Algorithms and Computation in Mathematics • Volume 1

Manuel Bronstein

Symbolic Integration I

Transcendental Functions

Second Edition

Springer

Manuel Bronstein

INRIA
2004 route des Lucioles – B.P. 93
06902 Sophia Antipolis Cedex, France
e-mail: Manuel.Bronstein@sophia.inria.fr

Mathematics Subject Classification (2000): 12F20, 12H05, 12Y05, 13N15, 28-04, 33B10, 33F10, 68W30

Library of Congress Control Number: 2004110974

ISSN 1431-1550

ISBN 3-540-21493-3 Springer Berlin Heidelberg New York
ISBN 3-540-60521-5 1st edition Springer-Verlag Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springeronline.com

Typesetting: by the author using a Springer LaTeX macro package
Production: LE-TeX Jelonek, Schmidt & Vöckler GbR, Leipzig
Cover design: *design & production* GmbH, Heidelberg

Printed on acid-free paper 46/3142YL - 5 4 3 2 1 0

Foreword

This book brings together two streams of research in mathematics and computing that were begun in the nineteenth century and made possible through results brought to fruition in the twentieth century.

Methods for indefinite integration have been important ever since the invention of the calculus in the 1700s. In the 1800s Abel and Liouville began the earliest mathematical research on algorithmic methods on integration in finite terms leading to what might be considered today as an early mathematical vision of a complete algorithmic solution for integrating elementary functions. In an 1842 publication Lady Ada Augusta, Countess of Lovelace, describing the capabilities of Babbage's analytical engine put forth the vision that computational devices could do algebraic as well as numerical calculations when she said that "[Babbage's Analytical Engine] can arrange and combine its numerical quantities exactly as if they were *letters* or any other *general* symbols; and in fact it might bring out its results in algebraical *notation* were provisions made accordingly." Thus these two visions set the stage for a century and a half of research that partially culminates in this book.

Progress in the mathematical realm continued through out the nineteenth and twentieth centuries. The Russian mathematician Mordukhai-Boltovskoi wrote the first two books on this subject in 1910 and 1913[1].

With the invention of electronic computers in the late 1930s and early 1940s, a new impetus was given to both the mathematical and computational streams of work. In the meantime in the mathematical world important progress had been made on algebraic methods of research. Ritt began to apply the new algebraic techniques to the problem of integration in finite terms, an approach that has proven crucially important. In 1948 he published the results of his research in a little book, *Integration in Finite Terms.* The use of these algebraic ideas were brought to further fruition by Kolchin, Rosenlicht, and, particularly for problems of symbolic integration, by three of Rosenlicht's Ph.D. students — Risch, Singer, and Bronstein[2].

[1] *On the Integration in Finite Terms of Linear Differential Equations.* Warsaw, 1910 (in Russian) and *On the Integration of Transcendental Functions.* Warsaw, 1913 (in Russian).

[2] Let me hasten to add that there have been important contributions by many others and it is not my intention to give a complete history of the field in this short paragraph, but to indicate some of main streams of work that have led to the current book.

On the computational side, matters rested until 1953 when two early programs were written, one by Kahrimanian at Temple University and another by Nolan at Massachusetts Institute of Technology, to do analytic differentiation — the inverse of indefinite integration. There was active research in the late 1950s and early 1960s on list processing packages and languages that laid the implementation foundations for today's computer algebra systems. Slagle's 1961 thesis was an early effort to write a program, in LISP, to do symbolic integration. With the advent of general computer algebra systems, some kind of symbolic integration facility was implemented in most. These integration capabilities opened the eyes of many early users of symbolic mathematical computation to the amazing potential of this form of computation. But yet none of the systems had a complete implementation of the full algorithm that Risch had announced in barest outline in 1970. There were a number of reasons for this. First and foremost, no one had worked out the many aspects of the problem that Risch's announcement left incomplete.

Starting with his Ph.D. dissertation and continuing in a series of beautiful and important papers, Bronstein set out to fill in the missing components of Risch's 1970 announcement. Meanwhile working at the IBM T. J. Watson Research Center, he carried out an almost complete implementation of the integration algorithms for elementary functions. It is the most complete implementation of symbolic integration algorithms to date.

In this book, Bronstein brings these mathematical and computational streams of research together in a highly effective manner. He presents the algorithmic details in pseudo-code that is easy to implement in most of the general computer algebra systems. Indeed, my students and I have implemented and tested many of the algorithms in MAPLE and MACSYMA. Bronstein's style and appropriate level of detail makes this a straightforward task, and I expect this book to be the standard starting place for future implementers of symbolic integration algorithms. Along with the algorithms, he presents the mathematics necessary to show that the algorithms work correctly. This is a very interesting story in its own right and Bronstein tells it well. Nonetheless, for those primarily interested in the algorithms, much of the mathematics can be skipped at least in a first study. But the full beauty of the subject is to be most appreciated by studying both aspects.

The full treatment of the subject is a long one and it is not finished in this volume. The longer and more difficult part involving the integration of algebraic functions must await a second volume. This volume serves as a good foundation to the topic of symbolic integration and as a nice introduction to the literature for integration of algebraic functions and for other aspects such as integration involving non-elementary functions. Study, learn, implement, and enjoy!

B. F. Caviness

Preface to the Second Edition

I have taken the opportunity of this second edition to add a chapter on parallel integration, a method that is used by several computer algebra systems, either before or in place of the complete integration algorithm. I have also added new references and exercises that expand on topics such as obtaining continuous integrals or the relations between special polynomials, Darboux polynomials and constants in monomial extensions.

I would like to thank all the readers of the first edition who have sent me various corrections and suggestions. While I have tried to incorporate all of them in this edition, I remain responsible for the remaining errors.

Sophia Antipolis, June 2004 *M. Bronstein*

Preface to the First Edition

The integration problem, which is as old as calculus and differentiation, can be informally stated very concisely: given a formula for a function $f(x)$, determine whether there is a formula for a differentiable function $F(x)$ satisfying

$$\frac{dF}{dx} = f(x)$$

and compute such an $F(x)$, which is called an antiderivative of $f(x)$ and is denoted

$$F(x) = \int f(x)dx$$

if it exists. Yet, while symbolic differentiation is a rather simple mechanical process, suitable as an exercise in a first course in analysis or computer programming, the inverse problem has been challenging scientists since the time of Leibniz and Newton, and is still a challenge for mathematicians and computer scientists today. Despite the many great strides made since the 19th century in showing that integration is in essence a mechanical process, although quite more complicated than differentiation, most calculus and analysis textbooks give students the impression that integration is at best a mixture of art and science, with flair in choosing the right change of variable or approach being an essential ingredient, as well as a comprehensive table of integrals.

