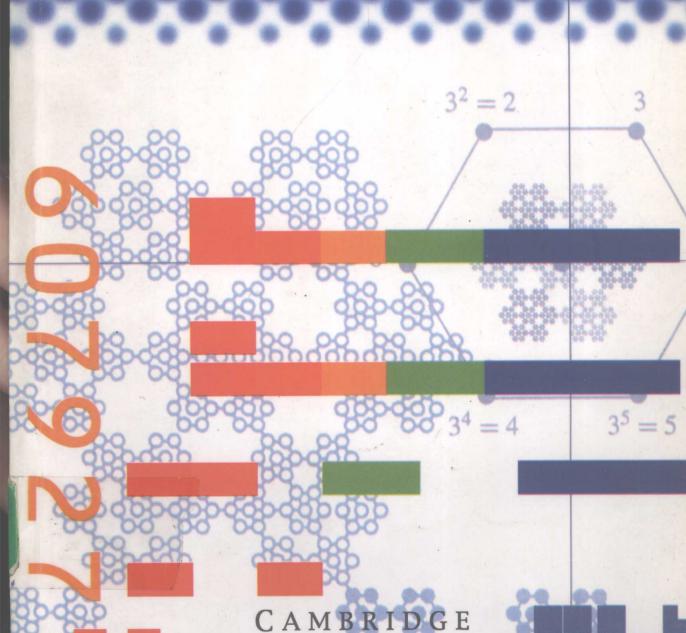
Modern Computer Algebra

Joachim von zur Gathen and Jurgen Gerhard

现代计算代数



老界图出出版:

1415926535897

Modern Computer Algebra

JOACHIM VON ZUR GATHEN and JÜRGEN GERHARD

Universität Paderborn

CAMBRIDGE UNIVERSITY PRESS

~ 界图出出版公司

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE The Pitt Building, Trumpington Street, Cambridge CB2 1RP

CAMBRIDGE UNIVERSITY PRESS

The Edinburgh Building, Cambridge CB2 2RU, UK http://www.cup.cam.ac.uk 40 West 20th Street, New York, NY 10011-4211, USA http://www.cup.org 10 Stamford Road, Oakleigh, Melbourne 3166, Australia

This book is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 1999

© Cambridge University Press 1999

British Library cataloguing in publication data available

ISBN 0 521 64176 4 hardback

This edition of Modern Computer Algebra by J.von zur Gathen and J.Gerhard is published by arrangement with the Syndicate of the Press of University of Cambridge, Cambridge, England.

Licensed edition for sale in the People's Republic of China only. Not for export elsewhere.

أَ وَ مَا مِنُ غَآبِبَةٍ فِي آلسَّمَآءِ وَ الْأَرْضِ اِلَّا فِي كِتَبٍ مُّبِيْنٍ The Holy Quran (732)

Les bons élèves font la gloire du maître.²

Joseph Liouville (1846)

Je prie les lecteurs de n'ajouter point du tout de foi à tout ce qu'ils trouveront ici écrit, mais seulement de l'examiner et n'en recevoir que ce que la force et l'évidence de la raison les pourra contraindre de croire.³

René Descartes (1647)

The subject is full of pitfalls. I have pointed out some mistakes made by others, but have no doubt that I have made new ones. It may be expected that any errors will be discovered and eliminated in due course.

Francis Sowerby Macaulay (1916)

Wherfore I trust thei that be learned, and happen to reade this worke, wil beare the moare with me, if thei finde any thyng, that thei doe mislike: Wherein if thei will use this curtesie, either by writynge to admonishe me thereof, either theim selfes to sette forthe a moare perfecter woorke, I will thynke them praise worthie.

Robert Recorde (1557)

There is a theory which states that if ever anyone discovers exactly what the Universe is for and why it is here, it will instantly disappear and be replaced by something even more bizarre and inexplicable. There is another theory which states that this has already happened.

Douglas Adams (1980)

¹ There is nothing hidden in heaven or on earth that is not in a clear book.

