (iii) follow immediately from the fact that $u_n = t \cdot X_n$ is the immediate lexicographical successor of t. Now let $j < n$, and assume that the conditions hold for u_{j+1}. It follows immediately from ($*$) that (i) continues to hold and that $u_j > u_{j+1} > t$, i.e., (ii) continues to hold too. Since the loop was entered with $i = j$, we must have had $s \,|\, u_{j+1}$ for some $s \in S$. It follows that $s \,|\, u'$ for each multiple u' of u_{j+1}. By the last lemma together with (i) for u_{j+1} and ($*$), the $\leq$-least term above u_{j+1} which is not a multiple of u_{j+1} is u_j. This together with (iii) for u_{j+1} implies (iii) for u_j.

Now if the **if**-condition is found to be true for u_k, then correctness of this call of the algorithm follows easily from (ii) and (iii) for u_k. If not, then we must have $k = 1$, and $u_1 = X_1^\nu$ by (i) for u_1. Furthermore, since the **if**-condition is not satisfied, there exists $s = X_1^\mu \in S$ with $\mu \leq \nu$. By (iii) for u_1, there is no possible **true**-output below u_1, u_1 itself is not good, and all terms u above u_1 obviously satisfy $\deg_{X_1}(u) \geq \nu$, so they are divisible by s and thus not good either. We see that M is indeed empty, as our output would have us believe. □

Exercise 9.10 Let $T = T(X, Y, Z)$, $\leq$ the lexicographical term order with $X \gg Y \gg Z$. Furthermore, let

$$S = \{X^4, Y^3, YZ, Z^2\}$$

and $t = X^3$. Use repeated calls of LMINTERM to make a lexicographically ascending list of those terms above t that are not divided by any element of S. Why is the list finite?

Exercise 9.11 Modify the algorithm CONVGRÖBNER in such a way that it converts a given Gröbner basis w.r.t. any term order to a Gröbner basis w.r.t. a total degree order of the same ideal. Do not require that this ideal must be zero-dimensional. (Hint: You can realize MINTERM in this case by simply going through all terms in ascending order and dumping those that are multiples of elements of H. This will work fine, except that it does not tell you when you're done, i.e., when the new Gröbner basis has been picked up completely. There is an easy way of checking this.)

9.2 Computing in Finitely Generated Algebras

Let I be a proper ideal in a multivariate polynomial ring $K[\underline{X}]$ over a field K. We have discussed how we can compute in the residue class ring $K[\underline{X}]/I$ (Theorem 5.55). Furthermore, we saw in Section 6.3 that the residue classes of the reduced terms (w.r.t. some term order) are a basis of the K-vector space $K[\underline{X}]/I$, and that this canonical term basis is finite and can be computed provided I is a zero-dimensional ideal. In order to fully capture this structural diversity and its computational aspects, we need to consider *finitely generated K-algebras*. Throughout, K will be a field.

A **(commutative) K-algebra** is a commutative ring A containing K as a subring. Natural examples are thus the polynomial rings over K. Moreover, if I is a proper ideal of such a polynomial ring $K[\underline{X}]$, then the canonical homomorphism

$$\begin{array}{ccc} K[\underline{X}] & \longrightarrow & K[\underline{X}]/I \\ f & \longmapsto & f + I \end{array}$$

must be injective when restricted to K, for otherwise I would contain an element of K. Identifying K with its canonical image in $K[\underline{X}]/I$, we may thus operate on the assumption that K is a subring of $K[\underline{X}]/I$ and hence that $K[\underline{X}]/I$ is a K-algebra.

The usual concepts in ring theory relativize in a natural way to K-algebras. Let A and B be K-algebras. Then B is called a **K-subalgebra** of A if B is a subring of A with $K \subseteq B$. A **K-ideal** in A is a ring ideal I of A with $I \cap K = \{0\}$; in other words I is a proper ideal of A. For any K-ideal I, the quotient ring A/I forms a K-algebra when the elements of K are identified with their residue classes modulo I. A **K-algebra homomorphism** $\varphi : A \longrightarrow B$ is a ring homomorphism $\varphi : A \longrightarrow B$ with $\varphi \restriction K = \mathrm{id}_K$. A subset C of A **generates** A as a K-algebra if $A = K[C]$, i.e., A equals the result of adjoining C to K within A in the sense of Definition 1.109. A is **finitely generated** if it is generated by some finite set.

Every K-algebra A forms a K-vector space under the scalar multiplication given by the multiplication between elements of K and elements of A. Note that in the example $K[\underline{X}]/I$ described above, this yields the standard scalar multiplication of Example 3.2 (iii). **Basis** and **dimension** of A are defined in terms of A as K-vector space.

Note that the statement "C generates A" as defined above does *not* mean that C is a generating system for A as a K-vector space in the sense of Definition 3.5 (ii). The difference becomes obvious if we consider the K-algebra $A = K[\underline{X}]$. Here, A is finitely generated because it is generated by the finite set $\{X_1, \dots, X_n\}$. On the other hand, there does not exist a finite generating system for A as a K-vector space: if this were the case, then A would have to be finite dimensional, which it cannot be because the infinite set of all terms is a basis in this case.

Now let C be a basis of A. Then every product $c \cdot c'$ of elements of C has a unique representation as a linear combination of the elements of C,

$$c \cdot c' = \sum_{c'' \in C} \alpha_{cc'c''} \cdot c''$$

with $\alpha_{cc'c''} \in K$. (The sum is of course formally infinite in general, but only finitely many summands are non-zero.) The $\alpha_{cc'c''}$ are called the **structure constants** of A w.r.t. C.

Lemma 9.12 Let A be a K-algebra and let C be a basis of A. Then multiplication in A is uniquely determined by the structure constants of A w.r.t. C.

Proof Let $\{ \alpha_{cc'c''} \mid (c, c', c'') \in C^3 \}$ be the family of structure constants of A, and let $d = \sum_{c \in C} d_c c$ and $e = \sum_{c' \in C} e_{c'} c'$ be arbitrary elements of A. Then

$$de = \sum_{c, c' \in C} d_c e_{c'} c c'$$

$$\begin{aligned} &= \sum_{c,c' \in C} d_c e_{c'} \sum_{c'' \in C} \alpha_{cc'c''} c'' \\ &= \sum_{c'' \in C} \Bigg(\sum_{c,c' \in C} d_c e_{c'} \alpha_{cc'c''} \Bigg) c''. \quad \square \end{aligned}$$

More generally, the next lemma shows that a K-algebra is determined up to isomorphism by its basis and its structure constants.

Lemma 9.13 Let A and B be K-algebras with bases C and D and structure constants $\alpha_{cc'c''}$ and $\beta_{dd'd''}$, respectively. Let $\varphi : A \longrightarrow B$ be a homomorphism (an embedding, an isomorphism) of K-vector spaces such that $\varphi \restriction K = \mathrm{id}_K$, and let

$$\{a_{cd} \mid c \in C, \ d \in D\} \subseteq K$$

be such that $\varphi(c) = \sum_{d \in D} a_{cd} d$ for all $c \in C$. Then φ is a homomorphism (an embedding, an isomorphism) of K-algebras iff

$$\sum_{d,d' \in D} a_{cd} a_{c'd'} \cdot \beta_{dd'd''} = \sum_{c'' \in C} a_{c''d''} \cdot \alpha_{cc'c''}$$

for all $c, c' \in C$ and $d'' \in D$.

Proof Clearly, $\varphi(1) = 1$ and $\varphi(a + b) = \varphi(a) + \varphi(b)$ for all a, $b \in A$. For multiplication, it suffices by the previous lemma to consider c, $c' \in C$. We have

$$\begin{aligned} \varphi(cc') &= \varphi\Bigg(\sum_{c'' \in C} \alpha_{cc'c''} c'' \Bigg) \\ &= \sum_{c'' \in C} \alpha_{cc'c''} \cdot \varphi(c'') \\ &= \sum_{c'' \in C} \alpha_{cc'c''} \sum_{d'' \in D} a_{c''d''} d'' \\ &= \sum_{d'' \in D} \Bigg(\sum_{c'' \in C} a_{c''d''} \cdot \alpha_{cc'c''} \Bigg) d'', \end{aligned}$$

and

$$\begin{aligned} \varphi(c) \cdot \varphi(c') &= \Bigg(\sum_{d \in D} a_{cd} d \Bigg) \cdot \Bigg(\sum_{d' \in D} a_{c'd'} d' \Bigg) \\ &= \sum_{d,d' \in D} a_{cd} a_{c'd'} dd' \\ &= \sum_{d'' \in D} \Bigg(\sum_{d,d' \in D} a_{cd} a_{c'd'} \cdot \beta_{dd'd''} \Bigg) d''. \quad \square \end{aligned}$$

We will now restrict our attention to finitely generated K-algebras $A = K[b_1, \ldots, b_n]$. It is easy to see that any non-trivial homomorphic image of the polynomial ring $K[X_1, \ldots, X_n]$ is such a K-algebra if we identify K with its image under the homomorphism φ in question. We then have

$$\varphi(K[X_1, \ldots, X_n]) = K[\varphi(X_1), \ldots, \varphi(X_n)].$$

Conversely, if $A = K[b_1, \ldots, b_n]$ is a K-algebra, then the map $X_i \longmapsto b_i$ for $1 \leq i \leq n$ extends in a unique way to a surjective K-algebra homomorphism

$$\begin{array}{llcl} \varphi: & K[X_1, \ldots, X_n] & \longrightarrow & A \\ & \sum\limits_{i=1}^{m} a_i X_1^{\nu_{i1}} \cdot \ldots \cdot X_n^{\nu_{in}} & \longmapsto & \sum\limits_{i=1}^{m} a_i b_1^{\nu_{i1}} \cdot \ldots \cdot b_n^{\nu_{in}} \end{array}$$

So by the homomorphism theorem for rings, A is isomorphic as a K-algebra to $B = K[X_1, \ldots, X_n]/I$, where $I = \ker(\varphi)$, via the canonical isomorphism that maps $X_i + I$ to b_i. We see that up to isomorphic images, finitely generated K-algebras are quotient algebras of the form $K[X_1, \ldots, X_n]/I$ (I a proper ideal), which is the type of K-algebra that we were interested in to begin with.

For the rest of this section, we let $K[\underline{X}] = K[X_1, \ldots, X_n]$ and $\leq$ a term order on $T = T(X_1, \ldots, X_n)$. I is a proper ideal in $K[\underline{X}]$, and for $f \in K[\underline{X}]$, the residue class $f + I \in K[\underline{X}]/I$ is denoted by $\overline{f}$. Recall that $\mathrm{HT}(I)$ is the set of all head terms of polynomials in I, and $\mathrm{RT}(I) = T \setminus \mathrm{HT}(I)$ is the set of reduced terms w.r.t. I and $\leq$. In Proposition 6.52, we saw that the map $t \longmapsto \overline{t}$ is a bijection between $\mathrm{RT}(I)$ and the canonical term basis $\{\, \overline{t} \mid t \in \mathrm{RT}(I)\,\}$ of $K[\underline{X}]/I$. In view of this, we will write $\alpha_{tt't''}$ for the structure constant $\alpha_{\overline{t}\,\overline{t'}\,\overline{t''}}$.

Let us recall how we compute in the residue class ring $K[\underline{X}]/I$. First, we need a Gröbner basis of I which can be computed provided I is given by a finite generating set. Elements of $K[\underline{X}]/I$ are represented by their unique normal forms modulo G. Addition is performed by combining like terms, which, as one easily sees, results in a normal form because no reducible terms can be produced in the process. Two elements of $K[\underline{X}]/I$ are multiplied by multiplying them out as in $K[\underline{X}]$ and then reducing the result—which may now well be reducible—back to normal form modulo G. Since every term in a normal form modulo G is an element of $\mathrm{RT}(I)$, a normal form can be viewed as a linear combination of elements of $\mathrm{RT}(I)$; so in terms of the vector space structure, we are actually representing elements of $K[\underline{X}]/I$ as linear combinations of the canonical term basis. Now if the structure constants of $K[\underline{X}]/I$ w.r.t. G are known to us, then we can use the formula given in the proof of Lemma 9.12 for multiplication. This will then eliminate the necessity of reducing back to normal form and thus potentially speed up computations.

Our goal is now to compute the structure constants for $K[\underline{X}]/I$ w.r.t. the canonical term basis $\mathrm{RT}(I)$ from a Gröbner basis of I. If $K[\underline{X}]/I$ is finite-dimensional as a K-vector space, i.e., if I is a zero-dimensional ideal (Theorem 6.54), then computing in $K[\underline{X}]/I$ will thus be completely determined relative to the field operations by the finite set of data given by $\mathrm{RT}(I)$ and the corresponding structure constants. For the general case, we will see how we can still—by means of reduction modulo a Gröbner basis—compute the structure constants needed for each specific instance of the multiplication formula. A computational gain, however, is achieved only in the zero-dimensional case, where the finitely many structure constants can be computed once and for all.

It is easy to see how any structure constant $\alpha_{tt't''}$ (where $t, t', t'' \in \mathrm{RT}(I)$) can be computed once a Gröbner basis of I is at hand. All we have to do is compute a normal form

$$h = \sum_{s \in T(h)} a_s s \qquad (a_s \in K)$$

of tt' modulo G. Then $\overline{tt'} = \overline{h}$ in $K[\underline{X}]/I$, and since h is in normal form modulo G, $\overline{h}$ can be viewed as a linear combination of elements of $\mathrm{RT}(I)$:

$$\overline{tt'} = \overline{h} = \overline{\sum_{s \in T(h)} a_s s} = \sum_{s \in T(h)} a_s \cdot \overline{s}\,.$$

It is now clear that

$$\alpha_{tt't''} = \begin{cases} a_{t''} & \text{if} \quad t'' \in T(h) \\ 0 & \text{otherwise.} \end{cases} \tag{$*$}$$

Two special cases are of particular importance. If tt' is itself in normal form, then $h = tt'$, and we have

$$\alpha_{tt't''} = \begin{cases} 1 & \text{if} \quad t'' = tt' \\ 0 & \text{otherwise.} \end{cases} \tag{$**$}$$

For the other special case, recall that the stairs $\mathrm{st}(I)$ of I is the unique minimal finite basis of the set $\mathrm{HT}(I)$ w.r.t. the divisibility relation on T. Recall further that the unique reduced Gröbner basis G of I (w.r.t. $\leq$) satisfies $\mathrm{HM}(G) = \mathrm{st}(I)$, and that

$$T(g) \setminus \{\mathrm{HT}(g)\} \subseteq \mathrm{RT}(I)$$

for all $g \in G$. Now if we compute a normal form of tt' where $tt' \in \mathrm{st}(I)$, then this computation will be one reduction step $tt' \xrightarrow[g]{} tt' - g$ where g is the unique element of G with $\mathrm{HM}(g) = tt'$. We see that

$$\alpha_{tt't''} = \begin{cases} \text{the coefficient of } t'' \text{ in } -g & \text{if} \quad t'' \in T(g) \\ 0 & \text{otherwise.} \end{cases} \tag{$***$}$$

We see that in the two special cases above, the structure constants can be obtained without any computation at all, whereas in the general case, a potentially lengthy reduction chain is required. Our aim is now to arrange the computation in such a way that the general case can be reduced to the special cases in a way that requires only modest computational effort. The following lemma will be instrumental.

Lemma 9.14 If $tt' \notin \mathrm{RT}(I)$, then $\alpha_{tt't''} = 0$ for all $t'' \geq tt'$.

Proof Let h be as in the discussion above. Then $h < tt'$ in the induced order on $K[\underline{X}]$, and $tt' \notin T(h)$, so we must have $\mathrm{HT}(h) < tt'$. The claim is now obvious from the formula $(*)$ for $\alpha_{tt't''}$. □

It is clear that whenever t, t', s, s', $t'' \in T$, with $tt' = ss'$, then $\alpha_{tt't''} = \alpha_{ss't''}$. An effective computation should of course make use of this fact. We therefore define, for $u \in \mathrm{RT}(I) \cdot \mathrm{RT}(I)$ and $v \in \mathrm{RT}(I)$, the **combined structure constant**

$$\beta_{uv} = \alpha_{tt'v} \qquad (t, t' \in \mathrm{RT}(I) \text{ with } tt' = u),$$

and the family of combined structure constants w.r.t. $\mathrm{RT}(I)$ as

$$\widehat{\beta} = \{\, \beta_{uv} \mid u \in \mathrm{RT}(I) \cdot \mathrm{RT}(I),\ v \in \mathrm{RT}(I) \,\}.$$

It now clearly suffices to compute the set of combined structure constants. The idea behind the following algorithm is to arrange the computation of the β_{uv} by increasing first subscript.

Proposition 9.15 *Assume that the ground field K is computable, $\leq$ is a decidable term order, the ideal I is zero-dimensional, and suppose the reduced Gröbner basis G of I w.r.t. $\leq$ and the set* $\mathrm{RT}(I)$ *of reduced terms have been computed. Then the algorithm* STRCONST *of Table* 9.5 *computes the family $\widehat{\beta}$ of combined structure constants w.r.t.* $\mathrm{RT}(I)$.

Proof Termination of the algorithm is trivial. The algorithm clearly considers all pairs (u, v) for which a combined structure constant β_{uv} exists, and correctness is immediate from the formulas $(**)$ and $(***)$ in the **if**- and **elsif**-case. For the remaining case, we first note that here, u is a proper multiple of some element of $\mathrm{HM}(G)$ and can thus be written in the indicated form. Next, we claim that $u \in \mathrm{RT}(I) \cdot \mathrm{RT}(I)$ and $u = u'X_i$ imply that $u' \in \mathrm{RT}(I) \cdot \mathrm{RT}(I)$ and $X_i \in \mathrm{RT}(I)$. Let $u = st$ with s and t reduced, and assume w.l.o.g. that $s = s'X_i$. The claim now follows immediately from the facts that $u' = s't$ and that a factor of a reduced term is again reduced. This shows that the combined structure constants $\beta_{u'w}$ and $\beta_{(wX_i)v}$ occurring in the sum are well-defined. Furthermore, they are available to us at that point of the computation: clearly, $u' < u$, and $w < u'$ implies

TABLE 9.5. Algorithm STRCONST

```
Specification: $\widehat{\beta} \leftarrow$ STRCONST(RT($I$), $G$)
Given: the reduced Gröbner basis $G$ of a zero-dimensional ideal $I$ w.r.t.
        a decidable term order $\leq$, and the set RT($I$) of reduced terms
Find: the family $\widehat{\beta}$ of combined structure constants
begin
$U \leftarrow \mathrm{RT}(I) \cdot \mathrm{RT}(I)$;   $V \leftarrow \mathrm{RT}(I)$
create a matrix $B$ with an entry $\beta_{uv}$ for each $u \in U$ and $v \in V$
while $U \neq \emptyset$ do
    $u \leftarrow$ the $\leq$-minimal element of $U$
    $U \leftarrow U \setminus \{u\}$
    if $u \in \mathrm{RT}(I)$ then
      for all $v \in V$ do
          $\beta_{uv} \leftarrow \begin{cases} 1 & \text{if } u = v \\ 0 & \text{otherwise} \end{cases}$
      end
    elsif $u \in \mathrm{HM}(G)$ then
      $g \leftarrow$ the element of $G$ with $\mathrm{HM}(g) = u$
      for all $v \in V$ do
          $\beta_{uv} \leftarrow \begin{cases} \text{the coefficient of } v \text{ in } -g & \text{if } v \in T(g) \\ 0 & \text{otherwise} \end{cases}$
      end
    else write $u = u'X_i$ for some $1 \leq i \leq n$ such that $u' \notin V$
        for all $v \in V$ do
            $\beta_{uv} \leftarrow \sum_{\substack{w \in V \\ w < u'}} \beta_{u'w} \cdot \beta_{(w \cdot X_i)v}$
        end
    end
end
end STRCONST
```

$wX_i < u'X_i = u$. Correctness now follows from the following equation which uses Lemma 9.14.

$$\begin{aligned} u &= u'X_i \\ &= \sum_{\substack{w \in \mathrm{RT}(I) \\ w < u'}} \beta_{u'w} \cdot wX_i \\ &= \sum_{\substack{w \in \mathrm{RT}(I) \\ w < u'}} \beta_{u'w} \cdot \left(\sum_{v \in \mathrm{RT}(I)} \beta_{(wX_i)v} \cdot v \right) \end{aligned}$$

$$= \sum_{v \in \mathrm{RT}(I)} \left(\sum_{\substack{w \in \mathrm{RT}(I) \\ w < u'}} \beta_{u'w} \cdot \beta_{(wX_i)v} \right) \cdot v \quad \square$$

Exercise 9.16 Show that from $\mathrm{RT}(I)$ and the structure constants w.r.t. $\overline{\mathrm{RT}(I)}$, one can reconstruct the stairs $\mathrm{st}(I)$ and those elements of the reduced Gröbner basis G of I whose head term is not of the form X_i. (Hint: Argue that every element of $\mathrm{st}(I)$ is either of the form X_i, or of the form sX_j with s, $X_j \in \mathrm{RT}(I)$. Form all products of the latter kind, weed out those that are still reduced, then those that are multiples of others. Use the results on computing structure constants to reconstruct the corresponding elements of G.)

Concluding this section, we consider a question that arises naturally in connection with finitely generated K-algebras and Gröbner bases. We have seen how Gröbner bases can be used to decide the ideal membership problem for I and hence the equivalence problem for $\equiv_I$, where I is an ideal in the special K-algebra $K[\underline{X}]$. We will now show how the same methods can be employed to decide the membership problem for ideals J in $K[\underline{X}]/I$ provided finite bases of I and J are given. In view of the fact that every finitely generated K-algebra is isomorphic to such a quotient algebra, we are solving the ideal membership problem for suitably presented finitely generated K-algebras. As a first step, we prove that ideals in these algebras are always finitely generated.

Lemma 9.17 Every finitely generated K-algebra A is noetherian as a ring.

Proof Let $A = K[\underline{X}]/I$ for some proper ideal I of $K[\underline{X}]$ and let J be an ideal in A. Then by Proposition 1.63, the "lifting"

$$J^{\sim} = \{ f \in K[\underline{X}] \mid f + I \in J \}$$

is an ideal in $K[\underline{X}]$ with $J^{\sim} \supseteq I$. By the Hilbert basis theorem, $J^{\sim}$ has a finite generating set H. It is now easy to see that $H' = \{ f + I \mid f \in H \}$ is a finite generating set of J. □

Exercise 9.18 Derive the statement of the lemma above from Proposition 3.32.

Suppose now I is given as $\mathrm{Id}(F)$ for some finite set F of polynomials in $K[\underline{X}]$. Assume further that J is an ideal of $K[\underline{X}]/I$. As a consequence of the lemma above, we may assume that J is given to us in the following way: H is a finite subset of $K[\underline{X}]$, $H' = \{ f + I \mid f \in H \}$, and $J = \mathrm{Id}(H')$. With this setup, the ideal membership problem for J can now be solved algorithmically as follows.

Theorem 9.19 *Let F and H finite subsets of $K[\underline{X}]$. Set $I = \mathrm{Id}(F)$ and*

$$J = \mathrm{Id}(\{ f + I \mid f \in H \}),$$

and let G be a Gröbner basis of $\mathrm{Id}(F \cup H)$ with respect to some term order. Then $f + I \in J$ iff $f \xrightarrow[G]{*} 0$ for all $f \in K[\underline{X}]$.

Proof It is easy to see from the definition of J that

$$f + I \in J \quad \text{iff} \quad f \in \mathrm{Id}(H \cup I).$$

This together with $\mathrm{Id}(H \cup I) = \mathrm{Id}(H \cup F)$ implies $f + I \in J$ iff $f \xrightarrow[G]{*} 0$. □

9.3 Dimensions and the Hilbert Function

As in the previous section, we let I be a proper ideal of the polynomial ring $K[\underline{X}] = K[X_1, \dots, X_n]$ over the field K, and we denote by A the K-algebra $K[\underline{X}]/I$. Moreover, we will denote by T the set $T(\underline{X})$ of all terms in the variables $X_1, \dots, X_n$. Whenever $\{U_1, \dots, U_r\}$ is a subset of $\{X_1, \dots, X_n\}$, then we again denote by $T(\underline{U})$ and $K[\underline{U}]$ the set of all terms $t \in T$ and of all polynomials $f \in K[\underline{X}]$, respectively, that contain only variables $X_i \in \{U_1, \dots, U_r\}$. Recall that if I is a proper ideal of $K[\underline{X}]$, then $\{U_1, \dots, U_r\}$ is independent modulo I if $K[\underline{U}] \cap I = \{0\}$, and the dimension $\dim(I)$ of I is the maximum of the cardinalities of independent sets modulo I.

The only way to compute the dimension $\dim(I)$ that we have thus far is to compute all independent sets modulo I by means of Gröbner basis computations and then determine what the largest cardinality is among all these. We do know, however, that deciding zero-dimensionality of I is much easier than that: all we have to do is look at any Gröbner basis of I and see if it contains a univariate head term in each variable (Theorem 6.54). It is true that essentially the same method works for arbitrary dimensions: the dimension of I is the greatest cardinality of any set $\{U_1, \dots, U_r\}$ of variables with $T(\underline{U}) \cap \mathrm{HT}(G) = \emptyset$. The aim of this section is to prove this result for Gröbner bases w.r.t. total degree orders. Since these tend to be the easiest to compute, this special case is the most relevant for actual computations of dimensions. The generalization of the proof to arbitrary term orders will be outlined in the Notes to this chapter on p. 451.

The case of dimension zero being settled, we are mainly interested in the case $\dim(I) > 0$. By Theorem 6.54, this means we are considering infinite-dimensional K-algebras $A = K[\underline{X}]/I$. Our goal is nevertheless to find a way to distinguish between the different "sizes" of these algebras. The idea is to consider, instead of all of A, subspaces A_m of A obtained by bounding the total degree of polynomials by $m \in \mathbb{N}$. Then each A_m is finite-dimensional, and we can measure the order of growth of $\dim_K(A_m)$ as $m \longrightarrow \infty$. This is the idea behind the *Hilbert function*, a concept that has numerous applications in the theory of polynomial ideals. For $m \in \mathbb{N}$, we set

$$T_m = \{\, t \in T \mid \deg(t) \leq m \,\} \quad \text{and} \quad A_m = \{\, \bar{f} \in A/I \mid \deg(f) \leq m \,\}.$$

Then A_m is closed under addition and under multiplication by elements of K and hence is a subspace of the K-vector space $K[\underline{X}]/I$. It is easy to see that the finite set $\overline{T_m}$ is a generating system for A_m and thus A_m is finite-dimensional. The function

$$\begin{array}{rccc} H_I: & \mathbb{N} & \longrightarrow & \mathbb{N} \\ & m & \longmapsto & \dim_K(A_m) \end{array}$$

is called the **Hilbert function** of the ideal I. It is clear from Theorem 6.54 that I is zero-dimensional iff the Hilbert function H_I is eventually constant. Note that $1 \leq H_I(m)$ for all $m \in \mathbb{N}$ since A_m is never empty.

Our first goal now is to find upper and lower bounds for $H_I(m)$. To this end, we need to know what $|T_m|$ is, i.e., how many terms there are of total degree $\leq m$.

Lemma 9.20 $|T_m| = \binom{m+n}{n}$ for all $m \in \mathbb{N}$.

Proof Let $B = \{0,1\}^{m+n}$, i.e., the set of all $(m+n)$-tuples with entries from $\{0,1\}$. We define a map $\varphi : T_m \longrightarrow B$ as follows. If $t = X_1^{\nu_1} \cdot \cdots \cdot X_n^{\nu_n} \in T_m$, then we set $\varphi(t) = (a_1, \ldots, a_{m+n})$ with

$$a_k = \begin{cases} 0 & \text{if } \; k = \nu_1 + \cdots + \nu_i + i \text{ for some } 1 \leq i \leq n \\ 1 & \text{otherwise.} \end{cases}$$

The definition of $\varphi(t)$ can be visualized as follows: write down ν_1 many ones, then a zero to mark the end of the first exponent, then ν_2 many ones, and so on through ν_n many ones, then another zero, and finally $m - \deg(t)$ many ones.

$$\varphi(t) = (\underbrace{1, \ldots, 1}_{\nu_1 \text{ times}}, 0, \underbrace{1, \ldots, 1}_{\nu_2 \text{ times}}, 0, \ldots, \underbrace{1, \ldots, 1}_{\nu_n \text{ times}}, 0, \underbrace{1, \ldots, 1}_{\substack{m-\deg(t) \\ \text{times}}})$$

It is now an easy exercise to prove that φ is injective, and that the image of φ consists of all those $(a_1, \ldots, a_{m+n}) \in B$ with $a_k = 0$ for exactly n different indices k. It is clear that there are exactly $\binom{m+n}{n}$ such tuples. □

Lemma 9.21 The Hilbert function H_I of I satisfies

$$\binom{m+d}{d} \leq H_I(m) \leq \binom{m+n}{n}$$

for all $m \in \mathbb{N}$, where $d = \dim(I)$.

Proof The second inequality is immediate from Lemma 9.20 and the fact that $\overline{T_m}$ is a generating system for the K-vector space A_m. The first inequality is trivial for $d = 0$ since $1 \leq H_I(m)$ always holds. Now let $d > 0$. Then there exists an independent set $\{U_1, \ldots, U_r\}$ modulo I with cardinality d. It is an easy consequence of Lemma 6.53 (ii) and the definition of independence that the set $\overline{T(\underline{U}) \cap T_m}$ is linearly independent in A_m. Furthermore, we must have

$$|T(\underline{U}) \cap T_m| = \left| \overline{T(\underline{U}) \cap T_m} \right|,$$

since otherwise I would contain a polynomial of the form $0 \neq t_1 - t_2$ with $t_1, t_2 \in T(\underline{U})$. Together with Lemma 9.20 applied to $T(\underline{U})$, we conclude that

$$\binom{m+d}{d} = |T(\underline{U}) \cap T_m| = \left|\overline{T(\underline{U}) \cap T_m}\right| \leq \dim_K(A_m) = H_I(m). \quad \square$$

Recall that when a term order $\leq$ has been fixed, we denote by $\mathrm{RT}(I)$ the set $T \setminus \mathrm{HT}(I)$ of reduced terms w.r.t. I.

Definition 9.22 Let $\leq$ be a term order on T. A subset $\{U_1, \ldots, U_r\}$ of $\{X_1, \ldots, X_n\}$ is called **strongly independent** modulo the ideal I (w.r.t. to $\leq$) if $T(\underline{U}) \cap \mathrm{HT}(I) = \emptyset$, i.e., if $T(\underline{U}) \subseteq \mathrm{RT}(I)$. The number

$$d = \max\{\, |U| \mid U \subseteq \underline{X} \text{ and } U \text{ is strongly independent mod } I\,\}$$

is called the **strong dimension** of I (w.r.t. $\leq$).

Note that according to its definition, the strong dimension of I appears to depend on the term order in question, while the dimension of I obviously does not. What we are going to prove here is that for total degree orders, strong dimension and dimension coincide. It can actually be shown (cf. Section "Notes" on p. 451) that this is true for arbitrary term order, so that really the strong dimension does not depend on the term order at all. The following lemma tells us, among other things, that the strong dimension of I w.r.t. $\leq$ can be found by inspecting the head terms of a single Gröbner basis of I w.r.t. $\leq$.

Lemma 9.23 Let d and d' be the dimension and strong dimension of I, respectively. Suppose G is a Gröbner basis of I w.r.t. any term order, and let $\{U_1, \ldots, U_r\} \subseteq \{X_1, \ldots, X_n\}$. Then the following hold:

(i) If $\{U_1, \ldots, U_r\}$ is strongly independent mod I, then it is independent mod I. The converse fails in general.

(ii) $d' \leq d$.

(iii) $\{U_1, \ldots, U_r\}$ is strongly independent mod I iff $T(\underline{U}) \cap \mathrm{HT}(G) = \emptyset$.

(iv) The strong dimension d' equals the maximum of the cardinalities of those subsets $\{U_1, \ldots, U_r\}$ of $\{X_1, \ldots, X_n\}$ that satisfy

$$T(\underline{U}) \cap \mathrm{HT}(G) = \emptyset.$$

Proof (i) If $\{U_1, \ldots, U_r\}$ is dependent mod I, then there exists $f \in K[\,\underline{U}\,] \cap I$ with $f \neq 0$. It follows that $\mathrm{HT}(f) \in T(\underline{U}) \cap \mathrm{HT}(I)$, and so $\{U_1, \ldots, U_r\}$ is not strongly independent mod I. To obtain a simple counterexample for the converse, consider the ideal $I = \mathrm{Id}(X_2 - X_1)$ of $K[X_1, X_2]$ with $X_1 < X_2$. Then $\{X_2\}$ is independent but not strongly independent mod I.

Statement (ii) is an immediate consequence of (i). Part (iii) follows easily from the fact that $\mathrm{HT}(I) = \mathrm{mult}(\mathrm{HT}(G))$, and this easily implies (iv). □

We need a few more technicalities before we can relate the Hilbert function, the dimension, and the strong dimension to obtain the main results of this section. Let $M \in \mathbb{N}$ and $t \in T$. Then we define

$$\mathrm{top}_M(t) = \{\, i \in \{1, \ldots, n\} \mid \deg_{X_i}(t) \geq M \,\},$$

i.e., $\mathrm{top}_M(t)$ is the set of indices where "t tops M." Furthermore, we set

$$\mathrm{sh}_M(t) = \prod_{i \in \mathrm{top}_M(t)} X_i^M \cdot \prod_{\substack{i=1 \\ i \notin \mathrm{top}_M(t)}}^{n} X_i^{\deg_{X_i}(t)},$$

i.e., $\mathrm{sh}_M(t)$ is "t shaved at M." It is easy to see that for $t \in T$ and $M \in \mathbb{N}$, we have

$$\mathrm{sh}_M\big(\mathrm{sh}_M(t)\big) = \mathrm{sh}_M(t).$$

With this observation in mind, the proof of the following lemma is elementary and straightforward.

Lemma 9.24 Let S be a subset of T and $M \in \mathbb{N}$. Then the following hold:

(i) The relation $\sim_M$ defined by

$$s \sim_M t \quad \text{iff} \quad \mathrm{sh}_M(s) = \mathrm{sh}_M(t)$$

is an equivalence relation on S.

(ii) If S is such that $\mathrm{sh}_M(s) \in S$ for all $s \in S$, then the set

$$R_M = \{\, t \in S \mid \mathrm{sh}_M(t) = t \,\}$$

of "already shaved terms" is a system of unique representatives for the partition of S into equivalence classes w.r.t. $\sim_M$. The set R_M can also be described as

$$R_M = \{\, t \in S \mid \deg_{X_i}(t) \leq M \text{ for } 1 \leq i \leq n \,\}.$$

(iii) With $[t]_{\sim_M}$ denoting the equivalence class of $t \in R_M$ w.r.t. $\sim_M$, we have

$$[t]_{\sim_M} = \{\, st \mid st \in S,\ \deg_{X_i}(s) = 0 \text{ for all } i \notin \mathrm{top}_M(t) \,\},$$

i.e., the elements of $[t]_{\sim_M}$ are obtained by raising those exponents in t that equal M in such a way that the result remains in S. □

Lemma 9.25 Let $\leq$ be a term order on T and G a Gröbner basis of I w.r.t. $\leq$. Let d' be the strong dimension of I w.r.t $\leq$, and set

$$M = \max\{ \deg_{X_i}(\mathrm{HT}(g)) \mid g \in G,\ 1 \leq i \leq n \}.$$

Then the following hold for all $t \in T$:

(i) Whenever $i \in \mathrm{top}_M(t)$ and $\nu \in \mathbb{N}$, then $t \in \mathrm{RT}(I)$ iff $t \cdot X_i^{\nu} \in \mathrm{RT}(I)$.

(ii) If $m \in \mathbb{N}$ with $m \geq n \cdot M$, then

$$\max\{ |\mathrm{top}_M(t)| \mid t \in \mathrm{RT}(I) \cap T_m \} = d'.$$

Proof (i) This is immediate from the definition of M: divisibility of t by an element of $\mathrm{HT}(G)$ is not affected by changing an exponent that exceeds M to something else exceeding M.

(ii) To prove the inequality "$\leq$," assume for a contradiction that there exists $t \in \mathrm{RT}(I) \cap T_m$ with $\deg_{X_i}(t) \geq M$ for more than d' many indices. Then there exists a subset $\{U_1, \dots, U_r\}$ of $\{X_1, \dots, X_n\}$ with more than d' many elements and a decomposition $t = t_1 \cdot t_2$ with $t_1 \in T(\underline{U})$ and $t_2 \in T(\underline{X} \setminus \underline{U})$ such that

$$\deg_{X_i}(t_1) \geq M \quad \text{for all} \quad X_i \in \{U_1, \dots, U_r\}.$$

We must have $t_1 \in \mathrm{RT}(I)$ because t_1 is a factor of t. On the other hand, the set $\{U_1, \dots, U_r\}$ cannot be strongly independent, and so there exists $g \in G$ with $\mathrm{HT}(g) \in T(\underline{U})$. From the fact that the degree in each variable of $\mathrm{HT}(g)$ is less than or equal to M, it now follows that $\mathrm{HT}(g) \mid t_1$ and thus $t \in \mathrm{HT}(I)$, a contradiction.

For the inequality "$\geq$" of (ii), let $i_1, \dots, i_{d'}$ be pairwise different indices such that $\{X_{i_1}, \dots, X_{i_{d'}}\}$ is strongly independent. Using the fact that $m \geq n \cdot M$, it is easy to see that

$$X_{i_1}^M \cdot \dots \cdot X_{i_{d'}}^M \in \mathrm{RT}(I) \cap T_m \quad \text{and} \quad \left|\mathrm{top}_M(X_{i_1}^M \cdot \dots \cdot X_{i_{d'}}^M)\right| = d'. \quad \square$$

We are now getting to a point where we need to specialize to total degree orders (see Example 5.8 (iii)).

Lemma 9.26 Suppose $\leq$ is a total degree term order on T, and let $m \in \mathbb{N}$. Then the following hold:

(i) The set $\overline{\mathrm{RT}(I) \cap T_m}$ is a basis of A_m, and it has as many elements as $\mathrm{RT}(I) \cap T_m$.

(ii) $H_I(m) = |\mathrm{RT}(I) \cap T_m|$ for all $m \in \mathbb{N}$.

Proof (i) It is an immediate consequence of Proposition 6.52 that

$$\overline{\mathrm{RT}(I) \cap T_m}$$

is linearly independent in A_m and has as many elements as $\mathrm{RT}(I) \cap T_m$. To see that it generates A_m, let $f \in K[\underline{X}]$ with $\deg(f) \leq m$. From the fact that $\leq$ is a total degree order it is easy to see that the unique normal form h of f modulo any Gröbner basis w.r.t. $\leq$ must satisfy the same degree bound

$$\deg(h) \leq m.$$

It follows that all residue classes of terms in the representation of $\overline{f}$ as a linear combination of elements of $\overline{\mathrm{RT}(I)}$ (see Lemma 6.53 (iv)) are actually in $\overline{\mathrm{RT}(I) \cap T_m}$.

(ii) This is immediate from (i). □

In the proof of the next theorem, which is the main theorem of this section, we will need the following elementary observation. If $k \in \mathbb{N}$, and we set

$$q = \frac{X \cdot (X-1) \cdot \cdots \cdot (X-k+1)}{k!},$$

then $q \in \mathbb{Q}[X]$ with $\deg(q) = k$, and for all $N \in \mathbb{N}$ with $N \geq k$, we obtain

$$q(N) = \binom{N}{k}.$$

It is in fact customary to write $q = \binom{X}{k}$ in this case.

Theorem 9.27 *Let $\leq$ be a total degree order on T and G a Gröbner basis of I w.r.t. $\leq$. Let d be the dimension of I and d' the strong dimension of I w.r.t $\leq$. Finally, set*

$$M = \max\{ \deg_{X_i}(\mathrm{HT}(g)) \mid g \in G,\ 1 \leq i \leq n \}.$$

Then the following hold:

(i) $d = d'$, so that in particular, the strong dimension does not depend on the choice of the term order.

(ii) There exists a unique polynomial $h \in \mathbb{Q}[X]$ of degree d such that $H_I(m) = h(m)$ for all $m \geq n \cdot M$. If the ground field K is computable, then h and the number $n \cdot M$ can be computed from any given basis of I.

Proof We begin by showing (ii) with d replaced by d'; the actual claim (ii) will then clearly follow from (i). By the previous lemma, the desired polynomial h must satisfy

$$h(m) = |\mathrm{RT}(I) \cap T_m| \qquad (*)$$

for all $m \geq n \cdot M$. We will arrive at such a polynomial by counting the elements of $\mathrm{RT}(I) \cap T_m$. To this end, we let $m \in \mathbb{N}$ with $m \geq n \cdot M$. Now $\mathrm{RT}(I) \cap T_m$ is the disjoint union of the equivalence classes w.r.t. the equivalence relation $\sim_M$ of (i) of Lemma 9.24. Using the set R_M of "shaved terms" of (ii) of that lemma as a system of unique representatives, we have

$$|\mathrm{RT}(I) \cap T_m| = \sum_{t \in R_M} \left|[t]_{\sim_M}\right|, \tag{**}$$

where of course $[t]_{\sim_M}$ is the equivalence class of t w.r.t. $\sim_M$. Note that $\deg(t) \leq n \cdot M$ for all $t \in R_M$. We need a more explicit expression for the summands on the right hand side of (**).

So given $t \in R_M$, what is the size of $[t]_{\sim_M}$, i.e., how many s are there in $\mathrm{RT}(I) \cap T_m$ with $\mathrm{sh}_M(s) = t$? It easy to see from Lemma 9.24 (iii) and Lemma 9.25 (i) that

$$\begin{aligned}[t]_{\sim_M} = \{\, st \mid {} & s \in T,\ \deg(s) \leq m - \deg(t), \text{ and} \\ & \deg_{X_i}(s) = 0 \text{ for all } i \notin \mathrm{top}_M(t) \,\},\end{aligned}$$

i.e., $[t]_{\sim_M}$ is obtained by multiplying t by terms of total degree less than or equal to $m - \deg(t)$ that contain only variables with indices where "t reaches M." In view of Lemma 9.20, this means that

$$\left|[t]_{\sim_M}\right| = \binom{m - \deg(t) + |\mathrm{top}_M(t)|}{|\mathrm{top}_M(t)|},$$

and we can thus improve (**) to

$$|\mathrm{RT}(I) \cap T_m| = \sum_{t \in R_M} \binom{m - \deg(t) + |\mathrm{top}_M(t)|}{|\mathrm{top}_M(t)|}.$$

According to the remark preceding the theorem, it is now clear that a polynomial in $\mathbb{Q}[X]$ satisfying (*) is given by $h = q(X - \deg(t) + |\mathrm{top}_M(t)|)$, where

$$q = \sum_{t \in R_M} \binom{X}{|\mathrm{top}_M(t)|}. \tag{***}$$

Moreover, a typical summand on the right hand side of (***) has degree $|\mathrm{top}_M(t)|$ and a positive head coefficient, and we see that

$$\begin{aligned}\deg(h) &= \max\{\, |\mathrm{top}_M(t)| \mid t \in R_M \,\} \\ &= \max\{\, |\mathrm{top}_M(t)| \mid t \in \mathrm{RT}(I) \cap T_m \,\}.\end{aligned}$$

Lemma 9.25 (ii) tells us that this maximum equals the strong dimension d'. The existence proof of the polynomial h that we have just given shows that the last statement of (ii) concerning computability is true. It is clear that there can be only one $h \in \mathbb{Q}[X]$ satisfying $H_I(m) = h(m)$ for infinitely many $m \in \mathbb{N}$: if there were two different ones, then their difference would be a non-zero polynomial with infinitely many zeroes.

In order to prove (i), we first recall from Lemma 9.23 (ii) that $d' \leq d$. Furthermore, (ii) and Lemma 9.21 tell us that

$$\binom{m+d}{d} \leq H_I(m) = h(m)$$

for all sufficiently large $m \in \mathbb{N}$. By the remark preceding the theorem, there exists a polynomial $f \in \mathbb{Q}[X]$ of degree d and with positive head coefficient such that

$$f(m) = \binom{m+d}{d}$$

for all $m \in \mathbb{N}$, and thus $f(m) \leq h(m)$ for all sufficiently large $m \in \mathbb{N}$. Now if d'—which equals $\deg(h)$—were strictly less than d, then $f - h$ would be a polynomial with positive head coefficent, and Lemma 8.113 would allow us to conclude that $f(m) > h(m)$ for all sufficiently large $m \in \mathbb{N}$, a contradiction. □

The polynomial described in (ii) of the theorem above is called the **Hilbert polynomial** of the ideal I.

Corollary 9.28 *Suppose $\leq$ is a total degree term order on T, and let U be a strongly independent set mod I of maximal cardinality. Then $|U| = \dim I$, and U is maximally independent mod I.*

Proof By definition, $|U|$ equals the strong dimension of I, which in turn equals $\dim(I)$ by the theorem above. U is independent mod I by Lemma 9.23 (i), and having the greatest possible cardinality $\dim(I)$, it must be maximally independent. □

In view of Lemma 9.23 (ii) and the corollary above, it is now clear that the dimension of I can be computed from a Gröbner basis w.r.t. a total degree order of I by determining the set

$$M = \{\, \{U_1, \ldots, U_r\} \subseteq \{X_1, \ldots, X_n\} \mid T(\underline{U}) \cap \mathrm{HT}(G) = \emptyset \,\}$$

of all strongly independent sets mod I and then finding the maximum of the cardinalities of elements of M. Furthermore, any element of M having this cardinality is maximally independent mod I. Let us emphasize again that by a similar but considerably deeper proof which is outlined in the Notes on p. 451, all this is true for arbitrary term orders. A *maximal strongly independent set mod I* is of course a strongly independent set mod I which is not properly contained in any strongly independent set mod I. The reader who has a background in computer science will recognize the algorithm of the following proposition as one that searches a finite tree for branches that are maximal with a certain property. The tree is $\mathcal{P}(\{X_1, \ldots, X_n\})$, partially ordered by the reflexive-transitive closure of the relation

$$U \; r \; V \quad \text{iff} \quad V = U \cup \{X_i\} \text{ with } \min\{\, j \mid X_j \in U \,\} < i,$$

and the property is being strongly independent mod I. The correctness proof of the algorithm provides the necessary comment to understand the action of the subalgorithm DIMREC.

Proposition 9.29 *Assume that G is a Gröbner basis of the proper ideal I of $K[\underline{X}]$ w.r.t. a decidable total degree order. Then the algorithm* DIMENSION *of Table* 9.6 *computes the set of all maximal strongly independent sets mod I, the dimension of I, and a maximally independent set mod I.*

TABLE 9.6. Algorithm DIMENSION

```
Specification: $(M, d, U) \leftarrow$ DIMENSION($G$)
Given: a Gröbner basis w.r.t. a decidable total degree order,
        with Id($G$) proper
Find: the dimension $d$ of Id($G$),
      the set $M$ of all maximal strongly independent sets mod $I$, and
      a maximally independent set $U$ mod Id($G$)

Subalgorithm DIMREC
Specification: $M' \leftarrow$ DIMREC($S, k, U, M$)
               where $M', M \subseteq \mathcal{P}(\{X_1, \ldots, X_n\})$, $S \subseteq T(X_1, \ldots, X_n)$,
               $k \in \mathbb{N}$, and $U \in \mathcal{P}(\{X_1, \ldots, X_n\})$
begin DIMREC
$M' \leftarrow M$
for $i = k$ to $n$ do
    if $T(U \cup \{X_i\}) \cap S = \emptyset$ then
        $M' \leftarrow$ DIMREC($S, i + 1, U \cup \{X_i\}, M'$) end
end
if $U$ is not contained in any $V \in M'$ then
    $M' \leftarrow M' \cup \{U\}$ end
end DIMREC

begin
$M \leftarrow$ DIMREC(HT($G$), 1, $\emptyset$, $\emptyset$)
$d \leftarrow \max\{\, |U| \mid U \in M \,\}$
$U \leftarrow$ an element of $M$ of cardinality $d$
return($M, d, U$)
end DIMENSION
```

Proof For simplicity, we call a (maximal) strongly independent set mod I an (m.)s.i. set. In view of the remarks preceding the proposition, it suffices to prove that DIMREC(HT(G), 1, $\emptyset$, $\emptyset$) is the set of all m.s.i. sets. This in turn is easily deduced from the following claim.