The goal of this book is to show that computing symbolic antiderivatives is in fact an algorithmic process, and that the integration procedure for transcendental functions can be carried out by anyone with some familiarity with polynomial arithmetic. The integration procedure we describe is also capable of deciding when antiderivatives are not elementary, and proving it as a byproduct of its calculations. For example the following classical nonelementary integrals

$$\int e^{x^2} dx\,, \qquad \int \frac{dx}{\log(x)}\,, \qquad \int \frac{\sin(x)dx}{x}\,,$$

can be proven nonelementary with minimal calculations.

The algorithmic approach, pioneered by Abel and Liouville in the past century, eventually succeeded in producing a mechanical procedure for deciding whether an elementary function has an elementary antiderivative, and for computing one if so. This procedure, which Risch described in a series of reports [73, 74, 75, 76], unfortunately not all of them published, forms the basis of most of the symbolic integration algorithms of the past 20 years, all of them loosely grouped under the appellation *Risch algorithm*. The procedure which we describe in this book also has its roots in the original Risch algorithm [75] and its improvements, our main sources besides Risch being [12, 13, 83, 89].

We have tried to keep the presentation as elementary as possible, with the minimal background for understanding the algorithm being an introductory course in algebra, where the topics rings and fields, polynomial greatest common divisors, irreducible polynomials and resultants are covered[3]. Some additional background in field theory, essentially algebraic and transcendental extensions, is occasionally used in the proofs associated with the algorithm. The reader willing to accept the algorithm without proof can skip those sections while learning the algorithm.

We have also generalized and extended the original Risch algorithm to a wider class of functions, thereby offering the following features, some of them new, to the reader already familiar with symbolic integration:

- The algorithms in this book use only rational operations, avoiding factorization of polynomials into irreducibles.
- Extensions by tangents and arc-tangents are treated directly, thereby real trigonometric functions are integrated without introducing complex exponentials and logarithms in the computations.
- Antiderivatives in elementary extensions can still be computed when arbitrary primitives are allowed in the integrand, *e.g.* $\mathrm{Erf}(x)$, rather than logarithms.
- Several subalgorithms are applicable to a large class of non-Liouvillian extensions, thereby allowing integrals to be computed for such functions.

The material in this book has been used in several courses for advanced undergraduates in mathematics or computer science at the Swiss Federal Institute of Technology in Zurich:

- In a one-semester course on symbolic integration, emphasizing the algorithmic and implementation aspects. This course covers Chap. 2 in depth, Chap. 3 and 4 superficially, then concentrates on Chap. 5, 6, 7 and 8.
- In the first part of a one-semester course on differential algebra. This course covers Chap. 3, 4 and 5 in depth, turning after Liouville's Theorem to other topics (e.g. differential Galois theory).
- In the last part of a one-semester introductory course in computer algebra, where some algorithms from Chap. 2 and 5 are presented, usually without proofs.

[3] Those topics are reviewed in Chap. 1.

In all those courses, the material of Chap. 1 is covered as and when needed, depending on the background of the students. Chap. 9 contains complete proofs of several structure theorems and can be presented independently of the rest of this book.

By presenting the algorithm in pseudocode in various "algorithm boxes" throughout the text, we also hope to make this book useful for programmers implementing symbolic integrators: by following the pseudocode, they should be able to write an integrator without studying in detail the associated theory.

The reader will notice that several topics in symbolic integration are missing from this book, the main one being the integration of algebraic functions. Including algorithms for integrating algebraic and mixed algebraic-transcendental functions would however easily double the size of this book, as well as increase the mathematical prerequisites, since those algorithms require prior familiarity with algebraic curves and functions. We have thus decided to cover algebraic functions in a second volume, which will hopefully appear in the near future. In the meantime, this book is an adequate preparation to the extensive literature on the integration of algebraic functions [8, 9, 11, 14, 29, 73, 74, 76, 91]. Another related topic is integration in nonelementary terms, *i.e.* with new special functions allowed in the antiderivatives. Here also, the reader should have no difficulty moving on to the research literature [5, 6, 21, 22, 52, 53, 70, 94] after completing this book.

Acknowledgements
I am thankful to several colleagues and students who have read and corrected many early drafts of this book. I am particularly grateful to Bob Caviness, Thom Mulders and Paul Zimmermann, who corrected many errors in the final text and suggested several improvements. Sergei Abramov, Cedric Bächler, Johannes Grabmeier, David Stoutemyer, Jacques-Arthur Weil and Clifton Williamson have also helped a great deal with their corrections and suggestions. Of course, I am fully responsible for any error that may remain.

Finally, I wish to thank Dr. Martin Peters and his staff at Springer-Verlag for their great patience with this project.

Zurich, November 1996 *M. Bronstein*

Contents

1

Algebraic Preliminaries

We review in this chapter the basic algebraic structures and algorithms that will be used throughout this book. This chapter is not intended to be a replacement for an introductory course in abstract algebra, and we expect the reader to have already encountered the definitions and fundamental properties of rings, fields and polynomials. We only recall those definitions here and describe some algorithms on polynomials that are not always covered in introductory algebra courses. Since they are well-known algorithms in computer algebra, we do not reprove their correctness here, but give references instead. For a comprehensive introduction to constructive algebra and algebraic algorithms, including more efficient alternatives for computing greatest common divisors of polynomials, we recommend consulting introductory computer algebra textbooks [2, 28, 39, 64, 97]. Readers with some background in algebra can skip this chapter and come back to it later as needed.

1.1 Groups, Rings and Fields

An algebraic structure is usually a set together with one or more operations on it, operations that satisfy some computation rules called axioms. In order not to always list all the satisfied axioms for a given structure, short names have been given to the most common structures. Groups, rings and fields are such structures, and we recall their definitions in this section.