² Good students are the teacher's glory.

³ I ask the readers to put no faith at all in anything they find written here, but just to examine it and to accept only whatever the strength and evidence of reason may oblige them to believe.

Contents

	Intro	oduction	1
1	Cycl	ohexane, cryptography, codes, and computer algebra	9
	1.1	Cyclohexane conformations	9
	1.2	The RSA cryptosystem	14
	1.3	Distributed data structures	16
	1.4	Computer algebra systems	17
I	Euc	lid	21
2	Fun	damental algorithms	27
	2.1	Representation and addition of numbers	27
	2.2	Representation and addition of polynomials	30
	2.3	Multiplication	32
	2.4	Division with remainder	35
		Notes	39
		Exercises	39
3	The	Euclidean Algorithm	43
	3.1	Euclidean domains	43
	3.2	The Extended Euclidean Algorithm	46
	3.3	Cost analysis for \mathbb{Z} and $F[x]$	50
		Notes	55
		Exercises	57
4	App	lications of the Euclidean Algorithm	63
	4.1	Modular arithmetic	63
	4.2	Modular inverses via Euclid	67
	4.3	Repeated squaring	69
	4.4	Modular inverses via Fermat	70
	4.5	Linear Diophantine equations	71

viii Contents

	4.6	Continued fractions and Diophantine approximation	73
	4.7	Calendars	77
	4.8	Musical scales	78
		Notes	81
		Exercises	84
5	Modu	ılar algorithms and interpolation	89
	5.1	Change of representation	92
	5.2	Evaluation and interpolation	93
	5.3	Application: Secret sharing	95
	5.4	The Chinese Remainder Algorithm	96
	5.5	Modular determinant computation	101
	5.6	Hermite interpolation	105
	5.7	Rational function reconstruction	106
	5.8	Cauchy interpolation	110
	5.9	Padé approximation	112
	5.10	Rational number reconstruction	116
	5.11	Partial fraction decomposition	
		Notes	122
		Exercises	
6	The	resultant and gcd computation	131
U	6. l	Coefficient growth in the Euclidean Algorithm	131
	6.2	Gauß' lemma	137
	6.3	The resultant	142
	6.4	Modular gcd algorithms	
	6.5	Modular gcd algorithm in $F[x,y]$	151
	6.6	Mignotte's factor bound and a modular gcd algorithm in $\mathbb{Z}[x]$	153
	6.7	Small primes modular gcd algorithms	157
	6.8	Application: intersecting plane curves	161
	6.9	Nonzero preservation and the gcd of several polynomials	165
	6.10	Subresultants	
	6.11	Modular Extended Euclidean Algorithms	
	6.12		180
	6.13	Implementations	
	0.10	Notes	
		Exercises	
-	A 1	lication: Decoding BCH codes	197
7	App	Notes	
		Exercises	
		Exercises	

Contents ix

H	Nev	wton	205
8	Fast	multiplication	209
	8.1	Karatsuba's multiplication algorithm	. 210
	8.2	The Discrete Fourier Transform and the Fast Fourier Transform	. 215
	8.3	Schönhage and Strassen's multiplication algorithm	. 225
	8.4	Multiplication in $\mathbb{Z}[x]$ and $R[x,y]$. 233
		Notes	. 234
		Exercises	. 235
9	Newf	ton iteration	243
	9.1	Division with remainder using Newton iteration	. 243
	9.2	Generalized Taylor expansion and radix conversion	. 250
	9.3	Formal derivatives and Taylor expansion	. 251
	9.4	Solving polynomial equations via Newton iteration	. 253
	9.5	Computing integer roots	. 257
	9.6	Valuations, Newton iteration, and Julia sets	. 259
	9.7	Implementations of fast arithmetic	. 263
		Notes	. 272
		Exercises	. 272
10	Fast polynomial evaluation and interpolation		
	10.1	Fast multipoint evaluation	. 279
	10.2	Fast interpolation	. 283
	10.3	Fast Chinese remaindering	. 285
		Notes	. 290
		Exercises	. 290
11	Fast	Euclidean Algorithm	295
	11.1	A fast Euclidean Algorithm for polynomials	. 295
	11.2	Subresultants via Euclid's algorithm	. 306
		Notes	. 310
		Exercises	. 310
12	Fast linear algebra		
	12.1	Strassen's matrix multiplication	. 313
	12.2	Application: fast modular composition of polynomials	. 316
	12.3	Linearly recurrent sequences	. 317
	12.4	Wiedemann's algorithm and black box linear algebra	. 323
		Notes	. 330
		Exercises	. 331