Claim: Suppose DIMREC is called with input (HT(G), k, U, M), where $1 \leq k \in \mathbb{N}$, U is an s.i. set, and M is a set of m.s.i. sets that already contains all m.s.i. sets V with $U \subseteq V$ with the possible exception of U itself and those with

$$\min\{\, l \mid X_l \in V \setminus U \,\} \geq k.$$

Then DIMREC outputs the union of M and the set of all m.s.i. sets that contain U.

Proof: Termination of DIMREC with any input is immediate from the fact that it calls itself only when $k \leq n$, and in that case, it does so at most $n - k + 1$ times, and with each recursive call, the value of k increases by 1.

To prove the correctness of DIMREC, we first consider the case $k > n$. Then M contains all m.s.i. sets V with $U \subseteq V$ with the possible exception of U itself. The **for**-loop is skipped, and it is easy to see that U itself is added to M if and only if it is an m.s.i. set. We see that the algorithm is correct in the sense of the claim.

Let now $k \leq n$. The case $k > n$ being settled, we may use a "downward induction" on k and assume that DIMREC runs correctly as claimed whenever its second argument is strictly greater than k. It is not hard to see that it now suffices to prove the following statement.

(∗) During the run $i = j$ through the **for**-loop, where $k \leq j \leq n$, M' is enlarged by all m.s.i. sets V with $U \subseteq V,\ \ U \neq V$, and

$$\min\{\, l \mid X_l \in V \setminus U \,\} = j,$$

and that at the end, DIMREC adds U to M' if and only if U itself is an m.s.i. set.

The second part of this statement is easily seen to be true once the first part has been proved. For the first part, we use induction on j, presenting the cases $j = k$ and $j > k$ simultaneously. So let us consider the run $i = j$ of the **for**-loop. If

$$T(U \cup \{X_j\}) \cap \mathrm{HT}(G) \neq \emptyset,$$

then $U \cup \{X_j\}$ is dependent and there is nothing to be added to M' in this run. We claim that else, the input

$$(\mathrm{HT}(G), j+1, U \cup \{X_j\}, M')$$

of the recursive call of DIMREC satisfies the premise of the claim. If V is an m.s.i. set that satisfies $U \cup \{X_j\} \subseteq V$, is not equal to $U \cup \{X_j\}$, and has the property that

$$\min\{\, l \mid X_l \in V \setminus (U \cup \{X_j\}) \,\} < j+1,$$

then it is easy to see that $U \subseteq V,\ \ U \neq V$, and

$$\min\{\, l \mid X_l \in V \setminus U \,\} < j.$$

But this implies that $V \in M'$: for the case $j = k$, it follows from our assumption on the input $(\mathrm{HT}(G), k, U, M)$, and for the induction step $j > k$, it follows from that assumption together with the fact that M' has already been enlarged by all m.s.i. sets V with $U \subseteq V$, $U \neq V$, and

$$k \leq \min\{\, l \mid X_l \in V \setminus U \,\} < j.$$

Since $j + 1 > k$, we may now use our "downward" induction hypothesis concerning k to conclude that the output of the recursive call

$$\mathrm{DIMREC}(\mathrm{HT}(G), j + 1, U \cup \{X_j\}, M')$$

is the union of M' and all m.s.i. sets V with $U \cup \{X_j\} \subseteq V$. Since M' already contained all m.s.i. sets V with $U \subseteq V$, $U \neq V$, and

$$\min\{\, l \mid X_l \in V \setminus U \,\} < j,$$

it is thus enlarged by all those with

$$\min\{\, l \mid X_l \in V \setminus U \,\} = j,$$

and we have proved statement $(*)$. □

Notes

The fact that residue class rings of polynomial rings are also vector spaces over the ground field, i.e., the fact that they are actually commutative algebras over the ground field, was at the center of Buchberger's interest when he developed the theory of Gröbner bases. Computing the canonical basis and the structure constants as discussed at the beginning of Section 9.1 and in Section 9.2 of this chapter was actually the topic of his doctoral dissertation. The computation of the vector space basis of the ideal itself can be found in Billera and Rose (1989). The algorithm UNIVPOL for the computation of the univariate polynomials in a zero-dimensional ideal using linear algebra is also due to Buchberger (see, e.g., Buchberger, 1985a). The conversion of a given Gröbner basis to a Gröbner basis w.r.t. the lexicographical term order appears in Faugère et. al (1990).

The Hilbert function was introduced by D. Hilbert in Hilbert (1890), §IV as the *characteristic function of a module*. Its connection with Gröbner bases was for the first time explored in Möller and Mora (1983). Our method to compute dimensions via strongly independent sets is in the spirit of Kredel and Weispfenning (1988), where, however, the focus is on lexicographical term orders.

The fact that the method works for arbitrary term orders can be proved as follows. It clearly suffices to prove that the strong dimension of an ideal w.r.t. any term order equals the dimension. To this end, one may repeat the theory of Section 9.3 through Theorem 9.27 with the ordinary total degree replaced by a *grading* (weighted total degree) with positive integer weights. (Section 10.2 has details on gradings.) The only difficulty is that counting the terms in n variables of weighted degree less than or equal to m is not as easy in general as for the ordinary total degree; it is, however, easy and entirely sufficient to bound this number from above and below by polynomials in m of degree n. This proves the claim for term orders that are compatible with a grading with positive integer weights. If $\leq$ is an arbitrary term order, then one can find a grading Γ with positive integer weights and a Γ-compatible term order $\leq'$ such that the reduced Gröbner basis G w.r.t. $\leq$ is also the reduced Gröbner basis w.r.t. $\leq'$, and $\mathrm{HT}(G)$ is the same w.r.t. $\leq$ and $\leq'$. (References and some further explanations are to be found in Section "Term Orders and Universal Gröbner Bases" of Chapter "Outlook on Advanced and Related Topics" at the end of this book.) It is clear that then the strong dimension of I w.r.t. $\leq$ equals that w.r.t. $\leq'$, and we already know that the latter equals the dimension of I. See also Carrà Ferro (1987a) and Bayer and Stillman (1992) on the subject of dimensions and the Hilbert function.

Methods for the computation of the dimension of an ideal also appear in Giusti (1984) and Kandri-Rody (1985). Sturmfels and White (1991) prove a conjecture of Kredel and Weispfenning (1988), namely, the fact that for a prime ideal I, every maximal strongly independent set modulo I has cardinality $\dim(I)$.

10

Variations on Gröbner Bases

10.1 Gröbner Bases over PID's and Euclidean Domains

The material in this section is not needed for the remaining sections of this chapter. Here, we will generalize the theory of Gröbner bases to polynomial rings over principal ideal domains. We will show that for every given finite subset F of such a polynomial ring, the equivalence problem for the ideal $\mathrm{Id}(F)$ is solvable by means of a Gröbner basis construction. The reduction relation will not in general allow the computation of unique normal forms, but it will be such that $f \in \mathrm{Id}(F)$ iff every normal form of f equals 0. This is good enough for the solution of the equivalence problem (cf. Theorem 5.55). For Euclidean domains that allow the computation of unique remainders, we will even obtain a reduction relation with unique normal forms.

Throughout this section, let R be a PID, $R[\underline{X}] = R[X_1, \dots, X_n]$, and $\leq$ a fixed term order on the set T of terms in $X_1, \dots, X_n$. Since $R[\underline{X}] \subseteq Q_R[X_1, \dots, X_n]$, we may assume that $R[\underline{X}]$ is endowed with the induced linear quasi-order of Theorem 5.12. For the same reason, we may regard every element of $R[\underline{X}]$ as an element of $Q_R[\underline{X}]$ and use the notation $T(f)$, $M(f)$, $\mathrm{HT}(f)$, $\mathrm{HC}(f)$, and $\mathrm{HM}(f)$ in the previously defined sense. We will make ample use of the results of Section 1.7.

Let $m_1 = a_1 t_1$ and $m_2 = a_2 t_2$ be monomials in $R[\underline{X}]$. We say that m_2 divides m_1 and write $m_2 \mid m_1$ if there is a monomial $m_3 \in R[\underline{X}]$ such that $m_1 = m_2 m_3$. Since the type of reduction that will be used makes sense over any domain, we will call it *D-reduction.*

Definition 10.1 Let f, g, $p \in R[\underline{X}]$. We say that f **D-reduces** to g **modulo** p and write $f \xrightarrow[p]{} g$, if there exists $m \in M(f)$ with $\mathrm{HM}(p) \mid m$, say $m = m' \cdot \mathrm{HM}(p)$, and $g = f - m'p$.

D-reduction modulo a finite subset of $R[\underline{X}]$, D-reducibilty, D-normal forms, and top-D-reduction are defined in the obvious way according to Definition 5.18. We will frequently use the notation for the various closures of $\xrightarrow[p]{}$ and $\xrightarrow[P]{}$ that was introduced in Definition 4.71. Recall that a field K is a PID since $\{0\} = \mathrm{Id}(0)$ and $K = \mathrm{Id}(1)$ are the only ideals. Now if R happens to be a field, then it is easy to see that D-reduction coincides with reduction as defined before. Moreover, it is always true that a D-reduction step in $R[X_1, \ldots, X_n]$ is an ordinary reduction step in $Q_R[X_1, \ldots, X_n]$ in the sense of Definition 5.18.

Lemma 10.2 Let P be a finite subset of $R[\underline{X}]$. Then the following hold:

(i) $f \xrightarrow[P]{} g$ implies $g < f$.

(ii) The relation $\xrightarrow[P]{}$ is a noetherian reduction relation.

(iii) $f \xleftrightarrow[P]{*} g$ implies $f - g \in \mathrm{Id}(P)$.

Proof The proof of (i) and (ii) is immediate from the fact that every D-reduction step is an ordinary reduction step in $Q_R[X_1, \ldots, X_n]$. The proof of (iii) is literally the same as the proof of "$\Longleftarrow$" in Lemma 5.26. □

It is clear that $\xrightarrow[P]{}$ will not in general have unique normal forms. Unfortunately, this will not even be the case if P is a Gröbner basis of the type that we will compute here. We will thus not be able to make use of the notion of confluence and Newman's lemma. We will have to rely on standard representations instead. Let $0 \neq f \in R[\underline{X}]$. As before, a **standard representation** of f w.r.t. a finite subset P of $R[\underline{X}]$ is a representation

$$f = \sum_{i=1}^{k} m_i p_i$$

with monomials m_i and $p_i \in P$ $(1 \leq i \leq k)$ such that $\mathrm{HT}(m_i p_i) \leq \mathrm{HT}(f)$ for $1 \leq i \leq k$. The next lemma shows that as with ordinary polynomial reduction over a field, D-reducibility to zero implies the existence of a standard representation.

Lemma 10.3 Let P be a finite subset of $R[\underline{X}]$, $0 \neq f \in R[\underline{X}]$, and assume that $f \xrightarrow[P]{*} 0$. Then f has a standard representation w.r.t. P.

Proof Let $0 \neq f \in R[\underline{X}]$ such that $f \xrightarrow[P]{*} 0$, but f does not have a standard representation. We may assume that f is minimal with this property. Since $f \xrightarrow[P]{*} 0$, there exists $h \in R[\underline{X}]$ with $f \xrightarrow[g]{} h$ for some $g \in P$, say $h = f - mg$. If $h = 0$, then $f = mg$ is a standard representation of f. If not, then h has a standard representation

$$h = \sum_{i=1}^{k} m_i p_i$$

w.r.t. P by the minimality of f. Using the fact that $\mathrm{HT}(mg)$ is a term in f, one easily sees that

$$f = mg + \sum_{i=1}^{k} m_i p_i$$

is a standard representation of f w.r.t. P. □

Recall that a D-normal form of f modulo P is an $h \in R[\underline{X}]$ which is not reducible modulo P with $f \xrightarrow[P]{*} h$. What we need to solve the equivalence problem for $\mathrm{Id}(P)$ is a finite subset G of $R[\underline{X}]$ such that $\mathrm{Id}(G) = \mathrm{Id}(P)$, and all D-normal forms modulo G of elements of $\mathrm{Id}(G)$ equal 0.

Definition 10.4 A **D-Gröbner basis** is a finite subset G of $R[\underline{X}]$ with the property that all D-normal forms modulo G of elements of $\mathrm{Id}(G)$ equal zero. If I is an ideal of $R[\underline{X}]$, then a **D-Gröbner basis of** I is a D-Gröbner basis that generates the ideal I.

Exercise 10.5 Let G be a finite subset of $R[\underline{X}]$. Show that the following are equivalent:

(i) G is a D-Gröbner basis.

(ii) Every $0 \neq f \in \mathrm{Id}(G)$ is D-reducible modulo G.

(iii) Every $0 \neq f \in \mathrm{Id}(G)$ is top-D-reducible modulo G.

(iv) For each $0 \neq f \in \mathrm{Id}(G)$, there exists $g \in G$ with $\mathrm{HM}(g) \mid \mathrm{HM}(f)$.

(v) The set of all monomial multiples of highest monomials of elements of G equals the set of all monomial multiples of highest monomials of elements of $\mathrm{Id}(G)$.

Exercise 10.6 Let I be an ideal of $R[\underline{X}]$ and G a finite subset of I. Show that G is a D-Gröbner basis of I if and only if for each $0 \neq f \in I$, there exists $g \in G$ with $\mathrm{HM}(g) \mid \mathrm{HM}(f)$.

As with ordinary Gröbner bases, it is much easier to give an abstract existence proof of D-Gröbner bases than it is to find an algorithm that constructs them. Note that the next proposition also provides a proof of a special case of the Hilbert basis theorem, namely, the fact that every polynomial ring over a PID is noetherian.

Proposition 10.7 *Assume that R is a PID and let I be an ideal of $R[\underline{X}]$. Then I has a D-Gröbner basis.*

Proof For each term $t \in T$, we set

$$I_t = \{\, a \in R \mid at = \mathrm{HM}(f) \text{ for some } f \in I \,\} \cup \{0\}.$$

From the fact that I is an ideal of $R[\underline{X}]$, one easily concludes that each I_t is an ideal of R, and that $s \mid t$ implies $I_s \subseteq I_t$ for all $s,\ t \in T$. We claim that the set

$$\{\, I_t \mid t \in T \,\}$$

is finite. Assume for a contradiction that there is a sequence $\{s_i\}_{i \in \mathrm{N}}$ of elements of T with $I_{s_i} \neq I_{s_j}$ for $i \neq j$. By Proposition 4.45, we may assume that s_i divides s_j whenever $i < j$, and it follows that $\{I_{s_i}\}_{i \in \mathrm{N}}$ is a strictly ascending sequence of ideals of R. Looking at a sequence of generators of these ideals, we see that we are contradicting Lemma 4.2. Let now t_1, $\dots$, $t_r \in T$ be such that

$$\{I_{t_1}, \dots, I_{t_r}\} = \{\, I_t \mid t \in T \,\} \setminus \{\, \{0\} \,\}.$$

For $1 \leq k \leq r$, we set

$$C_k = \{\, t \in T \mid I_t = I_{t_k} \,\},$$

and we let B_k be a finite basis of C_k w.r.t. divisibility, and $a_k \in R$ a generator of the ideal I_{t_k}. It is easy to see from the choice of the I_{t_k} that for all $1 \leq k \leq r$ and $t \in B_k$, there exists $f_{a_k t} \in I$ with $\mathrm{HM}(f) = a_k t$. We claim that

$$\{\, f_{a_k t} \mid 1 \leq k \leq r,\ t \in B_k \,\}$$

is a D-Gröbner basis of I. We verify the condition of the last exercise. Let $b \in R$ and $s \in T$ such that bs is the head monomial of some non-zero element of I. Then $b \in I_s$, and $I_s = I_{t_k}$ for some index $1 \leq k \leq r$. It follows that $a_k \mid b$, and $t \mid s$ for some $t \in B_k$. We see that the head monomial $a_k t$ of $f_{a_k t}$ divides bs. □

In the case of Gröbner basis theory over a field, a sufficient condition for G to be a Gröbner basis was that every non-zero polynomial in $\mathrm{Id}(G)$ have a standard representation w.r.t. G. The following lemma provides a similar criterion for D-Gröbner bases.

Lemma 10.8 Assume that G is a finite subset of $R[\underline{X}]$ satisfying the following two conditions.

(i) For all g_1, $g_2 \in G$ there exists $h \in G$ with

$$\mathrm{HT}(h) \mid \mathrm{lcm}\big(\mathrm{HT}(g_1), \mathrm{HT}(g_2)\big) \quad \text{and} \quad \mathrm{HC}(h) \mid \gcd\big(\mathrm{HC}(g_1), \mathrm{HC}(g_2)\big).$$

(ii) Every $f \in \mathrm{Id}(G)$ has a standard representation w.r.t. G.

Then G is a D-Gröbner basis.

Proof By Exercise 10.5 above, it suffices to show that every non-zero element of $\mathrm{Id}(G)$ is D-reducible. Let $0 \neq f \in \mathrm{Id}(G)$, and let

$$f = \sum_{i=1}^{k} m_i g_i$$

be a standard representation of f w.r.t. G. Let $N \subseteq \{1, \dots, k\}$ be the set of all indices with the property that $\mathrm{HT}(f) = \mathrm{HT}(m_i g_i)$. Then $\mathrm{HM}(f) = \sum_{i \in N} \mathrm{HM}(m_i g_i)$, and thus

$$\mathrm{lcm}\{\, \mathrm{HT}(g_i) \mid i \in N \,\} \mid \mathrm{HT}(f), \quad \text{and} \quad \gcd\{\, \mathrm{HC}(g_i) \mid i \in N \,\} \mid \mathrm{HC}(f).$$

It is an easy though slightly tedious exercise to conclude from assumption (i) that there exists $h \in G$ such that $\mathrm{HT}(h)$ divides the above lcm, and $\mathrm{HC}(h)$ divides the gcd. We see that $\mathrm{HM}(h) \mid \mathrm{HM}(f)$, and thus f is D-reducible modulo G. □

In order to obtain a D-Gröbner basis construction, we must ask ourselves how a finite subset P of $R[\underline{X}]$ can fail to be a D-Gröbner basis. First of all, there is the S-polynomial problem: the example given at the beginning of Section 5.3 can actually be viewed as an example in $\mathbb{Z}[X, Y, Z]$. Now let $R[\underline{X}] = \mathbb{Z}[X, Y]$, and $P = \{p_1, p_2\}$ with $p_1 = 5X$ and $p_2 = 3Y$. Then $XY = 2Yp_1 - 3Xp_2 \in \mathrm{Id}(P)$ is in D-normal form modulo P. Here, we have lifted the head terms to their lcm while combining the gcd of the head coefficients.

Definition 10.9 Let $0 \neq g_i \in R[\underline{X}]$ with $\mathrm{HC}(g_i) = a_i$ and $\mathrm{HT}(g_i) = t_i$. Let $a = b_i a_i = \mathrm{lcm}(a_1, a_2)$ with $b_i \in R$, and $t = s_i t_i = \mathrm{lcm}(t_1, t_2)$ with $s_i \in T$ for $i = 1, 2$. Then the **S-polynomial** of g_1 and g_2 is defined as

$$\mathrm{spol}(g_1, g_2) = b_1 s_1 g_1 - b_2 s_2 g_2.$$

Now let $c_1, c_2 \in R$ such that $\gcd(a_1, a_2) = c_1 a_1 + c_2 a_2$. Then we define the **G-polynomial** of g_1 and g_2 w.r.t. c_1 and c_2 as

$$\mathrm{gpol}_{(c_1, c_2)}(g_1, g_2) = c_1 s_1 g_1 + c_2 s_2 g_2.$$

Strictly speaking, S-polynomials are only defined up to unit factors. As usual, there will be no harm in speaking of *the* S-polynomial. If R happens to be a field, then any two non-zero elements are associated, and any $c \neq 0$ is an lcm of $0 \neq a, b \in R$. The new definition of S-polynomials thus coincides with the old one. The G-polynomial of $g_1, g_2 \in R[\underline{X}]$ depends heavily on the choice of c_1 and c_2. If, for example, $\mathrm{HM}(g_1) = \mathrm{HM}(g_2)$, then both g_1 and g_2 are G-polynomials of g_1 and g_2. We will from now on assume that for each pair $0 \neq a_1, a_2 \in R$, a fixed choice of a pair $c_1, c_2 \in R$ has been made such that $c_1 a_1 + c_2 a_2 = \gcd(a_1, a_2)$, and that G-polynomials are formed using this choice. The subscript (c_1, c_2) may then be suppressed. (It is of course advantageous to choose $c_1 = 0$ or $c_2 = 0$ whenever possible, i.e., whenever one of $\mathrm{HC}(g_1)$ and $\mathrm{HC}(g_2)$ divides the other.)

Exercise 10.10 Let K be a field, G a finite subset of $K[X_1, \ldots, X_n]$, and $g_1, g_2 \in G$ with $g_1, g_2 \neq 0$. Assume that $\mathrm{spol}(g_1, g_2) \xrightarrow[G]{*} 0$. Show that $\mathrm{gpol}(g_1, g_2) \xrightarrow[G]{*} 0$.

Note that condition (i) of Lemma 10.8 is equivalent to the G-polynomial of g_1 and g_2 being top-D-reducible modulo G. Our aim is to show that G is a D-Gröbner basis if all S-polynomials D-reduce to 0 and all G-polynomials are top-D-reducible modulo G. The above exercise shows that we will recover Gröbner basis theory over fields as a special case.

Theorem 10.11 *Let G be a finite subset of $R[\underline{X}]$. Assume that for all g_1, $g_2 \in G$,* $\mathrm{spol}(g_1, g_2)$ *equals zero or has a standard representation w.r.t. G, and* $\mathrm{gpol}(g_1, g_2)$ *is top-D-reducible modulo G. Then every $0 \neq f \in \mathrm{Id}(G)$ has a standard representation.*

Proof Assume for a contradiction that $0 \neq f \in \mathrm{Id}(G)$ does not have a standard representation. Let

$$f = \sum_{i=1}^{k} m_i g_i \tag{1}$$

with monomials $0 \neq m_i = a_i t_i$ and $g_i \in G$ for $1 \leq i \leq k$. We may assume that $s = \max\{\, \mathrm{HT}(m_i g_i) \mid 1 \leq i \leq k \,\}$ is minimal among all such representations of f. Then $\mathrm{HT}(f) < s$. For a contradiction, we will produce a representation

$$f = \sum_{i=1}^{k'} m_i' g_i'$$

of the same kind such that $s' = \max\{\, \mathrm{HT}(m_i' g_i') \mid 1 \leq i \leq k' \,\} < s$. We proceed by induction on the number n_s of indices i with $s = \mathrm{HT}(m_i g_i)$. The case $n_s = 1$ is impossible since s cancels out. Let $n_s = 2$, and assume w.l.o.g. that $\mathrm{HT}(m_1 g_1) = \mathrm{HT}(m_2 g_2) = s$. This means that

$$s = t_1 \cdot \mathrm{HT}(g_1) = t_2 \cdot \mathrm{HT}(g_2),$$

and so $\mathrm{lcm}(\mathrm{HT}(g_1), \mathrm{HT}(g_2)) \mid s$, say

$$s = u \cdot \mathrm{lcm}\big(\mathrm{HT}(g_1), \mathrm{HT}(g_2)\big)$$

with $u \in T$. Since $n_s = 2$, we must even have $\mathrm{HM}(m_1 g_1) = -\mathrm{HM}(m_2 g_2)$, and so

$$a_1 \cdot \mathrm{HC}(g_1) = -a_2 \cdot \mathrm{HC}(g_2).$$

It follows that there exists $a \in R$ with

$$a \cdot \mathrm{lcm}\big(\mathrm{HC}(g_1), \mathrm{HC}(g_2)\big) = a_1 \cdot \mathrm{HC}(g_1) = -a_2 \cdot \mathrm{HC}(g_2),$$

and it is now easy to see that

$$m_1 g_1 + m_2 g_2 = au \cdot \mathrm{spol}(g_1, g_2).$$

By assumption, $\mathrm{spol}(g_1, g_2) = 0$, or else it has a standard representation

$$\mathrm{spol}(g_1, g_2) = \sum_{i=1}^{k''} m_i'' g_i''$$

w.r.t. G. Substituting for $m_1 g_1 + m_2 g_2$ in (1), we obtain a representation

$$f = \sum_{i=3}^{k} m_i g_i + au \sum_{i=1}^{k''} m_i'' g_i'', \tag{2}$$

where the second sum is missing if the S-polynomial was zero. The maximum of the head terms occurring in the first sum is less than s by our assumption $n_s = 2$; the maximum s'' of the head terms in the second sum (if any) satisfies

$$s'' < u \cdot \operatorname{lcm}\bigl(\mathrm{HT}(g_1), \mathrm{HT}(g_2)\bigr) = s.$$

Together, we see that the maximum s' of the head terms in the representation (2) satisfies $s' < s$, which means that (2) is the s'-representation that we were looking for.

Now let $n_s > 2$. Again we may assume w.l.o.g. that

$$\mathrm{HT}(m_1 g_1) = \mathrm{HT}(m_2 g_2) = s.$$

Moreover, we trivially have

$$\mathrm{HC}(m_1 g_1) = a_1 \cdot \mathrm{HC}(g_1) \quad \text{and} \quad \mathrm{HC}(m_2 g_2) = a_2 \cdot \mathrm{HC}(g_2), \tag{3}$$

where as before, a_1 and a_2 are the coefficients of m_1 and m_2, respectively. Top-D-reducibility of $\operatorname{gpol}(g_1, g_2)$ modulo G means that there exists $h \in G$ with

$$\mathrm{HT}(h) \mid \operatorname{lcm}\bigl(\mathrm{HT}(g_1), \mathrm{HT}(g_2)\bigr) \quad \text{and} \quad \mathrm{HC}(h) \mid \gcd\bigl(\mathrm{HC}(g_1), \mathrm{HC}(g_2)\bigr).$$

Since s is a common multiple of $\mathrm{HT}(g_1)$ and $\mathrm{HT}(g_2)$, we may conclude that $\mathrm{HT}(h) \mid s$, and (3) shows that

$$\mathrm{HC}(h) \mid \mathrm{HC}(m_1 g_1) \quad \text{and} \quad \mathrm{HC}(h) \mid \mathrm{HC}(m_2 g_2).$$

We can thus find a term $v \in T$ and b_1, $b_2 \in R$ such that

$$\mathrm{HM}(m_1 g_1) = b_1 v \cdot \mathrm{HM}(h) \quad \text{and} \quad \mathrm{HM}(m_2 g_2) = b_2 v \cdot \mathrm{HM}(h). \tag{4}$$

We can now modify our representation (1) of f as follows:

$$f = m_1 g_1 - b_1 v h + m_2 g_2 - b_2 v h + (b_1 + b_2) v h + \sum_{i=3}^{k} m_i g_i .$$

The equations (4) tell us that the head monomials of the first two summands cancel, and so do the ones of the third and fourth. We may thus apply the induction hypothesis to the first two summands and also to the next group of two. In the remaining $k-1$ summands, the highest term s occurs at most $n_s - 1$ times: there are exactly $n_s - 2$ occurrences in $\sum_{i=3}^{k} m_i g_i$, and the summand $(b_1 + b_2) v h$ contributes exactly one occurrence unless it happens to vanish. We see that the induction hypothesis applies here too. If we now add up these three representations to obtain, say,

$$f = \sum_{i=1}^{k'} m_i' g_i' ,$$

then it is easy to see that we get $s' = \max\{\, \mathrm{HT}(m_i' g_i') \mid 1 \le i \le k' \,\} < s$ as desired. □

Corollary 10.12 *Let G be a finite subset of $R[\underline{X}]$, and assume that for all g_1, $g_2 \in G$,*

$$\operatorname{spol}(g_1, g_2) \xrightarrow[G]{*} 0$$

and $\operatorname{gpol}(g_1, g_2)$ *is top-D-reducible modulo G. Then G is a D-Gröbner basis.*

Proof By Lemma 10.3, all non-zero S-polynomials have standard representations. By the above theorem, it follows that every $0 \neq f \in \operatorname{Id}(G)$ has a standard representation w.r.t. G. As we have mentioned before, top-D-reducibility of $\operatorname{gpol}(g_1, g_2)$ modulo G means that condition (i) of Lemma 10.8 is satisfied. Hence the lemma applies, and thus G is a D-Gröbner basis. □

The above corollary provides a criterion for G to be a D-Gröbner basis which can be effectively tested. More importantly, we will now use it to construct, from a finite subset P of $R[\underline{X}]$, a D-Gröbner basis G with $\operatorname{Id}(P) = \operatorname{Id}(G)$.

Definition 10.13 A ring R is called a **computable PID** if it is a computable ring, a PID, and the following two conditions hold:

(i) There is an algorithm that, upon input of $0 \neq a$, $b \in R$, computes c, $d \in R$ such that $ca + db$ is a gcd of a and b.

(ii) There is an algorithm that, for a, $b \in R$, decides whether $b \mid a$ and if so, computes $c \in R$ with $a = bc$.

Note that in a computable PID, we can compute least common multiples according to Proposition 1.84.

The following algorithm D-GRÖBNER for the computation of D-Gröbner bases is a fairly obvious imitation of the Buchberger algorithm. It enlarges the input set by non-zero normal forms of S-polynomials and G-polynomials until all S-polynomials reduce to zero and all G-polynomials are top-D-reducible. It does, however, give preferential treatment to G-polynomials: after dealing with one S-polynomial, it runs through an inner **while**-loop which treats the G-polynomials of all critical pairs that are currently on the list. This will allow us to say, even before termination has been proved, that for any given point in time during computation, there is a point in the future where the G-polynomials of all critical pairs that were then on the list will have been looked at. The same effect could also be achieved by treating critical pairs in chronological order on a first-come first-go basis, but this would preclude the search for optimizing selection strategies.

Theorem 10.14 *Let R be a computable PID and assume that the term order is decidable. Then the algorithm* D-GRÖBNER *of Table* 10.1 *computes, for every finite subset F of $R[\underline{X}]$, a D-Gröbner basis G in $R[\underline{X}]$ such that* $\operatorname{Id}(G) = \operatorname{Id}(F)$.

TABLE 10.1. Algorithm D-GRÖBNER

```
Specification: $G \leftarrow$ D-GRÖBNER($F$)
                Construction of a D-Gröbner basis $G$ for Id($F$)
Given: $F$ = a finite subset of $R[\underline{X}]$
Find: $G$ = a finite subset of $R[\underline{X}]$ such that $G$ is a D-Gröbner basis
      in $R[\underline{X}]$ with $F \subseteq G$ and $\mathrm{Id}(G) = \mathrm{Id}(F)$
begin
$G \leftarrow F$
$B \leftarrow \{\{f_1, f_2\} \mid f_1, f_2 \in G,\ f_1 \neq f_2\}$
$D \leftarrow \emptyset$
$C \leftarrow B$
while $B \neq \emptyset$ do
    while $C \neq \emptyset$ do
        select $\{f_1, f_2\}$ from $C$
        $C \leftarrow C \setminus \{\{f_1, f_2\}\}$
        if there does not exist $g \in G$ with $\mathrm{HT}(g) \mid \mathrm{lcm}(\mathrm{HT}(f_1), \mathrm{HT}(f_2))$,
          $\mathrm{HC}(g) \mid \mathrm{HC}(f_1)$, and $\mathrm{HC}(g) \mid \mathrm{HC}(f_2)$ then
          $h \leftarrow \mathrm{gpol}(f_1, f_2)$
          $h_0 \leftarrow$ some D-normal form of $h$ modulo $G$
          $D \leftarrow D \cup \{\{g, h_0\} \mid g \in G\}$
          $G \leftarrow G \cup \{h_0\}$
        end
    end
    select $\{f_1, f_2\}$ from $B$
    $B \leftarrow B \setminus \{\{f_1, f_2\}\}$
    $h \leftarrow \mathrm{spol}(f_1, f_2)$
    $h_0 \leftarrow$ some D-normal form of $h$ modulo $G$
    if $h_0 \neq 0$ then
      $D \leftarrow D \cup \{\{g, h_0\} \mid g \in G\}$
      $G \leftarrow G \cup \{h_0\}$
    end
$B \leftarrow B \cup D$;  $C \leftarrow D$;  $D \leftarrow \emptyset$
end
end D-GRÖBNER
```

Proof *Correctness:* The following is an invariant of the outer **while**-loop: $\mathrm{spol}(f_1, f_2) \xrightarrow[G]{*} 0$ for all f_1, $f_2 \in G$ with $\{f_1, f_2\} \notin B$. For the inner **while**-loop, an invariant is given by: for all f_1, $f_2 \in G$ with $\{f_1, f_2\} \notin C \cup D$, $\mathrm{gpol}(f_1, f_2)$ is top-D-reducible modulo G. Upon termination, we have $B = C = D = \emptyset$. Correctness thus follows from the corollary to Theorem 10.11.

Termination: We first note that at the end of each run through the outer **while**-loop, the new pairs that have just been added to B are all in the set C, and all G-polynomials of pairs of elements of C are being treated during the next run. Now assume that the algorithm does not terminate. Let $\{h_n\}_{n\in\mathbb{N}}$ be the non-zero D-reduced G- and S-polynomials in the order that they are being added to G. For $n \in \mathbb{N}$, let $a_n = \mathrm{HC}(h_n)$, $s_n = \mathrm{HT}(h_n)$, $m_n = a_n s_n$, and

$$G_n = F \cup \{ h_i \mid i < n \}.$$

By the above remark, there is a function $\varphi : \mathbb{N} \longrightarrow \mathbb{N}$ such that for all $i, n \in \mathbb{N}$ with $i < n$, $\mathrm{gpol}(h_i, h_n)$ is top-D-reducible modulo $G_{\varphi(n)}$. Furthermore, h_n is in D-normal form modulo G_n for all $n \in \mathbb{N}$.

By Dickson's lemma and Proposition 4.45, there exists a strictly ascending sequence $\{n_i\}_{i\in\mathbb{N}}$ of natural numbers such that

$$s_{n_i} \mid s_{n_j} \quad \text{for all} \quad i < j \in \mathbb{N}. \tag{$*$}$$

It follows that $a_{n_i} \nmid a_{n_j}$ for all $i < j \in \mathbb{N}$. This will now lead to a contradiction due to the fact that we periodically treat all new G-polynomials. We will recursively define a sequence $\{k_i\}_{i\in\mathbb{N}}$ with the following properties.

(i) For all $i \in \mathbb{N}$, there exists $j \in \mathbb{N}$ with $s_{k_i} \mid s_{n_j}$.

(ii) a_{k_j} properly divides a_{k_i} for all $i < j \in \mathbb{N}$.

The second property is a contradiction by Lemma 4.2. Set $k_1 = n_1$. Now assume that $k_1, \ldots, k_i$ have been defined. Let $j \in \mathbb{N}$ such that $s_{k_i} \mid s_{n_j}$. By ($*$) above, we may assume that $k_i < n_j$ and thus $a_{k_i} \nmid a_{n_j}$. Now $\mathrm{gpol}(h_{k_i}, h_{n_j})$ is top-D-reducible modulo $G_{\varphi(n_j)}$. This means that there exists $n < \varphi(n_j)$ such that

$$m_n \mid \mathrm{HM}(\mathrm{gpol}(h_{k_i}, h_{n_j})) = \gcd(a_{k_i}, a_{n_j}) \cdot s_{n_j}. \tag{$**$}$$

Set $k_{i+1} = n$. Then both (i) and (ii) follow immediately from ($**$): $s_n \mid s_{n_j}$, and $a_n \mid \gcd(a_{k_i}, a_{n_j})$ which is a proper divisor of a_{k_i} since $a_{k_i} \nmid a_{n_j}$. □

It is clear from the definition of a D-Gröbner basis that we can now decide the equivalence problem for $\mathrm{Id}(F)$ whenever a finite subset F of $R[\underline{X}]$ is given (cf. Theorem 5.55). Even for a D-Gröbner basis G, however, $\xrightarrow[G]{}$ will not in general be adequate for $\equiv_{\mathrm{Id}(F)}$ in the sense of Definition 4.78: take for $R[\underline{X}]$ any polynomial ring over $\mathbb{Z}$, and let $G = \{2\}$. Then G is clearly a D-Gröbner basis. But if the constant coefficient of some polynomial in $R[\underline{X}]$ is odd, then D-reduction modulo G cannot change it, and so we cannot have $1 \overset{*}{\underset{G}{\longleftrightarrow}} 3$ although $1 \equiv_{(2)} 3$. Neither does $\xrightarrow[G]{}$ have unique normal forms in general: let $R[\underline{X}] = \mathbb{Z}[X]$ and $G = \{2X+1\}$. Then $f = 2X^2 + 2X$ has the two normal forms $h_1 = X$ and $h_2 = -X - 1$.

Exercise 10.15 Compute a D-Gröbner basis of $F = \{3X + 1, 5XY + X\}$ in $\mathbb{Z}[X, Y]$. Show that $f = 7XY + 2Y + 3X + 1$ is not in $\mathrm{Id}(F)$ and does not have a unique normal form modulo the D-Gröbner basis that you have computed.

We are now going to show how the theory can be improved for Euclidean domains in such a way that in addition, we obtain adequacy and unique normal forms. As before, $\leq$ is a fixed term order on T. To obtain unique normal forms, remainders in R will have to be unique in the following sense.

Definition 10.16 Let R be a Euclidean domain. We call R a **Euclidean domain with unique remainders** if for each pair a, $b \in R$ with $b \neq 0$, a unique remainder of a upon division by b (remainder in the sense of the definition of Euclidean domains) has been chosen such that the following conditions are satisfied:

(i) For fixed $0 \neq b \in R$, the set of all unique remainders occurring when dividing by b is a unique set of representatives for the partition

$$\{\, a + \mathrm{Id}(b) \mid a \in R \,\}$$

of R.

(ii) There is a well-ordered set W and a function $\psi : R \setminus \{0\} \longrightarrow W$ such that for all divisions with remainder $a = qb + r$ where a, b, q, $r \in R$ are all non-zero, we have $\psi(r) < \psi(a)$.

It is easy to see that $K[X]$ for any field K is a Euclidean domain with unique remainders if we take $W = \mathbb{N}$ and $\psi(f) = \deg(f)$ for all $f \in K[X]$ (Proposition 2.28). Now let $R = \mathbb{Z}$, and set $W = \mathbb{N} \cup \{\infty\}$ with $m < \infty$ for all $m \in \mathbb{N}$. Let $\psi : \mathbb{Z} \longrightarrow W$ be defined by

$$\psi(m) = \begin{cases} m & \text{if} \quad m \geq 0 \\ \infty & \text{otherwise.} \end{cases}$$

Then it is easy to see that $\mathbb{Z}$ with W and ψ becomes a Euclidean domain with unique remainders if we specify remainders upon division by $0 \neq m$ to be in the interval $[0, m)$.

It should be noted that the abstract degree function that comes with every Euclidean domain cannot in general serve as the ψ of (ii) above: from the definition of a Euclidean domain, one easily derives that 1 and -1 must have the same abstract degree, while ψ must sometimes distinguish between 1 and -1 as we just saw.

Exercise 10.17 Define a relation $\leq'$ on $\mathbb{Z}$ by setting $m \leq' n$ iff $m = n$, or $|m| < |n|$, or $|m| = |n|$ and $m < 0$ (i.e., we take the natural order on $\mathbb{N}$ and place negative integers right below their absolute value). Show that $\mathbb{Z}$ is a Euclidean domain with unique remainders if we specify remainders upon division by $0 \neq m$ to be in the interval $[-|m/2|, |m/2|)$ and take $W = \mathbb{Z}$ with $\leq'$ and $\psi = \mathrm{id}_{\mathbb{Z}}$.

For the rest of this section, R will be a Euclidean domain with unique remainders, W with well-order $\leq$ and ψ as in the above definition, and $R[\underline{X}]$ a polynomial ring over R. We now define a new type of reduction over Euclidean domains with unique remainders.

Definition 10.18 Let f, g, $p \in R[\underline{X}]$. We say that f **E-reduces** to g **modulo** p and write $f \xrightarrow[p]{} g$ if there exists a monomial $m = at \in T(f)$ such that $\mathrm{HT}(p) \mid t$, say $t = s \cdot \mathrm{HT}(p)$, and

$$g = f - qsp$$

where $0 \neq q \in R$ is the quotient of a upon division with unique remainder by $\mathrm{HC}(p)$.

As before, $\xrightarrow[p]{*}$ denotes the reflexive-transitive closure of $\xrightarrow[p]{}$. E-reduction modulo a finite subset of $R[\underline{X}]$, E-reducibilty, and E-normal forms are defined in the obvious way as before. It is clear that we still have $f \xrightarrow[P]{} g$ implies $(f - g) \in \mathrm{Id}(P)$. The proof of the following lemma is immediate from Lemma 2.29.

Lemma 10.19 E-reduction extends D-reduction, i.e., every D-reduction step is an E-reduction step. □

Our first goal is to show that E-reduction modulo a finite set is noetherian. This will be achieved by well-ordering $R[\underline{X}]$ in analogy to the case of polynomial rings over fields. We consider a lexicographical order on the set $W \times T$: we let $(v, s) \leq (w, t)$ iff either $s < t$, or $s = t$ and $v \leq w$. (Note that this is different from the product order of Theorem 4.46.) The following lemma simply shows that a lexicographical product of well-orders is again a well-order.

Lemma 10.20 $(W \times T)$ as defined above is a well-ordered set.

Proof Assume that $\{(w_i, t_i)\}_{i \in \mathbb{N}}$ is a strictly descending chain in $(W \times T)$. Then $t_i \geq t_j$ for all $i < j$. Since the term order $\leq$ is a well-order, there must be $n_0 \in \mathbb{N}$ with $t_i = t_j$ for all $n_0 \leq i < j$. We conclude that $w_i > w_j$ for all $n_0 \leq i < j$, a contradiction. □

We now define a map χ from the set of monomials to $(W \times T)$ by setting $\chi(at) = (\psi(a), t)$. Then we extend χ to a map

$$\begin{aligned} \chi' : \quad R[\underline{X}] &\longrightarrow \mathcal{P}_{\mathrm{fin}}\big((W \times T)\big) \\ f &\longmapsto \chi\big(M(f)\big). \end{aligned}$$

Finally, we let $\leq'$ be the induced order of Theorem 4.69 on $\mathcal{P}_{\mathrm{fin}}\big((W \times T)\big)$, and we define a linear quasi-order $\leq$ on $R[\underline{X}]$ by setting

$$f \leq g \quad \text{iff} \quad \chi'(f) \leq' \chi'(g).$$

Then Lemma 4.35 with $\psi = \chi'$ tells us that the quasi-order $\leq$ on $R[\underline{X}]$ is well-founded.

Now let f, g, $p \in R[\underline{X}]$ such that $f \xrightarrow[p]{} g$ is an E-reduction step. Then this either eliminates a term from f as in D-reduction, in which case $g < f$. Else, it replaces a coefficient by its remainder upon divison by $\mathrm{HC}(p)$ with non-zero quotient while leaving all higher coefficients unchanged. Then the ψ-value of that coefficient decreases, and we see that again $g < f$. Considering that the order on $R[\underline{X}]$ is well-founded, we have thus proved the following theorem.

Theorem 10.21 *E-reduction is noetherian.* □

To obtain the desired bases that allow the computation of unique normal forms, we do not need another Gröbner basis algorithm. It will suffice to take a D-Gröbner basis and E-reduce modulo G. This is the content of the next theorem, whose proof hinges on the following lemma.

Lemma 10.22 Let h_1, h_2, $g \in R[\underline{X}]$ such that $h_1 - h_2$ is D-reducible modulo g. Then h_1 or h_2 is E-reducible modulo g.

Proof Assume for a contradiction that both h_1 and h_2 are in E-normal form (and thus in D-normal form) modulo g. Let $m = at$ be a monomial in $M(h_1 - h_2)$ such that $\mathrm{HM}(g) \mid m$. Then $m = (a_1 - a_2)t$ with $a_1 \neq a_2$, and for $i = 1, 2$, either $a_i = 0$ or $a_i t \in M(h_i)$. If one of a_1 and a_2 equals 0, then m is in $M(h_1)$ or in $M(h_2)$, contradicting the fact that h_1 and h_2 are in D-normal form. If a_1, $a_2 \neq 0$, then they must both equal their own unique remainder upon division by $\mathrm{HC}(g)$. Since $a_1 \neq a_2$, and the remainders occurring when dividing by $\mathrm{HC}(g)$ form a unique set of representatives for the residue classes

$$b + \mathrm{Id}(\mathrm{HC}(g)) \qquad (b \in R),$$

it follows that $\mathrm{HC}(g) \nmid (a_1 - a_2) = a$, a contradiction. □

Theorem 10.23 *Let R be a Euclidean domain with unique remainders, and suppose $G \subseteq R[\underline{X}]$ is a D-Gröbner basis. Then the following hold:*

(i) $f \xrightarrow[G]{*} 0$ for all $f \in \mathrm{Id}(G)$, where $\xrightarrow[G]{}$ denotes E-reduction modulo G.

(ii) E-reduction modulo G is adequate for $\equiv_{\mathrm{Id}(G)}$.

(iii) E-reduction modulo G has unique normal forms.

Proof (i) This is immediate from the definition of D-Gröbner bases and the fact that E-reduction extends D-reduction.

(ii) We must show that $f \overset{*}{\underset{G}{\longleftrightarrow}} g$ iff $f - g \in \mathrm{Id}(G)$. The direction "$\Longrightarrow$" follows easily by induction on the length of the $\underset{G}{\longleftrightarrow}$-chain as in the proof of Lemma 5.26. For "$\Longleftarrow$," let $f, g \in R[\underline{X}]$ with $f - g \in \mathrm{Id}(G)$. Let h_1 and h_2 be E-normal forms of f and g modulo G. Then we write

$$(h_2 - h_1) + (f - g) = (f - h_1) - (g - h_2).$$

The right-hand side is in $\mathrm{Id}(G)$ by "$\Longrightarrow$," $f - g$ is in $\mathrm{Id}(G)$ by assumption, and thus we have $h_2 - h_1 \in \mathrm{Id}(G)$. Since G is a D-Gröbner basis, it follows that $h_1 - h_2$ is D-reducible modulo G, and Lemma 10.22 together with the fact that h_1 and h_2 are in E-normal form implies $h_1 = h_2$. This shows that $f \overset{*}{\underset{G}{\longleftrightarrow}} g$. If we repeat the argument with $f = g$, then we obtain a proof of statement (iii). □

For actual computations with E-reduction, we need of course computability of the unique remainder, in addition to computability of R as a ring. Examples of computable Euclidean domains with unique remainders are obviously $K[X]$ for computable field K, and $\mathbb{Z}$ with remainders specified as above. It is clear that in a computable Euclidean domain R with unique remainders we can compute gcd's (Euclidean algorithm) and thus lcm's, decide divisibility (zero remainder), and divide effectively. In particular, E-reduction of polynomials over R is decidable, and we can compute D-Gröbner bases from given finite sets of polynomials over R.

The Gröbner basis theory for PID's and Euclidean domains can be further developed and applied in a similar manner as Gröbner bases over fields. The following exercise provides an example.

Exercise 10.24 Let R be a PID and $G \subseteq R[\underline{X}]$ a D-Gröbner basis w.r.t. any term order. Show that $G \cap R$ generates the ideal $\mathrm{Id}(G) \cap R$ of R.

10.2 Homogeneous Gröbner Bases

Even with the improvements of Section 5.5, the time and space consumption of the algorithm GRÖBNER is often unsatisfactory. A natural attempt to improve the situation further is to look for degree bounds by means of which one could compute a partial Gröbner basis for certain limited purposes. Unfortunately, there is no obvious way of achieving this. S-polynomials of high degree that occur during a run of the algorithm GRÖBNER may, after reduction, contribute a polynomial of much lower degree or even a constant. It is by no means clear how this phenomenon could be controlled. In this section, we show that *homogeneous* polynomials behave nicely in this respect. The next section will explain how the results for the homogeneous case can be put to use via homogenization, although most of the beauty of the homogeneous theory is lost in the process. We can obtain more powerful results at no extra cost by considering an arbitrary *grading* instead of the regular degree. Throughout this section, K will be a field,

$$K[\underline{X}] = K[X_1, \ldots, X_n],$$

and T the set of terms in the variables $X_1, \ldots, X_n$.