Definition 1.1.1. *A* group $(G, \circ)$ *is a nonempty set* G*, together with an operation* $\circ : G \times G \to G$ *satisfying the following axioms:*

(i) (Associativity) $\forall a, b, c \in G, a \circ (b \circ c) = (a \circ b) \circ c$.
(ii) (Identity element) $\exists e \in G$ *such that* $\forall a \in G, e \circ a = a \circ e = a$.
(iii) (Inverses) $\forall a \in G, \exists a^{-1} \in G$ *such that* $a \circ a^{-1} = a^{-1} \circ a = e$.

In addition, $\circ$ *is called* commutative *(or* Abelian*) if* $a \circ b = b \circ a$ *for all* $a, b \in G$*, and* $(G, \circ)$ *is called a* commutative group *(or* Abelian group*) if it is a group and* $\circ$ *is commutative.*

Example 1.1.1. Let $G = GL(\mathbb{Q}, 2)$ be the set of all the 2 by 2 matrices with rational number coefficients and nonzero determinant, and let $\circ$ denote the usual matrix multiplication. $(G, \circ)$ is then a group: associativity can easily be checked, the identity element is the identity matrix, and the inverse of a matrix in G is given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

which is in G since the determinant of any element of G is nonzero. Note that $(G, \circ)$ is not a commutative group since

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \circ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$$

and

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \circ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}.$$

Example 1.1.2. Let $G = \mathcal{M}_{2,2}(\mathbb{Q})$ be the set of all the 2 by 2 matrices with rational number coefficients, and let $\circ$ denote the usual matrix addition. It can easily be checked that $(G, \circ)$ is a commutative group with the zero matrix as identity element.

Definition 1.1.2. *A* ring $(R, +, \cdot)$ *is a set R, together with two operations $+ : R \times R \to R$ and $\cdot : R \times R \to R$ such that:*

(i) $(R, +)$ is a commutative group.
(ii) (Associativity) $\forall a, b, c \in R, a \cdot (b \cdot c) = (a \cdot b) \cdot c$.
(iii) (Multiplicative identity) $\exists i \in R$ such that $\forall a \in R, i \cdot a = a \cdot i = a$.
(iv) (Distributivity)

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

$(R, +, \cdot)$ is called a commutative ring *if it is a ring and $\cdot$ is commutative. In addition, we define the* characteristic *of R to be 0 if $n\,i \neq e$ for any positive integer n, the smallest positive integer m such that $m\,i = e$ otherwise. Let R and S be rings. A map $\phi : R \to S$ is a* ring–homomorphism *if $\phi(e_R) = e_S$, $\phi(i_R) = i_S$, and $\phi(a+b) = \phi(a)+\phi(b)$ and $\phi(ab) = \phi(a)\cdot\phi(b)$ for any $a, b \in R$. A* ring–isomorphism *is a bijective ring–homomorphism.*

In the rest of this book, whenever $(R, +, \cdot)$ is a ring, we write 0 for the identity element of R with respect to $+$, 1 for the identity element of R with respect to $\cdot$, and for $a, b \in R$, we write ab instead of $a \cdot b$.

Example 1.1.3. Let $R = \mathcal{M}_{2,2}(\mathbb{Q})$ be the set of all the 2 by 2 matrices with rational number coefficients, and let $+$ denote matrix addition and $\cdot$ denote matrix multiplication. $(R, +, \cdot)$ is then a ring, but not a commutative ring (see example 1.1.1). Since

$$n\,i = n\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} = \begin{pmatrix}n & 0\\ 0 & n\end{pmatrix}$$

is nonzero for any positive integer n, R has characteristic 0.

Example 1.1.4. Let $R = \mathbb{Z}_6$ (the integers modulo 6) with $+$ and $\cdot$ being the addition and multiplication of integers modulo 6. $(R, +, \cdot)$ is then a commutative ring, and the map $\phi : \mathbb{Z} \to \mathbb{Z}_6$ defined by $\phi(n) = n \pmod 6$ is a ring-homomorphism. Since $1+1+1+1+1+1 = 0$ in $\mathbb{Z}_6$, and $n1 \neq 0$ for $0 < n < 6$, $\mathbb{Z}_6$ has characteristic 6. Note that $2 \cdot 3 = 0$ in $\mathbb{Z}_6$, while $2 \neq 0$ and $3 \neq 0$, so we cannot in general deduce from an equation $ab = 0$ that either a or b must be 0. Commutative rings where we can make this simplification are very useful and common, so they receive a special name.

Definition 1.1.3. *An* integral domain $(R, +, \cdot)$ *is a commutative ring where* $0 \neq 1$ *and*

$$\forall a, b \in R, a \cdot b = 0 \Longrightarrow a = 0 \text{ or } b = 0.$$

Example 1.1.5. Let $R = \mathbb{Z}[\sqrt{-5}] = \{a+b\sqrt{-5}; a, b \in \mathbb{Z}\}$ with $+$ and $\cdot$ denoting complex addition and multiplication. $(R, +, \cdot)$ is then an integral domain.

We now come to the problem of factoring, *i.e.* writing elements of an integral domain as a product of other elements.

Definition 1.1.4. *Let $(R, +, \cdot)$ be an integral domain, and $x, y \in R$. We say that x* divides *y, and write $x \mid y$, if $y = xt$ for some $t \in R$. An element $x \in R$ is called a* unit *if $x \mid 1$. The set of all the units of R is written R^*. We say that $z \in R$ is a* greatest common divisor (gcd) *of $x_1, \ldots, x_n$ and write $z = \gcd(x_1, \ldots, x_n)$ if:*

(i) $z \mid x_i$ *for* $1 \leq i \leq n$,
(ii) $\forall t \in R, t \mid x_i$ *for* $1 \leq i \leq n \Longrightarrow t \mid z$.