X Contents

13	Fouri	ier Transform and image compression	335	
	13.1	The Continuous and the Discrete Fourier Transform	335	
	13.2		339	
		Notes		
		Exercises		
III	Ga	uß	347	
14	Facto	ring polynomials over finite fields	353	
	14.1	Factorization of polynomials	353	
	14.2	Distinct-degree factorization	356	
	14.3	Equal-degree factorization: Cantor and Zassenhaus' algorithm	358	
	14.4	A complete factoring algorithm	365	
	14.5	Application: root finding	368	
	14.6	Squarefree factorization	369	
	14.7	The iterated Frobenius algorithm	373	
	14.8	Algorithms based on linear algebra	377	
	14.9	Testing irreducibility and constructing irreducible polynomials .		
	14.10	Cyclotomic polynomials and constructing BCH codes	387	
		Notes	393	
		Exercises		
15	Hensel lifting and factoring polynomials 4			
	15.1	Factoring in $\mathbb{Z}[x]$ and $\mathbb{Q}[x]$: the basic idea	407	
	15.2	A factoring algorithm		
	15.3	Frobenius' and Chebotarev's density theorems		
	15.4	Hensel lifting	418	
	15.5	Multifactor Hensel lifting		
	15.6	Factoring using Hensel lifting: Zassenhaus' algorithm		
	15.7	Implementations		
		Notes		
		Exercises		
16	Short	vectors in lattices	447	
	16.1	Lattices	447	
	16.2	Lenstra, Lenstra and Lovász' basis reduction algorithm	449	
	16.3	Cost estimate for basis reduction		
	16.4	From short vectors to factors	461	
	16.5	A polynomial-time factoring algorithm for $\mathbb{Q}[x]$	463	
	16.6	Factoring multivariate polynomials	467	
		Notes		
		Exercises	472	

17	Appli	cations of basis reduction	477
	17.1	Breaking knapsack-type cryptosystems	477
	17.2	Pseudorandom numbers	479
	17.3	Simultaneous Diophantine approximation	
	17.4	Disproof of Mertens' conjecture	482
		Notes	483
		Exercises	
IV	Fer	rmat	485
1 Q	Drim	ality testing	491
10	18.1	Multiplicative order of integers	491
	18.2	The Fermat test	493
	18.3	The strong pseudoprimality test	494
	18.4	Finding primes	497
	18.5	The Solovay and Strassen test	503
	18.6	The complexity of primality testing	504
	10.0	Notes	506
		Exercises	509
19	Facto	oring integers	515
-	19.1	Factorization challenges	515
	19.2	Trial division	518
	19.3	Pollard's and Strassen's method	
	19.4	Pollard's rho method	
	19.5	Dixon's random squares method	523
	19.6	Pollard's $p-1$ method	531
	19.7	Lenstra's elliptic curve method	531
		Notes	541
		Exercises	. 543
20) App	lication: Public key cryptography	547
	20.1	Cryptosystems	. 547
	20.2	The RSA cryptosystem	. 550
	20.3	The Diffie-Hellman key exchange protocol	. 552
	20.4	_	. 553
	20.5	Rabin's cryptosystem	. 553
	20.6	Elliptic curve systems	. 554
	20.7	Short vector cryptosystems	. 554
		Notes	. 555
		Evereigns	. 555