Definition 10.25 A **grading** Γ of $K[\underline{X}]$ is a monoid homomorphism

$$\Gamma : (T, 1, \cdot) \longrightarrow (\mathbb{N}, (0), +),$$

i.e., a map $\Gamma : T \longrightarrow \mathbb{N}$ such that $\Gamma(1) = 0$ and $\Gamma(s \cdot t) = \Gamma(s) + \Gamma(t)$ for $s, t \in T$. For $0 \neq f \in K[\underline{X}]$, we define the **$\Gamma$-degree** of f as

$$\max\{\, \Gamma(t) \mid t \in T(f) \,\}.$$

By an abuse of notation, we denote the Γ-degree of f by $\Gamma(f)$ too. A nonzero polynomial $f \in K[\underline{X}]$ is called **Γ-homogeneous** if $\Gamma(s) = \Gamma(t)$ for all s, $t \in T(f)$. A term order $\leq$ on T is **Γ-compatible** if $\Gamma(s) < \Gamma(t)$ implies $s < t$ for all $s, t \in T$.

Examples 10.26 Let $a_1, \ldots, a_n \in \mathbb{N}$ and define $\Gamma : T \longrightarrow \mathbb{N}$ by

$$\Gamma(X_1^{\nu_1} \cdot \cdots \cdot X_n^{\nu_n}) = a_1\nu_1 + \cdots + a_n\nu_n.$$

Then Γ is obviously a grading of $K[\underline{X}]$. Taking in particular $a_1 = \cdots = a_n = 1$, this yields the **grading by total degree** where $\Gamma(f)$ is simply $\deg(f)$. This example should be used to visualize the statements and proofs in this section. If we fix an index j with $1 \leq j \leq n$ and set $a_j = 1$ and $a_i = 0$ for $i \neq j$, then we obtain the **grading by degree in the variable** X_j. More generally, for any subset J of $\{1, \ldots, n\}$, we may set $a_i = 1$ for $i \in J$ and $a_i = 0$ for $i \notin J$. The resulting grading Γ yields the **total degree in the variables** $\{X_i \mid i \in J\}$. Taking $a_1 = \cdots = a_n = 0$, we get the **trivial grading** Γ_0 of $K[\underline{X}]$ where all degrees are zero and every non-zero polynomial is Γ-homogeneous. In fact *any* grading Γ of $K[\underline{X}]$ arises from a *linear form* in this manner: set $a_i = \Gamma(X_i) \in \mathbb{N}$ for $1 \leq i \leq n$. Then

$$\Gamma(X_1^{\nu_1} \cdot \cdots \cdot X_n^{\nu_n}) = a_1\nu_1 + \cdots + a_n\nu_n.$$

This is expressed by saying that Γ is determined by the **weights** a_i of the indeterminates X_i.

For the rest of this section, whenever a grading Γ occurs in connection with a computation or an algorithm, we will assume that the weights $\Gamma(X_i)$ are given for $1 \leq i \leq n$, so that we may actually compute $\Gamma(t)$ for arbitrary $t \in T$.

Lemma 10.27 Let Γ be a grading of $K[\underline{X}]$ and $\leq$ a term order on T.

(i) Define the relation $\leq'$ on T by setting $s \leq' t$ iff the following holds:

$$\begin{array}{ccll} \Gamma(s) & < & \Gamma(t), & \text{or} \\ \Gamma(s) & = & \Gamma(t) & \text{and} \quad s \leq t. \end{array}$$

Then $\leq'$ is a Γ-compatible term order on T.

(ii) $s \mid t$ implies $\Gamma(s) \leq \Gamma(t)$ for all s, $t \in T$.

(iii) $\Gamma(f + g) \leq \max(\Gamma(f), \Gamma(g))$ for f, $g \in K[\underline{X}]$ with f, $g \neq 0$ and $f + g \neq 0$.

(iv) $\Gamma(fg) = \Gamma(f) + \Gamma(g)$ for $0 \neq f, g \in K[\underline{X}]$.

(v) Let $0 \neq c \in K$ and let f, $g \in K[\underline{X}]$ be Γ-homogeneous with $\Gamma(f) = \Gamma(g)$ and f, g, $f+g \neq 0$. Then cf and $f+g$ are Γ-homogeneous with $\Gamma(cf) = \Gamma(f+g) = \Gamma(f)$.

(vi) Let $0 \neq f, g \in K[\underline{X}]$ be Γ-homogeneous. Then fg is Γ-homogeneous.

Proof Statements (i), (ii), (iii), (v), and (vi) are easy to verify. In order to prove (iv), we use (i) to find a Γ-compatible term order $\leq'$ on T. Then we apply Lemma 5.17 to the head terms of fg, f, and g w.r.t. $\leq'$ and obtain:

$$\begin{aligned}\Gamma(fg) &= \Gamma\big(\mathrm{HT}(fg)\big) = \Gamma\big(\mathrm{HT}(f)\cdot\mathrm{HT}(g)\big)\\ &= \Gamma\big(\mathrm{HT}(f)\big) + \Gamma\big(\mathrm{HT}(g)\big) = \Gamma(f) + \Gamma(g). \ \square\end{aligned}$$

Exercise 10.28 What is the relationship between the grading by total degree on $K[\underline{X}]$, the total degree-lexicographical order on T, and the lexicographical order on T?

For the rest of this section, we let Γ be a fixed grading and $\leq$ a fixed term order on T. "Homogeneous" will from now on mean "Γ-homogeneous." Note that our fixed term order $\leq$ need not be Γ-compatible. The following lemma is crucial to the theory of homogeneous Gröbner bases.

Lemma 10.29 Let $d \in \mathbb{N}$ and $0 \neq f, p, g \in K[\underline{X}]$ with $\Gamma(f) = d$. Suppose p is homogeneous and $f \xrightarrow[p]{} g$. Then $\Gamma(p)$, $\Gamma(g) \leq d$. If, in addition, f is homogeneous too, then g is homogeneous with $\Gamma(g) = d$.

Proof From $f \xrightarrow[p]{} g$ it follows that $g = f - mp$ for some monomial $m = at \in K[\underline{X}]$ such that $t \cdot \mathrm{HT}(p) \in T(f)$. It now follows from Lemma 10.27 (ii), (iv), and (vi) and the homogeneity of p that $\Gamma(p) \leq \Gamma(mp) \leq d$. From Lemma 10.27 (iii), we then see that $\Gamma(g) \leq d$. If f is homogeneous too, then

$$\Gamma(mp) = \Gamma\big(t \cdot \mathrm{HT}(p)\big) = d,$$

and Lemma 10.27 (v) implies homogeneity of g with $\Gamma(g) = d$. □

Exercise 10.30 Show that the previous proposition continues to hold with $\xrightarrow[p]{}$ replaced by $\xrightarrow[P]{*}$ where P is a set of homogeneous polynomials.

The following lemma is immediate from the definition of S-polynomials together with Lemma 10.27 (v).

Lemma 10.31 Let g_1, $g_2 \in K[\underline{X}]$ be homogeneous with $\mathrm{spol}(g_1, g_2) \neq 0$. Then $\mathrm{spol}(g_1, g_2)$ is homogeneous and

$$\Gamma\big(\mathrm{spol}(g_1, g_2)\big) = \Gamma\Big(\mathrm{lcm}\big(\mathrm{HT}(g_1), \mathrm{HT}(g_2)\big)\Big). \quad \square$$

Exercise 10.32 Go back to Theorem 5.53 and make sure that you thoroughly understand the algorithm GRÖBNER and the proof of its correctness and termination.

In our treatment of Gröbner bases thus far, we have considered the algorithm GRÖBNER only for computable fields and decidable term orders. The existence of Gröbner bases in the general case had been proved earlier by different means. The homogeneous case can be treated in the same manner. However, the theory becomes much more elegant if we note that an application of GRÖBNER to a finite set $F \subseteq K[\underline{X}]$ can also be viewed as an abstract mathematical construction. It amounts to defining a sequence $\{G_i\}_{i\in\mathbb{N}}$ such that $G_0 = F$, and for $i > 0$, $G_i = G_{i-1} \cup \{h\}$ where h is a non-zero normal form of an S-polynomial of two elements in G_{i-1} if such a normal form exists, $G_i = G_{i-1}$ otherwise. The termination proof of the algorithm shows that there exists $m \in \mathbb{N}$ with $G_i = G_m$ for all $i \geq m$, and the correctness proof shows that G_m is a Gröbner basis of $\mathrm{Id}(F)$. In the sequel, we will talk about variants of the algorithm GRÖBNER that treat only S-polynomials satisfying a certain degree bound. In the case of computable field and decidable term order, these are to be viewed as actual algorithms; else, they amount to mathematical existence proofs. It is not hard to translate arguments concerning the algorithm into abstract mathematics. If, for example, we say, "let g be the first polynomial with property P showing up during computation," then this translates into "let k be the least natural number such that G_k contains an element with property P, and let g be the element of $G_k \setminus G_{k-1}$."

If $d_1 \in \mathbb{N}$, and $d_2 \in \mathbb{N}$ or d_2 is the symbol ∞, then we define

$$K[\underline{X}]_{[d_1,d_2]} = \{\, f \in K[\underline{X}] \mid d_1 \leq \Gamma(f) \leq d_2 \,\},$$

with the understanding that $k < \infty$ for all $k \in \mathbb{N}$. With d_1 and d_2 as before, we let $[d_1, d_2]$-GRÖBNER be the algorithm GRÖBNER with the sole modification that it considers only those critical pairs $\{g_1, g_2\}$ that satisfy

$$d_1 \leq \Gamma\Big(\mathrm{lcm}\big(\mathrm{HT}(g_1), \mathrm{HT}(g_2)\big)\Big) \leq d_2.$$

Then $[d_1, d_2]$-GRÖBNER applied to any finite set of polynomials must terminate since an infinite loop would be an infinite loop of the algorithm GRÖBNER. It is also obvious that the output set of the algorithm is a superset of the input set that generates the same ideal as the latter.

Proposition 10.33 *Let $d_1 \in \mathbb{N}$, and $d_2 \in \mathbb{N}$ or $d_2 = \infty$. Let F be a finite subset of $K[\underline{X}]$ such that each $f \in F$ is homogeneous, and set*

$$G = [d_1, d_2]\text{-GRÖBNER}(F).$$

Then the following hold:

(i) Every $g \in G$ is homogeneous, and $G \setminus F \subseteq K[\underline{X}]_{[d_1,d_2]}$.

(ii) $\mathrm{spol}(g_1, g_2) \xrightarrow[G]{*} 0$ *for all g_1, $g_2 \in G$ that satisfy*

$$d_1 \leq \Gamma\Big(\mathrm{lcm}\big(\mathrm{HT}(g_1), \mathrm{HT}(g_2)\big)\Big) \leq d_2.$$

Proof (i) The elements of F are homogeneous by assumption. Now assume for a contradiction that there is $g \in G \setminus F$ that is not homogeneous or does not satisfy $d_1 \leq \Gamma(g) \leq d_2$. We may assume that g is the first such polynomial that shows up during computation. Then g is a normal form modulo a set of homogeneous polynomials of an S-polynomial of a pair $\{g_1, g_2\}$ of homogeneous polynomials with

$$d_1 \leq \Gamma\Big(\mathrm{lcm}\big(\mathrm{HT}(g_1), \mathrm{HT}(g_2)\big)\Big) \leq d_2.$$

By Lemmas 10.31 and 10.29, g is homogeneous and in $K[\underline{X}]_{[d_1,d_2]}$. Part (ii) is immediate from the fact that the algorithm terminates precisely when all S-polynomials of the indicated type reduce to zero. □

We do not claim that $[d_1, d_2]$-GRÖBNER(F) does anything meaningful in the way of reducing arbitrary members of $\mathrm{Id}(F)$ to zero; the only interesting case will be $d_1 = 0$. It will soon be obvious why we chose the more general definition. Also, note that those $f \in F$ with $\Gamma(f) > d_2$ play no role in the algorithm. They are carried along for no other reason than to preserve the generated ideal. Those $f \in F$ with $\Gamma(f) < d_1$ can of course occur in critical pairs and may also be used during reduction of S-polynomials.

Corollary 10.34 *Let F be a finite subset of $K[\underline{X}]$ with f homogeneous for each $f \in F$. Then* $[0, \infty)$-GRÖBNER(F) *is a Gröbner basis of* $\mathrm{Id}(F)$ *that consists entirely of homogeneous polynomials.* □

Lemma 10.35 Let d_1, $d_2 \in \mathbb{N}$, and $d_3 \in \mathbb{N}$ or $d_3 = \infty$, such that $d_1 \leq d_2 \leq d_3$. Let F be a finite subset of $K[\underline{X}]$ consisting of homogeneous polynomials, and set

$$G = [d_2 + 1, d_3]\text{-GRÖBNER}\big([d_1, d_2]\text{-GRÖBNER}(F)\big).$$

Then G consists again of homogeneous polynomials, and

$$[d_1, d_3]\text{-GRÖBNER}(G) = G.$$

Proof It is clear from the previous proposition that every $g \in G$ is homogeneous. We must show that

$$\mathrm{spol}(g_1, g_2) \xrightarrow[G]{*} 0$$

for all g_1, $g_2 \in G$ that satisfy $d_1 \leq d \leq d_3$, where

$$d = \Gamma\Big(\mathrm{lcm}\big(\mathrm{HT}(g_1), \mathrm{HT}(g_2)\big)\Big).$$

If $d_2 < d \leq d_3$, then this is immediate from Proposition 10.33 (ii) since G is an output of $[d_2 + 1, d_3]$-GRÖBNER. If $d_1 \leq d \leq d_2$, then we must have $\Gamma(g_1)$, $\Gamma(g_2) \leq d_2$. It follows that

$$g_1, g_2 \in [d_1, d_2]\text{-GRÖBNER}(F)$$

since the application of $[d_2 + 1, d_3]$-GRÖBNER did not bring in anything of degree less than $d_2 + 1$ by Proposition 10.33 (i). The claim now follows again from Proposition 10.33 (ii). □

Exercise 10.36 Show that the lemma above continues to hold if we modify the definition of G to

$$G = [d_1, d_2]\text{-GRÖBNER}\big([d_2 + 1, d_3]\text{-GRÖBNER}(F)\big).$$

Lemma 10.37 Let $d \in \mathbb{N}^+$, and let F be a finite subset of $K[\underline{X}]_{[d,\infty]}$ with every $f \in F$ homogeneous. Then

$$[d, d']\text{-GRÖBNER}(F) = [0, d']\text{-GRÖBNER}(F),$$

and

$$[d, d']\text{-GRÖBNER}(F) \cap K[\underline{X}]_{[0,d-1]} = \emptyset$$

whenever $d \leq d' \in \mathbb{N}$ or $d' = \infty$.

Proof It is clear that $[0, d-1]$-GRÖBNER$(F) = F$. The claim now follows from Lemma 10.35 applied to F and 0, $d-1$, d', and from Proposition 10.33 (i). □

Proposition 10.38 *Let $d \in \mathbb{N}$, F a finite subset of $K[\underline{X}]$ consisting of homogeneous polynomials, and set*

$$G_d = [0, d]\text{-GRÖBNER}(F).$$

Then there exists a Gröbner basis G of Id(F) *that consists entirely of homogeneous polynomials and satisfies*

$$G_d \supseteq G \cap K[\underline{X}]_{[0,d]}.$$

Proof It follows immediately from Lemma 10.35, Corollary 10.34, and Proposition 10.33 (i) that

$$G = [d, \infty]\text{-GRÖBNER}(G_d)$$

has the desired properties. □

We are now going to prove that the output of the algorithm $[0, d]$-GRÖBNER with homogeneous input has all the nice properties of a Gröbner basis for polynomials (not necessarily homogeneous) of degree less than or equal to d. For any finite set $P \subseteq K[\underline{X}]$, we denote by $\xrightarrow[P]{d}$ the restriction of $\xrightarrow[P]{}$ to $K[\underline{X}]_{[0,d]}$.

Theorem 10.39 *Let G be a finite subset of $K[\underline{X}]$ consisting of homogeneous polynomials, and let $d \in \mathbb{N}$. Then the following are all equivalent:*

(i) $\xrightarrow[G]{d}$ is locally confluent.

(ii) $\xrightarrow[G]{d}$ *is confluent.*

(iii) $\xrightarrow[G]{d}$ *has the Church–Rosser property.*

(iv) $\xrightarrow[G]{d}$ *has unique normal forms.*

(v) $\mathrm{spol}(g_1, g_2) \xrightarrow[G]{*} 0$ *for all* g_1, $g_2 \in G$ *with*

$$\Gamma\Big(\mathrm{lcm}\big(\mathrm{HT}(g_1), \mathrm{HT}(g_2)\big)\Big) < d.$$

(vi) Every $0 \neq f \in \mathrm{Id}(G) \cap K[\underline{X}]_{[0,d]}$ *is reducible modulo* G.

(vii) $f \xrightarrow[G]{*} 0$ *for all* $f \in \mathrm{Id}(G) \cap K[\underline{X}]_{[0,d]}$.

(viii) Every $0 \neq f \in \mathrm{Id}(G) \cap K[\underline{X}]_{[0,d]}$ *is top-reducible modulo* G.

(ix) For every $s \in \mathrm{HT}(\mathrm{Id}(G) \cap K[\underline{X}]_{[0,d]})$ *there exists* $t \in \mathrm{HT}(G)$ *with* $t \mid s$.

(x) $\mathrm{HT}(\mathrm{Id}(G) \cap K[\underline{X}]_{[0,d]}) \subseteq \mathrm{Mult}(\mathrm{HT}(G))$.

(xi) The polynomials $h \in K[\underline{X}]_{[0,d]}$ *that are in normal form w.r.t.* $\xrightarrow[G]{}$ *form a system of unique representatives for the partition*

$$\{\, (f + \mathrm{Id}(G)) \cap K[\underline{X}]_{[0,d]} \mid f \in K[\underline{X}]_{[0,d]} \,\}$$

of $K[\underline{X}]_{[0,d]}$.

Proof It is clear that the restriction $\xrightarrow[G]{d}$ of $\xrightarrow[G]{}$ to $K[\underline{X}]_{[0,d]}$ is still a noetherian reduction relation. The equivalence of (i), (ii), (iii), and (iv) is thus Newman's lemma.

The equivalence of (vi), (vii), (viii), (ix), (x), and (xi) is easy to prove (cf. the proof of Theorem 5.35) if we bear in mind that reduction modulo a homogeneous polynomial does not increase the Γ-degree.

(iv)$\Longrightarrow$(v): Assume for a contradiction that h is a non-zero normal form w.r.t $\xrightarrow[G]{}$ of $\mathrm{spol}(g_1, g_2)$ for a pair $g_1, g_2 \in G$ that satisfies the indicated condition on the Γ-degree. Then $\mathrm{spol}(g_1, g_2) \in K[\underline{X}]_{[0,d]}$. By Exercise 10.30, h and each intermediate polynomial in the reduction chain $\mathrm{spol}(g_1, g_2) \xrightarrow[G]{*} h$ are in $K[\underline{X}]_{[0,d]}$ too, so that in fact

$$\mathrm{spol}(g_1, g_2) \xrightarrow[G]{d*} h.$$

Let

$$\mathrm{spol}(g_1, g_2) = a_2 s_2 g_1 - a_1 s_1 g_2$$

with constants a_1, a_2 and terms s_1, s_2. Then $a_2 s_2 g_1$, $a_1 s_1 g_2 \in K[\underline{X}]_{[0,d]}$. We see that

$$\begin{tikzcd} & a_2 s_2 g_1 \arrow[dl, "d"', "g_1"] \arrow[dr, "g_2"', "d"] & & \\ 0 & & \mathrm{spol}(g_1, g_2) \arrow[r, "d*", "G"'] & h, \end{tikzcd}$$

and thus h and 0 are two different normal forms of $a_2 s_2 g_1$ w.r.t $\xrightarrow[G]{d}$.

(v)$\Longrightarrow$(vi): Let $0 \neq f \in \mathrm{Id}(G) \cap K[\underline{X}]_{[0,d]}$. By Proposition 10.38, there exists a Gröbner basis G' of $\mathrm{Id}(G)$ consisting of homogeneous polynomials with

$$G \supseteq G' \cap K[\underline{X}]_{[0,d]}.$$

Then f is reducible modulo G', and by Lemma 10.29, this reduction is modulo g for some $g \in G' \cap K[\underline{X}]_{[0,d]}$. We see that f is indeed reducible modulo G.

(vi)$\Longrightarrow$(iv): Let g_1 and g_2 be two different normal forms of f w.r.t $\xrightarrow[G]{d}$. Then g_1 and g_2 must be in $K[\underline{X}]_{[0,d]}$ by the definition of $\xrightarrow[G]{d}$. The difference $g_1 - g_2$ is in $\mathrm{Id}(G)$ by Lemma 5.26 and in $K[\underline{X}]_{[0,d]}$ by Lemma 10.27 (iii). So $g_1 - g_2$ is reducible modulo G. It follows that at least one of g_1 and g_2 is reducible modulo G, and this must in fact be a $\xrightarrow[G]{d}$-reduction step by Lemma 10.29, a contradiction. □

Definition 10.40 Let G be a finite subset of $K[\underline{X}]$ consisting of homogeneous polynomials, and let $d \in \mathbb{N}$. Then we call G a ***d*-Gröbner basis** (w.r.t. Γ and $\leq$) if it satisfies the equivalent conditions of Theorem 10.39.

We can now summarize the results of this section thus far as follows. Given a finite set F of polynomials each of which is homogeneous w.r.t. a given grading Γ, we can compute a d-Gröbner basis ($d \in \mathbb{N}$) of $\mathrm{Id}(F)$ by running on F a truncated algorithm GRÖBNER which ignores all S-polynomials of Γ-degree greater than d. The result is good enough to test for membership in $\mathrm{Id}(F)$ any polynomial f with $\Gamma(f) \leq d$. Moreover, if a d-Gröbner basis has already been computed, then this can be extended to a d'-Gröbner basis ($d < d'$) by applying an algorithm GRÖBNER which treats only those S-polynomials whose Γ-degree is in the interval $[d+1, d']$. All polynomials that are being added in the process will have a Γ-degree in the same interval. In particular, if the polynomials in the original set F satisfy a lower Γ-degree bound l, then nothing happens below Γ-degree l at all during any computation of a d-Gröbner basis.

A natural question that arises at this point is if we can use truncated versions of the algorithms GRÖBNERNEW1 and GRÖBNERNEW2 instead of GRÖBNER, i.e., if we can still eliminate superfluous critical pairs according to Buchberger's criteria. Buchberger's first criterion deletes critical pairs $\{g_1, g_2\}$ with disjoint head terms because they reduce to zero by means of g_1 and g_2 themselves. It is thus completely unaffected by any truncation of the algorithm.

The second criterion skips $\{g_1, g_2\}$ because of the presence of two other critical pairs, namely, $\{g_1, h\}$ and $\{h, g_2\}$ for some h with

$$\mathrm{HT}(h) \mid \mathrm{lcm}(\mathrm{HT}(g_1), \mathrm{HT}(g_2)).$$

We know that then

$$\operatorname{lcm}(\operatorname{HT}(g_1), \operatorname{HT}(h)) \mid \operatorname{lcm}(\operatorname{HT}(g_1), \operatorname{HT}(g_2)) \quad \text{and}$$
$$\operatorname{lcm}(\operatorname{HT}(h), \operatorname{HT}(g_2)) \mid \operatorname{lcm}(\operatorname{HT}(g_1), \operatorname{HT}(g_2)).$$

So if $\Gamma(\operatorname{lcm}(\operatorname{HT}(g_1), \operatorname{HT}(g_2)))$ was below the degree bound d, then so are the Γ-degrees of the other two lcm's, and thus the other two critical pairs will not be the victims of truncation. The only thing that must be observed is that when a d-Gröbner basis is extended to a d'-Gröbner basis for $d' > d$ by means of GRÖBNERNEW1, then the critical pairs that have already been treated in the first computation must be properly marked at the beginning of the second computation.

There is an obvious immediate application of d-Gröbner basis computations in practice. Suppose we know that each element in a given finite set of polynomials is homogeneous w.r.t. a grading Γ whose weights are known to us. Now if we wish to test a particular polynomial f for membership in $\operatorname{Id}(F)$, then it suffices to compute a $\Gamma(f)$-Gröbner basis of $\operatorname{Id}(F)$. If, however, we do not have any a priori information on homogeneity w.r.t. any grading, then it will not usually be obvious to the eye whether or not there exists a grading that makes every $f \in F$ homogeneous. We will now explain how this question can be decided, and how the weights of a suitable grading can be computed in case of a positive answer.

Let $a_1, \ldots, a_n$ be unknowns, and for any term $t = X_1^{\nu_1} \cdot \cdots \cdot X_n^{\nu_n}$, set

$$\Gamma_a(t) = \sum_{i=1}^{n} a_i \nu_i.$$

For each $f \in F$, pick a term $t_f \in T(f)$. It is clear that the n-tuples of weights whose corresponding gradings will make each $f \in F$ homogeneous are precisely the non-negative integer solutions of the system

$$\Gamma_a(t_f) = \Gamma_a(t) \qquad (f \in F,\ t_f \neq t \in T(f))$$

of linear equations. We have thus reduced the problem to deciding the solvability in the non-negative integers of a system of linear equations and computing a solution if it exists. This can be achieved by an algorithm known as *integer linear programming* (ILP). Although ILP is itself among the unpleasantly complex algorithms, experience has shown that checking for homogeneity takes only a neglible fraction of the time required for the full Gröbner basis computation.

We close this section with a discussion of the ideal theoretic aspects of homogeneity. As before, Γ is a fixed grading. If $f \in K[\underline{X}]$ and $d \in \mathbb{N}$, then we denote by $f_{(d)}$ the sum of all monomials of f whose Γ-degree equals d. It is clear that either $f_{(d)} = 0$ or $f_{(d)}$ is homogeneous with $\Gamma(f_{(d)}) = d$. In the latter case, $f_{(d)}$ is called the **d-homogeneous part** of f. It is now easy to see that every polynomial has a unique representation as the sum of its homogeneous parts by descending degree. An ideal I of $K[\underline{X}]$ is called **homogeneous** if $f_{(d)} \in I$ for all $f \in I$ and $d \in \mathbb{N}$.

Proposition 10.41 *(i) Suppose $F \subseteq K[\underline{X}]$ and all $f \in F$ are homogeneous. Then* $\mathrm{Id}(F)$ *is homogeneous.*

(ii) Every homogeneous ideal I in $K[\underline{X}]$ has a finite basis consisting of homogeneous polynomials.

Proof (i) Every $g \in \mathrm{Id}(F)$ is a sum of homogeneous polynomials of the form mf where m is a monomial and $f \in F$, say $g = \sum_{i=1}^{r} m_i f_i$. Then for $d \in \mathbb{N}$,

$$g_{(d)} = \sum\{ m_i f_i \mid 1 \leq i \leq r,\ \Gamma(m_i f_i) = d\},$$

and so $g_{(d)} \in \mathrm{Id}(F)$.

(ii) By the Hilbert basis theorem, there exists a finite set P of polynomials in $K[\underline{X}]$ such that $\mathrm{Id}(P) = I$; let $F = \{ p_{(d)} \mid p \in P,\ d \in \mathbb{N}\}$. Then F is finite, every $f \in F$ is homogeneous, and $F \subseteq I$. Moreover, every $p \in P$ is a sum of elements of F; so $I = \mathrm{Id}(P) = \mathrm{Id}(F)$. □

By a d-Gröbner basis of an ideal I we mean of course a d-Gröbner basis G with $\mathrm{Id}(G) = I$. If I is a homogeneous ideal and $d \in \mathbb{N}$, then I has a finite basis F of homogeneous polynomials, and $[0, d]$-GRÖBNER(F) is then a d-Gröbner basis of I. We have thus proved the following corollary.

Corollary 10.42 *Every homogeneous ideal has a d-Gröbner basis.* □

The concept of homogeneous parts can also be used to make a connection between d-Gröbner bases and standard representations.

Lemma 10.43 Let F be a finite subset of $K[\underline{X}]$ consisting of homogeneous polynomials, and let $f \in \mathrm{Id}(F)$, say

$$f = \sum_{i=1}^{k} m_i f_i$$

with monomials $0 \neq m_i$ and $f_i \in F$ for $1 \leq i \leq k$. Then it is possible to delete summands on the right-hand side such that equality is preserved and $\Gamma(m_i f_i) \leq \Gamma(f)$ for $1 \leq i \leq k$. If f is itself homogeneous, then one can even achieve $\Gamma(m_i f_i) = \Gamma(f)$ for $1 \leq i \leq k$.

Proof It is clear that each summand $m_i f_i$ is again homogeneous. It follows that for all $d \in \mathbb{N}$, $f_{(d)}$ equals the sum of all those summands $m_i f_i$ with $\Gamma(m_i f_i) = d$. We see that our goal is achieved if we drop each summand with $f_{(\Gamma(m_i f_i))} = 0$. □

Cutting off a degree overhang according to the lemma above does clearly not destroy the property of a representation $f = \sum_{i=1}^{k} m_i f_i$ to be a standard representation. With this and Lemmas 5.60 and 5.61 in mind, the reader will find it easy to prove the following proposition.

Proposition 10.44 *The following two conditions may be added to the equivalent conditions of Theorem* 10.39*:*

(xii) Every $f \in K[\underline{X}]_{[0,d]}$ has a standard representation w.r.t. G.

(xiii) Every $f \in K[\underline{X}]_{[0,d]}$ has a standard representation w.r.t. G in which the Γ-degrees of the summands do not exceed the Γ-degree of f. □

In Exercise 5.40, we saw that the head terms of a Gröbner basis of an ideal are a Gröbner basis of the ideal generated by the head terms of elements of the ideal. The following exercise provides a similar result w.r.t. gradings.

Exercise 10.45 If $0 \neq f \in K[\underline{X}]$, then the $\Gamma(f)$-homogeneous part of f is called the **Γ-highest form** of f, or **highest form** of f for short, and we will denote it by $\mathrm{HF}(f)$. The set of highest forms of elements of a subset F of $K[\underline{X}]$ will be written as $\mathrm{HF}(F)$. Show the following:

(i) If $m \in K[\underline{X}]$ is a monomial and $0 \neq f \in K[\underline{X}]$, then $m \cdot \mathrm{HF}(f) = \mathrm{HF}(mf)$.

(ii) If $f_1, \ldots, f_m \in K[\underline{X}]$ are all non-zero, all have the same Γ-degree, and satisfy
$$\mathrm{HF}(f_1) + \cdots + \mathrm{HF}(f_m) \neq 0,$$
then
$$\mathrm{HF}(f_1) + \cdots + \mathrm{HF}(f_m) = \mathrm{HF}(f_1 + \cdots + f_m).$$

(iii) If I is an ideal of $K[\underline{X}]$ and f is an element of the ideal $\mathrm{Id}(\mathrm{HF}(I)$, then every Γ-homogeneous part of f is an element of $\mathrm{HF}(I)$.

(iv) If $G \subseteq K[\underline{X}]$ is a Gröbner basis of the ideal I of $K[\underline{X}]$ w.r.t. some Γ-compatible term order $\leq$, then $\mathrm{HF}(G)$ is a Gröbner basis w.r.t. $\leq$ of the ideal $\mathrm{Id}(\mathrm{HF}(I))$ of $K[\underline{X}]$. In fact, if $f \in \mathrm{Id}(\mathrm{HF}(I))$, then every homogeneous part of f contains a term that is reducible modulo $\mathrm{HF}(G)$.

If one wishes to develop the theory of homogeneous Gröbner bases in perfect analogy to our general treatment of Gröbner bases in Sections 5.2 and 5.3, then one needs the following technical result.

Exercise 10.46 Let P be a finite subset of $K[\underline{X}]$ consisting of Γ-homogeneous polynomials. Show the following:

(i) Let $f, g, h \in K[\underline{X}]$, and assume
$$f \xleftrightarrow[P]{} g \xleftrightarrow[P]{} h,$$
where $g = f + asp$, $h = g + btq$ with $a, b \in K$, $s, t \in T$, $p, q \in P$, and $\Gamma(sp) \neq \Gamma(tq)$. Set $g_1 = f + btq$. Then
$$f \xleftrightarrow[P]{} g_1 \xleftrightarrow[P]{} h.$$

(ii) Let $d \in \mathbb{N}$, $f, g \in K[\underline{X}]_{[0,d]}$, and suppose $f \xleftrightarrow[P]{*} g$. Then there exists a chain $f \xleftrightarrow[P]{*} g$ in which every intermediate polynomial is in $K[\underline{X}]_{[0,d]}$.

10.3 Homogenization

The results on homogeneous Gröbner bases of the last section can be used to obtain a deeper understanding of the behavior of degrees in any non-homogeneous Gröbner basis computation. This is achieved by means of *homogenization* with an additional variable. The technical details have a tendency of looking messy, but the theory is really quite simple. Given a polynomial f in the variables $X_1, \ldots, X_n$, we look at its degree d, i.e., the maximum of the degrees of its terms. (Think of the ordinary total degree for the moment.) We then multiply each term of lesser degree by a suitable power of a new variable Z, thus bringing all degrees up to d. Setting $Z = 1$ will take us back to f. The whole point of this section is to find out what happens if, instead of computing a Gröbner basis of a given set of polynomials, we first homogenize them as described above, then compute the Gröbner basis, and finally set the homogenizing variable Z equal to 1.

Let $K[\underline{X}] = K[X_1, \ldots, X_n]$ and Γ a grading on $T(\underline{X}) = T(X_1, \ldots, X_n)$. It is recommended for the reader to use the special example of the grading by total degree to visualize the results below. For convenience, we set $\Gamma(0) = 0$. Now let Z be a new variable. Then we set

$$K[\underline{X}, Z] = K[X_1, \ldots, X_n, Z], \quad T(\underline{X}, Z) = T(X_1, \ldots, X_n, Z),$$

and we extend Γ to a grading Γ' of $T(\underline{X}, Z)$ by setting $\Gamma(Z) = 1$. Suppose $\Gamma(X_i) = a_i$ for $1 \leq i \leq n$. Since K is a subfield of $Q_{K[\underline{X}, Z]}$, the field of fractions of $K[\underline{X}, Z]$, the map $\varphi : K[\underline{X}] \longrightarrow Q_{K[\underline{X}, Z]}$ defined by

$$\varphi(f(X_1, \ldots, X_n)) = f\left(\frac{X_1}{Z^{a_1}}, \ldots, \frac{X_n}{Z^{a_n}}\right)$$

is an instance of the substitution homomorphism of Lemma 2.17 (i), satisfying $\varphi \restriction K = \mathrm{id}_K$. For $f \in K[\underline{X}]$ with $\Gamma(f) = d$, we now define $f^* = Z^d \cdot \varphi(f)$.

Lemma 10.47 Let $f \in K[\underline{X}]$ and $d = \Gamma(f)$. Then $f^* \in K[\underline{X}, Z]$, and f^* is Γ'-homogeneous with $\Gamma'(f^*) = d$. Moreover,

$$M(f) = \{ atZ^{d-d'} \mid at \in M(f),\ \Gamma(t) = d' \},$$

and the map $m \longmapsto m \cdot Z^{d - \Gamma(m)}$ is a bijection between $M(f)$ and $M(f^*)$.

Proof If $t \in T(f)$, then $\varphi(t) = t/Z^{\Gamma(t)}$ and $\Gamma(t) \leq d$, and so

$$Z^d \cdot \varphi(t) = tZ^{d-\Gamma(t)} \in T(\underline{X}, Z).$$

Since φ is a homomorphism acting as the identity on K, it follows that $f^* = Z^d \cdot \varphi(f) \in K[\underline{X}]$. Moreover,

$$\begin{aligned} \Gamma'(Z^d \cdot \varphi(t)) &= \Gamma'(tZ^{d-\Gamma(t)}) \\ &= \Gamma(t) + \Gamma'(Z^{d-\Gamma(t)}) \\ &= \Gamma(t) + d - \Gamma(t) \ = \ d \end{aligned}$$

for all $t \in T(f)$. We see that f^* is Γ'-homogeneous of degree d. It is clear that

$$t_1^* = t_1 Z^{d-\Gamma(t_1)} \neq t_2 Z^{d-\Gamma(t_2)} = t_2^*$$

whenever t_1, $t_2 \in T(f)$ with $t_1 \neq t_2$. So there will not be any like terms in the sum

$$f^* = \sum_{at \in M(f)} atZ^{d-\Gamma(t)},$$

and the rest of the lemma is easy to prove from this observation. □

f^* is called the **homogenization** of f in $K[\underline{X}, Z]$ (with respect to the **homogenizing variable** Z).

Exercise 10.48 Show that $(fg)^* = f^*g^*$ for all f, $g \in K[\underline{X}]$.

Now let $g \in K[\underline{X}, Z]$. Then we define $g_* \in K[\underline{X}]$ by setting

$$g_*(X_1, \ldots, X_n) = g(X_1, \ldots, X_n, 1).$$

Since K is a subfield of $K[\underline{X}]$, the map $g \longmapsto g_*$ is a substitution homomorphism from $K[\underline{X}, Z]$ to $K[\underline{X}]$ which acts as the identity on K. It is easy to see that it is in fact surjective. g_* is called the **dehomogenization** of g w.r.t. Z.

Lemma 10.49 If $g \in K[\underline{X}, Z]$, say $g = \sum_{m \in M(g)} m$, then

$$g_* = \sum_{m \in M(g)} m_* .$$

If in addition, g is Γ'-homogeneous, then the map $m \longmapsto m_*$ from $M(g)$ to $M(g_*)$ is bijective.

Proof The first statement is immediate from the fact that the dehomogenization map is a homomorphism. Now assume that g is Γ'-homogeneous, and let t_1, $t_2 \in T(g)$, say

$$t_1 = s_1 Z^{d_1} \neq s_2 Z^{d_2} = t_2$$

with s_1, $s_2 \in T(\underline{X})$. Then $s_1 = s_2$ would imply $d_1 = d_2$ because of the homogeneity of g, and so we must have

$$(t_1)_* = s_1 \neq s_2 = (t_2)_*.$$

We see that there will not be any like terms in the sum $\sum_{m \in M(g)} m_*$, and this clearly implies the second claim. □

Lemma 10.50 (i) $(f^*)_* = f$ for all $f \in K[\underline{X}]$.

(ii) Let $g \in K[\underline{X}, Z]$ be Γ'-homogeneous of degree d, and let $d' = \Gamma(g_*)$. Then $d' \leq d$, and $g = Z^{d-d'}(g_*)^*$.

Proof (i) Let $m = at \in M(f)$, and let $d = \Gamma(f)$ and $d' = \Gamma(t)$. Then $(atZ^{d-d'})_* = at = m$. The claim is now easy to prove using the bijections of monomials of Lemmas 10.47 and 10.49.

(ii) The inequality $d' \leq d$ is immediate from the definition of g_*. Now let $m \in M(g)$, say $m = asZ^i$ with $s \in T(\underline{X})$. Then the monomial in $M(g_*)$ corresponding to m is as. Since $\Gamma(s) = d - i$, the monomial in $M((g_*)^*)$ corresponding to m_* is $asZ^{d'-(d-i)}$. We see that $Z^{d-d'}(g_*)^* = g$. □

Whenever $F \subseteq K[\underline{X}]$ and $G \subseteq K[\underline{X}, Z]$, we set

$$F^* = \{ f^* \mid f \in F \} \quad \text{and} \quad G_* = \{ g_* \mid g \in G \}.$$

It is clear that f^* and g_* can be computed from $f \in K[\underline{X}]$ and $g \in K[\underline{X}, Z]$ as soon as we can compute in $K[\underline{X}]$ at all.

Exercise 10.51 Let $G \subseteq K[\underline{X}, Z]$. Show that $\mathrm{Id}(G_*) = (\mathrm{Id}(G))_*$ when the former ideal is taken in $K[\underline{X}]$. (All you need to use is the fact that dehomogenization is a homomorphism.)

Lemma 10.52 Let $F = \{f_1, \ldots, f_m\} \subseteq K[\underline{X}]$ and $f = \sum_{i=1}^{m} q_i f_i$ with $q_i \in K[\underline{X}]$ for $1 \leq i \leq m$. Set

$$d = \max\{ \Gamma(q_i f_i) \mid 1 \leq i \leq m \},$$

and $d' = \Gamma(f)$. Then $Z^{d-d'} f^* \in \mathrm{Id}(F^*)$.

Proof By Lemma 10.47, we have

$$d = \max\bigl\{ \Gamma\bigl((q_i f_i)^*\bigr) \bigm| 1 \leq i \leq m \bigr\}.$$

So if we set

$$\bar{f} = \sum_{i=1}^{m} (q_i f_i)^* = \sum_{i=1}^{m} q_i^* f_i^*$$

(Exercise 10.48), then $\bar{f} \in \mathrm{Id}(F^*)$, and $\bar{f}$ is Γ'-homogeneous with $d'' = \Gamma'(\bar{f}) \leq d$. Moreover,

$$\begin{aligned} \bar{f}_* &= \Bigl(\sum_{i=1}^{m} (q_i f_i)^*\Bigr)_* = \sum_{i=1}^{m} (q_i^*)_* (f_i^*)_* \\ &= \sum_{i=1}^{m} q_i f_i = f. \end{aligned}$$

Using Lemma 10.50 (ii), we may now conclude that

$$\bar{f} = Z^{d''-d'} (\bar{f}_*)^* = Z^{d''-d'} f^*.$$

Since $d'' \leq d$, we finally obtain

$$Z^{d-d'} f^* = Z^{d-d''} Z^{d''-d'} f^* = Z^{d-d''} \bar{f} \in \mathrm{Id}(F^*). \quad □$$

Lemma 10.53 Let F be a finite subset of $K[\underline{X}]$. Then $(\mathrm{Id}(F^*))_* = \mathrm{Id}(F)$.

Proof Let $f \in \mathrm{Id}(F)$. Then by the previous lemma, $Z^k f^* \in \mathrm{Id}(F^*)$ for some $k \in \mathbb{N}$, and so

$$f = (f^*)_* = (Z^k f^*)_* \in \bigl(\mathrm{Id}(F^*)\bigr)_*.$$

Conversely, if $g \in \mathrm{Id}(F^*)$, say $g = \sum_{i=1}^k q_i f_i^*$ with $f_i \in F$ and $q_i \in K[\underline{X}, Z]$, then

$$\begin{aligned} g_* &= \Bigl(\sum_{i=1}^m q_i f_i^*\Bigr)_* = \sum_{i=1}^k (q_i)_* (f_i^*)_* \\ &= \sum_{i=1}^k (q_i)_* f_i \in \mathrm{Id}(F). \quad \square \end{aligned}$$

Lemma 10.54 Let F be a finite subset of $K[\underline{X}]$, and let $G \subseteq K[\underline{X}, Z]$ be a basis of $\mathrm{Id}(F^*)$. Then $\mathrm{Id}(G_*) = \mathrm{Id}(F)$.

Proof We have $\mathrm{Id}(G_*) = (\mathrm{Id}(G))_*$ by Exercise 10.51. Moreover, by assumption, $\mathrm{Id}(G) = \mathrm{Id}(F^*)$, and finally $(\mathrm{Id}(F^*))_* = \mathrm{Id}(F)$ by the previous lemma. Combining these equalities, we see that

$$\mathrm{Id}(G_*) = \bigl(\mathrm{Id}(G)\bigr)_* = \bigl(\mathrm{Id}(F^*)\bigr)_* = \mathrm{Id}(F). \quad \square$$

Recall from the previous section that a d-Gröbner basis is a finite set G of homogeneous polynomials such that $\mathrm{spol}(g_1, g_2) \xrightarrow[G]{*} 0$ whenever g_1, $g_2 \in G$ and $\Gamma(\mathrm{spol}(g_1, g_2)) \leq d$ (all this for a fixed term order, a grading Γ and $d \in \mathbb{N}$). A d-Gröbner basis is good enough to reduce to zero every $f \in \mathrm{Id}(G)$ with $\Gamma(f) \leq d$. Furthermore, if F is any finite set of homogeneous polynomials, then a d-Gröbner basis of $\mathrm{Id}(F)$ can be computed by means of the algorithm $[0, d]$-GRÖBNER.

Theorem 10.55 *Let F be a finite subset of $K[\underline{X}]$, let $d \in \mathbb{N}$, and suppose $G \subseteq K[\underline{X}, Z]$ is a d-Gröbner basis of $\mathrm{Id}(F^*)$ w.r.t. Γ' and some term order on $T(\underline{X}, Z)$. Furthermore, let $p \in K[\underline{X}]$ with $\Gamma(p) = d'$. Then the following are equivalent:*

(i) There exist $q_f \in K[\underline{X}]$ such that $p = \sum_{f \in F} q_f f$ and

$$\max\{\, \Gamma(q_f f) \mid f \in F \,\} \leq d.$$

(ii) $Z^{d-d'} p^* \xrightarrow[G]{*} 0$.

Proof (i)$\Longrightarrow$(ii): If (i) holds, then $Z^{d-d'} p^* \in \mathrm{Id}(F^*)$ by Lemma 10.52. The claim is now obvious from the fact that

$$\Gamma'(Z^{d-d'} p^*) = \Gamma'(Z^{d-d'}) + \Gamma'(p^*) = (d - d') + \Gamma(p) = d.$$

(ii)$\Longrightarrow$(i): From (ii), it follows that $Z^{d-d'}p^* \in \mathrm{Id}(G) = \mathrm{Id}(F^*)$. By Lemma 10.43, we can write

$$Z^{d-d'}p^* = \sum_{i=1}^{k} m_i f_i^*$$

with $f_i \in F$, monomials $m_i \in K[\underline{X}, Z]$, and

$$\Gamma'(m_i f_i^*) \leq \Gamma'(Z^{d-d'}p^*) = d.$$

It follows that

$$\begin{aligned} p &= (Z^{d-d'}p^*)_* = \left(\sum_{i=1}^{k} m_i f_i^*\right)_* \\ &= \sum_{i=1}^{k} (m_i)_* (f_i^*)_* = \sum_{i=1}^{k} (m_i)_* f_i . \end{aligned}$$

Furthermore,

$$\begin{aligned} \Gamma\big((m_i)_* f_i\big) &= \Gamma\big((m_i)_*\big) + \Gamma(f_i) \\ &\leq \Gamma'(m_i) + \Gamma'(f_i^*) \\ &= \Gamma'(m_i f_i^*) \ \leq \ d. \end{aligned}$$

The desired representation is now easily obtained by combining summands in the last sum above. □

Note that both the d-Gröbner basis computation and the reduction of $Z^{d-d'}p^*$ in the theorem take place in $K[\underline{X}, Z]$. We are not yet saying that (ii) above has anything to do with $p \xrightarrow[G_*]{*} 0$; if and how this is the case will be discussed below. Let us first look at potential practical uses of the theorem. Suppose we wish to test $p \in K[\underline{X}]$ for membership in $\mathrm{Id}(F)$. Assume further that we have the a priori knowledge that if $p \in \mathrm{Id}(F)$ at all, then it must have a representation $p = \sum_{f \in F} q_f f$ where the Γ-degree of each summand does not exceed a certain bound $d \in \mathbb{N}$. We can then compute a d-Gröbner basis of F^* and the normal form h of $Z^{d-d'}p^*$ modulo G, where $d' = \Gamma(p)$. If $h = 0$, then the direction (ii)$\Longrightarrow$(i) of the theorem tells us immediately that $p \in \mathrm{Id}(F)$. If $h \neq 0$, then the other direction says that p cannot be written as a sum of multiples of elements of F with Γ-degree bound d, and thus $p \notin \mathrm{Id}(F)$ by our a priori information. It is fairly obvious that there will not be an easy way to obtain such information in general if we do not even know whether $f \in \mathrm{Id}(F)$ at all. There are, however, results for the special case $p = 1$. We will just state one here because not surprisingly, proofs are very difficult in the area of such bounds. (Cf. the Notes on p. 508.)

Theorem 10.56 (Effective Nullstellensatz) *Let K be a field, and let F be a finite subset of $K[X_1,\dots,X_n]$. If $1 \in \mathrm{Id}(F)$, then $1 = \sum_{f\in F} q_f f$ with $q_f \in K[X_1,\dots,X_n]$ such that*

$$d = \max\{\deg(q_f f) \mid f \in F\} \leq \begin{cases} D^n & \textit{if } n > 1 \textit{ and } D \geq 3 \\ 3^n & \textit{if } n > 1 \textit{ and } D = 2 \\ 2D-1 & \textit{if } n = 1, \end{cases}$$

where $D = \max\{\deg(f) \mid f \in F\}$.