In addition, we say that x and y are coprime *if there exists a unit $u \in R^*$, which is a gcd of x and y.*

Example 1.1.6. Let $R = \mathbb{Z}\left[\sqrt{-5}\right]$ as in example 1.1.5, $x = 6$ and $y = 2 + 2\sqrt{-5}$. A norm argument shows that x and y have no gcd in R. Let $N : R \to \mathbb{Z}$ be the map given by $N(a + b\sqrt{-5}) = a^2 + 5b^2$ for $a, b \in \mathbb{Z}$. It can easily be checked that $N(uv) = N(u)N(v)$ for any $u, v \in R$, so $u \mid v$ in R implies that $N(u) \mid N(v)$ in $\mathbb{Z}$. Suppose that $z \in R$ is a greatest common divisor of x and y, and let $n = N(z) \geq 0$. Then, $n \mid N(x) = 36$ and $n \mid N(y) = 24$, so $n \mid 12$ in $\mathbb{Z}$. We have $2 \mid x$ and $2 \mid y$ in R, so $4 = N(2) \mid n$ in $\mathbb{Z}$. In addition, $1 + \sqrt{-5} \mid y$ in R, and

$$6 = 2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5}) \tag{1.1}$$

so $1 + \sqrt{-5} \mid x$ in R, hence $6 = N(1 + \sqrt{-5}) \mid n$ in $\mathbb{Z}$. Thus, $12 \mid n$ in $\mathbb{Z}$, so $n = 12$. Writing $z = a + b\sqrt{-5}$ for some $a, b \in \mathbb{Z}$, this implies that $N(z) = a^2 + 5b^2 = 12$, hence that $a^2 \equiv 2 \pmod 5$. But the squares in $\mathbb{Z}_5$ are $0, 1$ and 4, so this equation has no solution, implying that x and y have no gcd in R.

Although gcd's do not always exist, whenever they exist, they are unique up to multiplication by units.

Theorem 1.1.1. *Let $(R,+,\cdot)$ be an integral domain, and $x, y \in R$. If z and t are both gcd's of x and y, then $z = ut$ and $t = vz$ for some $u, v \in R^*$.*

Proof. Suppose that both z and t are gcd's of x and y. Then, $t \mid z$ since $t \mid x$, $t \mid y$, and $z = \gcd(x,y)$. Thus, $z = ut$ for some $u \in R$. Similarly, $z \mid t$, so $t = vz$ for some $v \in R$. Hence $z = ut = uvz$, so $(1 - uv)z = 0$. If $z \neq 0$, then $1 = uv$, so $u, v \in R^*$. If $z = 0$, then $t = vz = 0$, so $z = 1t$ and $t = 1z$. □

Definition 1.1.5. *Let R be an integral domain. A nonzero element $p \in R \backslash R^*$ is called* prime *if for any $a, b \in R$, $p \mid ab \Longrightarrow p \mid a$ or $p \mid b$. A nonzero element $p \in R \setminus R^*$ is called* irreducible *if for any $a, b \in R$, $p = ab \Longrightarrow a \in R^*$ or $b \in R^*$.*

Example 1.1.7. Let $R = \mathbb{Z}\left[\sqrt{-5}\right]$ as in example 1.1.5, and check that $2, 3, 1 + \sqrt{-5}$ and $1 - \sqrt{-5}$ are all irreducible elements of R. Equation (1.1) then shows that the same element can have several different factorizations into irreducibles. Therefore, integral domains where such a factorization is unique receive a special name.

Definition 1.1.6. *A* unique factorization domain (UFD) *$(R,+,\cdot)$ is an integral domain where for any nonzero $x \in R \setminus R^*$, there are $u \in R^*$, co-prime irreducibles $p_1, \ldots, p_n \in R$ and positive integers $e_1, \ldots, e_n$ such that $x = u\,p_1^{e_1} \cdots p_n^{e_n}$. Furthermore, this factorization is unique up to multiplication of u and the p_i's by units and up to permutation of the indices.*

Example 1.1.8. Let $R = \mathbb{Q}[X, Y]$ be the set of all the polynomials in the variables X and Y and with rational number coefficients. It is a classical result ([54] Chap. V §6, [92] §5.4) that $(R,+,\cdot)$ is a unique factorization domain where $+$ and $\cdot$ denote polynomial addition and multiplication respectively.

In any integral domain, a prime is always irreducible. The converse is not always true, but it holds in unique factorization domains. Thus, we can use interchangeably "prime" or "irreducible" whenever we are in a unique factorization domain, so, "the prime factorization of x" and "the irreducible factorization of x" have the same meaning.

Theorem 1.1.2 ([54] Chap. II §4). *Let $(R,+,\cdot)$ be an integral domain. Then every prime $p \in R$ is irreducible. If R is a unique factorization domain, then every irreducible $p \in R$ is prime.*

In addition, gcd's always exist in UFD's, and can be obtained from the irreducible factorizations.

Theorem 1.1.3. *If R is a UFD, then any $x, y \in R$ have a gcd in R.*

Proof. Let $x, y \in R$, and suppose first that $x = 0$. Then $y \mid y$, $y \mid 0$, and any $t \in R$ that divides x and y must divide y, so y is a gcd of x and y. Similarly, x is a gcd of x and y if $y = 0$, so suppose now that $x \neq 0$ and $y \neq 0$, and let $x = u \prod_{p \in \mathcal{X}} p^{n_p}$ and $y = v \prod_{p \in \mathcal{Y}} p^{m_p}$ be the irreducible factorizations of x and y, where $\mathcal{X}$ and $\mathcal{Y}$ are finite sets of irreducibles. We choose the units u and v so that any irreducible dividing both x and y is in $\mathcal{X} \cap \mathcal{Y}$. Let then

$$z = \prod_{p \in \mathcal{X} \cap \mathcal{Y}} p^{\min(n_p, m_p)} \in R. \tag{1.2}$$

We have

$$x = z\, u \prod_{p \in \mathcal{X} \cap \mathcal{Y}} p^{n_p - \min(n_p, m_p)} \prod_{p \in \mathcal{X} \setminus \mathcal{Y}} p^{n_p}$$

so $z \mid x$. A similar formula shows that $z \mid y$. Suppose that $t \mid x$ and $t \mid y$ for some $t \in R$, and let $t = w \prod_{p \in \mathcal{T}} p^{e_p}$ be its irreducible factorization where $\mathcal{T}$ is a finite set of irreducibles. For $p \in \mathcal{T}$, we have $x = tb = p^{e_p} ab$ for some $a, b \in R$, so $sp \in \mathcal{X}$ for some $s \in R^*$. Replacing w by ws^{-e_p}, we can assume that $p \in \mathcal{X}$, and $e_p \le n_p$ by the unicity of the irreducible factorization. Similarly, we get $p \in \mathcal{Y}$ and $e_p \le m_p$ since $t \mid y$. Hence, $\mathcal{T} \subseteq \mathcal{X} \cap \mathcal{Y}$ and $e_p \le \min(n_p, m_p)$ for any $p \in \mathcal{T}$. Thus,

$$z = t\, w^{-1} \prod_{p \in \mathcal{T}} p^{\min(n_p, m_p) - e_p} \prod_{p \in (\mathcal{X} \cap \mathcal{Y}) \setminus \mathcal{T}} p^{\min(n_p, m_p)}$$

which means that $t \mid z$, hence that $z = \gcd(x, y)$. □

It is a classical result due to Gauss that polynomials can be factored uniquely into irreducibles.