Contents

V	Hill	pert .	559		
21	Gröbner bases				
	21.1	Polynomial ideals	. 565		
	21.2	Monomial orders and multivariate division with remainder	. 570		
	21.3	Monomial ideals and Hilbert's basis theorem	. 575		
	21.4	Gröbner bases and S-polynomials			
	21.5	Buchberger's algorithm			
	21.6	Geometric applications			
	21.7	The complexity of computing Gröbner bases	. 589		
		Notes			
		Exercises			
22	Symb	polic integration	597		
	22.1	Differential algebra	. 597		
	22.2	Hermite's method	. 599		
	22.3	The method of Rothstein and Trager	. 601		
	٠	Notes	. 606		
	•	Exercises	. 606		
23	Symbolic summation				
	23.1	Polynomial summation	. 609		
	23.2	Harmonic numbers	. 614		
	23.3	Greatest factorial factorization	. 617		
	23.4	Hypergeometric summation: Gosper's algorithm	. 622		
		Notes	. 633		
		Exercises	. 635		
24	Applications 6				
	24.1	Gröbner proof systems	. 641		
	24.2	Petri nets	. 643		
	24.3	Proving identities and analysis of algorithms	. 645		
	24.4	Cyclohexane revisited	. 649		
		Notes			
		Exercises			
Ap	pendi	ix	665		
25	Fund	amental concepts	667		
	25.1	Groups	. 667		
	25.2	Rings			
	25.3				

Contents xiii

25.4	Finite fields
25.5	Linear algebra
25.6	Finite probability spaces
25.7	'Big Oh" notation
25.8	Complexity theory
	Notes
Sourc	s of illustrations
Sourc	s of quotations
List o	algorithms
List o	figures and tables
Refer	nces
List o	notation
Index	

Keeping up to date

Addenda and corrigenda, comments, solutions to selected exercises, and ordering information can be found on the book's web page:

http://www-math.uni-paderborn.de/mca/

Introduction

In science and engineering, a successful attack on a problem will usually lead to some equations that have to be solved. There are many types of such equations: differential equations, linear or polynomial equations or inequalities, recurrences, equations in groups, tensor equations, etc. In principle, there are two ways of solving such equations: approximately or exactly. *Numerical analysis* is a well-developed field that provides highly successful mathematical methods and computer software to compute *approximate* solutions.

Computer algebra is a more recent area of computer science, where mathematical tools and computer software are developed for the exact solution of equations.

Why use approximate solutions at all if we can have exact solutions? The answer is that in many cases an exact solution is not possible. This may have various reasons: for certain (simple) ordinary differential equations, one can prove that no closed form solution (of a specified type) is possible. More important are questions of efficiency: any system of linear equations, say with rational coefficients, can be solved exactly, but for the huge linear systems that arise in meteorology, nuclear physics, geology or other areas of science, only approximate solutions can be computed efficiently. The exact methods, run on a supercomputer, would not yield answers within a few days or weeks (which is not really acceptable for weather prediction).

However, within its range of exact solvability, computer algebra usually provides more interesting answers than traditional numerical methods. Given a differential equation or a system of linear equations with a parameter t, the scientist gets much more information out of a closed form solution in terms of t than from several solutions for specific values of t.

Many of today's students may not know that the *slide rule* was an indispensable tool of engineers and scientists until the 1960s. *Electronic pocket calculators* made them obsolete within a short time. In the coming years, *computer algebra systems* will similarly replace calculators for many purposes. Although still bulky and expensive (hand-held computer algebra calculators are yet a novelty), these systems can easily perform exact (or arbitrary precision) arithmetic with numbers,

matrices, polynomials, etc. They will become an indispensable tool for the scientist and engineer, from students to the work place. These systems are now becoming integrated with other software, like numerical packages, CAD/CAM, and graphics.

The goal of this text is to give an introduction to the basic methods and techniques of computer algebra. Our focus is threefold:

- o complete presentation of the mathematical underpinnings,
- o asymptotic analysis of our algorithms, sometimes "Oh-free",
- development of asymptotically fast methods.