In order to decide whether $1 \in \mathrm{Id}(F)$, i.e., whether $\mathrm{Id}(F)$ is proper, we may thus compute a d-Gröbner basis G of F^* w.r.t. the grading by total degree and any term order on $T(\underline{X}, Z)$, with $d \in \mathbb{N}$ as described above. We will then have $1 \in \mathrm{Id}(F)$ iff $Z^d \xrightarrow[G]{*} 0$. A sufficient condition for this to happen is that $Z^{d'} \in G \cap K[Z]$ for some $d' \leq d$; If the term order is such that $Z < t$ for all $1 \neq t \in T(\underline{X})$, then it is easy to see that this is even an equivalent condition.

It is clear that the procedure described above is at least potentially an improvement over a full Gröbner basis computation in $K[\underline{X}]$: we do not have to do anything above degree d. The following discussion provides more insight into the connection between the computations in $K[\underline{X}, Z]$ of Theorem 10.55 and an ordinary computation of a Gröbner basis G' of $\mathrm{Id}(F)$ and subsequent reduction of f modulo G'. Γ will once again be an arbitrary grading on $T(\underline{X})$.

Let $\leq$ be a term order on $T(\underline{X})$. Then we extend $\leq$ to a term order $\leq'$ on $T(\underline{X}, Z)$ by setting

$$s_1 Z^i \leq' s_2 Z^j \quad \text{iff} \quad \begin{array}{l} s_1 < s_2, \text{ or} \\ s_1 = s_2 \text{ and } i \leq j \end{array}$$

for s_1, $s_2 \in T(\underline{X})$. In other words, we let $Z < t$ for all $1 \neq t \in T(\underline{X})$. For t_1, $t_2 \in T(\underline{X})$, we then have $t_1 \leq t_2$ iff $t_1 \leq' t_2$, so there will be no harm in using $\leq$ also for $\leq'$.

The next lemma collects some facts concerning head terms, reduction, S-polynomials, and Gröbner bases w.r.t. homogenization and dehomogenization.

Lemma 10.57 (i) Let t_1, $t_2 \in T(\underline{X}, Z)$ with $\Gamma'(t_1) = \Gamma'(t_2)$. Then $t_1 < t_2$ iff $(t_1)_* < (t_2)_*$.

(ii) Let $g \in K[\underline{X}, Z]$ be Γ'-homogeneous. Then $\mathrm{HT}(g_*) = (\mathrm{HT}(g))_*$.

(iii) Let $f \in K[\underline{X}]$. Then $\mathrm{HT}(f^*) = (\mathrm{HT}(f))^* \cdot Z^{d-d'}$ where $d = \Gamma(f)$ and $d' = \Gamma(\mathrm{HT}(f))$.

(iv) Let f, g, $h \in K[\underline{X}, Z]$ be Γ'-homogeneous such that $f \xrightarrow[g]{} h\ [t]$ in $K[\underline{X}, Z]$. Then $f_* \xrightarrow[g_*]{} h_*\ [t_*]$ in $K[\underline{X}]$.

(v) Let G be a finite subset of $K[\underline{X}, Z]$ consisting of Γ'-homogeneous polynomials, and let f, $h \in K[\underline{X}, Z]$ with $f \xrightarrow[G]{*} h$. Then $f_* \xrightarrow[G_*]{*} h_*$.

(vi) Let g, $h \in K[\underline{X}, Z]$ be Γ'-homogeneous. Then

$$\big(\mathrm{spol}(g,h)\big)_* = \mathrm{spol}(g_*, h_*).$$

(vii) Let F be a finite subset of $K[\underline{X}]$, and let $G \subseteq K[\underline{X}, Z]$ be a Gröbner basis of $\mathrm{Id}(F^*)$ consisting of Γ'-homogeneous polynomials. Then G_* is a Gröbner basis of $\mathrm{Id}(F)$.

Proof (i) Let $t_1 = s_1 Z^i$ and $t_2 = s_2 Z^j$ with $s_1, s_2 \in T(\underline{X})$. Then $(t_1)_* = s_1$ and $(t_2)_* = s_2$, and so $(t_1)_* = (t_2)_*$ implies $i = j$ by the hypothesis on the Γ'-degrees. Using these facts, the equivalence is now easily proved from the definition of the term order on $T(\underline{X}, Z)$.

(ii) This is an immediate consequence of (i).

(iii) Let $\mathrm{HT}(f) = t$. Then $s < t$ for all $s \in T(f)$ with $s \neq t$. It follows that $sZ^i < tZ^j$ for all i, $j \in \mathbb{N}$ and $s \in T(f)$ with $s \neq t$, and thus $tZ^{d-d'} = \mathrm{HT}(f^*)$.

(iv) Let $h = f - asg$ with $t = s \cdot \mathrm{HT}(g) \in T(f)$. Then

$$t_* \in T(f_*), \quad s_* \cdot \mathrm{HT}(g_*) = t_*, \quad \text{and} \quad h_* = f_* - as_*g_*,$$

and so $f_* \xrightarrow[g_*]{} h_*\ [t_*]$.

(v) This is an immediate consequence of (iv).

(vi) Let $\mathrm{HM}(g) = as$ and $\mathrm{HM}(h) = bt$ with $a, b \in K$ and $s, t \in T(\underline{X}, Z)$, and let

$$t' = \mathrm{lcm}(s,t) = us = vt.$$

Then $\mathrm{spol}(g,h) = bug - avh$, $\mathrm{HM}(g_*) = as_*$, $\mathrm{HM}(h_*) = bt_*$, and

$$\mathrm{lcm}(s_*, t_*) = (t')_* = u_*s_* = v_*t_*.$$

It follows that

$$\big(\mathrm{spol}(g,h)\big)_* = bu_*g_* - av_*h_* = \mathrm{spol}(g_*, h_*).$$

(vii) Lemma 10.54 tells us that G_* is a basis of $\mathrm{Id}(F)$, and so it remains to show that G_* is a Gröbner basis. Let g, $h \in G$. Then $\mathrm{spol}(g,h) \xrightarrow[G]{*} 0$ because G is a Gröbner basis, and so we may conclude from (v) and (vi) above that

$$\mathrm{spol}(g_*, h_*) = \big(\mathrm{spol}(g,h)\big)_* \xrightarrow[G_*]{*} 0. \quad \square$$

As an immediate consequence of (v) above, we see that with the present setup of term orders on $T(\underline{X})$ and $T(\underline{X}, Z)$, condition (ii) of Theorem 10.55 implies $p \xrightarrow[G_*]{*} 0$. The converse cannot be true for rather trivial reasons. Let

$K[\underline{X}] = \mathbb{Q}[X_1, X_2]$, Γ the grading by total degree, $F = \{X_1 + 1, X_1\}$, and $d = 1$. Here, $F^* = \{X_1 + Z, X_1\}$,

$$G = [0,1]\text{-GRÖBNER}(F^*) = \{X_1 + Z, X_1, Z\},$$

and $G_* = \{X_1 + 1, X_1, 1\}$. Then $X_2 \xrightarrow[G_*]{*} 0$ and $d' = \deg(X_2) = 1$, so now if (ii) of Theorem 10.55 were true, we would have to have

$$Z^{d-d'} X_2 = X_2 \xrightarrow[G]{*} 0,$$

which is obviously false.

Exercise 10.58 Show that in the situation of Theorem 10.55 and with the present setup of term orders, $p \xrightarrow[G_*]{*} 0$ still implies that $p \in \mathrm{Id}(F)$. Conclude that for the "ideal membership test with a priori degree bound" that is described following Theorem 10.55, we may test the condition $p \xrightarrow[G_*]{*} 0$ instead of (ii) of Theorem 10.55.

Finally, let us compare the computation $[0,d]$-GRÖBNER(F^*) that Theorem 10.55 calls for with the ordinary computation GRÖBNER(F). If d is greater than any degree occurring in the computation anyway, then the homogenized version is potentially worse than the regular one. To see this, suppose two polynomials g and h show up in the homogenized computation with $\mathrm{HT}(g) = sZ^i$ and $\mathrm{HT}(h) = tZ^j$, $s \mid t$, and $i > j$. Then the ordinary algorithm can top-reduce h modulo g, while the homogenized one can not. As d becomes smaller, the homogenized version will cut the top off the ordinary computation and therefore tend to run faster. But there is more to it than that. Suppose a critical pair $\{g, h\}$ is up for treatment such that $\mathrm{HT}(g) = sZ^i$, $\mathrm{HT}(h) = tZ^j$, $i > j$, and

$$\Gamma\big(\mathrm{lcm}(s,t)\big) < d < i + \Gamma\big(\mathrm{lcm}(s,t)\big) = \Gamma'\big(\mathrm{lcm}(sZ^i, tZ^j)\big).$$

Then the ordinary algorithm would treat the pair even if it were somehow aware of the degree bound d. The homogenized version, however, knows that the head terms that $\mathrm{spol}(g,h)$ might produce—although of low degree—will not be needed for the reduction to zero of polynomials p as in (i) of Theorem 10.55. Using the truncated homogenized version of the algorithm thus amounts not only to implementing a degree bound, but also a very intelligent criterion for the detection of superfluous S-polynomials.

The concept of homogenization is also instrumental in a certain strategy for the selection of critical pairs in ordinary, non-homogenized Buchberger algorithms. Here, one performs all reductions according to the ordinary rules. At the same time, one carries along a "phantom degree" which indicates what the total degree of an input polynomial or a normal form of an S-polynomial would have been had the input been homogenized first. One then selects critical pairs in such a way that lower phantom degrees are preferred and ties are broken by the normal strategy. Under the normal strategy, Gröbner basis computations w.r.t. lexicographical orders tend to be much slower than those w.r.t. total degree orders. Under the new strategy, especially when used in connection with GRÖBNERNEW2, this phenomenon tends to disappear.

10.4 Gröbner Bases for Polynomial Modules

Throughout this section, K will be a field and $K[\underline{X}] = K[X_1, \ldots, X_n]$. We will be using some of the definitions and results of Section 3.3. The m-fold Cartesian product $(K[\underline{X}])^m$, which consists of all m-tuples of elements of $K[\underline{X}]$, forms a $K[\underline{X}]$-module in a natural way according to Example 3.28 (iv): here, addition is performed componentwise, and so is scalar multiplication by an element of $K[\underline{X}]$. In this section, we will denote this $K[\underline{X}]$-module by M_{mn}^K, i.e.,

$$M_{mn}^K = \underbrace{K[X_1, \ldots, X_n] \times \cdots \times K[X_1, \ldots, X_n]}_{m \text{ times}}.$$

It is clear that for computable K, the module M_{mn}^K is computable in the sense that we can effectively perform addition and scalar multiplication. Our goal in this section is to provide the obvious analogue of Gröbner basis theory for modules of this type in case K is computable: we are looking for an algorithm, which, given an element $\boldsymbol{f}$ of M_{mn}^K and a finite generating system of a submodule N of M_{mn}^K, decides whether or not $\boldsymbol{f} \in N$. As a matter of fact, we will, just as with Gröbner bases for polynomial ideals, be able to compute a unique normal form of $\boldsymbol{f}$. In particular, this will allow us to decide whether or not $\boldsymbol{f} + N = \boldsymbol{g} + N$ for given $\boldsymbol{f}$, $\boldsymbol{g} \in N$ by testing the condition $\boldsymbol{f} - \boldsymbol{g} \in N$. This is the solution to the *equivalence problem for submodules of* M_{mn}^K. In view of the fact that the operations in the factor module M_{mn}^K/N are performed on representatives, this allows us to compute in such factor modules.

Before we show how this goal can be achieved, let us recall from Section 3.3 that M_{mn}^K is also called the free $K[\underline{X}]$-module of rank m over $K[\underline{X}]$, and that a basis—in the sense of Section 3.3—of M_{mn}^K is given by $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m\}$, where for $1 \leq j \leq m$, we have set $\boldsymbol{e}_j = (e_{j1}, \ldots, e_{jn})$ with

$$e_{ji} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, we know from Proposition 3.32 (iii) that M_{mn}^K is noetherian, i.e., every submodule of M_{mn}^K has a finite generating system. A second proof of this fact will be the subject of an exercise at the very end of this section. In view of Proposition 3.31, our solution to the equivalence problem for submodules of M_{mn}^K will allow us to compute not only in factor modules of M_{mn}^K, but in every finitely generated $K[\underline{X}]$-module provided the kernel of the homomorphism of Proposition 3.31 is known to us.

Our proposed Gröbner basis theory for polynomial modules is based on a combination of classical Gröbner basis theory with the relativization of Section 10.2. Let $Z_1, \dots, Z_m$ be new indeterminates, and let us denote by $K[\underline{X}, \underline{Z}]$ the polynomial ring over K in the old and new varaibles, i.e.,

$$K[\underline{X}, \underline{Z}] = K[X_1, \dots, X_n, Z_1, \dots, Z_m].$$

Let Γ be the unique grading on $K[\underline{X}, \underline{Z}]$ that satisfies

$$\Gamma(X_1) = \cdots = \Gamma(X_n) = 0 \quad \text{and} \quad \Gamma(Z_1) = \cdots = \Gamma(Z_m) = 1.$$

Let now $H_1(K[\underline{X}, \underline{Z}])$ be the set of all Γ-homogeneous polynomials in $K[\underline{X}, \underline{Z}]$ of Γ-degree 1, enlarged by 0, i.e.,

$$H_1(K[\underline{X}, \underline{Z}]) = \left\{ h_1 \cdot Z_1 + \cdots + h_m \cdot Z_m \;\middle|\; h_1, \dots, h_m \in K[\underline{X}] \right\}.$$

The proof of the following lemma is now straightforward from the definitions.

Lemma 10.59 With the notation introduced above, the following hold:

(i) $H_1(K[\underline{X}, \underline{Z}])$ is a $K[\underline{X}]$-module if we take for addition and scalar multiplication the restrictions of the respective operations in the ring $K[\underline{X}, \underline{Z}]$.

(ii) The map

$$\begin{array}{cccc} \varphi: & M_{mn}^K & \longrightarrow & H_1(K[\underline{X}, \underline{Z}]) \\ & (h_1, \dots, h_m) & \longmapsto & h_1 Z_1 + \cdots + h_m Z_m \end{array}$$

is an isomorphism of $K[\underline{X}]$-modules. □

If K is computable, then it is clear that the isomorphism φ is "constructive": we can effectively translate any given $\boldsymbol{h} \in M_{mn}^k$ into $\varphi(\boldsymbol{h}) \in H_1(K[\underline{X}, \underline{Z}])$ and any given $f \in H_1(K[\underline{X}, \underline{Z}])$ into $\varphi^{-1}(f) \in M_{mn}^K$. This together with the fact that φ is an isomorphism means that it suffices to solve the Gröbner basis problem outlined at the beginning of the section for the $K[\underline{X}]$-module $H_1(K[\underline{X}, \underline{Z}])$. This problem will be reduced to the computation of 1-Gröbner bases (as defined in Section 10.2) in $K[\underline{X}, \underline{Z}]$ by means of the next lemma.

Recall that for a subset F of a module M, $\text{lin}(F)$ denotes the linear span of F in M, i.e., the submodule of M that is generated by F.

Lemma 10.60 Let $F \subseteq H_1(K[\underline{X}, \underline{Z}])$. Then

$$\text{lin}(F) = \text{Id}(F) \cap H_1(K[\underline{X}, \underline{Z}]),$$

where the linear span is taken in the $K[\underline{X}]$-module $H_1(K[\underline{X}, \underline{Z}])$ and the ideal is taken in the ring $K[\underline{X}, \underline{Z}]$.

Proof If $g \in \operatorname{lin}(F)$, then we can write

$$g = \sum_{f \in F} q_f f$$

with $q_f \in K[\underline{X}] \subseteq K[\underline{X}, \underline{Z}]$ for all $f \in F$, and we see that $g \in \operatorname{Id}(F)$. Conversely, suppose $g \in \operatorname{Id}(F) \cap H_1(K[\underline{X}, \underline{Z}])$. Then there exist monomials $m_1, \ldots, m_s$ in $K[\underline{X}, \underline{Z}]$ such that

$$g = \sum_{i=1}^{s} m_i f_i$$

with $f_i \in F$, not necessarily pairwise different, for $1 \leq i \leq s$. Since g is homogeneous of Γ-degree 1, we may, by Lemma 10.43, assume that $\Gamma(m_i f_i) = 1$ for $1 \leq i \leq s$. Since $\Gamma(f) = 1$ for all $f \in F$, it follows that $\Gamma(m_i) = 0$ and thus $m_i \in K[\underline{X}]$ for $1 \leq i \leq s$, which shows that $g \in \operatorname{lin}(F)$. □

It is now clear how the submodule membership problem for the $K[\underline{X}]$-module $H_1(K[\underline{X}, \underline{Z}])$ can be solved in principle, provided that K is computable. Given a finite subset F of $H_1(K[\underline{X}, \underline{Z}])$, let G be a Gröbner basis of $\operatorname{Id}(F)$ in $K[\underline{X}, \underline{Z}]$; an element f of $H_1(K[\underline{X}, \underline{Z}])$ is then in the linear span $\operatorname{lin}(F)$ of F in the $K[\underline{X}]$-module $H_1(K[\underline{X}, \underline{Z}])$ if and only if $f \xrightarrow[G]{*} 0$. In view of the results of Section 10.2, however, we can do a lot better than that. Let F be a finite subset of $H_1(K[\underline{X}, \underline{Z}])$. Recall that a 1-Gröbner basis of $\operatorname{Id}(F)$ in $K[\underline{X}, \underline{Z}]$ is obtained by running on F a Buchberger algorithm that considers only those S-polynomials whose Γ-degree is less than or equal to 1. We know that such a 1-Gröbner basis is good enough to test for membership in $\operatorname{Id}(F)$ any polynomial $f \in K[\underline{X}, \underline{Z}]$—not necessarily homogeneous—whose Γ-degree is less than or equal to 1. In particular, this test will work for the elements of $H_1(K[\underline{X}, \underline{Z}])$ (which happen to be homogeneous). We have proved the following theorem.

Theorem 10.61 *Let F be a finite subset of $H_1(K[\underline{X}, \underline{Z}])$, and let G be a 1-Gröbner basis w.r.t. Γ (and any term order) of* $\operatorname{Id}(F)$ *in $K[\underline{X}, \underline{Z}]$. Then an element of $H_1(K[\underline{X}, \underline{Z}])$ is in the linear span of F, taken in the $K[\underline{X}]$-module $H_1(K[\underline{X}, \underline{Z}])$, if and only if $f \xrightarrow[G]{*} 0$.* □

The Gröbner basis construction of the theorem starts with a set of homogeneous polynomials whose Γ-degrees equal 1. One of the results of Section 10.2 was that then the output of the algorithm [0, 1]-GRÖBNER (which is GRÖBNER truncated at Γ-degree 1) will again consist entirely of homogeneous polynomials of Γ-degree 1. Moreover, it is easy to see that the monomials employed in the forming of S-polynomials as well as those occurring in reduction steps must all have Γ-degree 0, i.e., they must be in $K[\underline{X}]$. This means that the computations that the theorem above calls for all take place in the $K[\underline{X}]$-module $H_1(K[\underline{X}, \underline{Z}])$. The whole process may thus be viewed as a Gröbner basis computation in a polynomial module.

In view of these observations, it is not hard to see that instead of coding the elements of the original $K[\underline{X}]$-module M_{mn}^K into Γ-homogeneous elements of $K[\underline{X},\underline{Z}]$, one may also develop a formalism to imitate Gröbner basis theory directly in M_{mn}^K. To this end, one observes that every element $\boldsymbol{f}$ of M_{mn}^K can be written in the form

$$\boldsymbol{f} = \sum_{i=1}^{s} c_j \boldsymbol{t}_j ,$$

where the c_j are in K and each $\boldsymbol{t}_j$ is a *positional term* of the form

$$(0,\ldots,0,t,0,\ldots,0) \qquad (t \in T(\underline{X})).$$

Noting that in our setup, such an element of M_{mn}^K corresponds to the term $t \cdot Z_i$, where i is the position of t in $\boldsymbol{t}$, it is clear how one can now define positional term orders, induced order on elements of M_{mn}^K, reduction of elements of M_{mn}^K, etc., in such a way that one obtains the same algorithmic solution to the submodule membership problem as in the theorem above. For actual implementations, this positional-term viewpoint is perhaps more appropriate, but it is important to realize that mathematically speaking, the theory is nothing but a special case of homogeneous Gröbner basis theory in polynomial rings.

Finally, we mention that submodule membership is of course not the only computational problem that can be solved by means of the Gröbner basis theory for modules that we have explained in this section. A large number of applications of Gröbner bases in ideal theory have analogues for polynomial modules. In order to obtain these results, one may proceed as follows. First, translate the given module problem into a problem in ideal theory by means of Lemma 10.59. Then look at the solution of that problem by means of Gröbner bases and see if and how it continues to work when relativized to homogeneous input with Γ-degree 1. The details of this process are necessarily unpleasant because it involves ruminating possibly lengthy proofs; two examples, namely, the extended Gröbner basis algorithm and the computation of syzygies, will be commented on in the next section.

Exercise 10.62 Use the results of this section to give an alternate proof of the fact that the $K[\underline{X}]$-module M_{mn}^K is noetherian.

10.5 Systems of Linear Equations

Recall that by a ring, we always mean a commutative ring with unity. Let R be a ring. We will write $R[\underline{Y}]$ for $R[Y_1,\ldots,Y_n]$. Suppose $f_1,\ldots,f_m \in R[\underline{Y}]$ are such that for $1 \leq j \leq m$, the polynomial f_j is either the zero polynomial or has total degree less than or equal to 1, i.e.,

$$T(f_j) \subseteq \{Y_1,\ldots,Y_n,1\}$$

if f_j is not the zero polynomial. Then

$$f_j = 0 \qquad (1 \le j \le m)$$

is called a **system of linear equations** over the ring R. More explicitly, a system of linear equations can be written in the form

$$\begin{aligned} a_{11}Y_1 + a_{12}Y_2 + \cdots + a_{1n}Y_n + b_1 &= 0 \\ a_{21}Y_1 + a_{22}Y_2 + \cdots + a_{2n}Y_n + b_2 &= 0 \\ &\vdots \\ a_{m1}Y_1 + a_{m2}Y_2 + \cdots + a_{mn}Y_n + b_m &= 0 \end{aligned}$$

with a_{ji}, $b_j \in R$. A **solution** of such a system is a common zero of the f_j in R^n. The problem of solving the system is thus the problem of determining all common zeroes of the f_j in R^n. In this context, the variables $Y_1, \ldots, Y_n$ are referred to as **unknowns**. The system is called **homogeneous** if $b_j = 0$ for $1 \le j \le m$, **inhomogeneous** otherwise. In the latter case, the system obtained by replacing each b_j with 0 is referred to as the **associated homogeneous system.**

A solution of a system of linear equations over R is by definition an element of R^n. Viewing R^n as an R-module in the sense of Example 3.28 (iv), we obtain the following results on the set of solutions of the system. The proofs are straightforward from the definitions.

Lemma 10.63 Let $S \subseteq R^n$ be the set of all solutions of a given system of linear equations over a ring R, and let S_0 be the set of solutions of the associated homogeneous system. Then the following hold:

(i) S_0 is a submodule of R^n.

(ii) If $S \neq \emptyset$, then $S = \{\, c + d \mid d \in S_0 \,\}$ for every $c \in S$.

(iii) If $S \neq \emptyset$, then $|S| = 1$ iff $S_0 = \{0\}$. □

With the notation of the lemma, the problem of solving a system of linear equations can now be made more precise. We must first decide if S is non-empty, and if so, we must produce an element of S and a set of generators for the R-module S_0.

The main purpose of this section is to show how Gröbner bases can be employed to solve this problem when R is a polynomial ring over a field. Before doing so, we will discuss the "classical" case, where R is a field. We mention that the theory of Gröbner bases is, in a sense, inappropriate here: the theory of systems of linear equations over a field really belongs with the theory of finite-dimensional vector spaces, and actual implementations should be specifically tailored to the problem. However, it is interesting to see how Gröbner basis theory also provides the complete algorithmic solution. As a matter of fact, our results on polynomial reduction alone will suffice here.

In view of (iii) of the previous lemma, the set of all solutions of a homogeneous system of linear equations over a field is a subspace of the K-vector space K^n; we will refer to it as the *solution space* of the system. The following lemma basically says that the algorithm REDUCTION of Proposition 5.30, when applied to a set of polynomials of total degree at most 1, becomes the classical Gaussian elimination algorithm.

Lemma 10.64 Let K be a computable field, and suppose

$$F = \{f_1, \ldots, f_m\} \subseteq K[\underline{Y}]$$

is a set of non-zero polynomials of total degree less than or equal to 1. Let $\leq$ be a term order on $T(\underline{Y})$ that satisfies $Y_n < \cdots < Y_1$. Then with $G = \text{REDUCTION}(F)$, the following hold:

(i) F and G have the same zeroes in K^n. In particular, if $1 \in G$, then F does not have a zero in K^n.

(ii) Each $g \in G$ has total degree less than or equal to 1.

(iii) If $1 \notin G$, then G is of the form $\{g_1, \ldots, g_r\}$ with $r \leq n$, where for $1 \leq j \leq r$, we have $\text{HM}(g_j) = Y_i$ for some $1 \leq i \leq n$. Moreover, $\text{HM}(g_j) \neq \text{HM}(g_k)$ for $1 \leq j < k \leq r$.

(iv) If $1 \notin G$, then a zero of F in K^n can be read off from G as follows. For $1 \leq j \leq r$, let $g_j = a_{j1}Y_1 + \cdots + a_{jn}Y_n + b_j$. Let

$$N = \{i_1, \ldots, i_s\} \subseteq \{1, \ldots, n\}$$

be the set of those indices i for which $Y_i \notin \text{HT}(G)$. Then the element $c = (c_1, \ldots, c_n)$ of K^n defined by

$$c_i = \begin{cases} 0 & \text{if } \; i \in N \\ -b_j \text{ where } Y_i = \text{HT}(g_j) & \text{otherwise} \end{cases}$$

is a zero of F.

(v) If $1 \notin G$, then with the notation of (iv) above, a basis for the solution space of the homogeneous system associated with the system

$$f_j = 0 \qquad (1 \leq j \leq m)$$

of linear equations is given by $\{v_1, \ldots, v_s\}$, where $v_k = (v_{k1}, \ldots, v_{kn})$ with

$$v_{ki} = \begin{cases} 0 & \text{if } \; i \in N \text{ and } i \neq i_k \\ 1 & \text{if } \; i = i_k \\ -a_{ji_k} \text{ where } Y_i = \text{HT}(g_j) & \text{otherwise.} \end{cases}$$

Before we prove the lemma, we show how the implied algorithm may be visualized in case $1 \notin G$. Suppose we order the monomials of each $g \in G$ in descending order, then order the elements of G by descending head terms, and finally display G by writing the coefficients of each $g \in G$ in a row. We thus obtain an array of width $n+1$ and height r which would typically be of the form

$$\begin{array}{cccccccccc|c}
1 & & * & * & & & * & & \cdots & \cdots & * \\
 & 1 & * & * & & & * & & \cdots & \cdots & * \\
 & & & & 1 & & * & & \cdots & \cdots & * \\
 & & & & & 1 & * & & \cdots & \cdots & * \\
 & & & & & & & 1 & \cdots & \cdots & * \\
 & & & & & & & & \ddots & \vdots & \vdots
\end{array},$$

where an $*$ stands for a possibly non-zero entry, blank space stands for zeroes, and we have separated the constant coefficients from the linear ones by a vertical line. Next, we insert $n-r$ rows each of which has one entry 1 and all other entries 0 in such a way that the entries on the diagonal of the square to the left of the vertical line all equal 1. Moreover, we switch the sign on all entries except those on the diagonal. This yields the array

$$\begin{array}{cccccccccc|c}
1 & & -* & -* & & & -* & & \cdots & & -* \\
 & 1 & -* & -* & & & -* & & \cdots & & -* \\
 & & 1 & & & & & & \cdots & & \\
 & & & 1 & & & & & \cdots & & \\
 & & & & 1 & & -* & & \cdots & & -* \\
 & & & & & 1 & -* & & \cdots & & -* \\
 & & & & & & 1 & & \cdots & & \\
 & & & & & & & 1 & \cdots & & -* \\
 & & & & & & & & \ddots & & \vdots \\
 & & & & & & & & & 1 & -*
\end{array}.$$

Items (iv) and (v) of the lemma now tell us that the rightmost column is a special solution of the system

$$f_i = 0 \qquad (1 \leq i \leq r),$$

and that a basis of the solution space of the associated homogeneous system is given by the set of all those column vectors that share a diagonal element with one of the newly added rows.

We also mention that in case $1 \in G$, i.e., when there is no solution, the application of REDUCTION destroys all information because it outputs $G = \{1\}$. If one is still interested in the solution space of the associated homogeneous system in this case, then this collapsing can be prevented by homogenizing the input first by means of an additional variable which must be placed less than all others in the term order.

Proof of Lemma 10.64 (i) This is immediate from the fact that F and G generate the same ideal in $K[\underline{Y}]$.

(ii) If f, p, $g \in K[\underline{Y}]$ are such that the total degrees of f and p both equal 1 and $f \xrightarrow[p]{} g$, say $g = f - mp$ with a monomial m, then it is easy to see that m must be a constant. It follows easily that REDUCTION applied to a set of polynomials of total degree at most one will produce an output with the same property.

(iii) Suppose $1 \notin G$. Then $\mathrm{HT}(G) \subseteq \{Y_1, \dots, Y_n\}$ by (ii) above. Since G is reduced, no two elements of G can have the same variable as their head term, and this observation easily implies the claim.

(iv) By (i) above, it suffices to show that $g(\boldsymbol{c}) = 0$ for all $g \in G$. Let $1 \leq j \leq r$, and suppose $\mathrm{HM}(g_j) = Y_l$. Then $a_{jl} = 1$, and since G is reduced, we have $a_{ji} = 0$ whenever $i \neq l$ and $i \notin N$. It follows easily that

$$g(\boldsymbol{c}) = \sum_{i=1}^{n} a_{ji}c_i + b_j = 1 \cdot (-b_j) + b_j = 0.$$

(v) Let S_0 be the solution space in question, and let S_0' be the solution space of the homogeneous system

$$a_{j1}Y_1 + \cdots + a_{jn}Y_n = 0 \qquad (1 \leq j \leq r)$$

associated with $\{\, g_j = 0 \mid 1 \leq j \leq r \,\}$. We claim that $S_0 = S_0'$. With $\boldsymbol{c}$ as in (iv), we know from the previous lemma that the set of all zeroes of F in K^n equals $\boldsymbol{c} + S_0$, and the set of all zeroes of G in K^n equals $\boldsymbol{c} + S_0'$. From the fact that F and G have the same zeroes in K^n, one now easily concludes that indeed $S_0 = S_0'$. We must thus prove that $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_s\}$ is a basis of the solution space of

$$h_j = 0 \qquad (1 \leq j \leq r), \tag{$*$}$$

where we have set $h_j = a_{j1}Y_1 + \cdots + a_{jn}Y_n$. Note that $\mathrm{HT}(h_j) = \mathrm{HT}(g_j)$.

We begin by showing that $\boldsymbol{v}_1, \dots, \boldsymbol{v}_s$ are indeed solutions of the homogeneous system $(*)$. To this end, let $1 \leq j \leq r$ and $1 \leq k \leq s$, and suppose $\mathrm{HT}(h_j) = Y_l$. Recall that here, $a_{jl} = 1$, and $a_{ji} = 0$ whenever $i \neq l$ and $i \notin N$. Discussing the indices i between 1 and n according to whether they are in N or not, one easily proves that

$$h_j(\boldsymbol{v}_k) = \sum_{i=1}^{n} a_{ji}v_{ki} = 1 \cdot (-a_{ji_k}) + a_{ji_k} \cdot 1 = 0,$$

where the subindexing of i refers to $N = \{i_1, \dots, i_s\}$ as in the definition of $\boldsymbol{v}_k$. To see that $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_s\}$ are linearly independent, suppose

$$\sum_{k=1}^{s} \lambda_k \boldsymbol{v}_k = 0 \qquad (\lambda_k \in K).$$

This equation holds componentwise, and the i_k-th component of the equation reads $\lambda_k = 0$ for $1 \le k \le r$.

It remains to prove that $\{v_1, \ldots, v_s\}$ is a generating system for the solution space of $(*)$. Suppose $d = (d_1, \ldots, d_n) \in K^n$ is a solution of $(*)$. We claim that

$$d = \sum_{k=1}^{s} d_{i_k} v_k \,. \tag{$**$}$$

We first prove that this equation holds in the ith component for each $i \in N$. But if $i \in N$, then $i = i_k$ for some $1 \le k \le s$, and the ith component reads $d_{i_k} = d_{i_k} \cdot 1$. To see that the equation $(**)$ actually holds true in its entirety, we may assume w.l.o.g. that the indexing of the elements $g_1, \ldots, g_r$ of G is such that the head terms are in descending order. Then looking at the equations of the system $(*)$ from bottom up, one easily sees that for any solution e of $(*)$, those components e_i of e that satisfy $i \in N$ determine e completely: for $l \notin N$, we must have

$$e_l = - \sum_{i=l+1}^{n} a_{ji} e_i \,,$$

where j is such that $Y_l = \mathrm{HT}(g_j)$. But the right-hand side of $(**)$ is an element of the solution space of $(*)$ which agrees with d on the ith component for all $i \in N$, and so the two must be equal. □

As we have mentioned before in Section 5.5, the output of REDUCTION is already a Gröbner basis here because the head terms are pairwise disjoint; however, this fact is not really relevant in this context.

Exercise 10.65 Write an algorithm for the computation of the set of solutions of a system of linear equations over a computable field based on the last two lemmas.

We will now discuss a type of system of linear equations where Gröbner basis techniques are truly appropriate. It is the case where R is a polynomial ring over a field, say $R = K[X_1, \ldots, X_r]$, which, as usual, will be denoted by $K[\underline{X}]$. Here, we will move constant coefficients to the right-hand side, so that we are trying to solve a system of the form

$$\begin{array}{rcl} a_{11}Y_1 + a_{12}Y_2 + \cdots + a_{1n}Y_n & = & b_1 \\ a_{21}Y_1 + a_{22}Y_2 + \cdots + a_{2n}Y_n & = & b_2 \\ & \vdots & \\ a_{m1}Y_1 + a_{n2}Y_2 + \cdots + a_{mn}Y_n & = & b_m \end{array} \qquad (a_{ji}, b_j \in K[\underline{X}]).$$

The case $m = 1$ of one linear equation has been discussed already in Section 6.1. If $m > 1$, then we consider the $K[\underline{X}]$-module $(K[\underline{X}])^m$, which we will once again denote by M_{mr}^K. For $1 \le i \le n$, we define an element a_j of M_{mr}^K by setting

$$a_i = (a_{1i}, \ldots, a_{mi}),$$

and we let $\boldsymbol{b} = (b_1, \dots, b_m)$. Solving the system of linear equations above then amounts to finding values in $K[\underline{X}]$ for the unknowns $Y_1, \dots, Y_n$ such that the equation

$$Y_1 \cdot \boldsymbol{a}_1 + \cdots + Y_n \cdot \boldsymbol{a}_n = \boldsymbol{b} \tag{$*$}$$

holds in the module M^K_{mr}. Such values obviously exist if and only if $\boldsymbol{b}$ lies in the linear span $\operatorname{lin}(\boldsymbol{a}_1, \dots, \boldsymbol{a}_n)$ of $\boldsymbol{a}_1, \dots, \boldsymbol{a}_n$ in the module M^K_{mr}. Let us recall from the previous section how this condition can be decided. We let $Z_1, \dots, Z_m$ be new indeterminates and consider the polynomial ring

$$K[\underline{X}, \underline{Z}] = K[X_1, \dots, X_r, Z_1, \dots, Z_m].$$

We then let Γ be the grading on $K[\underline{X}, \underline{Z}]$ that assigns weight 0 to $X_1, \dots, X_r$ and weight 1 to $Z_1, \dots, Z_m$. We denote by $H_1(K[\underline{X}, \underline{Z}])$ the subset of $K[\underline{X}, \underline{Z}]$ consisting of 0 and all homogeneous polynomials of Γ-degree 1, i.e.,

$$H_1(K[\underline{X}, \underline{Z}]) = \{ h_1 \cdot Z_1 + \cdots + h_m \cdot Z_m \mid h_1, \dots, h_m \in K[\underline{X}] \}.$$

Then $H_1(K[\underline{X}, \underline{Z}])$ is a $K[\underline{X}]$-module under the obvious operations, and it is naturally isomorphic to M^K_{mr} under the map

$$\begin{array}{rccc} \varphi : & M^K_{mr} & \longrightarrow & H_1(K[\underline{X}, \underline{Z}]) \\ & (h_1, \dots, h_m) & \longmapsto & h_1 Z_1 + \cdots + h_m Z_m. \end{array}$$

In order to decide the condition

$$\varphi(\boldsymbol{b}) \in \operatorname{lin}(\varphi(\boldsymbol{a}_1), \dots, \varphi(\boldsymbol{a}_n)),$$

we must compute a 1-Gröbner basis G of $\operatorname{Id}(\varphi(\boldsymbol{a}_1), \dots, \varphi(\boldsymbol{a}_n))$ in $K[\underline{X}, \underline{Z}]$ and test whether $\varphi(\boldsymbol{b}) \xrightarrow[G]{*} 0$. A positive answer is equivalent to the solvability of $(*)$. If this is the case, then values for $Y_1, \dots, Y_n$ can be found as follows. For the computation of the 1-Gröbner basis $G = \{g_1, \dots, g_s\}$, we may employ a truncated version of the *extended* Gröbner basis algorithm so that for $1 \leq j \leq s$, we obtain representations

$$g_j = \sum_{i=1}^{n} c_{ji} \cdot \varphi(\boldsymbol{a}_i) \qquad (c_{ji} \in K[\underline{X}, \underline{Z}]).$$

Moreover, when reducing $\varphi(\boldsymbol{b})$ to 0 modulo G, we may let the algorithm REDPOL provide $q_1, \dots, q_s \in K[\underline{X}, \underline{Z}]$ with

$$\varphi(\boldsymbol{b}) = q_1 g_1 + \cdots + q_s g_s,$$

and it follows easily that

$$\sum_{j=1}^{s} q_j c_{j1} \cdot \varphi(\boldsymbol{a}_1) + \cdots + \sum_{j=1}^{s} q_j c_{jn} \cdot \varphi(\boldsymbol{a}_n) = \varphi(\boldsymbol{b}).$$

We have already mentioned in the previous section that all monomials that occur in the formation of S-polynomials and in reduction steps when computing the 1-Gröbner basis in the present situation are actually in $K[\underline{X}]$, and the same holds true for the monomials that occur in the reduction steps of $\varphi(\boldsymbol{b}) \xrightarrow[G]{*} 0$. This means that for $1 \leq i \leq n$, we have

$$\sum_{j=1}^{s} q_j c_{ji} \in K[\underline{X}],$$

and this together with the fact that φ is an isomorphism of $K[\underline{X}]$-modules shows that

$$Y_i = \sum_{j=1}^{s} q_j c_{ji} \qquad (1 \leq i \leq n)$$

solves the equation $(*)$.

It remains to discuss the solution of the equation

$$Y_1 \cdot \boldsymbol{a}_1 + \cdots + Y_n \cdot \boldsymbol{a}_n = 0,$$

which of course corresponds to the homogeneous system associated to our original system of linear equations over $K[\underline{X}]$. Once again, we use the fact that φ is an isomorphism of modules and investigate the equation

$$Y_1 \cdot \varphi(\boldsymbol{a}_1) + \cdots + Y_n \cdot \varphi(\boldsymbol{a}_n) = 0 \qquad (**)$$

instead. The $\varphi(\boldsymbol{a}_i)$ are elements of $K[\underline{X}, \underline{Z}]$. The results of Section 6.1 tell us how to find the set of all solutions of $(**)$ in $K[\underline{X}, \underline{Z}]$: this is precisely the module of syzygies of

$$(\varphi(\boldsymbol{a}_1), \ldots, \varphi(\boldsymbol{a}_n))$$

in $K[\underline{X}, \underline{Z}]$. What we are looking for is the intersection of this module with $H_1(K[\underline{X}, \underline{Z}])$. It is now a matter of retracing the proofs of Proposition 6.1 and Theorem 6.4 to see that the following holds. The set of solutions in $K[\underline{X}]$ of $(**)$ is obtained by essentially the same procedure that lead to Theorem 6.4. Instead of a full Gröbner basis computation, however, one uses the truncated one that computes the 1-Gröbner basis w.r.t. Γ. The elements of A of Theorem 6.4 will then come out to be in $K[\underline{X}]$. As for B^* of the same theorem, one must include in the set B of Proposition 6.1—which B^* is a transformation of—only those $\boldsymbol{r}_{ij}$ that correspond to S-polynomials of Γ-degree 1.

Exercise 10.66 Write an algorithm for the solution of systems of linear equations over a polynomial ring $K[\underline{X}]$, where K is a computable field.

10.6 Standard Bases and the Tangent Cone

Throughout this section, K will once again be a field, and we will write $K[\underline{X}]$ for $K[X_1, \ldots, X_n]$. The characteristic property of a Gröbner basis G in $K[\underline{X}]$ w.r.t. some term order is that the head term of each $f \in \mathrm{Id}(G)$ is divided by the head term of some $g \in G$. In this section, we show how for a certain type of term order, one may compute ideal bases in $K[\underline{X}]$ with a dual property.

Definition 10.67 Let $\leq$ be a term order on $T(\underline{X})$. For $0 \neq f \in K[\underline{X}]$, we call the $\leq$-least element of f the **lowest term** of f and denote it by $\mathrm{LT}_{\leq}(f)$. A **standard basis** (w.r.t. $\leq$) is a finite subset G of $K[\underline{X}]$ such that for each $0 \neq f \in \mathrm{Id}(G)$, there exists $g \in G$ with $\mathrm{LT}_{\leq}(g) \mid \mathrm{LT}_{\leq}(f)$. If I is an ideal of $K[\underline{X}]$, then a finite subset G of I is called a **standard basis** (w.r.t. $\leq$) **of** I if it is a standard basis w.r.t. $\leq$, and $I = \mathrm{Id}(G)$.

Throughout, we let Γ be a grading of $K[\underline{X}]$ that satisfies $\Gamma(X_i) > 0$ for $1 \leq i \leq n$. Moreover, $\leq$ will be a Γ-compatible term order on $T(\underline{X})$. This is the type of term order for which we will obtain existence and construction of standard bases. The results on homogenization of Section 10.3 will be instrumental. We let Z be a new indeterminate, and as usual, we denote by $T(\underline{X}, Z)$ the set of all terms in the variables $X_1, \ldots, X_n, Z$ and write

$$K[\underline{X}, Z] = K[X_1, \ldots, X_n, Z].$$

We extend Γ to a grading Γ' of $K[\underline{X}, Z]$ by setting $\Gamma'(Z) = 1$. For u, $v \in T(\underline{X}, Z)$, say $u = s \cdot Z^k$ and $v = t \cdot Z^m$ with s, $t \in T(\underline{X})$, we set

$$\begin{aligned} s \cdot Z^k \leq' t \cdot Z^m \quad \text{iff} \quad & k < m, \text{ or} \\ & k = m \text{ and } \Gamma(s) < \Gamma(t), \text{ or} \\ & k = m, \ \Gamma(s) = \Gamma(t), \text{ and } t \leq s. \end{aligned}$$

It is a straightforward exercise to show that $\leq'$ is a term order on $T(\underline{X}, Z)$. (The condition $\Gamma(X_i) > 0$ for $1 \leq i \leq n$ is needed to ensure that $1 \leq' t$ for all $t \in T(\underline{X}, Z)$.) The relevance of $\leq'$ for the problem of finding standard bases stems from the fact that it reverses $\leq$ in the sense of the following lemma.

Lemma 10.68 Let u, $v \in T(\underline{X}, Z)$, say $u = s \cdot Z^k$ and $v = t \cdot Z^m$ with s, $t \in T(\underline{X})$, and assume that $\Gamma'(u) = \Gamma'(v)$. Then $u \leq' v$ iff $t \leq s$.

Proof For the direction "$\Longrightarrow$," assume that $s \cdot Z^k \leq' t \cdot Z^m$. If $k < m$, then $\Gamma(s) > \Gamma(t)$ and so $t \leq s$ since $\leq$ is Γ-compatible. If $k = m$, then necessarily $\Gamma(s) = \Gamma(t)$ and so $t \leq s$ by the definition of $\leq'$. For the reverse implication, suppose $t \leq s$. If $\Gamma(t) < \Gamma(s)$, then we must have $k < m$ and thus $s \cdot Z^k \leq' t \cdot Z^m$. If $\Gamma(t) = \Gamma(s)$, then also $k = m$, and $t \leq s$ implies $s \cdot Z^k \leq' t \cdot Z^m$. □

The following lemma and theorem use the notation f^* for the homogenization w.r.t. Z of $f \in K[\underline{X}]$ and g_* for the dehomogenization w.r.t. Z of $g \in K[\underline{X}, Z]$ as defined in Section 10.3. Moreover, we denote the head term w.r.t. $\leq'$ of $g \in K[\underline{X}, Z]$ by $\mathrm{HT}_{\leq'}(g)$.

Lemma 10.69 If $g \in K[\underline{X}, Z]$ is Γ'-homogeneous, then

$$\left(\mathrm{HT}_{\leq'}(g)\right)_* = \mathrm{LT}_{\leq}(g_*).$$

Proof If $u \in T(\underline{X}, Z)$, say $u = t \cdot Z^k$ with $t \in T(\underline{X})$, then clearly $u_* = t$. The previous lemma can thus be restated as saying that for u, $v \in T(\underline{X}, Z)$ with $\Gamma'(u) = \Gamma'(v)$, we have $u \leq' v$ iff $v_* \leq u_*$. The claim now follows easily from the fact that for $h \in K[\underline{X}, Z]$, say

$$h = \sum_{i=1}^{m} a_i u_i \qquad (a_i \in K,\ u_i \in T(\underline{X}, Z)),$$

the dehomogenization is given by

$$h_* = \sum_{i=1}^{m} a_i (u_i)_* . \quad \square$$

The theorem below uses the fact that an ideal that is generated by a set of homogeneous polynomials has a Gröbner basis that consists entirely of homogeneous polynomials (Corollary 10.34).

Theorem 10.70 *Let F be a finite subset of $K[\underline{X}]$, and let $G \subseteq K[\underline{X}, Z]$ be a Gröbner basis of F^* w.r.t. $\leq'$ such that each $g \in G$ is Γ'-homogeneous. Then G_* is a standard basis of* $\mathrm{Id}(F)$ *in $K[\underline{X}]$ w.r.t. $\leq$.*

Proof We have $\mathrm{Id}(G_*) = \mathrm{Id}(F)$ by Lemma 10.54. It remains to show that G_* is a standard basis in $K[\underline{X}]$. To this end, let $f \in \mathrm{Id}(G_*)$. Then $f \in \mathrm{Id}(F)$ by the above equality, and so by Lemma 10.52, there exists $k \in \mathbb{N}$ with

$$Z^k \cdot f^* \in \mathrm{Id}(F^*) = \mathrm{Id}(G).$$

Since G is a Gröbner basis in $K[\underline{X}, Z]$ w.r.t. $\leq'$, it follows that there exists $g \in G$ such that

$$\mathrm{HT}_{\leq'}(g) \mid \mathrm{HT}_{\leq'}(Z^k \cdot f^*).$$

This divisibility is clearly preserved if we dehomogenize, i.e., set $Z = 1$, and the lemma preceding the theorem tells us that

$$\left(\mathrm{HT}_{\leq'}(g)\right)_* = \mathrm{LT}_{\leq}(g_*)$$

and

$$\left(\mathrm{HT}_{\leq'}(Z^k \cdot f^*)\right)_* = \mathrm{LT}_{\leq}\left((Z^k \cdot f^*)_*\right) = \mathrm{LT}_{\leq}(f).$$

In view of the trivial fact that $g_* \in G_*$, this proves the claim. □

It is clear from the last theorem that for computable K, we may actually compute standard bases from given finite ideal bases w.r.t. every decidable term order that is compatible with some grading that assigns non-zero weights only. An example would be the total degree-lexicographical term order.