Theorem 1.1.4 ([54] Chap. V §6, [92] §5.4). *If R is a UFD, then the polynomial ring $R[X_1, \dots, X_n]$ is a UFD.*

Definition 1.1.7. *Let $(G, \circ)$ be a group with identity element e. We say that $H \subseteq G$ is a* subgroup of $(G, \circ)$ *if:*

(i) $e \in H$.
(ii) $\forall a, b \in H, a \circ b \in H$.
(iii) $\forall a \in H, a^{-1} \in H$.

In practice, given a subset H of a group G, it is equivalent to check the above properties (i), (ii) and (iii), or that H is not empty and that $a \circ b^{-1} \in H$ for any $a, b \in H$.

Example 1.1.9. Let $G = GL(\mathbb{Q}, 2)$ as in example 1.1.1 with $\circ$ denoting matrix multiplication, and let $H = SL(\mathbb{Q}, 2)$ be the subset of G consisting of all the matrices whose determinant is equal to 1. The identity matrix is in H, so H is not empty, and for any $a, b \in H$, the determinant of $a \circ b^{-1}$ is the quotient of the determinant of a by the determinant of b, which is 1, so H is a subgroup of G.

Definition 1.1.8. *Let $(R,+,\cdot)$ be a commutative ring. A subset I of R is called an* ideal *if $(I,+)$ is a subgroup of $(R,+)$ and $xa \in I$ for any x in R and a in I. Let $x_1,\dots,x_n \in R$. The* ideal generated by *$\{x_1,\dots,x_n\}$ is the smallest ideal of R containing $\{x_1,\dots,x_n\}$, and is denoted $(x_1,\dots,x_n)$. An ideal $I \subseteq R$ is called* principal *if $I = (x)$ for some $x \in R$.*

In fact, the ideal generated by $\{x_1,\dots,x_n\}$ is just the set of all the linear combinations of the x_i's with coefficients in R.

Theorem 1.1.5. *Let $(R,+,\cdot)$ be a commutative ring, and $x_1,\dots,x_n \in R$. Then,*

$$(x_1,\dots,x_n) = \{a_1x_1 + \cdots + a_nx_n; a_1,\dots,a_n \in R\}.$$

Proof. Let $I = \{a_1x_1 + \cdots + a_nx_n, a_1,\dots,a_n \in R\}$. Then $x_i \in I$ for any i. Let $a = \sum_{i=1}^n a_ix_i \in I$ and $b = \sum_{i=1}^n b_ix_i \in I$. We have $a - b = \sum_{i=1}^n (a_i - b_i)x_i \in I$, so $(I,+)$ is a subgroup of $(R,+)$. For any $x \in R$, we have $xa = \sum_{i=1}^n (xa_i)x_i \in I$, so I is an ideal of R containing $\{x_1,\dots,x_n\}$. Let now J be any ideal of R containing $\{x_1,\dots,x_n\}$, and let $a = \sum_{i=1}^n a_ix_i \in I$. For each i, $x_i \in J$, so $a_ix_i \in J$ since $RJ \subseteq J$, so $a \in J$ since $(J,+)$ is a group. Hence $I \subseteq J$, so $I = (x_1,\dots,x_n)$. □

Example 1.1.10. Let $R = \mathbb{Q}[X,Y]$ as in example 1.1.8, and let $I = (X,Y)$. It can be checked that I is not principal, hence that not every ideal of R is principal. Naturally, this means that integral domains where every ideal is principal receive a special name.

Definition 1.1.9. *A* principal ideal domain (PID) *$(R,+,\cdot)$ is an integral domain where any ideal is principal.*

Example 1.1.11. Let $R = \mathbb{Q}[X]$ be the set of all the univariate polynomials in X with rational number coefficients. $(R,+,\cdot)$ is then a principal ideal domain ([54] Chap. V §4, [92] §3.7) where $+$ and $\cdot$ denote polynomial addition and multiplication respectively.

The last, and most useful, type of ring that we use in this book, is an integral domain in which Euclidean division can be carried out.

Definition 1.1.10. *A* Euclidean domain *$(R,+,\cdot)$ is an integral domain together with a map $\nu : R \setminus \{0\} \to \mathbb{N}$ such that:*

(i) $\forall a,b \in R \setminus \{0\}, \nu(ab) \geq \nu(a)$.
(ii) (Euclidean division) For any $a,b \in R$, $b \neq 0$, there are $q,r \in R$ such that $a = bq + r$ and either $r = 0$ or $\nu(r) < \nu(b)$.

The map ν is called the size function *of the Euclidean domain.*

Example 1.1.12. The ring $(\mathbb{Z},+,\cdot)$ of the integers with the usual addition and multiplication is a Euclidean domain with $\nu(a) = |a|$, a fact that was known to Euclid, and which is the origin of the name.

Even though the notions of principal ideal domains and Euclidean domains are defined for an arbitrary integral domain, there is in fact a linear hierarchy of integral domains.

Theorem 1.1.6 ([92] §3.7). *Every Euclidean domain is a PID.*

Theorem 1.1.7 ([54] Chap. II §4, [92] §3.8). *Every PID is a UFD.*

Since every PID is a UFD, and gcd's always exist in UFD's by Theorem 1.1.3, then gcd's always exist in PID's. We show that in PID's, the gcd of two elements generates the same ideal than them.