It is customary to give bounds on running times of algorithms (if any are given at all) in a "big-Oh" form (explained in Section 25.7), say as $O(n\log n)$ for the FFT. We often prove "Oh-free" bounds in the sense that we identify the numerical coefficient of the leading term, as $\frac{3}{2}n\log_2 n$ in the example; we may then add O(smaller terms). But we have not played out the game of minimizing these coefficients; the reader is encouraged to find smaller constants herself.

Many of these fast methods have been known for a quarter of a century, but their impact on computer algebra systems has been slight, partly due to an "unfortunate myth" (Bailey, Lee & Simon 1990) about their practical (ir)relevance. But their usefulness has been forcefully demonstrated in the last few years; we can now solve problems—for example, the factorization of polynomials—of a size that was unassailable a few years ago. We expect this success to expand into other areas of computer algebra, and indeed hope that this text may contribute to this development. The full treatment of these fast methods motivates the "modern" in its title. (Our title is a bit risqué, since even a "modern" text in a rapidly evolving discipline such as ours will obsolesce quickly.)

The basic objects of computer algebra are numbers and polynomials. Throughout the text, we stress the structural and algorithmic similarities between these two domains, and also where the similarities break down. We concentrate on polynomials, in particular univariate polynomials over a field, and pay special attention to finite fields.

We will consider arithmetic algorithms in some basic domains. The tasks that we will analyze include conversion between representations, addition, subtraction, multiplication, division, division with remainder, greatest common divisors, and factorization. The basic domains for computer algebra are the natural numbers, the rational numbers, finite fields, and polynomial rings.

Our three goals, as stated above, are too ambitious to keep up throughout. In some chapters, we have to content ourselves with sketches of methods and outlooks on further results. Due to space limitations, we sometimes have recourse to the lamentable device of "leaving the proof to the reader". Don't worry, be happy: solutions to the corresponding exercises are available on the book's web site.

Introduction 3

After writing most of the material, we found that we could structure the book into five parts, each named after a mathematician that made a pioneering contribution on which some (but, of course, not all) of the modern methods in the respective part rely. In each part, we also present selected applications of some of the algorithmic methods.

The first part EUCLID examines Euclid's algorithm for calculating the gcd, and presents the subresultant theory for polynomials. Applications are numerous: modular algorithms, continued fractions, Diophantine approximation, the Chinese Remainder Algorithm, secret sharing, and the decoding of BCH codes.

The second part NEWTON presents the basics of fast arithmetic: FFT-based multiplication, division with remainder and polynomial equation solving via Newton iteration, and fast methods for the Euclidean Algorithm and the solution of systems of linear equations. The FFT originated in signal processing, and we discuss one of its applications, image compression.

The third part GAUSS deals exclusively with polynomial problems. We start with univariate factorization over finite fields, and include the modern methods that make attacks on enormously large problems feasible. Then we discuss polynomials with rational coefficients. The two basic algorithmic ingredients are Hensel lifting and short vectors in lattices. The latter has found many applications, from breaking certain cryptosystems to Diophantine approximation.

The fourth part FERMAT is devoted to two integer problems that lie at the foundation of algorithmic number theory: primality testing and factorization. The most famous modern application of these classical topics is in public key cryptography.

The fifth part HILBERT treats three different topics which are somewhat more advanced than the rest of the text, and where we can only exhibit the foundations of a rich theory. The first topic is Gröbner bases, a successful approach to deal with multivariate polynomials, in particular questions about common roots of several polynomials. The next topic is symbolic integration, where we concentrate on the basic case of integrating rational functions. The final topic is symbolic summation; we discuss polynomial and hypergeometric summation.

The text concludes with an appendix that presents some foundational material in the language we use throughout the book: The basics of groups, rings, and fields, linear algebra, probability theory, asymptotic *O*-notation, and complexity theory.

Each of the first three parts contains an implementation report on some of the algorithms presented in the text. As case studies, we use two special purpose packages for integer and polynomial arithmetic: NTL by Victor Shoup and BIPOLAR by the authors.