The interest in standard bases actually stems from algebraic geometry. Let us consider a univariate polynomial ideal I generated by $f \in K[X]$, say

$$f = \sum_{i=0}^{k} a_i X^i \qquad (a_i \in K).$$

Let j be the least index with $a_i \neq 0$. The number j then describes the local behavior of f as a function from K to K at the point $0 \in K$: it is easy to see that j is the multiplicity of 0 as a zero of f; moreover, it also equals the least natural number ν with $f^{(\nu)}(0) \neq 0$. In algebraic geometry, one defines the *multipicity* of $(0, \dots, 0)$ as a zero of the multivariate polynomial $f \in K[\underline{X}]$ to be the least $d \in \mathbb{N}$ such that f contains a term of total degree d. Computing a standard basis of a multivariate ideal I may thus be viewed as an attempt to obtain information about the local behavior of I at the point $(0, \dots, 0) \in K^n$. (If c is any zero of I in K^n, then one may employ the same methods by passing from f to $f(X_1 - c_1, \dots, X_n - c_n)$.) What one is interested in is the ideal generated by the *lowest forms* of the ideal.

As before, Γ will be a grading with $\Gamma(X_i) > 0$ for $1 \leq i \leq n$, and $\leq$ will be a Γ-compatible term order. Since the extended term order $\leq'$ is no longer relevant here, we will write $\mathrm{LT}(f)$ instead of $\mathrm{LT}_{\leq}(f)$. Let $f \in K[\underline{X}]$, and suppose d is the least natural number such that $T(f)$ contains an element of Γ-degree d. Then the d-homogeneous part $f_{(d)}$ (i.e., the sum of all monomials of f of Γ-degree d) is called the **Γ-lowest form** of f, or **lowest form** of f for short, and we denote it by $\mathrm{LF}(f)$. Furthermore, we write

$$\mathrm{LF}(F) = \{\, \mathrm{LF}(f) \mid f \in F \,\}$$

whenever F is a subset of $K[\underline{X}]$. If I is an ideal of $K[\underline{X}]$ and Γ is the standard grading by total degree, then the variety of the ideal $\mathrm{Id}(\mathrm{LF}(I))$ in K^n is called the **tangent cone** of I. In algebraic geometry, it is usually assumed that K is algebraically closed, but the tangent cone is perhaps better visualized in the case $K = \mathbb{Q}$ or $K = \mathbb{R}$.

Exercise 10.71 Draw pictures of the varieties in $\mathbb{R}^2$ of the ideals $\mathrm{Id}(Y - X^2)$, $\mathrm{Id}(Y - X^2 - 2X)$, $\mathrm{Id}(Y - X^2 - 2X - 1)$, and $\mathrm{Id}((X-3)^2 + (Y-4)^2 - 25)$ as well as their respective tangent cones.

What we are going to prove is that for every ideal I of $K[\underline{X}]$, the ideal $\mathrm{Id}(\mathrm{LF}(I))$ is generated by $\mathrm{LF}(G)$ whenever G is a standard basis of I w.r.t. $\leq$, which—as it is important to keep in mind—is Γ-compatible. We could thus decide membership in $\mathrm{Id}(\mathrm{LF}(I))$ by computing a Gröbner basis of

$\mathrm{LF}(G)$ w.r.t. any term order. We will see that this is not necessary; membership in $\mathrm{Id}(\mathrm{LF}(I))$ can be decided by means of an "upside-down" reduction process using $\mathrm{LF}(G)$ as is. The theory below is in fact an upside-down version of the results of Exercise 10.45.

Let f, $p \in K[\underline{X}]$ be non-zero polynomials, and let $m \in K[\underline{X}]$ be a monomial. We say that f **LL-reduces** to $f - mp$ modulo p and write

$$f \xrightarrow[p]{} f - mp$$

if m is such that the product of m and the lowest monomial of p equals the lowest monomial of f. LL-reduction modulo a finite subset of $K[\underline{X}]$, LL-reducibility, and LL-normal forms are defined in the obvious way according to Definition 5.18.

The definition of a standard basis can now be rephrased as saying that every non-zero $f \in \mathrm{Id}(G)$ is LL-reducible modulo G. Unfortunately, however, LL-reduction is not noetherian in general: an infinite ascending chain is for example given by

$$X \xrightarrow[1+X]{} X^2 \xrightarrow[1+X]{} X^3 \xrightarrow[1+X]{} \cdots .$$

The next lemma and proposition show that all is well if we LL-reduce modulo homogeneous polynomials. The proof of the lemma should be easy for anybody who has studied other types of reduction.

Lemma 10.72 Let f, g, $p \in K[\underline{X}]$ such that f LL-reduces to g modulo p. Then the following hold:

(i) $\mathrm{LT}(f) < \mathrm{LT}(g)$.

(ii) If in addition, p is Γ-homogeneous, then every term in $T(g) \setminus T(f)$ has the same Γ-degree as $\mathrm{LT}(f)$. □

In the proof of the following proposition we will once again abuse the notation Γ by writing $\Gamma(f)$ for the maximum of the Γ-degrees of the terms of f. The proposition shows that for a standard basis G consisting of homogeneous polynomials, membership in $\mathrm{Id}(G)$ can be decided by means of LL-reduction.

Proposition 10.73 *If G is a set of Γ-homogeneous polynomials, then LL-reduction modulo G is a noetherian reduction relation. If in addition, G is a standard basis, then the following are equivalent for each $f \in K[\underline{X}]$:*

(i) $f \in \mathrm{Id}(G)$.

(ii) 0 is an LL-normal form of f modulo G.

(iii) Every LL-normal form of f modulo G equals 0.

Proof It follows immediately from (i) of the lemma above that LL-reduction is strictly antisymmetric. Now assume for a contradiction that $\{f_i\}_{i \in \mathbb{N}}$ is an infinite sequence of elements of $K[\underline{X}]$ such that f_i LL-reduces to f_{i+1} modulo G. From the previous lemma, one easily concludes that $\{\mathrm{LT}(f_i)\}_{i \in \mathbb{N}}$ is strictly ascending, and an easy induction on i shows that $\Gamma(t) \leq \Gamma(f_0)$ for all terms occurring anywhere in the f_i. We would thus have to have infinitely many different terms of Γ-degree less than or equal to $\Gamma(f_0)$, which is easily seen to be impossible in view of the fact that Γ assigns non-zero weights only.

Now assume that in addition, G is a standard basis, and let $f \in K[\underline{X}]$. Since LL-reduction modulo G is noetherian, there exists an LL-normal form of f modulo G. The direction "(iii)$\Longrightarrow$(ii)" is now trivial. Next, we note that clearly, the difference of any two polynomials that are connected by a chain of LL-reductions modulo G is in $\mathrm{Id}(G)$. The direction "(ii)$\Longrightarrow$(i)" is thus immediate from the fact that here, the difference of f and 0 lies in $\mathrm{Id}(G)$. For "(i)$\Longrightarrow$(iii)," let h be any normal form of $f \in \mathrm{Id}(G)$. Then h is an element of $\mathrm{Id}(G)$ in LL-normal form modulo G and must thus equal 0 since G is a standard basis. □

Corollary 10.74 *Let I be an ideal of $K[\underline{X}]$ and G a finite subset of I consisting of homogeneous polynomials only such that for each $0 \neq f \in I$, there exists $g \in G$ with $\mathrm{LT}(g) \mid \mathrm{LT}(f)$. Then G is a standard basis of I.*

Proof It follows immediately from the obvious inclusion $\mathrm{Id}(G) \subseteq I$ that G is a standard basis. To see that that the reverse inclusion holds as well, let $f \in I$. Then f has an LL-normal form h modulo G. The difference $f - h$ is in $\mathrm{Id}(G)$ and thus in I, and so h is in I. The assumption on divisibilities says that 0 is the only LL-normal form modulo G in the ideal I, and so we have $h = 0$. Using $f - h \in \mathrm{Id}(G)$ again, we see that $f \in \mathrm{Id}(G)$. □

We are now in a position to prove the main theorem on the connection between standard bases and the tangent cone. In view of the last proposition and the fact that lowest forms of polynomials are by definition homogeneous, the theorem below allows us to decide membership in $\mathrm{Id}(\mathrm{LF}(I))$ by means of LL-reduction in case K is computable.

Theorem 10.75 *Let G be a standard basis of the ideal I of $K[\underline{X}]$. Then $\mathrm{LF}(G)$ is a standard basis of the ideal $\mathrm{Id}(\mathrm{LF}(I))$ of $K[\underline{X}]$.*

Proof In view of the corollary above, it suffices to prove that for each $0 \neq f \in \mathrm{Id}(\mathrm{LF}(I))$, there exists $h \in \mathrm{LF}(G)$ with $\mathrm{LT}(h) \mid \mathrm{LT}(f)$. If $f \in \mathrm{Id}(\mathrm{LF}(I))$, then there exist monomials $m_1, \ldots, m_k$ and non-zero polynomials $f_1, \ldots, f_k \in I$ with

$$f = m_1\mathrm{LF}(f_1) + \cdots + m_k\mathrm{LF}(f_k) = \mathrm{LF}(m_1 f_1) + \cdots + \mathrm{LF}(m_k f_k).$$

Each summand on the right-hand side is Γ-homogeneous. Now if $f \neq 0$, then $\mathrm{LF}(f)$ is the sum of all those summands whose Γ-degree equals

$\Gamma(\mathrm{LF}(f))$, say

$$\mathrm{LF}(f) = \mathrm{LF}(m_{i_1} f_{i_1}) + \cdots + \mathrm{LF}(m_{i_r} f_{i_r}).$$

From the fact that the sum of the lowest forms of the polynomials $m_{i_1} f_{i_1}$, $\ldots$, $m_{i_r} f_{i_r}$ is not zero, one easily sees that

$$\mathrm{LF}(m_{i_1} f_{i_1}) + \cdots + \mathrm{LF}(m_{i_r} f_{i_r}) = \mathrm{LF}(m_{i_1} f_{i_1} + \cdots + m_{i_r} f_{i_r}).$$

Now $m_{i_1} f_{i_1} + \cdots + m_{i_r} f_{i_r} \in I$, and since the term order $\leq$ in question is Γ-compatible, the lowest term of this polynomial is the lowest term of its lowest form, i.e., the lowest term of $\mathrm{LF}(f)$. We may conclude that there exists $g \in G$ with

$$\mathrm{LT}(g) \mid \mathrm{LT}\big(\mathrm{LF}(f)\big).$$

By the same argument as above, we also have

$$\mathrm{LT}(g) = \mathrm{LT}\big(\mathrm{LF}(g)\big) \quad \text{and} \quad \mathrm{LT}(f) = \mathrm{LT}\big(\mathrm{LF}(f)\big).$$

The divisibility above can thus be writen as

$$\mathrm{LT}\big(\mathrm{LF}(g)\big) \mid \mathrm{LT}(f),$$

and thus $h = \mathrm{LF}(g)$ is the desired element of $\mathrm{LF}(G)$. □

Exercise 10.76 Let I be a zero-dimensional ideal of $K[\underline{X}]$. Show the following:

(i) The set $\{\, t \in T(\underline{X}) \mid \mathrm{LT}(f) \nmid t \text{ for all } f \in I \,\}$ is finite.

(ii) Suppose $(0, \ldots, 0)$ is a zero of I, and let Q be the primary component of I whose associated prime is $\mathrm{Id}(X_1, \ldots, X_n)$. Then $\mathrm{LT}(Q) = \mathrm{LT}(I)$ and $\mathrm{LF}(Q) = \mathrm{LF}(I)$.

10.7 Symmetric Functions

Throughout this section, we will use our usual notational conventions, where K is a field, $T(\underline{X})$ the set of all terms in X_1, $\ldots$, X_n, and $K[\underline{X}] = K[X_1, \ldots, X_n]$. If π is a permutation on the set $\{1, \ldots, n\}$, then according to Lemma 2.17 (i), the element

$$(X_{\pi(1)}, \ldots, X_{\pi(n)}) \in (K[\underline{X}])^n$$

gives rise to a homomorphism

$$\begin{array}{rccc} \varphi_\pi : & K[\underline{X}] & \longrightarrow & K[\underline{X}] \\ & f & \longmapsto & f(X_{\pi(1)}, \ldots, X_{\pi(n)}) \end{array}$$

satisfying $\varphi_\pi(X_i) = X_{\pi(i)}$ for $1 \leq i \leq n$, and and $\varphi_\pi \restriction K = \mathrm{id}_K$. It is easy to see that we have $\varphi_\pi \circ \varphi_{\pi^{-1}} = \varphi_{\pi^{-1}} \circ \varphi_\pi = \mathrm{id}_{K[\underline{X}]}$, and so φ_π is in fact an automorphism of $K[\underline{X}]$. It is equally easy to see that the set

$$F_\pi = \{\, f \in K[\underline{X}] \mid \varphi_\pi(f) = f \,\}$$

is a subring of $K[\underline{X}]$. Moreover, $K \subseteq F_\pi$, so that F_π is in fact a K-subalgebra of $K[\underline{X}]$. It follows that the intersection

$$S(K[\underline{X}]) = \bigcap_{\substack{\pi \text{ a permuta-} \\ \text{tion on } \{1,\ldots,n\}}} F_\pi$$

is again a K-subalgebra and thus, in particular, a subring of $K[\underline{X}]$. The elements of this subring are called **symmetric functions**, or, more precisely, **symmetric polynomials** (over K in $X_1, \ldots, X_n$). A polynomial $f \in K[\underline{X}]$ is thus symmetric if

$$f(X_1, \ldots, X_n) = f(X_{\pi(1)}, \ldots, X_{\pi(n)})$$

for every permutation π on $\{1, \ldots, n\}$. For $1 \leq i \leq n$, we define

$$\sigma_i = \sum_{1 \leq j_1 < \cdots < j_i \leq n} X_{j_1} \cdot \cdots \cdot X_{j_i}$$

and call σ_i the ith **elementary symmetric polynomial.** The polynomial σ_i is thus the sum of all possible distinct products of exactly i different variables, and one easily verifies that $\sigma_1, \ldots, \sigma_n$ are indeed symmetric polynomials. More explicitly, we have

$$\begin{aligned}
\sigma_1 &= X_1 + \cdots + X_n \\
\sigma_2 &= X_1X_2 + \cdots + X_1X_n + X_2X_3 + \cdots + X_2X_n + \cdots + X_{n-1}X_n \\
&\vdots \\
\sigma_n &= X_1 \cdot \cdots \cdot X_n.
\end{aligned}$$

The main non-algorithmic result of this section is that the ring $S(K[\underline{X}])$ of symmetric polynomials is generated by $\sigma_1, \ldots, \sigma_n$, i.e., that

$$S(K[\underline{X}]) = \{\, p(\sigma_1, \ldots, \sigma_n) \mid p \in K[\underline{X}] \,\}.$$

From an algorithmic point of view, we could thus employ the algorithm SUBRINGMEMTEST of Corollary 6.45 to compute, for symmetric $f \in K[\underline{X}]$, a polynomial $p \in K[\underline{X}]$ with $f = p(\sigma_1, \ldots, \sigma_n)$. (This procedure would also yield a method to decide whether or not a given polynomial is symmetric, but this can of course also achieved directly by inspection.) We will see that we can do much better than that. We will define a type of polynomial reduction modulo $\{\sigma_1, \ldots, \sigma_n\}$ in such a way that we can compute a normal form g of $f \in K[\underline{X}]$ with the property that g vanishes if and only if f is symmetric; moreover, the reduction process will provide $p \in K[\underline{X}]$ with $f = p(\sigma_1, \ldots, \sigma_n) + g$, so that in fact $f = p(\sigma_1, \ldots, \sigma_n)$ in case of symmetry.

For the rest of this section, we let $\leq$ be the lexicographical term order on $T(\underline{X})$, where $X_1 \gg \cdots \gg X_n$. A term $t \in T(\underline{X})$ is called **descending** if $t = X_1^{\nu_1} \cdot \cdots \cdot X_n^{\nu_n}$ with $\nu_1 \geq \cdots \geq \nu_n$.

Lemma 10.77 (i) If $\nu_1, \ldots, \nu_n \in \mathbb{N}$ and $g_1, \ldots, g_n \in K[\underline{X}]$, then

$$\mathrm{HT}(g_1^{\nu_1} \cdot \cdots \cdot g_n^{\nu_n}) = \big(\mathrm{HT}(g_1)\big)^{\nu_1} \cdot \cdots \cdot \big(\mathrm{HT}(g_n)\big)^{\nu_n}.$$

(ii) If $0 \neq f \in K[\underline{X}]$ is symmetric, then $\mathrm{HT}(f)$ is descending.

(iii) $\mathrm{HT}(\sigma_i) = X_1 \cdot \cdots \cdot X_i$ for $1 \leq i \leq n$.

(iv) If $t = X_1^{\nu_1} \cdot \cdots \cdot X_n^{\nu_n}$ is descending, then

$$t = \mathrm{HT}(\sigma_1^{\nu_1 - \nu_2} \cdot \cdots \cdot \sigma_{n-1}^{\nu_{n-1} - \nu_n} \cdot \sigma_n^{\nu_n}).$$

Proof (i) This follows easily from Lemma 5.17 together with induction on $\nu_1 + \cdots + \nu_n$.

(ii) Let $0 \neq f \in K[\underline{X}]$ be symmetric, and let $X_{i_1}^{\nu_1} \cdot \cdots \cdot X_{i_n}^{\nu_n}$ be the head term of f, where $\{i_1, \ldots, i_n\} = \{1, \ldots, n\}$, and the i_l are such that $\nu_1 \geq \cdots \geq \nu_n \geq 0$ and $i_j < i_k$ whenever $1 \leq j < k \leq n$ with $\nu_j = \nu_k$. In other words, we write the variables by decreasing exponents first and then break ties by increasing indices. We claim that $i_l = l$ for $1 \leq l \leq n$, so that t is indeed descending. Assume for a contradiction that this is not so. Let j be the least index such that $i_j \neq j$. Then $i_j > j$ because the map $j \longmapsto i_j$ is clearly bijective. Moreover, we have $j = i_k$ for some k with $j < k$. Now $i_k < i_j$ and $j < k$ together imply $\nu_j > \nu_k$ by our choice of the indexing. Let π be the permutation that does nothing but switch i_j and i_k. From the fact that

$$f(X_1, \ldots, X_n) = f(X_{\pi(1)}, \ldots, X_{\pi(n)})$$

one easily concludes that

$$s = X_{i_1}^{\nu_1} \cdot \cdots \cdot X_{i_{j-1}}^{\nu_{j-1}} \cdot X_{i_k}^{\nu_j} \cdot X_{i_{j+1}}^{\nu_{j+1}} \cdot \cdots \cdot X_{i_{k-1}}^{\nu_{k-1}} \cdot X_{i_j}^{\nu_k} \cdot X_{i_{k+1}}^{\nu_{k+1}} \cdot \cdots \cdot X_{i_n}^{\nu_n}$$

is a term in f. From the fact that $i_l = l$ for $1 \leq l < j$ together with $j = i_k$ and $\nu_j > \nu_k$ we see that $s > t$, contradicting the fact that $t = \mathrm{HT}(f)$.

(iii) This is immediate from (ii) together with the rather obvious fact that $X_1 \cdot \cdots \cdot X_i$ is the only descending term in σ_i.

(iv) Using (i) and (iii) above, we see that

$$\begin{aligned}
&\mathrm{HT}(\sigma_1^{\nu_1 - \nu_2} \cdot \cdots \cdot \sigma_{n-1}^{\nu_{n-1} - \nu_n} \cdot \sigma_n^{\nu_n}) \\
&= \big(\mathrm{HT}(\sigma_1)\big)^{\nu_1 - \nu_2} \cdot \cdots \cdot \big(\mathrm{HT}(\sigma_{n-1})\big)^{\nu_{n-1} - \nu_n} \cdot \cdots \cdot \big(\mathrm{HT}(\sigma_n)\big)^{\nu_n} \\
&= X_1^{\nu_1 - \nu_2} \cdot (X_1 X_2)^{\nu_2 - \nu_3} \cdot \cdots \cdot (X_1 \cdots X_{n-1})^{\nu_{n-1} - \nu_n} \cdot \cdots \cdot (X_1 \cdots X_n)^{\nu_n} \\
&= X_1^{\nu_1} \cdot \cdots \cdot X_n^{\nu_n}. \quad \square
\end{aligned}$$

If $t = X_1^{\nu_1} \cdot \cdots \cdot X_n^{\nu_n}$ is a descending term, then we write

$$p_t = \sigma_1^{\nu_1 - \nu_2} \cdot \cdots \cdot \sigma_{n-1}^{\nu_{n-1} - \nu_n} \cdot \sigma_n^{\nu_n}.$$

The polynomial p_t is thus symmetric with head term t.

Exercise 10.78 What is p_t if t is 1, $X_1 \cdot \dots \cdot X_i$, $X_1^\nu \cdot \dots \cdot X_i^\nu$?

We are now in a position to define the type of reduction that will lead to normal forms as described earlier in this section. Let f, $g \in K[\underline{X}]$. Then we say that f **σ-reduces** to g and write $f \xrightarrow[\sigma]{} g$ if there is a monomial at occurring in f such that t is descending and $g = f - ap_t$ with p_t as defined above. We define σ-reducibility and σ-normal forms in the obvious way as in Definition 5.18. The notations for the various closures of $\xrightarrow[\sigma]{}$ and $\xrightarrow[\sigma]{}$ will be as in Definition 4.71. It is clear that a σ-reduction step is a special case of an ordinary polynomial reduction step according to Definition 5.18. The following lemma is therefore immediate from Theorem 5.21.

Proposition 10.79 *The relation* $\xrightarrow[\sigma]{}$ *is a noetherian reduction relation.* □

It should be noted that the definition of σ-reduction makes no reference to the term order at all. The sole reason for specifying a term order $\leq$ is to turn σ-reduction into a restricted version of ordinary reduction w.r.t. $\leq$, so that results such as the above proposition need not be proved over again. The next lemma collects some results that are specific to σ-reduction as opposed to ordinary polynomial reduction.

Lemma 10.80 Let $f \in K[\underline{X}]$. Then the following hold:

(i) f is σ-reducible iff it contains a descending term.

(ii) If f is symmetric and non-zero, then it is σ-reducible.

(iii) If $g \in K[\underline{X}]$ such that $f \xrightarrow[\sigma]{*} g$, then there exists $p \in K[\underline{X}]$ with $f = p(\sigma_1, \dots, \sigma_n) + g$. In particular, $f - g$ is symmetric.

(iv) If K is computable, then σ-reduction is decidable. Moreover, the algorithm that performs σ-reduction can be devised in such a way that it computes a polynomial p as in (iii) when it σ-reduces f to g.

Proof Part (i) is immedate from the definition of σ-reduction, while (ii) follows from (i) together with the fact that the head term of a symmetric polynomial is descending by Lemma 10.77 (iii). For statement (iii), assume that $f \xrightarrow[\sigma]{*} g$. We first note that an easy induction on the length of the reduction chain shows that there exist descending terms $t_1, \dots, t_r$ with

$$g = f - \sum_{i=1}^{r} a_i p_{t_i} \qquad (a_1, \dots, a_r \in K).$$

If we define, for an arbitrary descending term $t = X_1^{\nu_1} \cdot \dots \cdot X_n^{\nu_n}$,

$$s_t = X_1^{\nu_1 - \nu_2} \cdot \dots \cdot X_{n-1}^{\nu_{n-1} - \nu_n} \cdot X_n^{\nu_n},$$

then it is easy to see from the definition of the p_{t_i} that the polynomial

$$p = \sum_{i=1}^{r} a_i s_{t_i}$$

has the desired property. The fact that $f - g$ is symmetric is now immediate from the fact that the symmetric polynomials form a subring of $K[\underline{X}]$. The statements of (iv) on decidability and computability are easy consequences of the proof of of (iii) and of the definitions. □

Proposition 10.81 *The relation $\xrightarrow[\sigma]{}$ has unique normal forms.*

Proof Let f, g_1, $g_2 \in K[\underline{X}]$ such that g_1 and g_2 are σ-normal forms of f, and assume for a contradiction that $g_1 \neq g_2$. Then by the previous lemma, both $f - g_1$ and $f - g_2$ are symmetric, and thus

$$g_1 - g_2 = (f - g_2) - (f - g_1)$$

is symmetric too. Lemma 10.77 (ii) now tells us that $\mathrm{HT}(g_1 - g_2)$ is descending, and since this must have been a term in g_1 or g_2, it follows with (ii) of the last lemma that g_1 or g_2 was σ-reducible, a contradiction. □

The main theorem of this section states that just as ordinary reducibility to zero modulo a Gröbner basis is equivalent to ideal membership, σ-reducibility to zero is equivalent to membership in the ring of symmetric polynomials. Moreover, the latter ring is generated by the elementary symmetric functions.

Theorem 10.82 *Let $f \in K[\underline{X}]$. Then the following are equivalent:*

(i) f is symmetric.

(ii) There exists $p \in K[\underline{X}]$ with $f = p(\sigma_1, \ldots, \sigma_n)$.

(iii) $f \xrightarrow[\sigma]{*} 0$.

Proof For "(i)⟹(iii)," let g be a σ-normal form of f. Then $g = (f - g) - f$ is symmetric by (iii) of the last lemma. Being in σ-normal form, it must thus equal zero by (ii) of the last lemma. The implication "(iii)⟹(ii)" is immediate from (iii) of the last lemma, and "(ii)⟹(i)" follows from the fact that the symmteric polynomials form a subring of $K[\underline{X}]$. □

Corollary 10.83 *The elementary functions generate the ring of symmetric polynomials, i.e.,*

$$S(K[\underline{X}]) = \{\, p(\sigma_1, \ldots, \sigma_n) \mid p \in K[\underline{X}] \,\}. \quad \square$$

The next corollary is immediate from the theorem above together with the proposition and the lemma preceding it.

Corollary 10.84 *Assume that K is computable. Then one can find an algorithm that computes, for arbitrary $f \in K[\underline{X}]$, polynomials p, $g \in K[\underline{X}]$ such that*

$$f = p(\sigma_1, \ldots, \sigma_n) + g,$$

the polynomial g is uniquely determined by f, and f is symmetric iff $g = 0$.

Exercise 10.85 Let $K = \mathbb{Q}$ and $n = 3$, and let $f = X_1^2X_2^2 + X_1^2X_3^2 + X_2^2X_3^2$. Use σ-reduction to confirm the obvious fact that f is symmetric, and compute $p \in K[X_1, X_2, X_3]$ with $f = p(\sigma_1, \sigma_2, \sigma_3)$.

Knowing that every symmetric polynomial f can be represented in the form $p(\sigma_1, \ldots, \sigma_n)$ with $p \in K[\underline{X}]$, we are now going to show that the polynomial p is uniquely determined by f.

Lemma 10.86 Let $\nu_1, \ldots, \nu_n \in \mathbb{N}$. Then

$$\mathrm{HT}(\sigma_1^{\nu_1} \cdot \cdots \cdot \sigma_n^{\nu_n}) = X_1^{\nu_1 + \cdots + \nu_n} \cdot X_2^{\nu_2 + \cdots + \nu_n} \cdot \cdots \cdot X_n^{\nu_n}.$$

Proof According to Lemma 10.77 (i) and (iii), we have

$$\begin{aligned} \mathrm{HT}(\sigma_1^{\nu_1} \cdot \cdots \cdot \sigma_n^{\nu_n}) &= X_1^{\nu_1} \cdot (X_1X_2)^{\nu_2} \cdot \cdots \cdot (X_1 \cdot \cdots \cdot X_n)^{\nu_n} \\ &= X_1^{\nu_1 + \cdots + \nu_n} \cdot X_2^{\nu_2 + \cdots + \nu_n} \cdot \cdots \cdot X_n^{\nu_n}. \quad \square \end{aligned}$$

Lemma 10.87 Let $\nu_1, \ldots, \nu_n, \mu_1, \ldots, \mu_n \in \mathbb{N}$ such that $\nu_i \neq \mu_i$ for at least one index i with $1 \leq i \leq n$. Then

$$\mathrm{HT}(\sigma_1^{\nu_1} \cdot \cdots \cdot \sigma_n^{\nu_n}) \neq \mathrm{HT}(\sigma_1^{\mu_1} \cdot \cdots \cdot \sigma_n^{\mu_n}).$$

Proof The claim follows easily from the last lemma together with the obvious fact that

$$\nu_i + \cdots + \nu_n \neq \mu_i + \cdots + \mu_n$$

if i is maximal with $1 \leq i \leq n$ and $\nu_i \neq \mu_i$. □

Lemma 10.88 If $0 \neq p \in K[\underline{X}]$, then $p(\sigma_1, \ldots, \sigma_n) \neq 0$.

Proof For each $t = X_1^{\nu_1} \cdot \cdots \cdot X_n^{\nu_n} \in T(p)$ we define

$$u_t = \mathrm{HT}(\sigma_1^{\nu_1} \cdot \cdots \cdot \sigma_n^{\nu_n}).$$

By the previous lemma, we have $u_{t_1} \neq u_{t_2}$ whenever t_1, $t_2 \in T(p)$ with $t_1 \neq t_2$. It is now easy to see that $u = \max\{\, u_t \mid t \in T(p) \,\}$ must be a term in $p(\sigma_1, \ldots, \sigma_n)$. □

Proposition 10.89 *If $f \in K[\underline{X}]$ is symmetric, then the polynomial $p \in K[\underline{X}]$ that satisfies $p(\sigma_1, \ldots, \sigma_n)$ is uniquely determined by f.*

Proof Let $p_1, p_2 \in K[\underline{X}]$ with $p_1(\sigma_1,\ldots,\sigma_n) = p_2(\sigma_1,\ldots,\sigma_n)$. Then

$$0 = p_1(\sigma_1,\ldots,\sigma_n) - p_2(\sigma_1,\ldots,\sigma_n) = (p_1 - p_2)(\sigma_1,\ldots,\sigma_n),$$

and so $p_1 = p_2$ by the lemma above. □

We see that if we reduce a polynomial $f \in K[\underline{X}]$ to its σ-normal form $g \in K[\underline{X}]$, then the resulting polynomial $p \in K[\underline{X}]$ with $f - g = p(\sigma_1,\ldots,\sigma_n)$ will always be the same, regardless of the particular strategy that was used in the (non-deterministic) σ-reduction algorithm.

In addition to being uniquely determined by f, the polynomial p also satisfies an interesting degree bound.

Proposition 10.90 *Let $0 \neq p \in K[\underline{X}]$ and $f = p(\sigma_1,\ldots,\sigma_n)$. Then the total degree of p equals the degree of f in X_i for $1 \leq i \leq n$.*

Proof From the fact that f is a symmetric polynomial, one easily concludes that $\deg_{X_i}(f) = \deg_{X_j}(f)$ for $1 \leq i, j \leq n$. If $t = X_1^{\nu_1} \cdot \cdots \cdot X_n^{\nu_n}$ is a term, then we write

$$t(\sigma_1,\ldots,\sigma_n) \quad \text{for} \quad \sigma_1^{\nu_1} \cdot \cdots \cdot \sigma_n^{\nu_n}.$$

With this notation, we have

$$f = \sum_{t \in T(p)} a_t t(\sigma_1,\ldots,\sigma_n) \qquad (a_t \in K).$$

Using Lemma 10.87, we may conclude that

$$\mathrm{HT}(f) = \max\{\, \mathrm{HT}\big(t(\sigma_1,\ldots,\sigma_n)\big) \bigm| t \in T(p) \,\}. \tag{$*$}$$

Looking at the left-hand side, we see that by the choice of the term order as the lexicographical one, this term has the same degree in X_1 as the entire polynomial f. For the same reason, it also has maximal degree in X_1 among all terms in the set on the right-hand side. But Lemma 10.86 tells us that

$$\deg_{X_1}\Big(\mathrm{HT}\big(t(\sigma_1,\ldots,\sigma_n)\big)\Big) = \deg(t).$$

We have proved that the degree in X_1 of the term in $(*)$ equals both the degree in X_1 of f and the total degree of p. □

Corollary 10.91 *Let $f \in K[\underline{X}]$, and let g be the σ-normal form of f. If $f \neq g$, then the total degree of the polynomial $p \in K[\underline{X}]$ that satisfies $f - g = p(\sigma_1,\ldots,\sigma_n)$ is less than or equal to the degree of f in X_1.*

Proof We have $g < f$, and so $\deg_{X_1}(g) \leq \deg_{X_1}(f)$. It follows that

$$\deg_{X_1}(f - g) \leq \deg_{X_1}(f),$$

and the claim is now immediate from the proposition above. □

Exercise 10.92 Use σ-reduction to give an alternate proof of the corollary above.

Exercise 10.93 (i) Let $t = X_1^{\nu_1} \cdot \cdots \cdot X_n^{\nu_n}$ be a term, and set $f = \sigma_1^{\nu_1} \cdot \cdots \cdot \sigma_n^{\nu_n}$. Show that

$$\sum_{i=1}^{n} i \cdot \nu_i = \deg(f).$$

Conclude that $\deg(f) < n$ implies that $\nu_i = 0$ for $i > \deg(f)$. (Hint: You want to use Lemmas 10.86 and the fact that the σ_i are homogeneous.)

(ii) Let $p \in K[\underline{X}]$ and $f = p(\sigma_1, \ldots, \sigma_n)$. Give an example showing that in contrast to (i) above, it may happen that

$$\sum_{i=1}^{n} i \cdot \deg_{X_i}(p) > \deg(f).$$

(iii) Let p and f be as in (ii) above. Show that

$$\max\left\{ \sum_{i=1}^{n} i \cdot \nu_i \;\middle|\; X_1^{\nu_1} \cdot \cdots \cdot X_n^{\nu_n} \in T(p) \right\} = \deg(f)$$

if $p \neq 0$. In particular, $\deg(f) < n$ implies that $\deg_{X_i}(p) = 0$ for $i > \deg(f)$. (Hint: You need Lemma 10.87 and the arguments that you used to prove (i) above.)

(iv) Let $f \in K[\underline{X}]$ with σ-normal form $g \in K[\underline{X}]$, and suppose $f, g \neq 0$. Show that $\deg(g) \leq \deg(f)$.

(v) Let $0 \neq f \in K[\underline{X}]$ with σ-normal form $g \in K[\underline{X}]$, and let $p \in K[\underline{X}]$ such that $f - g = p(\sigma_1, \ldots, \sigma_n)$. Show that

$$\max\left\{ \sum_{i=1}^{n} i \cdot \nu_i \;\middle|\; X_1^{\nu_1} \cdot \cdots \cdot X_n^{\nu_n} \in T(p) \right\} \leq \deg(f)$$

if $p \neq 0$. In particular, $\deg(f) < n$ implies that $\deg_{X_i}(p) = 0$ for $i > \deg(f)$.

Notes

Gröbner bases over PID's and Euclidean domains were investigated by Zacharias (1978), Kandri-Rody and Kapur (1984, 1988), and Pan (1989) (cf. also Pauer, 1992). Our approach most closely resembles that of Pan (1989), except that we employ a different technique to ensure termination of the algorithm. A systematic investigation of possible refinements and applications of the theory in analogy to ordinary Gröbner basis theory does not seem to have been undertaken yet; cf., however, Möller (1988), Gianni, et al. (1988), and Mark (1992).

Homogeneous polynomials—also referred to as *forms*—seem to have enjoyed more attention in the early days of modern algebra than they do now;

however, they continue to be an important object of mathematical research. In Gröbner basis theory, homogeneity and homogenization are mainly relevant in connection with complexity problems (Lazard, 1983; Möller and Mora, 1984). Our treatment in Sections 10.2 and 10.3 is hardly more than a systematic and mathematically rigorous account of what seems to belong to the folklore of the theory. The main reason why we have included a discussion of homogeneity and homogenization here is that these techniques allow an elegant treatment of Gröbner bases for modules and of the tangent cone algorithm. The degree bounds for the representation of 1 in terms of a given basis of an improper ideal that we have quoted were proved by Fitchas and Galligo (1988a).

Gröbner basis theory for modules and its application to systems of linear equations over the polynomial ring is perhaps the most important generalization of ordinary Gröbner basis theory. References include Bayer (1982), Lazard (1983), Möller and Mora (1986a), and Furukawa et al. (1986); see also Billera and Rose (1989) for an interesting application.

The tangent cone belongs to the classical concepts in algebraic geometry. A method for the computation of its equations by means of standard bases is due to Mora (1982); the corresponding algorithm is therefore also called the *Mora algorithm.* Our approach is technically different because we use homogenization in order to reduce the theory of standard bases to ordinary Gröbner basis theory as much as possible (see Schwartz, 1988).

Our treatment of symmetric function consists of classical results (see, e.g., Weber, 1898 or van der Waerden, 1966) prepared in the more contemporary style of reduction relations. The definitive work on the subject is Noether (1916), where rings of polynomials that are invariant under finite groups of linear transformations of the variables are treated in an algorithmic manner. More recent results can be found in Lauer (1976), Giusti et al. (1988), Valibouze (1989), and Göbel (1992).

Appendix: Outlook on Advanced and Related Topics

As we have pointed out in the preface, the treatment of computational algebra offered by this book is not comprehensive. The purpose of this chapter is to round off the picture at least in and around the area that is at the core of this book, namely, the theory of Gröbner bases. Each section will briefly outline a problem, give a rough idea of existing results, and get the reader started on the relevant literature.

Complexity of Gröbner Basis Constructions

The complexity of a mathematical algorithm can be measured in different ways. The most straightforward approach is to implement the algorithm, run it on some computer, and measure the time and space required for getting the output on a number of examples. This method should not be belittled. If one duly takes into account such factors as the speed of the particular computer that one uses, the details of the implementation such as the choice of datatypes to represent the mathematical objects, and the fact that the size of the examples is necessarily within a limited range, then this method will allow precisely the kind of evaluation of the quality of the algorithm that matters in practice.

The purpose of *complexity theory* is to provide a more precise concept of measurability for the comparison of different solutions to an algorithmic problem. The first step is to give a rigorous definition of a mathematically idealized computing machine on which all computations are to be performed, such as a Turing machine or a register machine. The mathematical objects in question are then coded into words over a finite alphabet. The computing machine is capable of holding the letters of the alphabet in individual *cells*, with at most one letter in each cell. Each configuration of letters in cells is a *state* of the machine, and the machine modifies these states in well-defined single steps. This way, the number of steps and the number of cells required to run a particular algorithm on a particular input becomes well-defined. It is clear that the manner in which mathematical objects are represented by words may have considerable influence on these numbers.

On the basis of this rigorous measurement of the time and space consumption of an application of an algorithm to a specific input, one may now ask for a way to evaluate the efficiency of the algorithm as such. One way to achieve this is to ask for upper bounds for the number of steps (*time complexity*) and the number of cells (*space complexity*) as functions of the size (number of cells required after coding) of the input. In case one obtains least upper bounds, this is also called the *asymptotic worst case complexity.* Apart from the fact that these bounds, by their nature, yield information only about the worst case, they also frequently contain multiplicative constants that are only insufficiently determined. It is therefore desirable to also obtain *average complexities* in the probabilistic sense. This requires of course a realistic probability measure of the input space. Examples can be found in Knuth (1969). For mathematically involved problems, the analysis of the propagation of probabilities that this approach calls for has been achieved only in rare cases (see, e.g., Borgwardt, 1980).

The type of complexity measure that we have described thus far is, for rather obvious reasons, also called the *bit complexity* of the algorithm. The mathematics that is required to to deal with this type of complexity tends to be very hard. One must therefore often be content with weaker results such as an upper bound for the size of the output as a function of the size of the input.

Turning to Gröbner bases in particular, let us first note that a natural system of parameters to measure the size of a finite set F of polynomials is given by the number n of indeterminates, the number $|F|$ of polynomials in the set, the maximal total degree $\operatorname{maxdeg}(F)$ of the polynomials in F, and the maximal size $\operatorname{maxcoeff}(F)$ of the coefficients of the polynomials under a given coding.

Now let us look at the output G of the Buchberger algorithm as a function of the size of the input F. In addition to the parameters explained above, we also consider the maximal degree D and the maximal size S of the coefficients of any polynomial occurring during computation. Then the following hold. Firstly, D as well as $|G|$ are bounded by recursive functions of n, $|F|$, and $\operatorname{maxdeg}(F)$. These functions are independent of the ground field, the term order and the size of the input coefficients. Secondly, the maximal size M of any coefficient appearing in the construction is bounded by a recursive function of n, $|F|$, $\operatorname{maxdeg}(F)$, and $\operatorname{maxcoeff}(F)$, again independently of the term order. If all coefficients are represented as rational expressions in the input coefficients, then this bound is also independent of the ground field. The proof of all this employs a coding of Gröbner basis constructions in formulas of first-order logic together with an application of the compactness theorem for first-order logic (see Weispfenning, 1986). The bounds obtained by such general principles are of course in no way explicit; they do, however, naturally extend to the construction of universal Gröbner bases and comprehensive Gröbner bases (see the next two sections of this chapter), and also to Gröbner bases over commutative regular rings (see Weispfenning, 1987b) and non-commutative polynomial rings of solvable type over fields (see Kandri-Rody and Weispfenning, 1990). The arguments also show that the computing time (i.e., the number of steps) required for a Gröbner basis construction is bounded by a recursive function of n, $|F|$, and $\mathrm{maxdeg}(F)$ when an arithmetic operation and an equality test in the ground field and a comparison of terms in the term order are counted as one step each. (Interestingly, this is not true for Gröbner bases over PID's.) When computations in the ground field are performed in polynomial time, then for fixed n, $|F|$, and $\mathrm{maxdeg}(F)$, the Gröbner basis G of $\mathrm{Id}(F)$ can be constructed in polynomial time in $\mathrm{maxcoeff}(F)$, i.e., the number of steps is bounded by a polynomial function of $\mathrm{maxcoeff}(F)$.

The search for explicit bounds concerning the size of Gröbner bases requires a much more refined algebraic and combinatorial analysis. For the case of two variables the following upper bounds on the degrees and the number of the polynomials in a reduced Gröbner basis G as functions of the size of the input F have been obtained in Buchberger (1983) and Lazard (1983). Let $\mathrm{mindeg}(F)$ denote the minimal degree of a polynomial in F; then $|G| \le \mathrm{mindeg}(F) + 1$, and for a total degree term order, $D \le 2 \cdot \mathrm{maxdeg}(F) - 1$, where D is as defined above. Both bounds are worst case optimal, i.e., they are least upper bounds.

For the case of three variables, Winkler (1984) and Möller and Mora (1984) provide the following bounds:

$$D \le (8 \cdot \mathrm{maxdeg}(F) + 1) \cdot 2^{\mathrm{mindeg}(F)}$$

for the total degree-lexicographical term order, and, under a conjecture of Lazard, $\mathrm{maxdeg}(G) \le (\mathrm{maxdeg}(F))^2$ for every total degree term order. Möller and Mora (1984) and Bayer (1982) also provide upper bounds for an arbitrary number of variables under further assumptions on the ideal considered, in particular information on the Hilbert polynomial of the ideal (cf. Kondrat'eva and Pankrat'ev, 1987) or an H-basis of the ideal (cf. Macaulay, 1916). Möller and Mora (1984) show that $\mathrm{maxdeg}(G)$ is bounded by a polynomial in $\mathrm{maxdeg}(F)$ for fixed number n of variables; under Lazard's conjecture, they give the explicit bound

$$\mathrm{maxdeg}(G) \le \big((n+1) \cdot (\mathrm{maxdeg}(F) + 1)\big)^{(n+1) \cdot 2^{\dim(\mathrm{Id}(F))+1}}. \qquad (*)$$

Giusti (1984) shows that for fixed number n of variables, there exists a polynomial f of degree a^n with $a \le \sqrt{3}$ such that

$$\mathrm{maxdeg}(G) \le f\big(\mathrm{maxdeg}(F)\big). \qquad (**)$$

Under the additional assumption that $\dim(I) \le 1$ and the term order is compatible with some weighted degree, Giusti (1985) shows that

$$\mathrm{maxdeg}(G) \le (n+1) \cdot \big(\mathrm{maxdeg}(F)\big)^n.$$

The upper bounds discussed above may not be least upper bounds. Concerning lower bounds for these upper bounds, the following is known. Möller and Mora (1984) and Huynh (1986) show, in essence, that the doubly exponential behavior in the number of variables that is exhibited in (*) and (**) cannot be improved. Their proof is based on a refinement of arguments of Mayr and Meyer (1982), where it is shown that the ideal membership problem in polynomial rings over arbitrary ground fields is exponential-space hard. This means that any algorithm that decides ideal membership requires space that grows exponentially in the input size. Huynh (1986a) shows that for fixed number of indeterminates, the ideal membership problem is NP-hard.

Finally, we mention that a bound on the maximal degree of a reduced Gröbner basis G automatically yields a bound on $|G|$, simply because the head terms of the elements of G are pairwise different.

Term Orders and Universal Gröbner Bases

The construction of a Gröbner basis from a given finite set F of polynomials in $K[\underline{X}] = K[X_1, \ldots, X_n]$ depends on the choice of the term order $\leq$ on the set T of terms in $X_1, \ldots, X_n$. A reduced Gröbner basis of $\mathrm{Id}(F)$ is in fact uniquely determined by F and the term order. This raises the following questions.

1. How can the possible term orders on T be characterized?
2. How many different reduced Gröbner bases can an ideal have?
3. Do ideals have bases that are Gröbner bases simultaneously w.r.t. every term order, and if so, can these be constructed?
4. Given a finite set of polynomials, is it possible to predict which term orders will yield fast or slow Gröbner basis computations?

It was shown in Lemma 4.64 how certain n-tuples of univariate polynomials with real coefficients induce admissible orders on $\mathbb{N}^n$ and thus, via the natural correspondence of Lemma 5.4, term orders on T. It is proved in Weispfenning (1987)—and, in a different setting, in Robbiano (1985)—that every admissible order on $\mathbb{N}^n$ and hence every term order arises in this way. This provides a satisfactory answer to the first question.

For an answer to the second and third questions, one first considers a modification of the first one. Suppose a finite set S of terms is given, and one wishes to find all orders on S that are restrictions of term orders on T. Since S is assumed to be finite, it is clear that there can be only finitely many different such restricted term orders on S. It can be shown that every such restriction is induced by an n-tuple of rational numbers in the same manner as global term orders are induced by n-tuples of univariate polynomials.

Moreover, it is possible to compute from S finitely many elements of $\mathbb{Q}^n$ such that these induce precisely all the different restrictions of term orders on S. Proofs of these facts can be found in Mora and Robbiano (1988), Ritter and Weispfenning (1992), and Weispfenning (1987).

The following construction shows that the answer to the third question is positive (cf. Weispfenning, 1987a). Given a finite subset F of $K[\underline{X}]$, compute all possible restrictions of term orders to the finite set $T(F)$ of all terms occuring in elements of F. W.r.t. each of these, form an S-polynomial of a pair of elements of F, and compute its normal form modulo F. (Note that for the computation of a normal form modulo F, the only information that is needed is what the head terms of the elements of F are.) Form a finite list of supersets of F, one for each restricted term order on F, by adding the respecective normal form of the S-polynomial to F unless the latter equals zero. For each set P in this list, compute the finitely many restricted term orders on the finite set $T(P)$ that extend the restricted term order on $T(F)$ which gave rise to P. Then do with P and these new term orders as before with F, making sure not to choose the same critical pair again for the formation of the S-polynomial. Continuing in this way as long as possible, one obtains a finitely branching tree, and each branch in this tree represents a possible course of the Buchberger algorithm w.r.t. at least one term order. From the fact that the Buchberger algorithm always terminates together with König's lemma (Theorem 4.55), we may conclude that the tree is finite. It is now easy to see that each leaf of the tree is a Gröbner basis of $\mathrm{Id}(F)$, that for each term order $\leq$ on T, there exists a leaf which is a Gröbner basis of $\mathrm{Id}(F)$ w.r.t. $\leq$, and that the union of the leaves is a *universal Gröbner basis* of $\mathrm{Id}(F)$, i.e., a simultaneous Gröbner basis of $\mathrm{Id}(F)$ w.r.t. every term order. A slight modification of the construction shows that every ideal of $K[\underline{X}]$ has only finitely many reduced Gröbner bases, and that these may be computed from any given ideal basis. The mathematical fact behind these results is the compactness of the space of term orders on T w.r.t. to a certain topology; this can also be used to obtain an abstract proof of the existence of universal Gröbner bases (Schwartz, 1988).