Theorem 1.1.8. *If R is a PID, then $(x, y) = (\gcd(x, y))$ for any $x, y \in R$.*

Proof. Let $x, y \in R$ and $z \in R$ be a generator of the ideal (x, y), *i.e.* $(z) = (x, y)$. Then, $x \in (z)$, so $x = zu$ for some $u \in R$, which means that $z \mid x$. Similarly, $y \in (z)$, so $z \mid y$. In addition, $z \in (x, y)$, so $z = ax + by$ for some $a, b \in R$. Let $t \in R$ be such that $t \mid x$ and $t \mid y$. Then $x = ct$ and $y = dt$ for some $c, d \in R$. Hence, $z = act + bdt = (ac + bd)t$ so $t \mid z$, which implies that $z = \gcd(u, v)$. □

We finally recall some important definitions and results about fields.

Definition 1.1.11. *A* field $(F, +, \cdot)$ *is a commutative ring where* $(F \setminus \{0\}, \cdot)$ *is a group, i.e. every nonzero element is a unit* $(F^* = F \setminus \{0\})$.

Example 1.1.13. Let $F = \mathbb{Z}_5$ (the integers modulo 5) with $+$ and $\cdot$ being the addition and multiplication of integers modulo 5. $(F, +, \cdot)$ is then a field.

Example 1.1.14. Let R be an integral domain and define the relation $\sim$ on $R \times R \setminus \{0\}$ by $(a, b) \sim (c, d)$ if $ad = bc$. It can easily be checked that $\sim$ is an equivalence relation on $R \times R \setminus \{0\}$ and that the set of equivalence classes is a field with the usual operations

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd} \quad \text{and} \quad \frac{a}{b}\frac{c}{d} = \frac{ac}{bd}$$

where a/b denotes the equivalence class of (a, b). This field is called the *quotient field* of R. For example, the quotient field of $\mathbb{Z}$ is $\mathbb{Q}$ and the quotient field of the polynomial ring $D[x]$ is the rational function field $D(x)$ when D is an integral domain.

Definition 1.1.12. *Let $F \subseteq E$ be fields. An element $\alpha \in E$ is called* algebraic *over F if $p(\alpha) = 0$ for some nonzero polynomial $p \in F[X]$,* transcendental *over F otherwise. E is called an* algebraic extension *of F if all the elements of E are algebraic over F.*

Definition 1.1.13. *A field F is called* algebraically closed *if for every polynomial $p \in F[X] \setminus F$ there exists $\alpha \in F$ such that $p(\alpha) = 0$. A field E is called an* algebraic closure *of F if E is an algebraically closed algebraic extension of F.*

Note that if F is algebraically closed, then any $p \in K[X] \setminus K$ factors linearly as $p = c\prod_{i=1}^{n}(X - \alpha_i)^{e_i}$ over F: p must have one root α in F by definition, and $p/(X - \alpha)$ factors linearly over F by induction. The fundamental result about algebraic closures is a result of E. Steinitz which states that they exist and are essentially unique.

Theorem 1.1.9 ([54] Chap. VII §2, [92] §10.1). *Every field F has an algebraic closure, and any two algebraic closures of F are isomorphic.*

In view of the above theorem, we can refer to *the* algebraic closure of a field F, and we denote it $\overline{F}$. The last result we mention in this section is Hilbert's Nullstellensatz, which is not needed in the algorithm, but is needed in order to eliminate the possibility of new transcendental constants appearing in antiderivatives. We present it here in both its classical forms.

Theorem 1.1.10 (Weak Nullstellensatz, [92] §16.5**).** *Let F be an algebraically closed field, I an ideal of the polynomial ring $F[X_1, \dots, X_n]$ and $V(I)$ be the subset of F^n given by*

$$V(I) = \{(x_1, \dots, x_n) \in F^n \text{ s.t. } p(x_1, \dots, x_n) = 0 \text{ for all } p \in I\}. \tag{1.3}$$

Then, $V(I) = \emptyset \iff 1 \in I$.

Theorem 1.1.11 (Nullstellensatz, [54] Chap. X §2, [92] §16.5**).** *Let F be an algebraically closed field, I an ideal of the polynomial ring $F[X_1, \dots, X_n]$ and $V(I)$ be given by (1.3). For any $p \in F[X_1, \dots, X_n]$, if $p(x_1, \dots, x_n) = 0$ for every $(x_1, \dots, x_n) \in V(I)$, then $p^m \in I$ for some integer $m > 0$.*

1.2 Euclidean Division and Pseudo-Division

Let K be a field and x be an indeterminate over K. We first describe the classical polynomial division algorithm ([92] §3.4), which, given $A, B \in K[x]$, $B \neq 0$, produces unique $Q, R \in K[x]$ such that $A = BQ + R$ and either $R = 0$ or $\deg(R) < \deg(B)$. This shows that the polynomial ring $K[x]$ is a Euclidean domain with the degree for size function when K is field. Q and R are called the *quotient of A by B*, and the *remainder of A modulo B* respectively.

```
PolyDivide($A, B$)    (* Euclidean Polynomial Division *)

    (* Given a field $K$ and $A, B \in K[x]$ with $B \neq 0$, return $Q, R \in K[x]$ such
    that $A = BQ + R$ and either $R = 0$ or $\deg(R) < \deg(B)$. *)

    $Q \leftarrow 0$, $R \leftarrow A$
    while $R \neq 0$ and $\delta \leftarrow \deg(R) - \deg(B) \geq 0$ do
        $T \leftarrow \frac{\mathrm{lc}(R)}{\mathrm{lc}(B)} x^\delta$, $Q \leftarrow Q + T$, $R \leftarrow R - BT$
    return($Q, R$)
```

Example 1.2.1. Here is the Euclidean division of $A = 3x^3 + x^2 + x + 5$ by $B = 5x^2 - 3x + 1$ in $\mathbb{Q}[x]$:

Q	R	δ	T
0	$3x^3 + x^2 + x + 5$	1	$\frac{3}{5}x$
$\frac{3}{5}x$	$\frac{14}{5}x^2 + \frac{2}{5}x + 5$	0	$\frac{14}{25}$
$\frac{3}{5}x + \frac{14}{25}$	$\frac{52}{25}x + \frac{111}{25}$	-1	