Most chapters end with some bibliographical and historical notes or supplementary remarks, and a variety of exercises. The latter are marked according to their difficulty: exercises with a * are somewhat more advanced, and the few marked with ** are more difficult or may require material not covered in the text. Laborious (but not necessarily difficult) exercises are marked by a long arrow \rightarrow . The

4 Introduction

book's web page http://www-math.uni-paderborn.de/mca/ has some hints and solutions.

This book presents foundations for the mathematical engine underlying any computer algebra system, and we give substantial coverage-often, but not always, up to the state of the art-for the material of the first three parts, dealing with Euclid's algorithm, fast arithmetic, and the factorization of polynomials. But we hasten to point out some unavoidable shortcomings. For one, we cannot cover completely even those areas that we discuss, and our treatment leaves out major interesting developments in the areas of computational linear algebra, sparse multivariate polynomials, combinatorics and computational number theory, quantifier elimination and solving polynomial equations, and differential and difference equations. Secondly, some important questions are left untouched at all; we only mention computational group theory, parallel computation, computing with transcendental functions, isolating real and complex roots of polynomials, and the combination of symbolic and numeric methods. Finally, a successful computer algebra system involves much more than just the mathematical engine: efficient data structures, a fast kernel and a large compiled or interpreted library, user interface, graphics capability, clever marketing, etc. These issues are highly technologydependent, and there is no single good solution for them.

The present book can be used as the textbook for a one-semester or a twosemester course in computer algebra. The basic arithmetic algorithms are discussed in Chapters 2 and 3, and Sections 4.1-4.4, 5.1-5.5, 8.1-8.2, 9.1-9.4, 14.1-14.6, and 15.1-15.2. In addition, a one-semester undergraduate course might be slanted towards computational number theory (9.5, 18.1–18.4, and parts of Chapter 20), geometry (21.1-21.6), or integration (4.5, 5.11, 6.2-6.4, and Chapter 22), supplemented by fun applications from 4.6-4.8, 5.6-5.9, 6.8, 9.6, Chapter 13, and Chapters 1 and 24. A two-semester course could teach the "basics" and 6.1-6.7, 10.1-10.2, 15.4-15.6, 16.1-16.5, 18.1-18.3, 19.1-19.2, 19.4, 19.5 or 19.6-19.7, and one or two of Chapters 21-23, maybe with some applications from Chapters 17, 20, and 24. A graduate course can be more eclectic. We once taught a course on "factorization", using parts of Chapters 14-16 and 19. Another possibility is a graduate course on "fast algorithms" based on Part II. For any of these suggestions, there is enough material so that an instructor will still have plenty of choice of which areas to skip. The logical dependencies between the chapters are given in Figure 1.

The prerequisite for such a course is linear algebra and a certain level of mathematical maturity; particularly useful is a basic familiarity with algebra and analysis of algorithms. However, to allow for the large variations in students' background, we have included an appendix that presents the necessary tools. For that material, the border between the boring and the overly demanding varies too much to get it right for everyone. If those notions and tools are unfamiliar, an instructor may have to expand beyond the condensed description in the appendix. Otherwise,