Answers to the fourth question are still more or less at a heuristic level. Perhaps the best result for the time being is a dynamic version of the Buchberger algorithm which periodically modifies the term order during computation (Sturmfels, 1989). More precisely, it uses the Hilbert function of the head term ideal in order to determine the term order relative to which the current basis is "closest" to being a Gröbner basis.

Comprehensive Gröbner Bases

Classical Gröbner basis theory solves certain problems concerning ideals—given by finite bases—in multivariate polynomial rings over fields. There are a number of circumstances under which theses same problems arise for ideal bases containing parameters, in the following sense. Suppose F is a finite set of polynomials in the variables $X_1, \ldots, X_n$ over a coefficient ring which is itself a polynomial ring over a domain R in the variables $U_1, \ldots, U_n$, i.e.,

$$F \subseteq R[U_1, \ldots, U_r][X_1, \ldots, X_n] = R[\underline{U}][\underline{X}].$$

Given a term order $\leq$ on $T(\underline{X})$, one may now ask if there exists a finite subset G of $R[\underline{U}][\underline{X}]$ such that for every homomorphism from $R[\underline{U}]$ to some field K, the image in $K[\underline{X}]$ of G under the induced homomorphism is a Gröbner basis w.r.t. $\leq$ in $K[\underline{X}]$ and generates the same ideal as the image of F under the same homomorphism. Here, one would naturally call $U_1, \ldots, U_n$ *parameters* and $X_1, \ldots, X_n$ the *main variables.* A set G as described above is called a *comprehensive Gröbner basis* for F. A comprehensive Gröbner basis for F is thus a subset of $R[\underline{U}][\underline{X}]$ which, under every "specialization in a field" of the coefficients, becomes a Gröbner basis of the ideal generated by the same specialization of F. The simplest, natural case would be the one where R is already a field, and "specialization" means nothing but substitution for the parameters in some extension field of R; the more general situation is the one where R is a domain which is not a field, and "specialization" actually involves homomorphic mapping of elements of R, e.g., from $\mathbb{Z}$ to $\mathbb{Z}/p\mathbb{Z}$ for some prime p.

To see where the non-triviality of the problem lies, one must first understand that the ordinary Gröbner basis property is not in general preserved under substitution for one or several of the variables. To see this, consider the subset $G = \{X+1, UY+X\}$ of $\mathbb{Q}[U, X, Y]$. Then G is clearly a Gröbner basis in $\mathbb{Q}[U, X, Y]$ w.r.t. every term order satisfying $X < Y$. Now let us view U as a parameter. If we take for our specialization the natural inclusion of $\mathbb{Q}[U]$ in $\mathbb{Q}(U)$, then G is still a Gröbner basis in $\mathbb{Q}(U)[X, Y]$ for term orders with $X < Y$. The same holds if we specialize to $\mathbb{Q}$ by setting $U = 1$. However, if we set $U = 0$, then the Gröbner basis property is lost.

The example shows that any attempt to construct a comprehensive Gröbner basis must take into account different cases according to whether or not certain coefficients vanish, simply because the vanishing or non-vanishing of coefficents determines what the head term of a polynomial will be. Rather surprisingly, it is possible to construct—and thus to actually compute in case R is computable—a comprehensive Gröbner basis from any given finite subset of $R[\underline{U}][\underline{X}]$. Simplifying only slightly, the construction can be described as follows. For every polynomial in F, one considers all possibilities of what its head term could be, due to the vanishing of higher coefficients. One then considers the combined possibilities for all of F and starts out a Buchberger algorithm for each possibility, forming and reducing to normal form the S-polynomial of some critical pair. This is possible on the basis of knowing what the head terms are. After this first step has been performed, the situation is similar to the one after the first step of the construction of the universal Gröbner basis of the previous section: if the normal form of the S-polynomial does not vanish, then this gives rise to a further branching of the tree of Buchberger algorithms, because one has to consider all possibilities of what the head term of the new arrival could be. Continuing the process in this way, one obtains a finitely branching tree. The existence of an infinite branch of this tree would contradict Dickson's lemma, because *none* of the terms of the polynomial that is being added when branching at a node is divisible by any term that has already been declared a head term in that branch. The König tree lemma allows us to conclude that the construction must terminate after finitely many steps. It is perhaps noteworthy that this is true despite the fact that a large number of branches may be "virtual" Buchberger algorithms that are really nonsensical, because the corresponding choice of the head terms is based on an assumption like "$U = 0$ and $U^2 \neq 0$."

The point of the above construction is that the union G of the leaves of the tree is supposed to be a comprehensive Gröbner basis for the input set. For this to happen, we must make two minor modifications to the procedure. First of all, polynomial reduction as we have defined it requires division by certain head coefficients, meaning that we were really passing to $Q_{R[\underline{U}]}[\underline{X}]$. This is not acceptable because that way, we would obtain coefficients that may become undefined under certain specializations. However, the problem is easily amended by using a modified, denominator-free version of polynomial reduction, where one multiplies the polynomial that is being reduced by the head coefficient of the one that it is being reduced by. The second point is that by declaring a certain term of a polynomial to be the head term and dropping all higher monomials on the basis of the case assumption that their coefficients vanish, we will in general be leaving the ideal generated by F in $R[\underline{U}][\underline{X}]$. As a consequence, the ideal generated by G under a specialization may be larger than the one generated by F under that specialization. This problem is overcome by the following "schizophrenic" attitude. Whenever an S-polynomial is formed or a reduction step is performed under a certain case assumption on head terms, then in each polynomial, the monomials above the presumed head monomial are carried along as dummies in the computation despite the fact that they are assumed not to be present at all. This does clearly not affect the termination of the procedure. With these two modifications to the procedure, it is indeed true that the union of the leaves of the tree that was described above is a comprehensive Gröbner basis for the input set. If one keeps track of the assumptions on the vanishing of coefficients that were made during computation, then each leaf of the tree provides a set of conditions together with a set of polynomials such that this set is a Gröbner basis of the input ideal under every specialization that meets these conditions. The set of all leaves of the tree is then called a *Gröbner system* for the input set.

Comprehensive Gröbner bases can be used to solve parametric versions of virtually all the problems to which ordinary Gröbner bases provide solutions. A typical problem of this kind is the question of properness of an ideal of $K[\underline{X}]$, which we have seen to be equivalent to the existence of a zero of the ideal in the algebraic closure $\overline{K}$ of the ground field K. The parametric version of this problem is known as the *elimination problem:* given a finite subset F of $K[\underline{U}][\underline{X}]$, for which values in $\overline{K}$ of the parameters $\underline{U}$ do the polynomials in F have a common zero in $\overline{K}^n$? In other words, we wish to decide for which values of the parameters the ideal generated by F is proper. Using comprehensive Gröbner bases, the problem can be solved as follows. Let G be a comprehensive Gröbner basis for F with respect to some fixed term order. Then for any specialization $\sigma : K[\underline{U}] \longrightarrow L$ into some algebraically closed field L, the polynomial system $\sigma(F)$ obtained from F by specializing the coefficients according to σ has a common zero in L^n iff the following condition holds for every $g \in G$: whenever $\sigma(a) = 0$ for all monomials $a \cdot t$ of g with $t \neq 1$, then $\sigma(b) = 0$ for the coefficient b of 1 in g. This can also be viewed geometrically as the computation of projections of varieties. (Recall from Section 7.6 that the problem of finding the smallest variety containing such a projection is much easier: it requires no more than a single Gröbner basis computation.)

Classically, the elimination problem is solved by means of resultants together with the technique of introducing Kronecker variables (see van der Waerden, 1931). The approach via comprehensive Gröbner bases is more direct and—at least in the case of several main variables—computationally easier. In the case where F consists of two univariate polynomials, the resultant is always part of a comprehensive Gröbner basis for F. The theory of comprehensive Gröbner bases is due to Weispfenning (1992); see also Sit (1991, 1992) on the subject of solving parametric systems.

Gröbner Bases and Automatic Theorem Proving

Automatic theorem proving in mathematics is concerned with the problem of designing algorithms that prove or disprove conjectures about classes of mathematical structures. More precisely, such an algorithm is to decide whether or not formulas of a certain specified type in some first-order language L hold in a specified class of L-structures. Yet in other words, one is looking for a *decision method* for the validity of a certain type of formula in a class of structures. Examples are the word problems that were treated in the theorems of Section 6.4. There, the decision could be achieved by means of Gröbner basis computations involving sets of polynomials that were implicit in the given formula.

As an example, let us recall Theorem 6.59. Here, it is shown that for a given field K and polynomial expressions $f, g_1, \ldots, g_m$ in $x_1, \ldots, x_n$ over

K, the validity of the formula

$$\forall x_1 \cdots \forall x_n \left(\bigwedge_{i=1}^{m} g_i(\underline{x}) = 0 \longrightarrow f(\underline{x}) = 0 \right) \qquad (*)$$

in the class of all extension fields of K is equivalent to membership of f in the radical of the ideal generated by the g_i, a condition that can be algorithmically decided according to Corollary 6.41. The Hilbert Nullstellensatz tells us that membership of f in the radical of the ideal generated by the g_i is in fact equivalent to the validity of (*) in any one algebraically closed extension field of K, e.g., in $\mathbb{C}$ in case $K = \mathbb{Q}$. This decision method has interesting applications in elementary real geometry. There, one is typically dealing with configurations consisting of finitely many lines, circles, ellipses, hyperplanes, spheres, and the like. Given a conjecture concerning such a configuration, one can often, using Cartesian coordinates, find a sentence of the type (*) above with rational parameters such that the conjecture is true if and only if the sentence holds true in $\mathbb{R}$. What we have, according to the above, is an algorithm to decide whether or not the statement holds in $\mathbb{C}$. Since $\mathbb{C}$ is an extension field of $\mathbb{R}$ and (*) is a universal statement, the decision method thus provides a *sufficient* condition for the conjecture to hold in $\mathbb{R}$. Rather surprisingly, this sufficient condition is satisfied for many natural and classical geometric problems.

One way to use the method described above is to look for valid geometrical theorems by trial and error, adding or removing hypotheses such as non-triviality conditions until a true statement has been found. A more ambitious goal is to algorithmically generate additional conditions that turn an invalid statement into a valid one. This can be achieved by a method due to Wu (1984, 1986) that is based on a variant of Tarski's quantifier elimination method for algebraically closed fields, which will be discussed below. Wu's method is based on the concept of *characteristic sets* which is originally due to Ritt (1950) (see the section on characteristic sets below). A different way of finding missing hypotheses for a conjectured theorem of the form (*) is provided by the computation of comprehensive Gröbner bases. One views some of the variables of the problem as parameters and then solves the corresponding elimination problem as discussed in the previous section.

Let us now return to the general decision problem for the validity of formulas of a first-order language L in a class Σ of L-structures. Suppose that we have an effective *quantifier elimination* procedure, i.e., an algorithm that computes, for every formula φ of L, a quantifier-free formula ψ such that φ and ψ are equivalent in Σ. Due to the absence of quantifiers, it is often possible to decide the validity of ψ in Σ by inspection. This yields a solution to the decision problem w.r.t. Σ for *all* formulas of L. If, for example, L is the language of rings, then quantifier-free formulas are disjunctions of conjunctions of simple formulas such as $1 = 1$, $1 \neq 1$, $1 = 0$,

$x + x = 0$. The validity in a given class of rings such as the class of fields, or the class of fields of a given characteristic, is then clearly decidable.

Unfortunately, the possibility of quantifier elimination is not exactly a frequent phenomenon (see Weispfenning, 1983, for an overview). In the 1930s, Tarski has given explicit but highly complex quantifier elimination methods for the class of algebraically closed fields and the class of real closed fields. Comprehensive Gröbner bases provide a quantifier elimination procedure for the algebraically closed case that is much more efficient than Tarski's original algorithm. On the basis of elementary logical equivalences, one easily sees that it suffices to treat formulas of the form

$$\exists x_1 \cdots \exists x_n (f_1 = 0 \wedge \cdots \wedge f_r = 0 \wedge g_1 \neq 0 \wedge \cdots \wedge g_s \neq 0),$$

where the f_i and g_i are polynomial expressions over $\mathbb{Q}$ in the variables x_1, $\ldots$, x_n and possibly further free variables. Setting $g = g_1 \cdot \cdots \cdot g_s$, this formula is equivalent to

$$\exists x_1 \cdots \exists x_n (f_1 = 0 \wedge \cdots \wedge f_r = 0 \wedge g \neq 0).$$

Introducing the new variable z, the latter formula becomes equivalent to

$$\exists x_1 \cdots \exists x_n \exists z (f_1 = 0 \wedge \cdots \wedge f_m = 0 \wedge (z \cdot g - 1) = 0).$$

This is an instance of the elimination problem as described in the previous section which can be dealt with using comprehensive Gröbner bases.

Characteristic Sets and Wu–Ritt Reduction

The concept of a characteristic set was originally introduced by Ritt (1950) as a tool for studying solutions of algebraic differential equations (see the section on Gröbner bases and differential algebra below). Wu (1984, 1986) transferred the algorithmic aspects of Ritt's method to ordinary polynomial rings with the intent of finding an effective method for automatic theorem proving in elementary geometry. His approach has since been developed extensively; see, e.g., Chou (1988). The problem studied by Wu is essentially the same as the one stated at the beginning of the previous section: given polynomial expressions f, f_1, $\ldots$, f_m in the variables x_1, $\ldots$, x_n over $\mathbb{Q}$, decide the validity of the formula

$$\forall x_1 \cdots \forall x_n \left(\bigwedge_{i=1}^{m} f_i(\underline{x}) = 0 \longrightarrow f(\underline{x}) = 0 \right) \qquad (*)$$

in $\mathbb{R}$. Just like the Gröbner basis approach, Wu's method actually decides the validity of $(*)$ in $\mathbb{C}$ and thus provides no more than a sufficient criterion for its validity in $\mathbb{R}$; as we have noted before, however, this sufficient criterion is satisfied surprisingly often for theorems in geometry.

The basic features of Wu's theory are as follows. Let K be a field and $K[\underline{X}] = K[X_1, \ldots, X_n]$. If $0 \neq f \in K[\underline{X}]$, then the *class* of f is defined as the maximal k such that the degree of f in X_k is not 0. The head coefficient of f viewed as a polynomial in X_k is then called the *initial* of f. If f, g are non-zero polynomials in $K[\underline{X}]$ and g has class k and initial p, then one may, by a rather obvious process called *pseudo-division*, find a polynomial r in $K[\underline{X}]$ that is either zero or whose degree in X_k is less than that of g such that $r = p^d f - qg$, where $q \in K[\underline{X}]$ and $d = \max\{0, \deg_{X_k}(f) - \deg_{X_k}(g) + 1\}$. Taking such a pseudo-division as a single reduction step, it should be clear how one defines *Wu–Ritt reduction* of a polynomial modulo a set of polynomials. It is easy to see that the pseudoremainder of a single reduction step is less than the dividend in the quasi-order on $K[\underline{X}]$ that is induced by the inverse lexicographical term order; it follows that Wu–Ritt reduction is noetherian.

A finite sequence $(f_1, \ldots, f_m)$ is called an *ascending set* if the corresponding sequence of the classes of the f_i is strictly increasing and f_i is Wu–Ritt-reduced modulo $\{f_1, \ldots, f_{i-1}\}$. One may now define a quasi-order on the set of ascending sets by comparing the entries of two sequences from left to right according to the quasi-order on $K[\underline{X}]$ that is induced by the inverse lexicographical term order and letting the first strict inequality decide. If no such inequality is encountered, then the ascending set of greater length, if any, is declared less (sic!) than the other one. It is not hard to see that this quasi-order is well-founded. It follows that among the ascending sets that can be made up from the elements of a given subset F of $K[\underline{X}]$, there must be a minimal one; such a minimal ascending set is then called a *characteristic set* of F. If F is finite, then one may actually construct a characteristic set of F by starting with an element of minimal class and then add, as long as this is possible, minimal elements of higher class than and reduced modulo the previous ones.

The centerpiece of the theory is the following completion procedure. Starting with a finite subset F of $K[\underline{X}]$, one constructs a characteristic set of F and then enlarges F by all Wu–Ritt normal forms—modulo the set of polynomials occurring in the characteristic set—of elements of F. The resulting enlarged set is clearly again finite, and so the procedure can be repeated. It is not hard to see that the next characteristic set is less than the previous one in the quasi-order defined above, and one concludes that the process must terminate after finitely many iterations. If the resulting set is G and the last characteristic set is B, then clearly G generates the same ideal as F, and every element of G Wu–Ritt reduces to zero modulo the underlying set of B.

The idea now is to apply this completion procedure to $F = \{f_1, \ldots, f_m\}$ and use Wu–Ritt reduction to decide the validity of $(*)$. Let G and B be as in the previous paragraph. If it is found that f Wu–Ritt reduces to zero modulo the underlying set P of B, then all we may conclude is that $mf \in \mathrm{Id}(F)$ for some power product of initials of elements of P, and this does rather obviously not allow us to draw the desired conclusion that f vanishes on the variety of F. This problem is overcome by a further extension of the completion procedure which corresponds to a decomposition of the variety of G (which of course equals that of F). We will define a finitely branching tree of pairs (H, C), where H is a finite subset of $K[\underline{X}]$, C is a characteristic set of H, and every $h \in H$ Wu–Ritt reduces to zero modulo the underlying set of C (i.e., H cannot be further completed in the sense of the previous paragraph). The root of the tree is the result (G, B) of the completion procedure of the previous paragraph applied to the original set F. To obtain the first level of the tree, we set, for each $g \in B$,

$$G_g = (G \setminus \{g\}) \cup \{p, \mathrm{red}(g)\},$$

where p is the initial of g and $\mathrm{red}(g)$ is the reductum of g as a polynomial in X_k where k is its class. To each G_g, we apply the aforementioned completion procedure, and we take for the first level of our tree the set of all pairs (H, C) thus obtained. It is easy to see that now the variety of G equals the union of the varieties $V(H)$ of H, where H runs through the first level, and the set $V(B) \setminus V(\{p\})$, where p is the product of the initials of the entries of B. Moreover, each characteristic set C in the first level is less than B in the quasi-order on ascending sets because if nothing else, the entry g of B that gave rise to C can be replaced by its reductum. The higher levels of the tree are now obtained in the obvious way by doing with each leaf as we just did with (G, B). We cannot have an infinite branch because of the well-foundedness of the quasi-order on ascending sets. It follows that the tree has finitely many elements, say $(H_1, B_1), \ldots, (H_k, B_k)$. It is now easy to prove that

$$V(F) = V(G) = \bigcup_{i=1}^{k} \left(V(B_i) \setminus V(\{p_i\})\right),$$

where p_i is the product of the initials of the entries of B_i. A sufficient condition for $f \in K[\underline{X}]$ to vanish on $V(F)$ is then obviously given by "f Wu–Ritt reduces to 0 modulo B_i for $1 \leq i \leq k$." This condition is not in general a necessary one, as exemplified by the trivial example $F = \{X^2\}$, where the polynomial X will not be recognized as vanishing on $V(F)$.

Theoretically, the method can be pushed to the point where it does provide a decision method for the validity of the statement $(*)$ in all extension fields of $\mathbb{Q}$. By intertwining the splitting procedure described above with successive factorization of polynomials in ascending sets over suitable extension fields of $\mathbb{Q}$, one can arrive at a a finite system of characteristic sets satisfying certain irreducibility conditions. These *irreducible* characteristic sets then provide a necessary and sufficient condition for $f \in K[\underline{X}]$ to vanish on $V(F)$. In practice, however, this last step of the method has turned out to be extremely time and space consuming.

As a sufficient criterion for the validity of certain theorems in elementary real geometry, the Wu–Ritt method has been tested on some 150 examples in (see Chou, 1988). When successful, it tends to perform slightly better than Gröbner basis methods.

An interesting connection between characteristic sets and Gröbner bases has been noted in Kandri-Rody (1984): if G is a reduced Gröbner basis in $K[\underline{X}]$ w.r.t. the inverse lexicographical term order, then every characteristic set of G is in fact a characteristic set of $\mathrm{Id}(G)$.

Term Rewriting

In this section, the word "term" is used in the sense of its definition in logic and model theory, where it means a well-formed formal expression involving variables and function symbols of a first-order language. If, for example, L is the language of rings consisting of the constants 0 and 1 and the binary function symbols "$+$," "$-$," and "$\cdot$," and we use infix notation with the usual rules of preference, then $(x \cdot y - 1) \cdot ((1 + x) + 0 \cdot y)$ is an L-term.

The problem of term rewriting has been posed in a number of variants; see, e.g., Jantzen (1988) for a comprehensive treatment. In its most basic form, the problem is as follows. Suppose we are given a first-order language L and finitely many equations between L-terms, say $u_i = v_i$ for $1 \le i \le m$. What one is looking for is a finite set of rules for rewriting L-terms such that for each L-term s and each rule, it can be decided whether the rule is applicable to s, and if so, application of the rule to s produces a term t such that the implication

$$\forall x_1 \cdots \forall x_n \Big(\bigwedge_{i=1}^{m} u_i = v_i \Big) \longrightarrow \forall x_1 \cdots \forall x_n (s = t)$$

holds in the class of all L-structures, where the variables occurring in s, t, and the u_i and v_i are among $x_1, \ldots, x_n$. Equivalently, one could say that the equation $s = t$ holds in the *equational class* of all L-structures in which the equations $u_i = v_i$ hold for $1 \le i \le m$.

The point of doing all this is that one hopes to find a set of rules such that the relation

$$s \longrightarrow t \quad \text{iff} \quad t \text{ is the result of an application of a rule to } s$$

on the set of L-terms satisfies the following two conditions.

(i) $\longrightarrow$ is a noetherian, confluent reduction relation.

(ii) $\longrightarrow$ is adequate for the equational class Σ in question in the sense that $s \overset{*}{\longleftrightarrow} t$ if and only if $s = t$ holds in Σ, i.e., is valid in every structure in Σ.

If this is the case, then one may, according to Newman's lemma, compute unique normal forms of L-terms w.r.t. $\longrightarrow$. Moreover, one can then decide whether $s = t$ holds in Σ: this will be the case if and only if the normal forms of s and t agree. One has thus found a decision method in the sense of the section on automaic theorem proving for formulas of the form $s = t$ w.r.t. Σ.

In the section on automatic theorem proving, we have focused on Gröbner basis methods as a tool for solving decision problems. The approach via term rewriting systems, which will be further described below, is not really an application of Gröbner bases; what makes it interesting in this context is the fact that it uses a critical pair completion procedure which makes it resemble the Buchberger algorithm. As we have already pointed out in the notes to Chapter 5, the concept of critical pair completion was actually found independently by Buchberger (1965) for his Gröbner basis algorithm and by Knuth and Bendix (1970) in connection with term rewriting.

More precisely, the Knuth–Bendix approach proceeds as follows. One starts with an "orientation" of the given equations, i.e., a set of pairs

$$R = \{\, (u_i, v_i) \mid 1 \leq i \leq m \,\},$$

the set of *rewrite rules*, such that $\{\, u_i = v_i \mid 1 \leq i \leq m \,\}$ is the set of given equations that define the class Σ. Then R induces a relation $\longrightarrow$ on the set of L-terms as follows: $s \longrightarrow t$ iff there exists a rule $(u, v) \in R$ and a substitution σ of terms for variables such that the term $\overline{\sigma}(u)$ obtained from u by the substitution σ has some occurrence in s, and t is obtained from s by replacing this specific occurrence of $\overline{\sigma}(u)$ in s by $\overline{\sigma}(v)$. Unfortunately, $\longrightarrow$ will not in general be a noetherian reduction relation: it is easy to see that for example, the presence of an equation expressing commutativity of a function will prevent $\longrightarrow$ from being strictly antisymmetric. There are, however, many interesting instances where R is such that the relation R is indeed a noetherian reduction relation which is adequate for the class Σ. The hard part is to ensure confluence of $\longrightarrow$. This is where the concept of critical pair completion comes in.

Roughly speaking, a critical pair is a pair $(s \longrightarrow t_1, s \longrightarrow t_2)$ of reductions given by specializations of two rules via two substitutions that act on *nested* subterms s_1 and s_2 of s. One shows that if for every such ciritcal pair, one has $t_1 \downarrow t_2$, then the reduction is locally confluent and thus confluent by Newman's lemma. In analogy to the Buchberger algorithm, the strategy to achieve this situation is to complete the set R of rewrite rules as follows. For every critical pair $(s \longrightarrow t_1, s \longrightarrow t_2)$, one computes normal forms t_1^* of t_1 and t_2^* of t_2. If these do not agree, then one forces $t_1 \downarrow t_2$ by adding to R the rule (t_1^*, t_2^*) or (t_2^*, t_1^*). There are two obvious problems to this procedure. Firstly, the new rules have to be selected in such a way that the induced relation $\longrightarrow$ is still a noetherian reduction relation. Secondly, one must be able to prove termination. If and how this can be achieved will of course depend on the class Σ that one is talking about; in most cases, the analogy to the Buchberger algorithm will come to an end at this point. For a more detailed discussion of the relationship between Knuth–Bendix algorithm and the Buchberger algorithm, we refer the reader to Winkler (1984, 1989).

Of the many variants of term rewriting that one encounters in mathematics and computer science, we mention a more general version of the problem that we have just discussed: here, one modifies the definition of adequacy of $\longrightarrow$ in such a way that $s \overset{*}{\longleftrightarrow} t$ is equivalent not to the validity of $s = t$, but of

$$\forall x_1 \cdots \forall x_n \left(\bigwedge_{i=1}^{r} s_i = t_i \longrightarrow s = t \right)$$

in the equational class Σ, where the s_i and t_i are L-terms. It is easy to see that this amounts to solving the word problem for Σ by means of term rewriting and normal forms.

Standard Bases in Power Series Rings

At the end of Section 2.1, we explained how one may define a formal power series over a ring R as a "polynomial with infinitely many monomials." We also saw how addition, subtraction and multiplication of power series can be defined in close analogy to the operations in the polynomial ring, and that the set of all formal power series over R forms a ring under these operations. This ring is commonly denoted by $R[[X_1, \ldots, X_n]]$, or $R[[\underline{X}]]$ for short. We have also mentioned in the Notes to Chapter 5 that there exists an analogy to Gröbner bases in power series rings over fields, namely, the concept of *standard bases.* Standard bases in power series rings are actually more directly analogous to standard bases in polynomial rings as described in Section 10.6; for easier reference, however, we will describe standard bases in power series rings in relationship to Gröbner bases in polynomial rings. Standard bases were introduced by Hironaka (1964) as a tool in connection with the resolution of singularities in analytic spaces. Hironaka's work took place independently of Buchberger's development of Gröbner basis theory; it was not until the seventies that the analogy was brought to light (Briançon, 1973; Galligo, 1974, 1979; Robbiano, 1986; Mora 1988a; Becker 1990, 1990a).

There cannot of course be a plain analogue to Gröbner basis theory in power series rings: since a power series may have infinitely many terms, it does not in general have a head term w.r.t. any term order. It does, however, have a lowest term w.r.t. every term order, simply because term orders are always well-orders. The theory of standard bases in power series rings can now be described in a hand-waving manner as follows. Formulate the theory of Gröbner bases in polynomial rings in a non-algorithmic way, i.e., without any reference to polynomial reduction as a terminating process, and "turn it upside down" in the following sense: replace polynomials by power series and head terms by lowest terms, and turn around all inequalities in connection with term orders. This description of the results of the theory is surprisingly accurate, although most of the proofs are substantially different from the polynomial case.

For a more explicit description of the theory, we let K be a field. The first step is to prove a result that corresponds to the fact that every polynomial has a normal form modulo any finite set of polynomials. This result is obtained by stripping Proposition 5.22 of its algorithmic content and turning it upside down: if f is an element and G is a finite subset of $K[[\underline{X}]]$, then there exists a normal form $r \in K[[\underline{X}]]$ of f modulo G in the sense that

$$f = \sum_{g \in G} q_g g + r \qquad (q_g \in K[[\underline{X}]]) \qquad (*)$$

such that for all $g \in G$, the lowest term of g does not divide any term of r, and the minimum of the lowest terms of the summands on the right-hand side equals the lowest term of f. The set G is called a *standard basis* if 0 is such a normal form modulo G of every $f \in \mathrm{Id}(G)$, a condition which, in analogy to Gröbner basis terminology, is also expressed by saying that f has a *standard representation* modulo G. The normal form of $(*)$ is then uniquely determined for every $f \in K[[\underline{X}]]$. An equivalent characterization of standard bases is that for every $f \in \mathrm{Id}(G)$, there exists $g \in G$ such that the lowest term of g divides the lowest term of f. This yields an easy existence proof for standard bases, i.e., a proof of the fact that every ideal has a basis which is a standard basis: take a Dickson basis of the set of lowest terms of elements of the ideal and then choose elements of the ideal with these lowest terms. The resulting set is clearly a standard basis, and an easy additional argument shows that it is also a basis of the given ideal. These basics of the theory settled, one then obtains further analogues to results from Gröbner basis theory such as uniqueness of the reduced standard basis and existence of universal standard bases.

Power series being infinitary objects by nature, problems concerning actual computations are necessarily more difficult than in the polynomial case. It can be proved that with a natural definition of *S-series* which is modelled after the definition of S-polynomials, a finite subset G of $K[[\underline{X}]]$ is a standard basis if and only if all S-series of pairs of elements of G have standard representations modulo G. On the basis of this result, one may then, for term orders of order type ω, compute initial segments of the elements of a standard basis of $\mathrm{Id}(G)$ up to any prescribed total degree.

Non-Commutative Gröbner Bases

Let us recall from Section 2.1 that the polynomial ring in $X_1, \ldots, X_n$ over a field K was defined as the monoid ring over K and the additive monoid $\mathbb{N}^n$. It can be proved that the construction of monoid rings generalizes naturally to non-Abelian monoids; however, the result will be a non-commutative ring. It is easy to see that the set of all tuples—of arbitrary but finite length—of elements of the set $\{1, \ldots, n\}$ forms a non-Abelian monoid M under concatenation, with the empty tuple as the neutral element. If K is a field, then one may consider the monoid ring KM. As in the commutative case, K can naturally be viewed as a subfield of KM. One may then introduce a notation that is similar to the one used in the commutative case. Here, X_k is the element of KM that takes value 1 at the one-tuple (k) and value 0 everywhere else. It is not hard to see that with this notation, every element of KM has a representation of the form $a_1 t_1 + \cdots + a_m t_m$, where the $a_\imath$ are in K and the t_i are "non-commutative terms" that are of the form $X_{k_1}^{\nu_1} \cdot \cdots \cdot X_{k_r}^{\nu_r}$. Moreover, the ring operations are, loosely speaking, performed as in the commutative polynomial ring except that variables no longer commute with each other. The ring KM is therefore called the *non-commutative polynomial ring in* X_1, ..., X_n *over* K, and it is denoted by $K\langle X_1, \ldots, X_n \rangle$.

The definition of an ideal in a non-commutative ring is similar to the commutative case; however, one must distinguish left, right, and two-sided ideals, depending on whether I is closed under left, right, or two-sided multiplication with ring elements. It turns out that in contrast to the commutative case, $K\langle X_1, \ldots, X_n \rangle$ is not noetherian for $n \geq 2$. Nevertheless, one may try to design algorithms that decide the membership of a polynomial $f \in K\langle X_1, \ldots, X_n \rangle$ in a finitely generated left, right, or two-sided ideal of $K\langle X_1, \ldots, X_n \rangle$. The attempt to imitate Gröbner basis theory in the non-commutative case works fine up to the point where the termination of the analogue to the Buchberger algorithm is to be proved. It turns out that due to the lack of a Dickson lemma for non-commutative terms, this "non-commutative Buchberger algorithm" does actually fail to terminate in general (see Mora, 1986, 1988). Worse still, the ideal membership problem for two-sided ideals is algorithmically unsolvable already in $\mathbb{Q}\langle X_1, X_2 \rangle$. In other words, not only does the theory of Gröbner bases fail, there cannot be any other algorithm solving this problem. This can be proved by showing how a solution would give rise to a solution of the halting problem for Turing machines (see Kandri-Rody and Weispfenning, 1990, and the references given there).

A class of non-commutative rings for which the construction of finite Gröbner bases for arbitrary ideals is possible is the class of *solvable algebras.* It comprises many algebras arising in mathematical physics such as Weyl algebras, enveloping algebras of finite-dimensional Lie algebras, and iterated skew polynomial rings. Gröbner bases in these algebras were studied for special cases by Apel and Lassner (1985) and in full generality by Kandri-Rody and Weispfenning (1990). Solvable algebras can be viewed as certain residue class rings of non-commutative polynomial rings; alternatively, they can be described as ordinary, commutative polynomial rings

$K[\underline{X}]$ furnished with a new non-commutative multiplication "$*$" which is such that $K[\underline{X}]$ becomes a ring when "$*$" replaces the commutative multiplication "$\cdot$," and that "$*$" yields the same result as "$\cdot$" except in the case where the first factor contains a variable with an index that is higher than the index of some variable occuring in the second factor. The relationship between "$*$" and "$\cdot$" for a product of this type is determined by the requirement that

$$X_j * X_i = c_{ij} X_i \cdot X_j + p_{ij},$$

where $0 \neq c_{ij} \in K$, and $p_{ij} \in K[\underline{X}]$ is such that $p_{ij} = 0$ or $\mathrm{HT}(p_{ij}) < X_i X_j$ with respect to some fixed term order $\leq$. A $*$-product $f * g$ of two non-zero polynomials f, $g \in K[\underline{X}]$ is then always of the form

$$f * g = c \cdot f \cdot g + p,$$

where $0 \neq c \in K$ and $p \in K[\underline{X}]$ with $p = 0$ or $\mathrm{HT}(p) < \mathrm{HT}(f \cdot g)$. This is essentially what makes Gröbner basis theory work for left and right ideals in these algebras. Roughly speaking, this is because one may, at any point during polynomial reduction or the forming of an S-polynomial, switch variables at the cost of introducing lower terms, and these do not cause trouble in the end due to the fact that the key arguments used in the theory depend on the head terms only. For the construction of a two-sided Gröbner basis one has to intertwine the construction of a one-sided Gröbner basis with iterated multiplication of the polynomials in the current ideal basis by all indeterminates from the other side.

Many applications of Gröbner basis theory such as ideal membership test, computation of intersections of ideals, computation of syzygies, and computation in residue class rings and modules can be transferred readily from the commutative case to solvable algebras. Other applications such as the subring membership test fail to work. Rather surprisingly, the construction of comprehensive Gröbner bases can be carried out in solvable algebras, even to the extent that the commutator relations may contain parameters (see Kredel and Weispfenning, 1990). This opens the way for developing a rudimentary algebraic geometry concerning varieties of "solvable polynomials" over skewfields (see Kredel, 1992).

Another class of non-commutative algebras that admit the construction of finite Gröbner bases for finitely generated ideals is studied in Weispfenning (1992a).

Gröbner Bases and Differential Algebra

Differential algebra arises from the study of differential equations in much the same way as algebra arises from the study of polynomial equations. The theory is largely due to Ritt (1950) and to Kolchin (1973). Just as the algebraic theory of equations studies solutions of polynomial equations as abstract objects in algebraic structures such as rings or fields, differential algebra studies solvability and solutions of differential equations in *differential rings* and *differential fields.* A differential ring is a commutative ring R together with one or several maps from R to itself, the so-called *differential operators*, that satisfy the usual differentiation rules w.r.t. addition and multiplication. The concept of a polynomial ring is replaced by the concept of a *differential polynomial ring* $R\{X_1, \ldots, X_n\}$ over the differential ring R. The elements of $R\{X_1, \ldots, X_n\}$, which are called differential polynomials, are polynomials with coefficients in R in the indeterminates X_i and further indeterminates that are obtained from the X_i by iterated application of the differential operators. $R\{X_1, \ldots, X_n\}$ is thus a polynomial ring over R in infinitely many indeterminates; in particular, it is no longer noetherian, even if R is a field: the ideal generated by all indeterminates, for example, is not finitely generated.

From the differential viewpoint, however, noetherianity of the differential polynomial ring is not the point. To see why, let us consider the special case of a differential polynomial ring over a differential field K extending $\mathbb{Q}$ with only one differential operator D. This is the situation arising from the study of ordinary algebraic differential equations. A *system of algebraic differential equations* is now a system of equations of the form

$$f_1 = 0, \ldots, f_m = 0,$$

with $f_i \in K\{X_1, \ldots, X_n\}$ for $1 \leq i \leq m$. Possible solutions of such a system are n-tuples of elements of some differential extension field L of K. Since $D(0) = 0$ holds in every differential ring, the variety of solutions in L of a system of differential equations is closed under differentiation, i.e., under application of D. Accordingly, one is led to the study of *differential ideals* in $K\{X_1, \ldots, X_n\}$, meaning ideals that are closed under differentiation. Similarly, radical differential ideals are defined as the analogue to radical ideals. It turns out that the counterpart of the Hilbert basis theorem (the Ritt–Raudenbush basis theorem) holds for radical differential ideals, but not for arbirary differential ideals: every radical differential ideal is finitely generated as a radical differential ideal, but not in general as a differential ideal.

What is required next is a suitable Nullstellensatz. The following weak version (cf. Theorem 6.61) is due to Ritt: the implication

$$\forall x_1 \cdots \forall x_n \left(\bigwedge_{i=1}^{m} f_i(\underline{x}) = 0 \longrightarrow f_0(\underline{x}) = 0 \right)$$

holds in the class of all differential field extensions of K if and only if f_0 is in the radical differential ideal generated by $f_1, \ldots, f_m$. Taking $f_0 = 1$, we see that the system $f_1 = 0, \ldots, f_m = 0$ has a solution in some differential extension field of K if and only if the differential ideal generated by $f_1, \ldots, f_m$ does not contain 1. In its strong form, the Nullstellensatz is much more subtle. It requires the construction of differentially closed fields to replace the algebraically closed fields in the Hilbert Nullstellensatz (see Macintyre, 1977, for an overview and references).

Deciding the solvability of a system of algebraic differential equations is thus equivalent to the problem of deciding whether 1 is in a differential ideal that is given by a finite basis. This problem may be viewed both as a special case of the radical differential ideal membership problem and of the differential ideal membership problem. The former has found an algorithmic solution in Seidenberg's elimination method for differential algebra (Seidenberg, 1956), which proceeds by eliminating one variable at a time. The resulting complexity of the procedure renders the method useless for all practical purposes. Unfortunately, despite several partial results (Carrà Ferro, 1987), the attempt to imitate Gröbner basis methods in the context of differential ideals and radical differential ideals has been unsuccessful to date. The main obstacle is a missing a priori bound on the order (i.e., the number of iterated applications of D) of the differential polynomials involved in such a construction. Such a bound is needed to guarantee termination. As a matter of fact, it is not even known whether the problem of deciding membership in the differential ideal has an algorithmic solution at all.

One potential application of Gröbner basis theory in differential algebra is as follows. If one considers the algebra that is generated by the differential operators and the differentiation variables, viewed as operators, then one obtains the commutator relations $D_i \cdot X_i - X_i \cdot D_i = 1$ and $D_i \cdot X_j - X_j \cdot D_i = 0$ for $i \neq j$. The resulting algebra is in fact a Weyl algebra, and one may apply the Gröbner basis theory that was described in the section on noncommutative Gröbner bases above.

Finally, methods that are reminiscent of Gröbner basis techniques have been used in partial differential algebra for a long time, namely, in what is known as Riquier–Janet theory (Riquier, 1910; Janet, 1929). Here, one is concerned with the transformation of a given system of partial differential equations into a special form from which power series solutions can be obtained easily. The role of terms as multipliers in Gröbner basis theory is taken over by expressions obtained by composition of the differential operators. The relation between these methods and Gröbner basis theory has not been clarified in a satisfactory manner to date.

Selected Bibliography

The bibliography is divided into three parts: conference proceedings, books and monographs, and articles. Proceedings are listed by name of conference or acronym thereof, while books and articles are ordered by author/editor. The focus is on publications pertaining to Gröbner basis theory; only a few major works concerning other areas of algebra and computational algebra are listed.

Conference Proceedings

Advances in Robot Kinematics, Lenarcic, J. and Stifter, S. (eds.), Springer-Verlag, New York, 1991.

AAECC-2. Poli, A. (ed.), *Applied Algebra, Algorithmics, and Error-Correcting Codes.* 2nd International Conference, Toulouse, France, October 1984, Springer LNCS **228**.

AAECC-3. Calmet, J. (ed.), *Algebraic Algorithms and Error-Correcting Codes.* 3rd International Conference, Grenoble, France, July 1985, Springer LNCS **229**.

AAECC-4. Beth, T. and Clausen, M. (eds.), *Applicable Algebra, Error-Correcting Codes, Combinatorics, and Computer Algebra.* 4th International Conference, Karlsruhe, FRG, September 1986, Springer LNCS **307**.

AAECC-5. Huguet, L. and Poli, A. (eds.), *Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes.* 5th International Conference, Menorca, Spain, June 1987, Springer LNCS **356**.

AAECC-6. Mora, T. (ed.), *Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes.* 6th International Conference, Rome, Italy, July 1988, Springer LNCS **357**.

AAECC-8. Sakata, S. (ed.), *Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes.* 8th International Conference, Tokyo, Japan, August 1990, Springer LNCS **508**.

AAECC-9. Mattson, H.F., Mora, T., and Rao, T.R.N. (eds.), *Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes.* 9th International Symposium, New Orleans, LA, October 1991, Springer LNCS **539**.

Computers and Mathematics, Kaltofen, E. and Watt, S.M. (eds.), Springer-Verlag, New York, 1989.

IV International Conference on Computer Algebra in Physical Research, Dubna, USSR, 22-26 May 1990, Gerdt, V.P., Rostovtsev, V.A., and Shirkov, D.V. (eds.), World Scientific Publishing Co., Singapore.

EUROCAL '83. van Hulzen, J.A. (ed.), *European Computer Algebra Conference.* London, England, March 1983, Springer LNCS **162**.

EUROCAL '85. Buchberger, B. (ed.), *European Conference on Computer Algebra*, Vol. I. Linz, Austria, April 1985, Springer LNCS **203**.

EUROCAL '85. Caviness, B.F. (ed.), *European Conference on Computer Algebra*, Vol. II. Linz, Austria, April 1985, Springer LNCS **204**.

EUROCAL '87. Davenport, J.H. (ed.), *European Conference on Computer Algebra.* Leipzig, GDR, June 1987, Springer LNCS **378**.

EUROCAM '82. Calmet, J. (ed.), *European Computer Algebra Conference.* Marseille, France, April 1982, Springer LNCS **144**.

EUROSAM '79. Ng, E.W. (ed.), *An International Symposium on Symbolic and Algebraic Manipulation.* Marseille, France, June 1979, Springer LNCS **72**.

EUROSAM '84. Fitch, J. (ed.), *International Symposium on Symbolic and Algebraic Manipulation.* Cambridge, England, July 1984, Springer LNCS **174**.

ISSAC '88. Gianni, P. (ed.), *International Symposium on Symbolic and Algebraic Computation.* Rome, Italy, July 1988, Springer LNCS **358**.

ISSAC '89. *International Symposium on Symbolic and Algebraic Computation.* Portland, Oregon, July 1989, ACM Press, New York.

ISSAC '90. Watanabe, S. and Nagata, M. (eds.), *International Symposium on Symbolic and Algebraic Computation.* Tokyo, Japan, August 1990, ACM Press, New York.

ISSAC '91. Watt, S.M. (ed.), *International Symposium on Symbolic and Algebraic Computation.* Bonn, FRG, July 1991, ACM Press, New York.

Logic and Machines: Decision Problems and Complexity. Proc. Symposium Rekursive Kombinatorik, Münster May 1985, Börger, E., Hasenjaeger, G., and Rödding, D. (eds.), 1984, Springer LNCS **171**.

MACSYMA '77. *Proc. of the 1977 MACSYMA Users' Conference*, Washington D.C., 1977, NASA CP-2012.

MACSYMA '84. *Third MACSYMA Users' Conference*, Schenectady, NY, July 1984.

MEGA '90. Mora, T. and Traverso, C. (eds.), *Effective Methods in Algebraic Geometry.* Castiglioncello, Livorno, Italy, April 1990, Progress in Mathematics **94**, Birkhäuser Verlag, Basel.

Rewriting Techniques and Applications, Dijon, May 1985. Jouannaud, J.P. (ed.), Springer LNCS **202**.

Rewriting Techniques and Applications, Bordeaux, May 1987. Lescanne, P. (ed.), Springer LNCS **256**.

Rewriting Techniques and Applications, Chapel Hill, NC, April 1989. Dershowitz, N. (ed.), Springer LNCS **355**.

Rewriting Techniques and Applications, Como, April 1991. Book, R.V. (ed.), Springer LNCS **488**.

The Second International Symposium on Symbolic and Algebraic Computation by Computers, August 1984, RIKEN, Wako-Shi, Saitama, 351-01, Japan.

SYMSAC '71. Petrick, S.R. (ed.), *Second Symposium on Symbolic and Algebraic Manipulation*, ACM Press, New York, 1971.

SYMSAC '76. Jenks, R.D. (ed.), *1976 ACM Symposium on Symbolic and Algebraic Computation*, ACM Press, New York, 1976.

SYMSAC '81. Wang, P.S. (ed.), *1981 ACM Symposium on Symbolic and Algebraic Computation*, ACM Press, New York, 1981.

SYMSAC '86. Char, B.W. (ed.), *1986 ACM Symposium on Symbolic and Algebraic Computation*, University of Waterloo, Ontario, 1986.

TCA '87. Janßen, R. (ed.), *Trends in Computer Algebra.* International Symposium Bad Neuenahr, FRG, May 1987, Springer LNCS **296**.

Books and Monographs

Akritas, A.G. (1989). *Elements of Computer Algebra with Applications.* John Wiley, New York.

Apel, J. (1988). *Gröbnerbasen in nicht-kommutativen Algebren und ihre Anwendungen.* Doctoral Dissertation, Univ. Leipzig.

Atiyah, M.F. and MacDonald, J.G. (1969). *Introduction to Commutative Algebra.* Addison-Wesley, Reading, MA.

Bayer, D. (1982). *The Division Algorithm and the Hilbert Scheme.* Ph.D. Thesis, Harvard University, Order-Nr. 82-22588, University Microfilms International, 300 N. Zeeb Rd., Ann Arbor, MI 48106.

Bjork, I.E. (1979). *Rings of Differential Operators.* North Holland, Amsterdam.

Bochnak, J., Coste, M., and Roy, M.F. (1987). *Géometrie Algébrique Réelle.* Ergebnisse der Mathematik und ihrer Grenzgebiete, Springer-Verlag, Berlin.

Borgwardt, K.H. (1980). *The Simplex Method: A Probabilistic Analysis.* Springer-Verlag, New York.

Borodin, A. and Munroe, I. (1975). *The Computational Complexity of Algebraic and Numeric Problems.* Elsevier, New York.

Buchberger, B., Collins, G.E., and Loos, R. (eds.) (1982). *Computer Algebra: Symbolic and Algebraic Computation.* Springer-Verlag, New York.

Chou, S.C. (1988). *Mechanical Geometry Theorem Proving.* Reidel, Dordrecht.

Cox, D., Little, J., and O'Shea, D. (1992). *Ideals, Varieties, and Algorithms. An Introduction to Computational Algebraic Geometry and Commutative Algebra.* Springer-Verlag, New York.

Davenport, J.H., Siret, Y., and Tournier, E. (1988). *Computer Algebra: Systems and Algorithms for Algebraic Computation.* Academic Press, London.

Dixmier, J. (1974). *Algèbres Enveloppantes.* Bordas (Gauthier-Villars), Paris.

El From, Y. (1983). *Sur les Algèbres de Type Résoluble.* Thèse de 3e cycle, Université de Pierre et Marie Curie, Paris VI.

Fraîssé, R. (1986). *Theory of Relations.* Studies in Logic and the Foundations of Mathematics **118**, North-Holland, Amsterdam.