Thus,

$$A = B\left(\frac{3}{5}x + \frac{14}{25}\right) + \left(\frac{52}{25}x + \frac{111}{25}\right).$$

This algorithm requires the coefficients to be from a field because it makes the quotient in K of the two leading coefficients. If K is an integral domain, the leading coefficient of B does not always divide exactly the leading coefficient of A, so Euclidean division is not always possible. For example it is not possible in the above example to do a Euclidean division of A by B in $\mathbb{Z}[x]$. But it is possible to apply **PolyDivide** to $25A$ and B in $\mathbb{Z}[x]$ since all the divisions in $\mathbb{Z}$ will then be exact. In general, given an integral domain D and $A, B \in D[x]$, applying **PolyDivide** to $b^{\delta+1}A$ and B where $b = \operatorname{lc}(B)$ and $\delta = \max(-1, \deg(A) - \deg(B))$ only generates exact divisions in D, and the Q and R it returns are respectively called the *pseudo-quotient of A by B* and *pseudo-remainder of A modulo B*. They satisfy $b^{\delta+1}A = BQ + R$ and either $R = 0$ or $\deg(R) < \deg(B)$. We write $\operatorname{pquo}(A, B)$ and $\operatorname{prem}(A, B)$ for the pseudo-quotient and pseudo-remainder of A and B. It is more efficient in practice to multiply A by b iteratively, as is done in the algorithm below, rather than once by $b^{\delta+1}$.

```
PolyPseudoDivide($A, B$)      (* Euclidean Polynomial Pseudo-Division *)

    (* Given an integral domain $D$ and $A, B \in D[x]$ with $B \neq 0$, return
    pquo$(A, B)$ and prem$(A, B)$. *)

    $b \leftarrow \operatorname{lc}(B)$, $N \leftarrow \deg(A) - \deg(B) + 1$, $Q \leftarrow 0$, $R \leftarrow A$
    while $R \neq 0$ and $\delta \leftarrow \deg(R) - \deg(B) \geq 0$ do
        $T \leftarrow \operatorname{lc}(R)x^\delta$, $N \leftarrow N - 1$, $Q \leftarrow bQ + T$, $R \leftarrow bR - TB$
    return($b^N Q, b^N R$)
```

Example 1.2.2. With A and B as in example 1.2.1, we get $b = 5$, $N = 2$, and

Q	R	δ	T	N
0	$3x^3 + x^2 + x + 5$	1	$3x$	1
$3x$	$14x^2 + 2x + 25$	0	14	0
$15x + 14$	$52x + 111$	-1		

so $25A = B(15x + 14) + (52x + 111)$.

1.3 The Euclidean Algorithm

Let D be a Euclidean domain and $\nu : D \setminus \{0\} \to \mathbb{N}$ its size function. The Euclidean division in D can be used to compute the greatest common divisor of any two elements of D. The basic idea, which goes back to Euclid who used it to compute the gcd of two integers, is that if $a = bq + r$, then $\gcd(a, b) = \gcd(b, r)$. Since $\gcd(x, 0) = x$ for any $x \in D$, the last nonzero element in the sequence $(a_i)_{i \geq 0}$ defined by

$$a_0 = a, \quad a_1 = b, \quad \text{and} \quad (q_i, a_i) = \textbf{EuclideanDivision}(a_{i-2}, a_{i-1}) \text{ for } i \geq 2$$

is then a gcd of a and b. Since for $a_i \neq 0$ and $i \geq 1$, either $a_{i+1} = 0$ or $\nu(a_{i+1}) < \nu(a_i)$, that sequence can only have a finite number of nonzero elements. This yields an algorithm for computing $\gcd(a, b)$ by repeated Euclidean divisions.

```
Euclidean($a, b$)      (* Euclidean algorithm *)

    (* Given a Euclidean domain $D$ and $a, b \in D$, return $\gcd(a, b)$. *)

    while $b \neq 0$ do
        $(q, r) \leftarrow$ EuclideanDivision($a, b$)          (* $a = bq + r$ *)
        $a \leftarrow b$
        $b \leftarrow r$
    return $a$
```

Example 1.3.1. Applying the Euclidean algorithm to

$$a = x^4 - 2x^3 - 6x^2 + 12x + 15 \quad \text{and} \quad b = x^3 + x^2 - 4x - 4$$

in $D = \mathbb{Q}[x]$ gives:

a	b	q	r
$x^4 - 2x^3 - 6x^2 + 12x + 15$	$x^3 + x^2 - 4x - 4$	$x - 3$	$x^2 + 4x + 3$
$x^3 + x^2 - 4x - 4$	$x^2 + 4x + 3$	$x - 3$	$5x + 5$
$x^2 + 4x + 3$	$5x + 5$	$\frac{1}{5}x + \frac{3}{5}$	0
$5x + 5$	0		

so $5x + 5$ is a gcd of a and b in $\mathbb{Q}[x]$.

The Euclidean algorithm can be easily extended to return not only a gcd of a and b, but also elements s and t in D such that $sa + tb = \gcd(a, b)$. Such elements always exist since $\gcd(a, b)$ belongs to the ideal generated by a and b by Theorem 1.1.8.

```
**ExtendedEuclidean**$(a, b)$    (* Extended Euclidean algorithm *)

  (* Given a Euclidean domain $D$ and $a, b \in D$, return $s, t, g \in D$ such that
  $g = \gcd(a, b)$ and $sa + tb = g$. *)

  $a_1 \leftarrow 1$, $a_2 \leftarrow 0$, $b_1 \leftarrow 0$, $b_2 \leftarrow 1$
  while $b \neq 0$ do
    $(q, r) \leftarrow$ EuclideanDivision$(a, b)$                    $(* \ a = bq + r \ *)$
    $a \leftarrow b$, $b \leftarrow r$
    $r_1 \leftarrow a_1 - qb_1$, $r_2 \leftarrow a_2 - qb_2$
    $a_1 \leftarrow b_1$, $a_2 \leftarrow b_2$, $b_1 \leftarrow r_1$, $b_2 \leftarrow r_2$
  return$(a_1, a_2, a)$
```