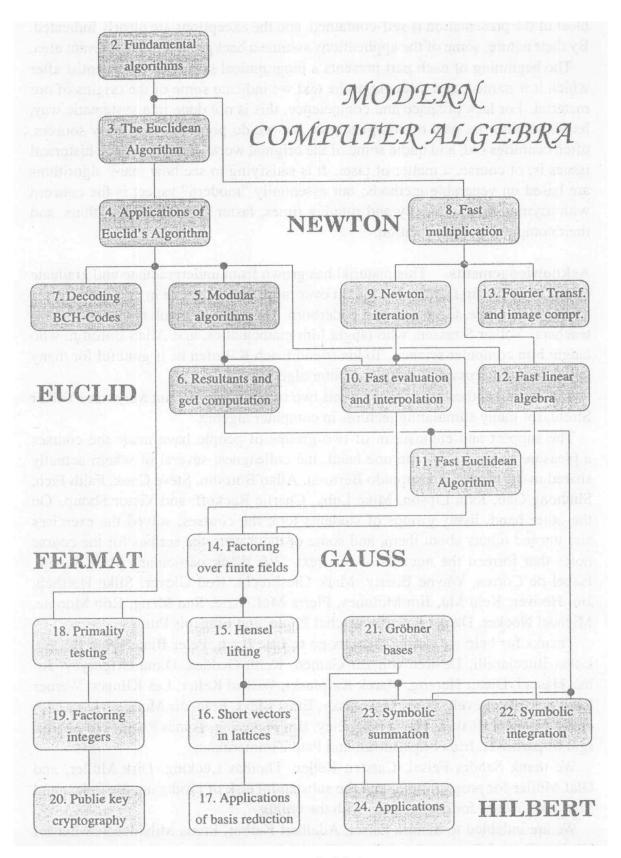


FIGURE 1: Leitfaden.

most of the presentation is self-contained, and the exceptions are clearly indicated. By their nature, some of the applications assume a background in the relevant area.

The beginning of each part presents a biographical sketch of the scientist after which it is named, and throughout the text we indicate some of the origins of our material. For lack of space and competence, this is not done in a systematic way, let alone with the goal of completeness, but we do point to some early sources, often centuries old, and quote some of the original work. Interest in such historical issues is, of course, a matter of taste. It is satisfying to see how many algorithms are based on venerable methods; our essentially "modern" aspect is the concern with asymptotic complexity and running times, faster and faster algorithms, and their computer implementation.

Acknowledgements. This material has grown from undergraduate and graduate courses that the first author has taught over more than a decade in Toronto, Zürich, Santiago de Chile, Canberra, and Paderborn. He wants to thank above all his two teachers: Volker Strassen, who taught him mathematics, and Allan Borodin, who taught him computer science. To his friend Erich Kaltofen he is grateful for many enlightening discussions about computer algebra.

The second author wants to thank his two supervisors, Helmut Meyn and Volker Strehl, for many stimulating lectures in computer algebra.

The support and enthusiasm of two groups of people have made the courses a pleasure to teach. On the one hand, the colleagues, several of whom actually shared in the teaching: Leopoldo Bertossi, Allan Borodin, Steve Cook, Faith Fich, Shuhong Gao, John Lipson, Mike Luby, Charlie Rackoff, and Victor Shoup. On the other hand, lively groups of students took the courses, solved the exercises and tutored others about them, and some of them were the scribes for the course notes that formed the nucleus of this text. We thank particularly Paul Beame, Isabel de Correa, Wayne Eberly, Mark Giesbrecht, Rod Glover, Silke Hartlieb, Jim Hoover, Keju Ma, Jim McInnes, Pierre McKenzie, Sun Meng, Rob Morenz, Michael Nöcker, Daniel Panario, Michel Pilote, and François Pitt.

Thanks for help on various matters go to Eric Bach, Peter Blau, Wieb Bosma, Louis Bucciarelli, Désirée von zur Gathen, Keith Geddes, Dima Grigoryev, Johan Håstad, Dieter Herzog, Marek Karpinski, Wilfrid Keller, Les Klinger, Werner Krandick, Ton Levelt, János Makowsky, Ernst Mayr, François Morain, Gerry Myerson, Michael Nüsken, David Pengelley, Bill Pickering, Tomás Recio, Jeff Shallit, Igor Shparlinski, Irina Shparlinski, and Paul Zimmermann.

We thank Sandra Feisel, Carsten Keller, Thomas Lücking, Dirk Müller, and Olaf Müller for programming and the substantial task of producing the index, and Marianne Wehry for tireless help with the typing.

We are indebted to Sandra Feisel, Adalbert Kerber, Preda Mihăilescu, Michael Nöcker, Daniel Panario, Peter Paule, Daniel Reischert, Victor Shoup, and Volker Strehl for carefully proofreading parts of the draft.