Fulton, W. (1969). *Algebraic Curves: An Introduction to Algebraic Geometry.* W.A. Benjamin, Reading, MA.

Geddes, K.O., Czapor, S.R., and Labahn, G. (1992). *Algorithms for Computer Algebra.* Kluwer Academic Publishers, Boston.

Göbel, M. (1992). Reduktion G-symmetrischer Polynome für beliebige Permutationsgruppe G. Diploma Thesis, Univ. Passau.

Goodearl, K.R. and Warfield, R.B. (1989). *An Introduction to Noncommutative Noetherian Rings.* London Math. Soc. Stud. Texts. **16**, Cambridge Univ. Press, Cambridge.

Gröbner, W. (1968). *Algebraische Geometrie*, Vol. I. Bibliographisches Institut, Mannheim.

Gröbner, W. (1970). *Algebraische Geometrie*, Vol. II. Bibliographisches Institut, Mannheim.

Hartshorne, R. (1977). *Algebraic Geometry.* Graduate Texts in Mathematics **52**, Springer-Verlag, New York.

Hense, K. (1908). *Theorie der Algebraischen Zahlen.* B.G. Teubner, Leipzig.

Herstein, I.N. (1964). *Topics in Algebra.* Blaisdell, Waltham, MA.

Hrbacek, K. and Jech, Th. (1984). *Introduction to Set Theory.* Marcel Dekker, New York.

Jacobson, N. (1979). *Lie Algebras.* Dover, New York.

Jacobson, N. (1985). *Basic Algebra*, Vols. I, II. Freeman, New York.

Janet, M. (1929). *Leçons sur les Systèmes d'Equations aux Derivées Partielles.* Gauthier-Villars, Paris.

Jantzen, M. (1988). *Confluent String Rewriting.* EATCS Monographs on Theoretical Computer Science **14**, Springer-Verlag, New York.

Kandri-Rody, A. (1984). *Effective Methods in the Theory of Polynomial Ideals.* Ph.D. Thesis, Rensselaer Polytechnic Institute, Troy, NY.

Kaplansky, I. (1968). *Commutative Rings.* Queen Mary College Math. Notices, London.

Kaplansky, I. (1974). *Commutative Rings.* University of Chicago Press, Chicago.

Kline, M. (1985). *Mathematics and the Search for Knowledge.* Oxford University Press, Oxford.

Knuth, D.E. (1969). *The Art of Computer Programming*, Vol. 2. Addison-Wesley, Reading, MA.

König, D. (1936). *Theorie der endlichen und unendlichen Graphen.* Akademische Verlagsgesellschaft, Leipzig.

König, J. (1903). *Einleitung in die allgemeine Theorie der algebraischen Grössen.* B.G. Teubner, Leipzig.

Kolchin, E.R. (1973). *Differential Algebra and Algebraic Groups.* Academic Press, New York.

Kredel, H. (1992). *Solvable Polynomial Rings.* Doctoral Dissertation, Univ. Passau.

Krull, W. (1935). *Idealtheorie.* Ergebnisse der Mathematik und ihrer Grenzgebiete, Springer-Verlag, Berlin. Reprint Chelsea Publ. Co., New York (1950).

Kunen, K. (1983). *Set Theory: An Introduction to Independence Proofs.* North Holland, Amsterdam.

Lang, S. (1971). *Algebra.* Addison-Wesley, Reading, MA.

Lausch, H. and Noebauer, W. (1973). *Algebra of Polynomials.* North Holland, Amsterdam, 1973.

Lipson, J.D. (1981). *Elements of Algebra and Algebraic Computing.* Addison-Wesley, Reading, MA.

Macaulay, F.S. (1916). *Algebraic Theory of Modular Systems.* Cambridge Tracts in Mathematics **19**, Cambridge University Press, Cambridge.

Marden, M. (1949). *The Geometry of the Zeros of a Polynomial in a Complex Variable.* AMS Math. Surv. **3**.

Marden, M. (1966). *Geometry of Polynomials.* 2nd ed., American Mathematical Society, Providence, RI.

Mark, W. (1992). Gröbnerbasen über Hauptidealringen und euklidischen Ringen. Diploma Thesis, Univ. Passau.

Matsumura, H. (1980). *Commutative Algebra.* Benjamin-Cummings, Reading, MA.

Mines, R., Richman, F., and Ruitenburg, W. (1988). *A Course in Constructive Algebra.* Springer-Verlag, New York.

Moore, G.H. (1982). *Zermelo's Axiom of Choice: Its Origins, Development, and Influence.* Springer-Verlag, New York.

Morgan, A. (1987). *Solving Polynomial Systems Using Continuation for Engineering and Scientific Problems.* Prentice-Hall, Englewood Cliffs, NJ.

Nagata, M. (1977). *Field Theory.* Marcel Dekker, New York.

Nastasescu, C. and van Oystaeyen, F. (1982). *Graded Ring Theory.* North-Holland, Amsterdam.

Netto, E. (1896/1900). *Vorlesungen über Algebra,* Vols. I, II. B.G. Teubner, Leipzig.

Perron, O. (1951). *Algebra. Die Grundlagen,* Vol. 1. de Gruyter, Berlin.

Pommaret, J.F. (1978). *Systems of Partial Differential Equations and Lie Pseudogroups.* Gordon and Breach, New York.

Renschuch, B. (1976). *Elementare und praktische Idealtheorie.* VEB Deutscher Verlag der Wissenschaften, Berlin.

Riquier, F. (1910). *Les Systèmes d'Equations aux Derivées Partielles.* Gauthier-Villars, Paris.

Ritt, J.F. (1950). *Differential Algebra.* AMS, New York.

Rosenstein, J.G. (1982). *Linear Orderings.* Pure and Applied Mathematics **98**, Academic Press, London.

Rubin, H. and Rubin, J.E. (1963). *Equivalents of the Axiom of Choice.* North Holland, Amsterdam.

Schaller, S. (1979). *Algorithmic Aspects of Polynomial Residue Class Rings.* Ph.D. Thesis, Department of Computer Science, University of Wisconsin, Comp. Sci. Tech. Rep. 370.

Schinzel, A. (1982). *Selected Topics on Polynomials.* University of Michigan Press, Ann Arbor, MI.

Schrader (1976). *Zur konstruktiven Idealtheorie.* Diplomarbeit, Mathematisches Institut II, Universität Karlsruhe.

Schreyer, F.O. (1980). *Die Berechnung von Syzygien mit dem verallgemeinerten Weierstrassschen Divisionssatz und eine Anwendung auf analytische Cohen-Macaulay Stellenalgebren minimaler Multiplizität.* Diplomarbeit am Fachbereich Mathematik der Universität Hamburg, 1980.

Seidenberg, A. (1968). *Elements of the Theory of Algebraic Curves.* Addison-Wesley, Reading, MA.

Shafarevich, I.R. (1977). *Basic Algebraic Geometry.* Springer-Verlag, New York.

Sharp, R.Y. (1990). *Steps in Commutative Algebra.* Cambridge University Press, Cambridge.

Sims, C.C. (1984). *Abstract Algebra: A Computational Approach.* John Wiley, New York.

Specker, E. and Strasser, V. (eds.) (1976). *Komplexität von Entscheidungsproblemen.* Springer LNCS **43**.

Steinitz, E. (1910). *Algebraische Theorie der Körper.* Journal f. reine und angew. Mathematik **137**, 167–309. Reprint Chelsea Publ. Co., New York (1950).

Tangora, M.C. (1988). *Computers in Algebra.* Marcel Dekker, New York.

Tarski, A. (1951). *A decision method for elementary algebra and geometry.* 2nd rev. ed., University of California Press, Berkeley.

Trotter, P.G. (1969). *A Canonical Basis for Ideals of Polynomials in Several Variables and with Integer Coefficients.* Ph.D. Thesis, University of New South Wales.

Uspensky, J.V. (1948). *Theory of Equations.* McGraw-Hill, New York.

Ueberberg, J. (1992). *Einführung in die Computer Algebra mit REDUCE.* BI Wiss. Verl. Mannheim.

van der Waerden, B.L. (1931). *Moderne Algebra*, Vols. I, II. Springer-Verlag, Heidelberg.

van der Waerden, B.L. (1966). *Modern Algebra*, Vols. I, II. Frederick Ungar Publishing Co., New York.

van der Waerden (1967/1971). *Algebra*, Vols. I, II. Springer-Verlag, Heidelberg.

Weber, H. (1898/1899/1908). *Lehrbuch der Algebra*, Vols. I, II, III. 2. Aufl., Braunschweig. Reprint Chelsea Publ. Comp., New York, 1961.

Welsh, D.J. (1976). *Matroid Theory.* L.M.S. Monographs, Academic Press, London.

White, N. (ed.) (1986). *Theory of Matroids.* Cambridge University Press, Cambridge.

Winkler, F. (1984). *The Church–Rosser Property in Computer Algebra and Special Theorem Proving: An Investigation of Critical-Pair Completion Algorithms.* Doctoral Dissertation, University of Linz.

Wüthrich, H.R. (1977). *Ein schnelles Quantoreneliminationsverfahren für die Theorie der algebraisch abgeschlossenen Körper.* Ph.D. Thesis, Univ. Zürich.

Zacharias, G. (1978). *Generalized Gröbner Bases in Commutative Polynomial Rings.* Bachelor Thesis, Lab. for Computer Science, MIT.

Zariski, O. and Samuel, P. (1958/1960). *Commutative Algebra*, Vols. I, II. Van Nostrand, Princeton, NJ. Reprint Springer-Verlag, New York, 1975/1979.

Articles

Abbott, J.A., Bradford, R.J., and Davenport, J.H. (1988). Factorisation of polynomials: Old ideas and recent results. In: Janßen, R. (ed.), *Trends in Computer Algebra*, Springer LNCS **296**, 81–91.

Abhyankar, S. and Li, W. (1989). On the Jacobian conjecture: A new approach via Gröbner bases. *J. Pure Appl. Algebra* **61**, 211–222.

Adams, W.W. and Boyle, A.K. (1992). Some results on Gröbner bases over commutative rings. *J. Symb. Comp.* **13**/5, 473–484.

Akritas, A.G. (1990). The two classical subresultant PRS methods. In: Gerdt, V.P., Rostovtsev, V.A., and Shirkov, D.V. (eds.), *IV International Conference on Computer Algebra in Physical Research*, World Scientific Publishing Co., Singapore, 225–229.

Alonso, M.E., Mora, T., and Raimondo, M. (1990). On the complexity of algebraic power series. In: Sakata, S. (ed.), *Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes (AAECC 8)*, Springer LNCS **508**, 197–207.

Alonso, M.E., Mora, T., and Raimondo, M. (1990a). Local decomposition algorithms. In: Sakata, S. (ed.), *Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes (AAECC 8)*, Springer LNCS **508**, 208–221.

Apel, J. and Lassner, W. (1985). An algorithm for calculations in enveloping fields of Lie algebras. In: *Proc. International Conference on Computer Algebra and Its Applications in Theoretical Physics*, JINR D11-85-791, Dubna, 231–141.

Apel, J. and Lassner, W. (1987). Computation and simplification in Lie fields. In: Davenport, J.H. (ed.), *EUROCAL '87, European Conference on Computer Algebra*, Springer LNCS **378**, 468–478.

Apel, J. and Lassner, W. (1988). An extension of Buchberger's algorithm and calculations in enveloping fields of Lie Algebras. *J. Symb. Comp.* **6**/2,3, 361–370.

Apel, J. and Klaus, U. (1990). Implementational aspects for non-commutative domains. In: Gerdt, V.P., Rostovtsev, V.A., and Shirkov, D.V. (eds.), *IV International Conference on Computer Algebra in Physical Research*, World Scientific Publishing Co., Singapore, 127–132.

Armbruster, D. and Kredel, H. (1986). Constructing universal unfoldings using Gröbner bases. *J. Symb. Comp.* **2**/4, 383–388.

Arnon, D.S. (1988). A bibliography of quantifier elimination for real closed fields. *J. Symb. Comp.* **5**/1,2, 267–274.

Audoly, S., Bellu, G., Buttu, A., and d'Angio, L. (1991). Procedures to investigate injectivity of polynomial maps and to compute the inverse. *AAECC* **2**, 91–103.

Auzinger, W. and Stetter, H.J. (1988). An elimination algorithm for the computation of all zeros of a system of multivariate polynomial equations. In: Agarwal, R.P., Chow, Y.M., and Wilson, S.J. (eds.), *Numerical Mathematics*, International Series of Numerical Mathematics **86**, Birkhäuser Verlag, 11–30.

Ayoub, C.W. (1982). The decomposition theorem for ideals in polynomial rings over a domain. *J. Algebra* **76**, 99–110.

Ayoub, C.W. (1983). On constructing bases for ideals in polynomial rings over the integers. *J. Number Th.* **17**, 204–225.

Bachmair, L. and Buchberger, B. (1980). A simplified proof of the characterization theorem for Gröbner bases. *ACM SIGSAM Bulletin* **14**/4, 29–34.

Bachmair, L. and Dershowitz, N. (1988). Critical pair criteria for completion. *J. Symb. Comp.* **6**/1, 1–18.

Backelin, L. and Froeberg, R. (1991). How we proved that there are exactly 924 cyclic 7-roots. In: Watt, S.M. (ed.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '91)*, ACM Press, New York, 103–111.

Ballantyne, A.M. and Lankford, D.S. (1981). New decision algorithms for finitely presented commutative semigroups. *J. Comp. Math. Appls.* **7**, 159–165.

Bareiss, E.H. (1968). Sylvester's identity and multistep integer-preserving Gaussian elimination. *Math. Comput.* **22**, 565–578.

Bayer, D.A. and Morrison, I. (1988). Standard bases and geometric invariant theory: I. Ideals and state polytopes. *J. Symb. Comp.* **6**/2,3, 209–217.

Bayer, D.A. and Stillman, M. (1986). The design of Macaulay: A system for computing in algebraic geometry and commutative algebra. In: Char, B.W. (ed.), *1986 ACM Symposium on Symbolic and Algebraic Computation*, University of Waterloo, Ontario, 157–162.

Bayer, D.A. and Stillman, M. (1987). A theorem on refining division orders by the reverse lexicographic order. *Duke Math. J.* **55**/2, 321–328.

Bayer, D.A. and Stillman, M. (1988). On the complexity of computing syzygies. *J. Symb. Comp.* **6**/2,3, 135–147.

Bayer, D.A. and Stillman, M. (1992). Computation of Hilbert Functions. *J. Symb. Comp.* **14**/1, 31–50.

Beck, R.E. and Kolman, B. (1986). Symbolic algorithms for Lie algebra computation. In: Char, B.W. (ed.), *1986 ACM Symposium on Symbolic and Algebraic Computation*, University of Waterloo, Ontario, 85–87.

Becker, T. (1990). Standard bases and some computations in rings of power series. *J. Symb. Comp.* **10**/2, 165–178.

Becker, T. (1990a). Stability and Buchberger criterion for standard bases in power series rings. *J. Pure Appl. Algebra* **66**, 219–227.

Becker, T. (1991). Homogeneity, pseudo-homogeneity, and Gröbner basis computations. In: Mattson, H.F., Mora, T., and Rao, T.R.N. (eds.), *Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes (AAECC 9)*, Springer LNCS **539**, 65–73.

Becker, T. and Weispfenning, V. (1991). The Chinese remainder problem, multivariate interpolation, and Gröbner bases. In: Watt, S.M. (ed.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '91)*, ACM Press, New York, 64–69.

Beckmann, P. (1982). Umwandlung von Rechts- in Linksquotienten über nichtkommutativen Ringen. *Wiss. Z. Karl-Marx-Univ. (Leipzig) Math. Natur. Reihe* **31**, 11–23.

Bergman, G.M. (1978). The diamond lemma for ring theory. *Adv. Math.* **29**, 178–218.

Berlekamp, E.R. (1967). Factoring polynomials over finite fields. *Bell Syst. Tech. J.* **46**, 1853–1859.

Berlekamp, E.R. (1970). Factoring polynomials over large finite fields. *Math. Comput.* **24**, 713–735.

Bertini, E. (1889). Zum Fundamentalsatz aus der Theorie der algebraischen Funktionen. *Math. Ann.* **34**, 447–449.

Billera, C.J. and Rose, L.L. (1989). Gröbner basis methods for multivariate splines. *Rutgers Research Report* **1**/89.

Bini, D. and Pan, V. (1990). Parallel polynomial computations by recursive processes. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 294.

Birkhoff, G. (1971). The role of modern algebra in computing. In: Birkhoff, G. and Marshall, H. (eds.), *Computers in Algebra and Number Theory*, SIAM-AMS Proc. IV, 1–47.

Björk, G. and Fröberg, F. (1991). A faster way to count the solutions of inhomogeneous sytems of algebraic equations, with applications to cyclic n-roots. *J. Symb. Comp.* **12**, 329–336..

Blass, A. and Gurevich, Y. (1984). Equivalence relations, invariants and normal forms. In: Börger, E., Hasenjaeger, G., and Rödding, D. (eds.), *Logic and Machines: Decision Problems and Complexity*, Springer LNCS **171**, 24–42.

Böge, W., Gebauer, R., and Kredel, H. (1986). Some examples for solving systems of algebraic equations by calculating Gröbner bases. *J. Symb. Comp.* **2**/1, 83–98.

Bradford, R. (1990). A parallelization of the Buchberger algorithm. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 296.

Briançon, J. (1973). Weierstrass préparé à la Hironaka. *Astérisque* **7-8**, 67–76.

Bronstein, M. (1986). Gsolve: A faster algorithm for solving systems of algebraic equations. In: Char, B.W. (ed.), *1986 ACM Symposium on Symbolic and Algebraic Computation*, University of Waterloo, Ontario, 247–249.

Brownawell, W.D. (1987). Bounds for the degrees in the Nullstellensatz. *Ann. Math.* **126**/3, 577–591.

Brundu, M. and Rossi, F. (1988). On the computation of generalized standard bases. *J. Symb. Comp.* **6**/2,3, 323–343.

Buchberger, B. (1965). Ein Algorithmus zum Auffinden der Basiselemente des Restklassenringes nach einem nulldimensionalen Polynomideal, (An algorithm for finding a basis for the residue class ring of a zero-dimensional polynomial ideal). Doctoral Dissertation Math. Inst. University of Innsbruck, Austria.

Buchberger, B. (1970). Ein algorithmisches Kriterium für die Lösbarkeit eines algebraischen Gleichungssystems. *Aequ. Math.* **4**/3, 374–383.

Buchberger, B. (1976). A theoretical basis for the reduction of polynomials to canonical form. *ACM SIGSAM Bulletin* **10**/3, 19–29.

Buchberger, B. (1976a). Some properties of Gröbner bases for polynomial ideals. *ACM SIGSAM Bulletin* **10**/4, 19–24.

Buchberger, B. (1979). A criterion for detecting unnecessary reductions in the construction of Gröbner bases. In: Ng, E.W. (ed.), *EUROSAM '79, An International Symposium on Symbolic and Algebraic Manipulation*, Springer LNCS **72**, 3–21.

Buchberger, B. (1983). A note on the complexity of constructing Gröbner bases. In: van Hulzen, J.A. (ed.), *EUROCAL '83, European Computer Algebra Conference*, Springer LNCS **162**, 137–145.

Buchberger, B. (1983a). Gröbner bases: A method in symbolic mathematics. In: *Conference on Systems and Techniques of Analytical Computing and Their Applications in Theoretical Physics*, Dubna.

Buchberger, B. (1984). A critical-pair-completion algorithm for finitely generated ideals in rings. In: Börger, E., Hasenjaeger, G., and Rödding, D. (eds.), *Logic and Machines: Decision Problems and Complexity*, Springer LNCS **171**, 137–161.

Buchberger, B. (1985). A survey on the method of Gröbner bases for solving problems in connection with systems of multivariate polynomials. In: *The Second International Symposium on Symbolic and Algebraic Computation by Computers*, RIKEN, Wako-Shi, Saitama, 351-01, Japan, 7-1–7-15.

Buchberger, B. (1985a). Gröbner bases: An algorithmic method in polynomial ideal theory. In: Bose, N.K. (ed.), *Multidimensional Systems Theory*, Reidel, Dordrecht, 184–232.

Buchberger, B. (1985b). Basic features and development of the critical-pair completion procedure. In: Jouannaud, J.P. (ed.), *Rewriting Techniques and Applications*, Springer LNCS **202**, 1–45.

Buchberger, B. (1987). History and basic features of the critical-pair completion procedure. *J. Symb. Comp.* **3**/1,2, 3–38.

Buchberger, B. (1987a). Applications of Gröbner bases in non-linear computational geometry. In: Janßen, R. (ed.), *Trends in Computer Algebra*, Springer LNCS **296**, 52–80.

Burdik, C., Zaruba, M., and Lassner, W. (1983). Formula manipulations with polynomials in annihilation and creation operators. In: *Conference on Systems and Techniques of Analytical Computing and Their Applications in Theoretical Physics*, Dubna.

Caniglia, L., Galligo, A., and Heintz, J. (1988). Some new effectivity bounds in computational geometry. In: Mora, T. (ed.), *Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes (AAECC 6)*, Springer LNCS **357**, 131–151.

Caniglia, L., Galligo, A., and Heintz, J. (1988a). Borne simple exponentielle pour les degrées dans le theoréme des zéros sur un corps de characteristique quelconque. *C. R. Acad. Sci. (Paris)* **307**/I, 255–258.

Caniglia, L., Guccione, J.A., and Guccione, J.J. (1990). Local membership problems for polynomial ideals. In: Mora, T. and Traverso, C. (eds.), *Effective Methods in Algebraic Geometry*, Progress in Mathematics **94**, Birkhäuser Verlag, Basel, 31–45.

Canny, J.F., Kaltofen, E., and Yagati, L. (1989). Solving systems of non-linear polynomial equations faster. In: *International Symposium on Symbolic and Algebraic Computation (ISSAC '89)*, ACM Press, New York, 121–128.

Cantor, D.G. and Zassenhaus, H. (1981). A new algorithm for factoring polynomials over finite fields. *Math. Comput.* **36**, 587–592.

Carrà Ferro, G. (1985). Some upper bounds for the multiplicity of an autoreduced subset of $\mathbb{N}^m$ and their applications. In: Calmet, J. (ed.), *Algebraic Algorithms and Error-Correcting Codes (AAECC 3)*, Springer LNCS **229**, 306–315.

Carrà Ferro, G. (1987). Gröbner bases and differential algebra. In: Huguet, L. and Poli, A. (eds.), *Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes (AAECC 5)*, Springer LNCS **356**, 129–140.

Carrà Ferro, G. (1987a). Some properties of lattice points and their application in differential algebra. *Preprint.*

Carrà Ferro, G. (1988). Gröbner bases and Hilbert schemes I. *J. Symb. Comp.* **6**/2,3, 219–230.

Caviness, B.F. (1970). On canonical forms and simplification. *J. ACM* **17**/2, 385–396.

Caviness, B.F. (1985). Computer algebra: Past and future. In: Buchberger, B. (ed.), *EUROCAL '85, European Conference on Computer Algebra*, Vol. I, Springer LNCS **203**, 1–18.

Chardin, M. (1990). Un algorithme pour le calcul des résultants. In: Mora, T. and Traverso, C. (eds.), *Effective Methods in Algebraic Geometry*, Progress in Mathematics **94**, Birkhäuser Verlag, Basel, 47–62.

Chen, G. (1990). An algorithm for computing the formal solutions of differential systems in the neighborhood of an irregular singular point. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 231–235.

Chistov, A.L. and Grigor'ev, D.Y. (1984). Complexity of quantifier elimination in the theory of algebraically closed fields. In: Chytil, M.P. and Koubek, V. (eds.), *Proc. 11th Symposium MFCS*, Springer LNCS **176**, 17–31.

Chou, S.C. (1990). Automated reasoning in geometries using the characteristic set method and Gröbner basis method. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 255–260.

Chou, S.C. and Gao, X.S. (1990). Methods for mechanical geometry formula deriving. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 265–270.

Collins, G.E. (1973). Computer algebra of polynomials and rational functions. *Am. Math. Mon.* **80**/7, 725–755.

Collins, G.E. (1975). Quantifier elimination for the elementary theory of real closed fields by cylindrical algebraic decomposition. Springer LNCS **33**, 134–183.

Collins, G.E. and Loos, R. (1980). SAC-2—Symbolic and algebraic computation version 2, a computer algebra system, ALDES—Algorithm description language. *ACM SIGSAM Bulletin* **14**/2.

Collins, G.E. and Loos, R. (1982). Real zeroes of polynomials. In: Buchberger, B., Collins, G.E., and Loos, R. (eds.), *Computer Algebra, Symbolic and Algebraic Computation*, Springer-Verlag, New York, 83–94.

Conti, P. and Traverso, C. (1991). Buchberger algorithm and integer programming. In: Mattson, H.F., Mora, T., and Rao, T.R.N. (ed.), *Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes (AAECC 9)*, Springer LNCS **539**, 130–139.

Cooperman, G., Finkelstein, L., and Sarawagi, N. (1990). A random base change algorithm for permutation groups. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 161–168.

Cordero, P. and Chirardi, G.C. (1972). Realization of Lie algebras and the algebraic treatment of quantum problems. *Fortschr. Phys.* **20**, 105–133.

Cowell, R.G. (1992). Application of ordered standard bases to catastrophe theory. *J. Symb. Comp.* **13**/1, 101–115.

Czapor, S.R. and Geddes, K.O. (1986). On implementing Buchberger's algorithm for Gröbner bases. In: Char, B.W. (ed.), *1986 ACM Symposium on Symbolic and Algebraic Computation*, University of Waterloo, Ontario, 233–238.

Czapor, S.R. (1987). Solving algebraic equations via Buchberger's algorithm. In: Davenport, J.H. (ed.), *EUROCAL '87, European Conference on Computer Algebra*, Springer LNCS **378**, 260–269.

Czapor, S.R. (1989). Solving algebraic equations: Combining Buchberger's algorithm with multivariate factorization. *J. Symb. Comp.* **7**/1, 49–53.

Davenport, J.H. (1979). The computerisation of algebraic geometry. In: Ng, E.W. (ed.), *EUROSAM '79, An International Symposium on Symbolic and Algebraic Manipulation*, Springer LNCS **72**, 119–133.

Davenport, J.H. (1988). The world of computer algebra. *New Sci.* **1629**, 71–72.

Davenport, J.H. and Trager, B.M. (1981). Factorization over finitely generated fields. In: Wang, P.S. (ed.), *1981 ACM Symposium on Symbolic and Algebraic Computation*, ACM Press, New York, 200–205.

Della Dora, J., Dicrescenzo, C., and Duval., D. (1985). About a new method for computing in algebraic number fields. In: Caviness, B.F. (ed.), *EUROCAL '85, European Conference on Computer Algebra*, Vol. II, Springer LNCS **204**, 289–290.

Demazure, M. (1985). Notes informelles de calcul formel. *Prepublications du Centre de Mathematiques de l'Ecole Polytechnique* **3**.

Dickenstein, A.M. Sessa, C. (1990). Duality methods for the membership problem. In: Mora, T. and Traverso, C. (eds.), *Effective Methods in Algebraic Geometry*, Progress in Mathematics **94**, Birkhäuser Verlag, Basel, 89–103.

Dickson, L.E. (1913). Finiteness of the odd perfect and primitive abundant numbers with n distinct prime factors. *Am. J. Math.* **35**, 413–422.

Dress, A. and Schiffels, G. (1987). Noetherian quasi orders and canonical ideal bases. *Preprint.*

Dubé, T.W. (1990). The structure of polynomial ideals and Gröbner bases. *SIAM J. Comput.* **19**/4, 750–773.

Duval, D. (1989). Simultaneous computations in fields of different characteristics. In: Kaltofen, E. and Watt, S.M. (eds.), *Computers and Mathematics*, Springer-Verlag, New York, 321–326.

Ebert, G.L. (1983). Some comments on the modular approach to Gröbner bases. *ACM SIGSAM Bulletin* **17**/2, 28–32.

Eilenberg, S. and Schützenberger, M.P. (1969). Rational sets in commutative monoids. *J. Algebra* **13**, 173–191.

Eisenbud, D., Huneke, C., and Vasconcelos, W. (1992). Direct methods for primary decomposition. *Inventiones Mathem.* **110**/2, 207–236.

van den Essen, A. (1990). A criterion to decide if a polynomial map is invertible and to compute the inverse. *Comm. Algeb.* **18**/10, 3183–3186.

Evans, T. (1951). On multiplicative systems defined by generators and relations I. Normal form theorems. *Math. Proc. Camb. Philos. Soc.* **47**, 637–649.

Evans, T. (1951a). The word problem for abstract algebras. *J. Lond. Math.* **26**, 64–71.

Faugère, J.C., Gianni, P., Lazard, D., and Mora, T. (1990). Efficient computation of zero-dimensional Gröbner bases by change of ordering. To appear in: *J. Symb. Comp.*

Ferro, A. and Gallo, G. (1987). Gröbner bases, Ritt's algorithm, and decision procedures for algebraic theories. In: Huguet, L. and Poli, A. (eds.), *Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes (AAECC 5)*, Springer LNCS **356**, 230–237.

Fitchas, N., Galligo, A., and Morgenstern, J. (1987). Algorithmes rapides en sequentiel et en parallel pour l'élimination de quantificateurs en géometrie élémentaire. *Seminaire Structures Algébriques Ordonnées, UER des Math., Publ. Univ. Paris VII, 1990.*

Fitzpatrick, P. and Flynn, J. (1992). A Gröbner basis technique for Padé approximation. *J. Symb. Comp.* **13**/2, 133–138.

Fitchas, N. and Galligo, A. (1990). Nullstellensatz effectif et conjecture de Serre (theorème de Quillen-Suslin) pour le calcul formel. *Math. Nachr.* **149**, 231–253.

Fröhlich, A. and Shepherdson, J.C. (1955). Effective procedures in field theory. *Philos. Trans. R. Soc. London Ser. A* **248**, 407–432.

Furukawa, A., Sasaki, T., and Koboyashi, H. (1986). Gröbner bases of a module over $K[x_1, \ldots, x_n]$ and polynomial solutions of a system of linear equations. In: Char, B.W. (ed.), *1986 ACM Symposium on Symbolic and Algebraic Computation*, University of Waterloo, Ontario, 222–224.

Galligo, A. (1974). A propos du théorème de préparation de Weierstrass. In: *Fonctions de Plusieurs Variables Complexes. Seminaire F. Norguet (Oct. 1970-Dec. 1973)*, Springer Lec. Notes Math. **409**, 543–579

Galligo, A. (1979). Théorème de division et stabilité en géometrie analytique locale. *Ann. I. Four.* **XXIX**, 107–184.

Galligo, A. (1985). Some algorithmic questions on ideals of differential operators. In: Caviness, B.F. (ed.), *EUROCAL '85, European Conference on Computer Algebra*, Vol. II, Springer LNCS **204**, 413–421.

Galligo, A. (1990). Exemples d'ensembles de points en position uniforme. In: Mora, T. and Traverso, C. (eds.), *Effective Methods in Algebraic Geometry*, Progress in Mathematics **94**, Birkhäuser Verlag, Basel, 105–117.

Galligo, A., Pottier, L., and Traverso, C. (1988). Greater easy common divisor and standard basis completion algorithms. In: Gianni, P. (ed.), *International Symposium on Symbolic and Algebraic Computation (ISSAC '88)*, Springer LNCS **358**, 162–176.

Galligo, A., Pottier, L., and Traverso, C. (1988a). Appendix to: Greater easy common divisor and standard basis-completion algorithms. *Rapports de Recherche*, INRIA, Centre de Sophia Antipolis.

Galligo, A. and Traverso, C. (1989). Practical determination of the dimension of an algebraic variety. In: Kaltofen, E. and Watt, S.M. (eds.), *Computers and Mathematics*, Springer-Verlag, New York, 46–52.

Gallo, G., Mishra, B., and Ollivier, F. (1990). Efficient algorithms and bounds for Wu–Ritt characteristic sets. In: Mora, T. and Traverso, C. (eds.), *Effective Methods in Algebraic Geometry*, Progress in Mathematics **94**, Birkhäuser Verlag, Basel, 119–142.

Gallo, G. and Mishra, B. (1991). Some constructions in rings of differential polynomials. In: Mattson, H.F., Mora, T., and Rao, T.R.N. (ed.), *Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes (AAECC 9)*, Springer LNCS **539**, 171–182.

Gao, X.S. and Chou, S.C. (1990). Computations with parametric equations. In: Watt, S.M. (ed.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '91)*, ACM Press, New York, 122–127.

Gatermann, K. (1990). Symbolic solution of polynomial equation systems with symmetry. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 112–119.

von zur Gathen, J. (1984). Parallel algorithms for algebraic problems. *SIAM J. Comput.* **13**, 802–824.

von zur Gathen, J. (1990). Polynomials over finite fields with large images. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 140–144.

Gateva-Ivanova, T. and Latyshev, V. (1988). On recognisable properties of associative algebras. *J. Symb. Comp.* **6**/2,3, 371–388.

Gateva-Ivanova, T. (1990). Noetherian properties and growth of some associative algebras. In: Mora, T. and Traverso, C. (eds.), *Effective Methods in Algebraic Geometry*, Progress in Mathematics **94**, Birkhäuser Verlag, Basel, 143–158.

Gebauer, R. (1985). *A collection of examples for Gröbner basis calculations.* IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598.

Gebauer, R. and Kredel, H. (1983). *Common distributive polynomial system 04/1983, Distributive integral polynomial system 06/1983, Distributive rational polynomial system 08/1983, Distributive arbitrary domain polynomial system 12/1983.* Institut für Angewandte Mathematik, Universität Heidelberg.

Gebauer, R. and Kredel, H. (1984). An algorithm for constructing Gröbner bases of polynomial ideals. *ACM SIGSAM Bulletin* **18**/1.

Gebauer, R. and Kredel, H. (1984a). *Real solution system for algebraic equations.* Technical Report, Institut für Angewandte Mathematik, Universität Heidelberg.

Gebauer, R. and Möller, H.M. (1986). Buchberger's algorithm and staggered linear bases. In: Char, B.W. (ed.), *1986 ACM Symposium on Symbolic and Algebraic Computation*, University of Waterloo, Ontario, 218–221.

Gebauer, R. and Möller, H.M. (1988). On an installation of Buchberger's algorithm. *J. Symb. Comp.* **6**/2/3, 275–286.

Gerdt, V.P. and Zharkov, A.Y. (1990). Computer generation of necessary integrability conditions for polynomial-nonlinear evolution systems. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 250–254.

Gerdt, V.P., Khutornoy, N.V., and Zharkov, A.Y. (1990). Solving algebraic systems which arise as necessary integrability conditions for polynomial-nonlinear evolution equations. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 299.

Gianni, P. (1987). Properties of Gröbner bases under specializations. In: Davenport, J.H. (ed.), *EUROCAL '87, European Conference on Computer Algebra*, Springer LNCS **378**, 293–297.

Gianni, P., Miller, V., and Trager, B. (1988). Decomposition of algebras. In: Gianni, P. (ed.), *International Symposium on Symbolic and Algebraic Computation (ISSAC '88)*, Springer LNCS **358**, 300–308.

Gianni, P. and Mora, T. (1989). Algebraic solution of systems of polynomial equations using Gröbner bases. In: Huguet, L. and Poli, A. (eds.), *Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes (AAECC 5)*, Springer LNCS **356**, 247–257.

Gianni, P. and Trager, B. (1985). GCD's and factoring multivariate polynomials using Gröbner bases. In: Caviness, B.F. (ed.), *EUROCAL '85, European Conference on Computer Algebra*, Vol. II, Springer LNCS **204**, 409–410.

Gianni, P., Trager, B., and Zacharias, G. (1988). Gröbner bases and primary decomposition of polynomial ideals. *J. Symb. Comp.* **6**/2,3, 149–167.

Giovini, A., Mora,T., Niesi, G., Robbiano, L., and Traverso, C. (1991). "One sugar cube, please," or Selection strategies in the Buchberger algorithm. In: Watt, S.M. (ed.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '91)*, ACM Press, New York, 49–54.

Giusti, M. (1984). Some effectivity problems in polynomial ideal theory. In: Fitch, J. (ed.), *EUROSAM '84, International Symposium on Symbolic and Algebraic Computation*, Springer LNCS **174**, 159–171.

Giusti, M. (1985). A note on the complexity of constructing standard bases. In: Caviness, B.F. (ed.), *EUROCAL '85, European Conference on Computer Algebra*, Vol. II, Springer LNCS **204**, 411–412.

Giusti, M. (1987). Complexity of standard bases in projective dimension zero. In: Davenport, J.H. (ed.), *EUROCAL '87, European Conference on Computer Algebra*, Springer LNCS **378**, 333–335.

Giusti, M. (1988). Combinatiorial dimension theory of algebraic varieties. *J. Symb. Comp.* **6**/2,3, 249–265.

Giusti, M. (1990). Complexity of standard bases in projective dimension zero II. In: Sakata, S. (ed.), *Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes (AAECC 8)*, Springer LNCS **508**, 322–328.

Giusti, M. and Heintz, J. (1990). Algorithmes—disons rapides—pour la décomposition d'une variété algébrique en composantes irréducibles et équidimensionelles. In: Mora, T. and Traverso, C. (eds.), *Effective Methods in Algebraic Geometry*, Progress in Mathematics **94**, Birkhäuser Verlag, Basel, 169–194.

Giusti, M. and Lazard, D. (1986). Complexity of standard basis computations, related algebraic problems and their common double exponential behaviour. *Lecture Notes for the Conference "Computer and Commutative Algebra,"* Genova, Italy, 1986. Prepublication du Centre de Mathematiques de l'Ecole Polytechnique.

Giusti, M., Lazard, D., and Valibouze, A. (1988). Algebraic transformations of polynomial equations, symmetric polynomials and elimination. In: Gianni, P. (ed.), *International Symposium on Symbolic and Algebraic Computation (ISSAC '88)*, Springer LNCS **358**, 309–314.

Grieco, M. and Zucchetti, B. (1989). How to decide whether a polynomial ideal is primary or not. In: Huguet, L. and Poli, A. (eds.), *Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes (AAECC 5)*, Springer LNCS **356**, 258–268.

Grigor'ev, D.Y. (1987). Computational complexity in polynomial algebra. In: Gleason, A.M. (ed.), *Proc. International Congress of Mathematicians*, Vol. 2, Berkeley, CA, 1452–1460.

Grigor'ev, D.Y. (1990). Complexity of solving systems of linear equations over the rings of differential operators. In: Mora, T. and Traverso, C. (eds.), *Effective Methods in Algebraic Geometry*, Progress in Mathematics **94**, Birkhäuser Verlag, Basel, 195–202.

Grigor'ev, D.Y. (1990a). How to test in subexponential time whether two points can be connected by a curve in a semialgebraic set. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 104–105.

Grigor'ev, D.Y. (1990b). Complexity of irreducibility testing for a system of linear ordinary differential equations. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 225–230.

Grigor'ev, D.Y. and Chistov, A.L. (1984). Fast decomposition of polynomials into irreducible ones and the solution of systems of algebraic equations. *Soviet Math. Dokl.* **29**/2, 380–383.

Grigor'ev, D.Y. and Vorobjov, N.N. (1988). Solving systems of polynomial inequalities in subexponential time. *J. Symb. Comp.* **5**/1,2, 37–64.

Gröbner, W. (1950). Über die Eliminationstheorie. *Monats. Math.* **54**, 71–78.

Habicht, W. (1948). Zur inhomogenen Eliminationstheorie. *Comm. Math. Helv.* **21**, 79–98.

Hamel, G. (1905). Eine Basis aller Zahlen und die unstetigen Lösungen der Funktionalgleichung: $f(x+y) = f(x) + f(y)$. *Math. Ann.* **60**, 459–462.

Heintz, J. (1976). Untere Schranken für die Komplexität logischer Entscheidungsprobleme. In: Specker, E. and Strasser, V. (eds.), *Komplexität von Entscheidungsproblemen*, Springer-Verlag Heidelberg, 127–137.

Heintz, J. (1983). Definability and fast quantifier elimination in algebraically closed fields. *Theor. Comput. Sci.* **24**, 239–277; Russian transl. in: *Kyberneticeskij Sbornik Novaja Serija Vyp.* **22**, Mir Moscow, 1985, 113–158.

Heintz, J. and Wüthrich, R. (1975). An efficient quantifier elimination algorithm for algebraically closed fields of any characteristic. *SIGSAM Bulletin* **9**/4, 11.

Herford, W. and Penz, H. (1991). A new notion of reduction: generating universal Gröbner bases of ideals in $K[x,y]$. *J. Symb. Comp.* **12**/6, 585–605.

Hermann, G. (1926). Die Frage der endlich vielen Schritte in der Theorie der Polynomideale. *Math. Ann.* **95**, 736–788.

Herzog, J. and Trung, N.V. (1992). Gröbner bases and multiplicity of determinantal and Pfaffian ideals. *Adv. Math.* **96**, 1–37.

Higman, G. (1952). Ordering by divisibility in abstract algebras. *Proc. London Math. Soc.* **3**/2, 326–336.

Hilbert, D. (1890). Über die Theorie der algebraischen Formen. *Math. Ann.* **36**, 473–534.

Hilbert, D. (1893). Über die vollen Invariantensysteme. *Math. Ann.* **42**, 313-373.

Hironaka, H. (1964). Resolution of singularities of an algebraic variety over a field of characteristic zero. *Ann. Math.* **79**, 109-326.

Hong, H. (1990). An improvement of the projection operator in cylindrical algebraic decomposition. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 261-264.

Hopcroft, J.E. and Krafft, D.B. (1985). The challenge of robotics for computer science. In: Schwartz, J.T. and Yap, C.K. (eds.), *Advances in Robotics: Algorithmic and Geometric Aspects of Robotics*, Vol. 1, 7–42.

Hsiang, J. (1985). Two results in term rewriting theorem proving. In: Jouannaud, J.P. (ed.), *Rewriting Techniques and Applications*, Springer LNCS **202**, 301–324.

Huet, G. (1978). An algorithm to generate the basis of solutions to homogeneous linear diophantine equations. *Inf. Process. Lett.* **7**/3, 144–147.

Huet, G. (1980). Confluent reductions: Abstract properties and applications to term rewriting systems. *J. ACM* **27**/4, 797–821.

Huet, G. and Oppen, D.C. (1980). Equations and rewrite rules: a survey. In: Book, R. (ed.) *Formal Languages. Perspectives and Open Problems*, Academic Press, London.

Huynh, D.T. (1986). A superexponential lower bound for Gröbner bases and Church–Rosser commutative Thue systems. *Inf. Contr.* **68**/1, 196–206.

Huynh, D.T. (1986a). The complexity of the membership problem for two subclasses of polynomial ideals. *SIAM J. Comput.* **15**, 581–594.

Kalkbrener, M. (1987). Solving systems of algebraic equations by using Gröbner bases. In: Davenport, J.H. (ed.), *EUROCAL '87, European Conference on Computer Algebra*, Springer LNCS **378**, 282–292.

Kalkbrener, M. (1990). Solving systems of bivariate algebraic equations by using primitive polynomial remainder sequences. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 295.

Kalkbrener, M. (1990a). Implicitization of rational parametric curves and surfaces. In: Sakata, S. (ed.), *Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes (AAECC 8)*, Springer LNCS **508**, 249–259.

Kaltofen, E. (1982). Factorization of polynomials. In: Buchberger, B., Collins, G.E., and Loos, R. (eds.), *Computer Algebra, Symbolic and Algebraic Computation*, Springer-Verlag, New York, 95–113.

Kaltofen, E., Lakshman, Y.N., and Wiley, J.M. (1990). Modular rational sparse multivariate polynomial interpolation. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 135–139.

Kandri-Rody, A. (1985). Dimension of ideals in polynomial rings. In: Avenhaus, J. and Madlener, K. (eds.), *Proc. Combinatorial Algorithms in Algebraic Structures, Otzenhausen 1985*, Fachbereich Informatik, University of Kaiserslautern.

Kandri-Rody, A. and Kapur, D. (1983). *On the relationship between Buchberger's Gröbner basis algorithm and the Knuth–Bendix completion procedure.* TIS Report No. 83CRD286, General Electric Research and Development Center, Schenectady, NY.

Kandri-Rody, A. and Kapur, D. (1984). Algorithms for computing Gröbner bases of polynomial ideals over various Euclidean rings. In: Fitch, J. (ed.), *EUROSAM '84, International Symposium on Symbolic and Algebraic Computation*, Springer LNCS **174**, 195–206.

Kandri-Rody, A. and Kapur, D. (1988). Computing a Gröbner basis of a polynomial ideal over a Euclidean domain. *J. Symb. Comp.* **6**/1, 37–57.

Kandri-Rody, A. and Saunders, B.D. (1984). Primality of ideals in polynomial rings. In: *Third MACSYMA Users' Conference*, Schenectady, NY, July 1984.

Kandri-Rody, A., Kapur, D., and Narendran, P. (1986). An ideal-theoretic approach to word problems and unification problems over finitely presented commutative algebras. In: Jouannaud, J.P. (ed.), *Rewriting Techniques and Applications*, Springer LNCS **202**, 345–364.

Kandri-Rody, A. and Weispfenning, V. (1990). Non-commutative Gröbner bases in algebras of solvable type. *J. Symb. Comp.* **9**/1, 1–26.

Kannan, R. (1985). Solving systems of linear equations over polynomials. *Theor. Comput. Sci.* **39**/1, 69–88.

Kannan, R. and Bachem, A. (1979). Polynomial algorithms for computing the Smith and Hermite normal forms of an integer matrix. *SIAM J. Comput.* **8**/4, 499–507.

Kapur, D. (1986). Geometry theorem proving using Hilbert's Nullstellensatz. In: Char, B.W. (ed.), *1986 ACM Symposium on Symbolic and Algebraic Computation*, University of Waterloo, Ontario, 202–208

Kapur, D. (1986a). Using Gröbner bases to reason about geometry problems. *J. Symb. Comp.* **2**/4, 1986, 399–408.

Kapur, D. and Madlener, K. (1989). A completion procedure for computing a canonical basis for a k-subalgebra. In: Kaltofen, E. and Watt, S.M. (eds.), *Computers and Mathematics*, Springer-Verlag, New York, 1–11.

Kapur, D. and Narendran, P. (1985). Existence and construction of a Gröbner basis for a polynomial ideal. Presented at a workshop on combinatorial algorithms in algebraic structures, Europäische Akademie, Otzenhausen, FRG.

Kapur, D. and Wan, H.K. (1990). Refutational proofs of geometry theorems via characteristic set computation. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 277–284.

Keller, O.H. (1939). Ganze Cremona Transformationen. *Monatsh. Math. Phys.* **47**, 299–306.

Knuth, D.E. and Bendix, P.B. (1970). Simple word problems in universal algebras. In: Leech, J. (ed.), *Computational Problems in Abstract Algebra*, Pergamon Press, 263–297.

Kobayashi, H. Fujise, T., and Furukawa, A. (1988). Solving systems of algebraic equations by a general elimination method. *J. Symb. Comp.* **5**, 303–320.

Kobayashi, H., Furukawa, A., and Sasaki, T. (1986). Gröbner bases of ideals of convergent power series. In: Char, B.W. (ed.), *1986 ACM Symposium on Symbolic and Algebraic Computation*, University of Waterloo, Ontario, 225–227.

Kobayashi, H., Moritsugu, S., and Hogan, R.W. (1988). Solving systems of algebraic equations. In: Gianni, P. (ed.), *International Symposium on Symbolic and Algebraic Computation (ISSAC '88)*, Springer LNCS **358**, 139–149.

Kobayashi, H., Moritsugu, S., and Hogan R.W. (1989). On radical zero-dimensional ideals. *J. Symb. Comp.* **8**/6, 545–552.

Kodaira, H. and Toshima, H. (1985). Gini coefficient of wealth in life cycle model. In: *The Second International Symposium on Symbolic and Algebraic Computation by Computers*, RIKEN, Wako-Shi, Saitama, 351-01, Japan, 11-1–11-32.

Kohno, M. (1990). Reduction problems in the theory of differential equations. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 244–249.

Kollreider, C. (1978). *Polynomial reduction: the influence of the ordering of terms on a reduction algorithm.* Technical Report Nr. 124, Universität Linz.

Kollreider, M. and Buchberger, B. (1978). *An improved algorithmic construction of Gröbner bases for polynomial ideals.* Technical Report Nr. 170, Universität Linz.

Kolman, B. (1976). Computers in the study of Lie algebras. In: Jenks, R.D. (ed.), *1976 ACM Symposium on Symbolic and Algebraic Computation*, ACM Press, New York, 300–311.

Kolyada, S.V. (1990). Systems for symbolic computations in Boolean algebra. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 291.

Kondrat'eva, M.V. and Pankrat'ev, E.V. (1987). A recursive algorithm for the computation of the Hilbert polynomial. In: Davenport, J.H. (ed.), *EUROCAL '87, European Conference on Computer Algebra*, Springer LNCS **378**, 365–375.

Kovacs, P. (1991). Minimum degree solutions for the inverse kinematics problem by application of the Buchberger algorithm. In: Lenarcic, J. and Stifter, S. (eds.), *Advances in Robot Kinematics*, Springer-Verlag, New York, 326–334.

Kredel, H. (1987). Primary ideal decomposition. In: Davenport, J.H. (ed.), *EUROCAL '87, European Conference on Computer Algebra*, Springer LNCS **378**, 270–281.

Kredel, H. (1988). Admissible term orderings used in computer algebra systems. *ACM SIGSAM Bulletin* **22**/1, 28–31.

Kredel. H. (1988a). Real roots of zero-dimensional ideals. *MIP-8824*, Universität Passau.

Kredel, H. (1990). MAS: Modula-2 algebra system. In: Gerdt, V.P., Rostovtsev, V.A., and Shirkov, D.V. (eds.), *IV International Conference on Computer Algebra in Physical Research*, World Scientific Publishing Co., Singapore, 31–34.

Kredel, H. (1990a). Computing in polynomial rings of solvable type. In: Gerdt, V.P., Rostovtsev, V.A., and Shirkov, D.V. (eds.), *IV International Conference on Computer Algebra in Physical Research*, World Scientific Publishing Co., Singapore, 211–221.

Kredel, H. and Weispfenning, V. (1988). Computing dimension and independent sets for polynomial ideals. *J. Symb. Comp.* **6**/2,3, 231–247.

Kredel, H. and Weispfenning, V. (1990). Parametric Gröbner bases for non-commutative polynomials. In: Gerdt, V.P., Rostovtsev, V.A., and Shirkov, D.V. (eds.), *IV International Conference on Computer Algebra in Physical Research*, World Scientific Publishing Co., Singapore, 236–246.

Krick, T. and Logar, A. (1991). An algorithm for the computation of the radical of an ideal in the ring of polynomials. In: Mattson, H.F., Mora, T., and Rao, T.R.N. (ed.), *Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes (AAECC 9)*, Springer LNCS **539**, 195–205.

Krick, T. and Logar, A. (1990). Membership problem, representation problem and the computation of the radical for one-dimensional ideals. In: Mora, T. and Traverso, C. (eds.), *Effective Methods in Algebraic Geometry*, Progress in Mathematics **94**, Birkhäuser Verlag, Basel, 203–216.

Kronecker, L. (1882). Grundzüge einer arithmetischen Theorie der algebraischen Grössen. *Crelle, J. Reine Angew. Math.* **92**, 1–122.

Krull, W. (1928). Primidealketten in allgemeinen Ringbereichen. *Sitzungsberichte Heidelberger Akad.*, 7. Abh.

Kruskal, J.B. (1972). The theory of well-quasi-ordering: A frequently discovered concept. *J. Comb. Th. A* **13**, 297–305.

Kutzler, B. and Stifter, S. (1986). New approaches to computerized proofs of geometry theorems. In: Davenport, J.H. and Gebauer, R. (eds.), *Proc. Computers and Mathematics*, Stanford.

Kutzler, B. and Stifter, S. (1986a). Automated geometry theorem proving using Buchberger's algorithm. In: Char, B.W. (ed.), *1986 ACM Symposium on Symbolic and Algebraic Computation*, University of Waterloo, Ontario, 209–214.

Kutzler, B. and Stifter, S. (1986b). On the application of Buchberger's algorithm to automated geometry theorem proving. *J. Symb. Comp.* **2**/4, 389–397.

Labonté, G. (1990). An algorithm for the construction of matrix representations for finitely presented non-commutative algebras. *J. Symb. Comp.* **9**/1, 27–38.

Labonté, G. (1990a). On the automatic construction of representations of noncommutative algebras. In: Gerdt, V.P., Rostovtsev, V.A., and Shirkov, D.V. (eds.), *IV International Conference on Computer Algebra in Physical Research*, World Scientific Publishing Co., Singapore, 122–126.

Lakshman, Y.N. (1990). A single exponential bound on the complexity of computing Gröbner bases of zero dimensional ideals. In: Mora, T. and Traverso, C. (eds.), *Effective Methods in Algebraic Geometry*, Progress in Mathematics **94**, Birkhäuser Verlag, Basel, 227–234.

Lakshman, Y.N. and Lazard, D. (1990). On the complexity of zero-dimensional algebraic systems. In: Mora, T. and Traverso, C. (eds.), *Effective Methods in Algebraic Geometry*, Progress in Mathematics **94**, Birkhäuser Verlag, Basel, 217–225.

Landau, S. (1985). Factoring polynomials over algebraic number fields. *SIAM J. Comput.* **14**/1, 184–195.

Landau, S. and Miller, G.L. (1985). Solvability by radicals is in polynomial time. *J. Comput. Syst. Sci.* **30**/2, 179–208.

Langemyr, L. (1990). Algorithms for a multiple algebraic extension. In: Mora, T. and Traverso, C. (eds.), *Effective Methods in Algebraic Geometry*, Progress in Mathematics **94**, Birkhäuser Verlag, Basel, 235–248.

Langemyr, L. (1991). Algorithms for a multiple algebraic extension II. In: Mattson, H.F., Mora, T., and Rao, T.R.N. (ed.), *Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes (AAECC 9)*, Springer LNCS **539**, 224–233.

Lankford, D. (1989). Generalized Gröbner bases: Theory and applications. A condensation. In: Dershowitz, N. (ed.), *Rewriting Techniques and Applications*, Springer LNCS **355**, 203–221.

Lasker, E. (1905). Zur Theorie der Moduln und Ideale. *Math. Ann.* **60**, 20–116.

Lassner, W. (1980). Noncommutative algebras prepared for computer calculations. In: *Conference on Systems and Techniques of Analytical Computing and Their Applications in Theoretical Physics*, Dubna, 58–74.

Lassner, W. (1985). Symbol representations of noncommutative algebras. In: Caviness, B.F. (ed.), *EUROCAL '85, European Conference on Computer Algebra*, Vol. II, Springer LNCS **204**, 99–115.

Lauer, M. (1976). Algorithms for symmetrical polynomials. In: Jenks, R.D. (ed.), *1976 ACM Symposium on Symbolic and Algebraic Computation*, ACM Press, New York, 242–247.

Lauer, M. (1976a). Canonical representatives for the residue classes of a polynomial ideal. In: Jenks, R.D. (ed.), *1976 ACM Symposium on Symbolic and Algebraic Computation*, ACM Press, New York, 339–345.

Lazard, D. (1977). Elimination non linéaire. In: *Symbolic Computational Methods and Applications, St. Maximin 1977*, 284–286.

Lazard, D. (1977a). Algèbre linéaire sur $K[X_1, ..., X_n]$ et élimination. *B. S. Math. Fr.* **105**, 165–190.

Lazard, D. (1979). Systems of algebraic equations. In: Ng, E.W. (ed.), *EUROSAM '79, An International Symposium on Symbolic and Algebraic Manipulation*, Springer LNCS **72**, 88–94.

Lazard, D. (1981). Resolution des systèmes d'équations algébriques. *Theor. Comput. Sci.* **15**, 77–110.

Lazard, D. (1982). Commutative algebra and computer algebra. In: Calmet, J. (ed.), *EUROCAM '82, European Computer Algebra Conference*, Springer LNCS **144**, 40–48.

Lazard, D. (1982a). On polynomial factorization. In: Calmet, J. (ed.), *EUROCAM '82, European Computer Algebra Conference*, Springer LNCS **144**, 126–134.

Lazard, D. (1983). Gröbner bases, Gaussian elimination, and resolution of systems of algebraic equations. In: van Hulzen, J.A. (ed.), *EUROCAL '83, European Computer Algebra Conference*, Springer LNCS **162**, 146–156.

Lazard, D. (1985). Ideal bases and primary decomposition: Case of two variables. *J. Symb. Comp.* **1**/3, 261–270.

Lazard D. (1992). Solving zero-dimensional algebraic systems. *J. Symb. Comp.* **13**/2, 117–131.

Lazard D. (1992a). A note on upper bounds for ideal-theoretic problems. *J. Symb. Comp.* **13**/3, 231–233.

Lehmer, D.H. (1961). A machine method for solving polynomial equations. *J. ACM* **8**/2, 151–162.

Lenstra, A.K. (1984). Polynomial factorization by root approximation. In: Fitch, J. (ed.), *EUROSAM '84, International Symposium on Symbolic and Algebraic Manipulation*, Springer LNCS **174**, 272–276.

Lenstra, A.K. (1987). Factoring multivariate polynomials over algebraic number fields. *SIAM J. Comput.* **16**/3, 591–598.

Lenstra, A.K., Lenstra, H.W., and Lovász, L. (1982). Factoring polynomials with rational coefficients. *Math. Ann.* **261**, 515–534.

Liu, Z-J. (1990). An algorithm for finding all isolated zeros of polynomial systems. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 300.

Lipson, J.D. (1971). Chinese remainder and interpolation algorithms. In: Petrick, S.R. (ed.), *Second Symposium on Symbolic and Algebraic Manipulation*, ACM Press, New York, 372–391.

LLovet, J. and Sendra, J.R. (1990). A modular approach to the computation of the number of real roots. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 298.

Logar, A. (1988). A computational proof of the Noether normalization lemma. In: Mora, T. (ed.), *Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes (AAECC 6)*, Springer LNCS **357**, 259–273.

Loos, R. (1982). Computing in algebraic extensions. In: Buchberger, B., Collins, G.E., and Loos, R. (eds.), *Computer Algebra, Symbolic and Algebraic Computation*, Springer-Verlag, New York, 173–187.

Lüneburg, H. (1987). On the computation of the Smith normal form. In: Janßen, R. (ed.), *Trends in Computer Algebra*, Springer LNCS **296**, 156–157.

Lüroth, P. (1876). Beweis eines Satzes über rationale Curven. *Math. Ann.* **9**, 163–165.

Macaulay, F.S. (1902). Some formulae in elimination theory. *Proc. London Math. Soc.* (1) **35**, 3–27.

Macaulay, F.S. (1927). Some properties of enumeration in the theory of modular systems. *Proc. London Math. Soc.* **26**, 531–555.

Macintyre, A. (1977). Model completeness. In: Barwise, J. (ed.), *Handbook of Mathematical Logic*, North-Holland, Amsterdam, 139-180.

Malashonok, G.I. (1990). Algorithms for the solution of systems of linear equations in commutative rings. In: Mora, T. and Traverso, C. (eds.), *Effective Methods in Algebraic Geometry*, Progress in Mathematics **94**, Birkhäuser Verlag, Basel, 289–298.

Manocha, D. (1990). Regular curves and proper parametrizations. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 271–276.

Malle, G. and Trinks, W. (1984). *Zur Behandlung algebraischer Gleichungssysteme mit dem Computer.* Mathematisches Institut, Universität Karlsruhe, unpublished manuscript.

Marinari, M.G. and Möller, H.M., and Mora. T. (1991). Gröbner bases of ideals given by dual bases. In: Watt, S.M. (ed.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '91)*, ACM Press, New York, 55–63.

Matzat, B.H. (1987). Computational methods in constructive Galois theory. In: Janßen, R. (ed.), *Trends in Computer Algebra*, Springer LNCS **296**, 137–155.

Mayr, E.W. and Meyer, A.R. (1982). The complexity of the word problems for commutative semigroups and polynomial ideals. *Adv. Math.* **46**, 305–329.

McClellan, M.T. (1971). The exact solution of systems of linear equations with polynomial coefficients. In: Petrick, S.R. (ed.), *Second Symposium on Symbolic and Algebraic Manipulation*, ACM Press, New York, 399–414.

Melenk, H. (1990). Practical application of Gröbner bases for the solution of polynomial equation systems. In: Gerdt, V.P., Rostovtsev, V.A., and Shirkov, D.V. (eds.), *IV International Conference on Computer Algebra in Physical Research*, World Scientific Publishing Co., Singapore, 230–235.

Melenk, H., Möller, H.M., and Neun, W. (1989). Symbolic solution of large stationary chemical kinetics problems. *IMPACT of Computing in Science and Engineering* **1**, 138–167.

Mignotte, M. (1976). Some problems about polynomials. In: Jenks, R.D. (ed.), *1976 ACM Symposium on Symbolic and Algebraic Computation*, ACM Press, New York, 227–228.

Mignotte, M. (1982). Some useful bounds. In: Buchberger, B., Collins, G.E., and Loos, R. (eds.), *Computer Algebra, Symbolic and Algebraic Computation*, Springer-Verlag, New York, 259–263.

Mignotte, M. and Pasteur, U.L. (1981). Some inequalities about univariate polynomials. In: Wang, P.S. (ed.), *1981 ACM Symposium on Symbolic and Algebraic Computation*, ACM Press, New York, 195–199.

Miola, A. and Mora, T. (1988). Constructive lifting in graded structures: A unified view of Buchberger and Hensel methods. *J. Symb. Comp.* **6**/2,3, 305–322.

Mishra, B. and Yap, C. (1989). Notes on Gröbner bases. *Inf. Sci.* **48**, 219–252.

Mishra, B. and Pedersen, P. (1990). Arithmetic with real algebraic numbers is in NC. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 120–126.

Möller, H.M. (1985). A reduction strategy for the Taylor resolution. In: Caviness, B.F. (ed.), *EUROCAL '85, European Conference on Computer Algebra*, Vol. II, Springer LNCS **204**, 526–534.

Möller, H.M. (1988). On the construction of Gröbner bases using syzygies. *J. Symb. Comp.* **6**/2,3, 345–359.

Möller, H.M. (1990). Computing syzygies a la Gauss–Jordan. In: Mora, T. and Traverso, C. (eds.), *Effective Methods in Algebraic Geometry*, Progress in Mathematics **94**, Birkhäuser Verlag, Basel, 335–345.

Möller, H.M. and Buchberger, B. (1982). The construction of multivariate polynomials with preassigned zeros. In: Calmet, J. (ed.), *EUROCAM '82, European Computer Algebra Conference*, Springer LNCS **144**, 24–31.

Möller, H.M. and Mora, T. (1983). The computation of the Hilbert fuction. In: van Hulzen, J.A. (ed.), *EUROCAL '83, European Computer Algebra Conference*, Springer LNCS **162**, 157–167.

Möller, H.M. and Mora, T. (1984). Upper and lower bounds for the degree of Gröbner bases. In: Fitch, J. (ed.), *EUROSAM '84, International Symposium on Symbolic and Algebraic Computation*, Springer LNCS **174**, 172–183.

Möller, H.M. and Mora, T. (1984a). Computational aspects of reduction strategies to construct resolutions of monomial ideals. In: Poli, A. (ed.), *Applied Algebra, Algorithmics, and Error-Correcting Codes (AAECC 2)*, Springer LNCS **228**, 182–197.

Möller, H.M. and Mora, T. (1986a). New constructive methods in classical ideal theory. *J. Algebra* **100**, 138–178.

Möller, H.M., Mora, T., and Traverso, C. (1992). Gröbner basis computation using syzygies. *Preprint.*

Moenck, R.T. (1977). On the efficiency of algorithms for polynomial factoring. *Math. Comput.* **31**, 235–250.

Mora, T. (1982). An algorithm to compute the equations of tangent cones. In: Calmet, J. (ed.), *EUROCAM '82, European Computer Algebra Conference*, Springer LNCS **144**, 158–165.

Mora, T. (1983). A constructive characterization of standard bases. In: *Algebra e Geometria*, Bolletino U.M.I., Serie VI, II-D, 41–50.

Mora, T. (1986). Gröbner bases for non-commutative polynomial rings. In: Calmet, J. (ed.), *Algebraic Algorithms and Error-Correcting Codes (AAECC 3)*, Springer LNCS **229**, 353–362.

Mora, T. (1988). Standard bases and non-Noetherianity: non-commutative polynomial rings. In: Beth, T. and Clausen, M. (eds.), *Applicable Algebra, Error-Correcting Codes, Combinatorics, and Computer Algebra (AAECC 4)*, Springer LNCS **307**, 98–109.

Mora, T. (1988a). Seven Variations on Standard Bases. *Preprint*, Dipartimento di Matematica, Università di Genova, Via L.B. Alberti 4, 16132 Genova (Italy).

Mora, T. (1988b). Gröbner bases in non-commutative algebras. In: Gianni, P. (ed.), *International Symposium on Symbolic and Algebraic Computation (ISSAC '88)*, Springer LNCS **358**, 150–161.

Mora, T. and Robbiano, L. (1988). The Gröbner fan of an ideal. *J. Symb. Comp.* **6**/2,3, 183–208.

Moreno Socías, G. (1991). An Ackermannian polynomial ideal. In: Mattson, H.F., Mora, T., and Rao, T.R.N. (ed.), *Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes (AAECC 9)*, Springer LNCS **539**, 269–280.

Mortisugu, S. and Inada, N. (1985). Symbolic Newton iteration and its applications. In: *The Second International Symposium on Symbolic and Algebraic Computation by Computers*, RIKEN, Wako-Shi, Saitama, 351-01, Japan, 10-1–10-13.

Moses, J. (1971). Algebraic simplification: A guide for the perplexed. In: Petrick, S.R. (ed.), *Second Symposium on Symbolic and Algebraic Manipulation*, ACM Press, New York, 282–304 and *C. ACM* **14/8**, 527–537.

Noether, E. (1916). Der Endlichkeitssatz der Invarianten endlicher Gruppen. *Math. Ann.* **77**, 81–92.

Noether, E. (1921). Idealtheorie in Ringbereichen. *Math. Ann.* **83**, 24–66.

Noether, E. (1923). Eliminationstheorie und allgemeine Idealtheorie. *Math. Ann.* **90**, 229–261.

Noether, E. and Schmeidler, W. (1920). Moduln in nichtkommutativen Bereichen, insbesondere aus Differential- und Differenzausdrücken. *Math. Z.* **8**, 1–35.

Noether, M. (1873). Über einen Satz aus der Theorie der algebraischen Funktionen. *Math. Ann.* **6**, 351–359.

Norman, A.C. (1990). A critical-pair completion based integration algorithm. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 201–205.

Okubo, K. (1990). Global theory of ordinary differential equations and formula manipulation. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 193–200.

Ollivier, F. (1990). Canonical bases: Relations with standard bases, finiteness conditions and application to tame automorphisms. In: Mora, T. and Traverso, C. (eds.), *Effective Methods in Algebraic Geometry*, Progress in Mathematics **94**, Birkhäuser Verlag, Basel, 379–400.

Ollivier, F. (1990a). Standard bases of differential ideals. In: Sakata, S. (ed.), *Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes (AAECC 8)*, Springer LNCS **508**, 304–321.

Pan, L. (1989). On the D-bases of polynomial ideals over principal ideal domains. *J. Symb. Comp.* **7/1**, 55–69.

Pankrat'ev, E.V. (1989). Computations in differential and difference modules. *Acta Applicandae Mathematicae* **16**, 167–189.

Pauer, F. (1992). On lucky ideals for Gröbner basis computations. To appear in: *J. Symb. Comp.*

Pavelle, R., Rothstein, M., and Fitch, J. (1981). Computer algebra. *Sci. Amer.* **245**, 136–152.

Pfister, G. (1990). The tangent cone algorithm and some applications to local algebraic geometry. In: Mora, T. and Traverso, C. (eds.), *Effective Methods in Algebraic Geometry*, Progress in Mathematics **94**, Birkhäuser Verlag, Basel, 401–409.

Pohst, M. and Yun, D.Y.Y. (1981). On solving systems of algebraic equations via ideal bases and elimination theory. In: Wang, P.S. (ed.), *1981 ACM Symposium on Symbolic and Algebraic Computation*, ACM Press, New York, 206–211.

Rabin, M. (1977). Decidable theories. In: Barwise, J. (ed.), *Handbook of Mathematical Logic*, North-Holland, Amsterdam, 595–629.

Rabinowitsch, J.L. (1930). Zum Hilbertschen Nullstellensatz. *Math. Ann.* **102**, 520.

Renegar, J. (1992). On the computational complexity and geomerty of the first-order theory of the reals, Parts I–III. *J. Symb. Comp.* **13**/3, 255–352.

Richman, F. (1974). Constructive aspects of Noetherian rings. *P. AMS* **44**/2, 436–441.

Ritter, G. and Weispfenning, V. (1991). On the number of term orders. *AAECC* **2**, 55–79.

Robbiano, L. (1985). Term orderings on the polynomial ring. In: Caviness, B.F. (ed.), *EUROCAL '85, European Conference on Computer Algebra*, Vol. II, Springer LNCS **204**, 513–517.

Robbiano, L. (1986). On the theory of graded structures. *J. Symb. Comp.* **2**/2, 139–170.

Robbiano, L. (1988). Computer and commutative algebra. In: Mora, T. (ed.), *Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes (AAECC 6)*, Springer LNCS **357**, 31–44.

Robbiano, L. (1990). Bounds for degrees and number of elements in Gröbner bases. In: Sakata, S. (ed.), *Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes (AAECC 8)*, Springer LNCS **508**, 292–303.

Robbiano, L. and Sweedler, M. (1988). Subalgebra bases. In: Bruns, W. and Simis, A. (eds.), *Proc. Commutative Algebra Salvador*, Springer Lecture Notes in Mathematics **1430**, 61–87.

Robbiano, L. and Valla, G. (1980). On the equations defining tangent cones. *Math. Proc. Camb. Philos. Soc.* **88**, 281–297.

Rutman, E.W. (1992). Gröbner bases and primary decomposition of modules. *J. Symb. Comp.* **14**/5, 483–503.

Sakata, S. (1988). N-dimensional Berlekamp–Massey algorithm for multiple arrays and construction of multivariate polynomials with preassigned zeros. In: Mora, T. (ed.), *Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes (AAECC 6)*, Springer LNCS **357**, 356–376.

Sakata, S. (1989). Synthesis of two-dimensional linear feedback shift registers and Gröbner bases. In: Huguet, L. and Poli, A. (eds.), *Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes (AAECC 5)*, Springer LNCS **356**, 394–407.

Sarges, H. (1976). Ein Beweis des Hilbertschen Basissatzes. *J. reine Angew. Math.* **283/84**, 436–437.

Schemmel, K.-P. (1987). An extension of Buchberger's algorithm to compute all reduced Gröbner bases of a polynomial ideal. In: Davenport, J.H. (ed.), *EUROCAL '87, European Conference on Computer Algebra*, Springer LNCS **378**, 300–310.

Schwartz, N. (1988). Stability of Gröbner bases. *J. Pure Appl. Algebra* **53**, 171–186.

Seidenberg, A. (1956). An elimination theory for differential algebra. *University of California Publ. in Math.* New Series **3**/2, 31-66.

Seidenberg, A. (1971). On the length of a Hilbert ascending chain. *P. AMS* **29**, 443–450.

Seidenberg, A. (1972). Constructive proof of Hilbert's theorem on ascending chains. *Trans. Amer. Math. Soc.* **174**, 443–450.

Seidenberg, A. (1974). Constructions in algebra. *Trans. Amer. Math. Soc.* **197**, 272–313.

Seidenberg, A. (1978). Constructions in a polynomial ring over the ring of integers. *Am. J. Math.* **100**, 685–703.

Seidenberg, A. (1984). On the Lasker–Noether decomposition theorem. *Am. J. Math.* **106**, 611–638.

Shannon, D. and Sweedler, M. (1988). Using Gröbner bases to determine algebra membership, split surjective algebra homomorphisms determine birational equivalence. *J. Symb. Comp.* **6**/2,3, 267–273.

Shirayanagi, K. (1990). On the isomorphism problem for finite-dimensional binomial algebras. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 106–111.

Simmons, H. (1970). The solution of a decision problem for several classes of rings. *Pac. J. Math.* **34**/2, 547–557.

Sims, C.C. (1990). Implementing the Baumslag–Cannonito–Miller polycyclic quotient algorithm. *J. Symb. Comp.* **9**/5,6, 707–723.

Sit, W.Y. (1991). A theory for parametric linear systems. In: Watt, S.M. (ed.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '91)*, ACM Press, New York, 112–121.

Sit, W.Y. (1992). An algorithm for solving parametric linear systems. *J. Symb. Comp.* **13**/4, 353-394.

Smedley, T.J. (1990). Detecting algebraic dependencies between unnested radicals. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 292–293.

Spangher, W. (1988). On the computation of Hilbert–Samuel series and multiplicity. In: Mora, T. (ed.), *Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes (AAECC 6)*, Springer LNCS **357**, 407–414.

Spear, D.A. (1977). A constructive approach to commutative ring theory. In: *Proc. MACSYMA Users' Conference*, NASA CP-2012, 369–376.

Stafford, J.T. (1978). Module structure of Weyl algebras. *J. London Math.* **18**, 429–442.

Stifter, S. (1987). A generalization of reduction rings. *J. Symb. Comp.* **4**/3, 351–364.

Stillman, M. (1990). Methods for computing in algebraic geometry and commutative algebra. In: Strickland, E. and Piacentini Cattaneo, M.G. (eds.), *Topics in Computational Algebra*, Kluwer, Boston, 77–103.

Sturmfels, B. (1989). Dynamic versions of the Buchberger algorithm. *Preprint.*

Sturmfels, B. and White, N. (1991). Computing combinatorial decompositions of rings. *Combinatorica* **11**/3, 275–293.

Szekeres, G. (1952). A canonical basis for the ideals of a polynomial domain. *Am. Math. Mon.* **59**/6, 379–386.

Takayama, N. (1990). Gröbner basis, integration and transcendental functions. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 152–156.

Takayama, N. (1990a). An algorithm of constructing the integral of a module—An infinite dimensional analog of Gröbner basis. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 206–211.

Takayama, N. (1992). An approach to the zero recognition problem by Buchberger algorithm. *J. Symb. Comp.* **14**/2,3, 265–282.

Traverso, C. (1988). Gröbner trace algorithms. In: Gianni, P. (ed.), *International Symposium on Symbolic and Algebraic Computation (ISSAC '88)*, Springer LNCS **358**, 125–138.

Traverso, C. and Leombattista, D. (1989). Experimenting the Gröbner basis algorithm with the ALPI system. In: *International Symposium on Symbolic and Algebraic Computation (ISSAC '89)*, ACM Press, New York, 192–198.

Trevisan, G. (1953). Classificazione dei semplici ordinamenti di un gruppo libero commutativo con N generatori. *Rend. Sem. Mat. Univ. Padova* **22**, 143–156.

Trinks, W. (1978). Über Buchbergers Verfahren, Systeme algebraischer Gleichungen zu lösen. *J. Number Th.* **10**/4, 475–488.

Trinks, W. (1985). On Improving approximate results of Buchberger's algorithm. In: Caviness, B.F. (ed.), *EUROCAL '85, European Conference on Computer Algebra*, Vol. II, Springer LNCS **204**, 608–612.

Ulmer, F. and Calmet, J. (1990). On Liouvillian solutions of homogeneous linear differential equations. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 236–243.

Valibouze, A. (1989). Résolvantes et fonctions symétriques. In: *International Symposium on Symbolic and Algebraic Computation (ISSAC '89)*, ACM Press, New York, 390–399.

Wall, B. (1989). On the computation of syzygies. *ACM SIGSAM Bulletin* **23**/4, 5–14.

Wang, P.S. (1978). An improved multivariate polynomial factoring algorithm. *Math. Comput.* **32**, 1215–1231.

Wang, P.S. (1990). Parallel univariate polynomial factorization on shared-memory multiprocessors. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 145–151.

Weispfenning, V. (1983). Aspects of quantifier elimination. In: Burmeister, P. (ed.), *Universal Algebra and its Links with Logic*, Heldermann Verlag, Berlin, 85–105.

Weispfenning, V. (1986). Some bounds for the construction of Gröbner bases. In: Beth, T. and Clausen, M. (eds.), *Applicable Algebra, Error-Correcting Codes, Combinatorics, and Computer Algebra (AAECC 4)*, Springer LNCS **307**, 195–201.

Weispfenning, V. (1987). Admissible orders and linear forms. *ACM SIGSAM Bulletin* **21**/2, 16–18.

Weispfenning, V. (1987a). Constructing universal Gröbner bases. In: Huguet, L. and Poli, A. (eds.), *Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes (AAECC 5)*, Springer LNCS **356**, 408–417.

Weispfenning, V. (1987b). Gröbner bases for polynomial ideals over commutative regular rings. In: Davenport, J.H. (ed.), *EUROCAL '87, European Conference on Computer Algebra*, Springer LNCS **378**, 336–347.

Weispfenning, V. (1988). The complexity of linear problems in fields. *J. Symb. Comp.* **5**/1,2, 3–27.

Weispfenning, V. (1990). The complexity of almost linear diophantine problems. *J. Symb. Comp.* **10**/5, 395–403.

Weispfenning, V. (1992). Comprehensive Gröbner bases. *J. Symb. Comp.* **14**/1, 1–29.

Weispfenning, V. (1992a). Finite Gröbner bases in non-noetherian skew polynomial rings. Proc. ISSAC '92.

Weispfenning, V. (1992b). Differential-term orders. Preprint, Univ. Passau.

Whitney, H. (1935). On the abstract properties of linear dependence. *Am. J. Math.* **57**, 504–533.

Winkler, F. (1983). An algorithm for constructing detaching bases in the ring of polynomials over a field. In: van Hulzen, J.A. (ed.), *EUROCAL '83, European Computer Algebra Conference*, Springer LNCS **162**, 168–179.

Winkler, F. (1984). On the complexity of the Gröbner bases algorithm over $K[x, y, z]$. In: Fitch, J. (ed.), *EUROSAM '84, International Symposium on Symbolic and Algebraic Manipulation*, Springer LNCS **174**, 184–194.

Winkler, F. (1986). Solution of equations I: Polynomial ideals and Gröbner bases. In: Jenks, R.D. (ed.), *Conference on Computers and Mathematics, Stanford University, 1986*, Series in Computational Mathematics, Springer-Verlag, New York.

Winkler, F. (1988). A p-adic approach to the computation of Gröbner bases. *J. Symb. Comp.* **6**/2,3, 287–304.

Winkler, F. (1988a). A geometrical decision algorithm based on the Gröbner bases algorithm. In: Gianni, P. (ed.), *International Symposium on Symbolic and Algebraic Computation (ISSAC '88)*, Springer LNCS **358**, 356–363.

Winkler, F. (1989). Knuth–Bendix procedure and Buchberger algorithm—a synthesis. In: *International Symposium on Symbolic and Algebraic Computation (ISSAC '89)*, ACM Press, New York, 55–67.

Winkler, F. (1990). Representation of algebraic curves. In: Gerdt, V.P., Rostovtsev, V.A., and Shirkov, D.V. (eds.), *IV International Conference on Computer Algebra in Physical Research*, World Scientific Publishing Co., Singapore, 185–189.

Wu, W.T. (1984). Basic principles of mechanical theorem proving in elementary geometries. *J. of System Science and Mathematical Science* **4**/3, 1984, 207–235.

Wu, W.T. (1986). Basic principles of mechanical theorem proving in elementary geometry. *J. of Automated Reasoning* **2**, 219–252.

Yokoyama, K., Noro, M., and Takeshima, T. (1989). Computing primitive elements of extension fields. *J. Symb. Comp.* **8**/6, 553–580.

Yokoyama, K., Noro, M., and Takeshima, T. (1990). On determining the solvability of polynomials. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 127–134.

Yokoyama, K., Noro, M., and Takeshima, T. (1990a). On factoring multivariate polynomials over algebraically closed fields. In: Watanabe, S. and Nagata, M. (eds.), *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC '90)*, ACM Press, New York, 297.

Yokoyama, K., Noro, M., and Takeshima, T. (1992). Solutions of systems of algebraic equations and linear maps on residue class rings. *J. Symb. Comp.* **14**/4, 399–417.

Yun, D.Y.Y. (1977). On the equivalence of polynomial GCD and squarefree factorization problems. In: *Proc. MACSYMA Users' Conference*, NASA CP-2012, 65–70.

Zassenhaus, H. (1969). On Hensel factorization, I. *J. Number Th.* **1**, 291–311.

Zermelo, E. (1904). Beweis, dass jede Menge wohlgeordnet werden kann. (Aus einem an Herrn Hilbert gerichteten Briefe.) *Math. Ann.* **59**, 514–516.

Zorn, M. (1935). A remark on a method in transfinite algebra. *B. AMS* **41**, 667–670.

List of Symbols

$\mathbb{N}^+$	set of positive natural numbers, 1
$A_1 \times \cdots \times A_n$	Cartesian product of sets, 5
$\prod_{i=1}^n A_i$	Cartesian product of sets, 5
A^n	Cartesian product of set with itself, 5
$a \longmapsto \varphi(a)$	a maps to $\varphi(a)$, 5
B^A	set of functions (maps) from A to B, 5
$\restriction$	restriction of map, 5
id_A	identity on A, 7
$\|A\|$	cardinality of set, 9
$\mathrm{C}(\boldsymbol{I},\mathbb{R})$	ring of continuous real-valued functions, 16
$\triangle$	symmetric set difference, 17
$\mathbb{Z}_p$	localization of integers at p, 20
U_R	group of units of ring, 21
$\simeq$	isomorphic to, 24
$\ker(\varphi)$	kernel of homomorphism, 24, (135)
$\mathrm{Id}(a)$	ideal generated by a, 26
$\mathrm{Id}(a_1,\ldots,a_n)$	ideal generated by a_1, ..., a_n, 26
$\mathrm{Id}(A)$	ideal generated by A, 26
$a+I$	residue class modulo ideal, 27
R/I	residue class ring modulo ideal, 27, 31
$\equiv \mathrm{mod} I$	congruence modulo ideal, 32
$\gcd(a,b)$	greatest common divisor, 39
$\mathrm{lcm}(a,b)$	least common multiple, 45, (211)
$R[M]$	ring adjunction, 52
$\prod_{i=1}^n R_i$	direct product of rings, 53
$R_1 \times \cdots \times R_n$	direct product of rings, 53

R^n	direct product of ring with itself, 53
R_M	quotient ring, 55
Q_R	field of fractions, 55
I^{e}	extension ideal, 57
J^{c}	contraction ideal, 57
$(M, 1, \cdot)$	multiplicative monoid, 62
$(\mathbb{N}^n, 0, +)$	additive monoid $\mathbb{N}^n$, 62
$\mathrm{supp}(f)$	support of function, 63
RM	monoid ring, 63
$T(X_1, \ldots, X_n)$	set of terms, 70
T	set of terms, 70
$\deg(t)$	degree of term, 70
$R[X_1, \ldots, X_n]$	polynomial ring over ring, 71
$R[\underline{X}]$	polynomial ring over ring, 71
$M(f)$	set of monomials of polynomial, 71, (193)
$T(f)$	set of terms of polynomial, 71, (193)
$\deg(f)$	degree of polynomial, 71
$C(f)$	set of coefficients of polynomial, 71, (193)
$\deg_{X_i}(f)$	degree in X_i of polynomial, 74
$f(c_1, \ldots, c_n)$	polynomial evaluation, 75
$f(\mathbf{c})$	polynomial evaluation, 75
$\mathrm{c}(f)$	content of polynomial, 93
$\mathrm{pp}(f)$	primitive part of polynomial, 93
$K(X_1, \ldots, X_n)$	rational function field, 94
f'	derivative of polynomial, 101
$\dim_K(V)$	dimension of vector space, 131
$N \leq M$	submodule relation, 135
$\ker(\varphi)$	kernel of homomorphism, 135, (24)
$\mathrm{lin}(B)$	linear span, 135
$\mathrm{syz}(a_1, \ldots, a_n)$	syzygies, 136
M/N	factor module, 136
$\mathrm{rad}(I)$	radical of ideal, 147

$\Delta(M)$	diagonal of set, 149
r^{-1}	inverse relation, 149
$s \circ r$	product of relations, 150
$[a]$	equivalence class of a, 152
r^+	transitive closure of relation, 154
r^*	reflexive-transitive closure of relation, 154
r_{s}	strict part of relation, 155
$(M \times N, \preceq)$	direct product of quasi-ordered sets, 163
U_a	upper set of a, 165
$\mathcal{P}_{\mathrm{fin}}$	set of finite subsets, 170
$\longrightarrow$	reduction relation, 174
$\xrightarrow{*}$	reflexive-transitive closure of $\longrightarrow$, 174
$\longleftrightarrow$	symmetric closure of $\longrightarrow$, 174
$\overset{*}{\longleftrightarrow}$	symmetric closure of reflexive-transitive closure, 174
$\xrightarrow{n}$	reduction chain, 174
$\overset{n}{\longleftrightarrow}$	back-and-forth reduction chain, 174
$\downarrow$	reduce to common element, 175
$M(f)$	set of monomials of polynomial, 193, (71)
$T(f)$	set of terms of polynomial, 193, (71)
$C(f)$	set of coefficients of polynomial, 193, (71)
$\mathrm{HT}(f)$	head term of polynomial, 194
$\mathrm{HM}(f)$	head monomial of polynomial, 194
$\mathrm{HC}(f)$	head coefficient of polynomial, 194
$K[\underline{X}]$	polynomial ring over field, 195
$f \xrightarrow[p]{} g\ [t]$	polynomial reduction, 195
$f \xrightarrow[p]{} g$	polynomial reduction, 195
$f \xrightarrow[P]{} g$	polynomial reduction, 196
$\equiv_I$	congruence modulo ideal, 201
$\mathrm{HT}(P)$	set of head terms, 206
$\mathrm{mult}(T)$	set of multiples, 206
$\mathrm{lcm}(s,t)$	least common multiple, 211, (45)
$\mathrm{spol}(g_1, g_2)$	S-polynomial, 211, (457)

$T(\underline{X})$	set of terms, 256
$I_{\underline{U}}$	elimination ideal, 256
$\underline{U} \ll \underline{X} \setminus \underline{U}$	"lexicographically" less, 256
$\underline{X} \ll \underline{Y}$	"lexicographically" less, 259
I_a	vanishing ideal, 263
$I : F$	ideal quotient, 264
$I : f^\infty$	union of all $I : f^s$, 266
$\dim(I)$	dimension of ideal, 271
$\dim_K(K[\underline{X}]/I)$	vector space dimension of residue class ring, 272
$\mathrm{RT}(I)$	set of reduced terms, 272
$K(A)$	field adjunction, 293
$K(a_1, \ldots, a_n)$	field adjunction, 293
$\overline{K}$	algebraic closure, 309
$V_L(I)$	variety of I in L, 327
$\mathrm{d}(I)$	depth of ideal, 323
$\mathrm{h}(I)$	height of ideal, 323
$\mathrm{Id}(I_1 \cdot \cdots \cdot I_r)$	ideal product, 335
$\mathrm{Id}(I^\nu)$	ideal power, 336
$\mathrm{dist}(\boldsymbol{I}, \boldsymbol{J})$	distance between intervals, 414
$\mathrm{dist}(\alpha, \boldsymbol{I})$	distance between point and interval, 414
$\mathrm{st}(I)$	stairs of ideal, 424
T_m	terms of degree less than m, 441
A_m	residue classes of "degree" less than m, 441
H_I	Hilbert function of I, 442
$\mathrm{top}_M(t)$	set of indices where t tops M, 444
$\mathrm{sh}_M(t)$	t shaved at M, 444
$\mathrm{spol}(g_1, g_2)$	S-polynomial, 457, (211)
$\mathrm{gpol}(g_1, g_2)$	G-polynomial, 457
$\Gamma(f)$	degree of polynomial w.r.t. grading, 466
$K[\underline{X}]_{[d_1,d_2]}$	slice of polynomial ring w.r.t. grading, 469
$f_{(d)}$	d-homogeneous part of polynomial, 474
$\mathrm{HF}(f)$	highest form of f, 476

M_{mn}^{K}	$K[X_1,\dots,X_n]$-module $(K[X_1,\dots,X_n])^m$, 485
$H_1(K[\underline{X},\underline{Z}])$	homogeneous degree-1 elements of $K[\underline{X},\underline{Z}]$, 486
$\mathrm{LT}_{\leq}(f)$	$\leq$-lowest term of f, 496
$\mathrm{LF}(f)$	lowest form of f, 498
$S(K[\underline{X}])$	ring of symmetric functions, 502
$\xrightarrow[\sigma]{}$	σ-reduction, 504

Index

Graduate Texts in Mathematics

continued from page ii

61 Whitehead. Elements of Homotopy Theory.
62 Kargapolov/Merlzjakov. Fundamentals of the Theory of Groups.
63 Bollobas. Graph Theory.
64 Edwards. Fourier Series. Vol. I 2nd ed.
65 Wells. Differential Analysis on Complex Manifolds. 2nd ed.
66 Waterhouse. Introduction to Affine Group Schemes.
67 Serre. Local Fields.
68 Weidmann. Linear Operators in Hilbert Spaces.
69 Lang. Cyclotomic Fields II.
70 Massey. Singular Homology Theory.
71 Farkas/Kra. Riemann Surfaces. 2nd ed.
72 Stillwell. Classical Topology and Combinatorial Group Theory. 2nd ed.
73 Hungerford. Algebra.
74 Davenport. Multiplicative Number Theory. 2nd ed.
75 Hochschild. Basic Theory of Algebraic Groups and Lie Algebras.
76 Iitaka. Algebraic Geometry.
77 Hecke. Lectures on the Theory of Algebraic Numbers.
78 Burris/Sankappanavar. A Course in Universal Algebra.
79 Walters. An Introduction to Ergodic Theory.
80 Robinson. A Course in the Theory of Groups. 2nd ed.
81 Forster. Lectures on Riemann Surfaces.
82 Bott/Tu. Differential Forms in Algebraic Topology.
83 Washington. Introduction to Cyclotomic Fields. 2nd ed.
84 Ireland/Rosen. A Classical Introduction to Modern Number Theory. 2nd ed.
85 Edwards. Fourier Series. Vol. II. 2nd ed.
86 van Lint. Introduction to Coding Theory. 2nd ed.
87 Brown. Cohomology of Groups.
88 Pierce. Associative Algebras.
89 Lang. Introduction to Algebraic and Abelian Functions. 2nd ed.
90 Brøndsted. An Introduction to Convex Polytopes.
91 Beardon. On the Geometry of Discrete Groups.
92 Diestel. Sequences and Series in Banach Spaces.
93 Dubrovin/Fomenko/Novikov. Modern Geometry—Methods and Applications. Part I. 2nd ed.
94 Warner. Foundations of Differentiable Manifolds and Lie Groups.
95 Shiryaev. Probability. 2nd ed.
96 Conway. A Course in Functional Analysis. 2nd ed.
97 Koblitz. Introduction to Elliptic Curves and Modular Forms. 2nd ed.
98 Bröcker/Tom Dieck. Representations of Compact Lie Groups.
99 Grove/Benson. Finite Reflection Groups. 2nd ed.
100 Berg/Christensen/Ressel. Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions.
101 Edwards. Galois Theory.
102 Varadarajan. Lie Groups, Lie Algebras and Their Representations.
103 Lang. Complex Analysis. 3rd ed.
104 Dubrovin/Fomenko/Novikov. Modern Geometry—Methods and Applications. Part II.
105 Lang. $SL_2(\mathbf{R})$.
106 Silverman. The Arithmetic of Elliptic Curves.
107 Olver. Applications of Lie Groups to Differential Equations. 2nd ed.
108 Range. Holomorphic Functions and Integral Representations in Several Complex Variables.
109 Lehto. Univalent Functions and Teichmüller Spaces.
110 Lang. Algebraic Number Theory.
111 Husemöller. Elliptic Curves.
112 Lang. Elliptic Functions.
113 Karatzas/Shreve. Brownian Motion and Stochastic Calculus. 2nd ed.
114 Koblitz. A Course in Number Theory and Cryptography. 2nd ed.
115 Berger/Gostiaux. Differential Geometry: Manifolds, Curves, and Surfaces.
116 Kelley/Srinivasan. Measure and Integral. Vol. I.
117 Serre. Algebraic Groups and Class Fields.
118 Pedersen. Analysis Now.

119 ROTMAN. An Introduction to Algebraic Topology.

120 ZIEMER. Weakly Differentiable Functions: Sobolev Spaces and Functions of Bounded Variation.

121 LANG. Cyclotomic Fields I and II. Combined 2nd ed.

122 REMMERT. Theory of Complex Functions. *Readings in Mathematics*

123 EBBINGHAUS/HERMES et al. Numbers. *Readings in Mathematics*

124 DUBROVIN/FOMENKO/NOVIKOV. Modern Geometry—Methods and Applications. Part III.

125 BERENSTEIN/GAY. Complex Variables: An Introduction.

126 BOREL. Linear Algebraic Groups. 2nd ed.

127 MASSEY. A Basic Course in Algebraic Topology.

128 RAUCH. Partial Differential Equations.

129 FULTON/HARRIS. Representation Theory: A First Course. *Readings in Mathematics*

130 DODSON/POSTON. Tensor Geometry.

131 LAM. A First Course in Noncommutative Rings.

132 BEARDON. Iteration of Rational Functions.

133 HARRIS. Algebraic Geometry: A First Course.

134 ROMAN. Coding and Information Theory.

135 ROMAN. Advanced Linear Algebra.

136 ADKINS/WEINTRAUB. Algebra: An Approach via Module Theory.

137 AXLER/BOURDON/RAMEY. Harmonic Function Theory.

138 COHEN. A Course in Computational Algebraic Number Theory.

139 BREDON. Topology and Geometry.

140 AUBIN. Optima and Equilibria. An Introduction to Nonlinear Analysis.

141 BECKER/WEISPFENNING/KREDEL. Gröbner Bases. A Computational Approach to Commutative Algebra.

142 LANG. Real and Functional Analysis. 3rd ed.

143 DOOB. Measure Theory.

144 DENNIS/FARB. Noncommutative Algebra.

145 VICK. Homology Theory. An Introduction to Algebraic Topology. 2nd ed.

146 BRIDGES. Computability: A Mathematical Sketchbook.

147 ROSENBERG. Algebraic K-Theory and Its Applications.

148 ROTMAN. An Introduction to the Theory of Groups. 4th ed.

149 RATCLIFFE. Foundations of Hyperbolic Manifolds.

150 EISENBUD. Commutative Algebra with a View Toward Algebraic Geometry.

151 SILVERMAN. Advanced Topics in the Arithmetic of Elliptic Curves.

152 ZIEGLER. Lectures on Polytopes.

153 FULTON. Algebraic Topology: A First Course.

154 BROWN/PEARCY. An Introduction to Analysis.

155 KASSEL. Quantum Groups.

156 KECHRIS. Classical Descriptive Set Theory.

157 MALLIAVIN. Integration and Probability.

158 ROMAN. Field Theory.

159 CONWAY. Functions of One Complex Variable II.

160 LANG. Differential and Riemannian Manifolds.

161 BORWEIN/ERDÉLYI. Polynomials and Polynomial Inequalities.

162 ALPERIN/BELL. Groups and Representations.

163 DIXON/MORTIMER. Permutation Groups.

164 NATHANSON. Additive Number Theory: The Classical Bases.

165 NATHANSON. Additive Number Theory: Inverse Problems and the Geometry of Sumsets.

166 SHARPE. Differential Geometry: Cartan's Generalization of Klein's Erlangen Program.

167 MORANDI. Field and Galois Theory.

168 EWALD. Combinatorial Convexity and Algebraic Geometry.

169 BHATIA. Matrix Analysis.

170 BREDON. Sheaf Theory. 2nd ed.

171 PETERSEN. Riemannian Geometry.

172 REMMERT. Classical Topics in Complex Function Theory.

173 DIESTEL. Graph Theory.

174 BRIDGES. Foundations of Real and Abstract Analysis.

175 LICKORISH. An Introduction to Knot Theory.

176 LEE. Riemannian Manifolds.

177 NEWMAN. Analytic Number Theory.

178 CLARKE/LEDYAEV/STERN/WOLENSKI. Nonsmooth Analysis and Control Theory.