Example 1.3.2. Using the same a and b as in example 1.3.1:

a	b	q	r
$x^4 - 2x^3 - 6x^2 + 12x + 15$	$x^3 + x^2 - 4x - 4$	$x - 3$	$x^2 + 4x + 3$
$x^3 + x^2 - 4x - 4$	$x^2 + 4x + 3$	$x - 3$	$5x + 5$
$x^2 + 4x + 3$	$5x + 5$	$\frac{1}{5}x + \frac{3}{5}$	0
$5x + 5$	0		

a_1	a_2	b_1	b_2
1	0	0	1
0	1	1	$-x + 3$
1	$-x + 3$	$-x + 3$	$x^2 - 6x + 10$
$-x + 3$	$x^2 - 6x + 10$	$\frac{1}{5}x^2 - \frac{4}{5}$	$-\frac{1}{5}x^3 + \frac{3}{5}x^2 + \frac{3}{5}x - 3$

Thus, $5x + 5$ is a gcd of a and b in $\mathbb{Q}[x]$, and

$$(-x + 3)a + (x^2 - 6x + 10)b = 5x + 5\,. \tag{1.4}$$

If only one of the coefficients s or t is needed, a variant of the extended Euclidean algorithm that computes only that coefficient can be used:

```
**HalfExtendedEuclidean**$(a, b)$    (* Half extended Euclidean algorithm *)

  (* Given a Euclidean domain $D$ and $a, b \in D$, return $s, g \in D$ such that
  $g = \gcd(a, b)$ and $sa \equiv g \pmod{b}$. *)

  $a_1 \leftarrow 1$, $b_1 \leftarrow 0$
  while $b \neq 0$ do
    $(q, r) \leftarrow$ EuclideanDivision$(a, b)$                    $(* \ a = bq + r \ *)$
    $a \leftarrow b$, $b \leftarrow r$
    $r_1 \leftarrow a_1 - qb_1$, $a_1 \leftarrow b_1$, $b_1 \leftarrow r_1$
  return$(a_1, a)$
```

This "half" variant of the algorithm is also used as a more efficient alternative to the extended Euclidean algorithm, since the second coefficient can be obtained from the first via

$$t = \frac{g - sa}{b}$$

where the division is always exact.

```
ExtendedEuclidean($a, b$)
(* Extended Euclidean algorithm – "half/full" version *)

    (* Given a Euclidean domain $D$ and $a, b \in D$, return $s, t, g \in D$ such that
    $g = \gcd(a, b)$ and $sa + tb = g$. *)

    $(s, g) \leftarrow$ HalfExtendedEuclidean($a, b$)        (* $sa \equiv g \pmod{b}$ *)
    $(t, r) \leftarrow$ EuclideanDivision($g - sa, b$)         (* $r$ must be 0 *)
    return($s, t, g$)
```

Example 1.3.3. Recomputing the extended gcd of the a and b of example 1.3.1, we get:

1. $(s, g) =$ **HalfExtendedEuclidean**$(a, b) = (-x + 3, 5x + 5)$
2. $g - sa = x^5 - 5x^4 + 30x^2 - 16x$
3. $(t, r) =$ **PolyDivide**$(g - sa, b) = (x^2 - 6x + 10, 0)$

so we recover (1.4).

The extended Euclidean algorithm can also be used to solve the diophantine equation

$$sa + tb = c \tag{1.5}$$

where $a, b, c \in D$ are given and $s, t \in D$ are the unknowns. For (1.5) to have a solution, it is necessary and sufficient that c be in the ideal generated by a and b, *i.e.* that c be a multiple of $\gcd(a, b)$ in D. The extended Euclidean algorithm first solves the equation $sa + tb = \gcd(a, b)$, and there remains only to multiply the solutions by $c/\gcd(a, b)$ to get a solution of (1.5). It should be noted that when c is in the ideal generated by a and b, then (1.5) has as many solutions as the number of elements of D (when a and b are nonzero), since $sa + tb = (s + bd)a + (t - ad)b$ for any $d \in D$. Since there can be no confusion with the previous extended Euclidean algorithm, which has only two parameters, we also call this algorithm the "extended Euclidean algorithm". As before, a half-extended version exists when only one of the coefficients is needed. We remark that the versions of the algorithm that we present here, and use extensively in the sequel, all return a solution s or (s, t) such that either $s = 0$ or $\nu(s) < \nu(b)$. An important consequence of this in polynomial rings (where $\nu(p) = \deg(p)$) is that if $\deg(c) < \deg(a) + \deg(b)$, then we also get either $t = 0$ or $\deg(t) < \deg(a)$. Indeed, if we had $\deg(s) < \deg(b)$ and $\deg(t) \geq \deg(a)$, then we would have $\deg(c) = \deg(sa + tb) = \deg(tb) = \deg(t) + \deg(b) \geq \deg(a) + \deg(b)$.

```
**ExtendedEuclidean**$(a, b, c)$
(* Extended Euclidean algorithm – diophantine version *)

    (* Given a Euclidean domain $D$ and $a, b, c \in D$ with $c \in (a, b)$, return
    $s, t \in D$ such that $sa + tb = c$ and either $s = 0$ or $\nu(s) < \nu(b)$. *)

    $(s, t, g) \leftarrow$ ExtendedEuclidean$(a, b)$            $(* \ g = sa + tb \ *)$
    $(q, r) \leftarrow$ EuclideanDivision$(c, g)$               $(* \ c = gq + r \ *)$
    if $r \neq 0$ then error "$c$ is not in the ideal generated by $a$ and $b$"
    $s \leftarrow qs$, $t \leftarrow qt$
    if $s \neq 0$ and $\nu(s) \geq \nu(b)$ then
        $(q, r) \leftarrow$ EuclideanDivision$(s, b)$           $(* \ s = bq + r \ *)$
        $s \leftarrow r$, $t \leftarrow t + qa$
    return$(s, t)$
```

Example 1.3.4. Suppose that we want to solve $sa + tb = x^2 - 1$ in $\mathbb{Q}[x]$ with the a and b of example 1.3.1. Applying **ExtendedEuclidean** we get:

1. $(s, t, g) =$ **ExtendedEuclidean**$(a, b) = (-x + 3, x^2 - 6x + 10, 5x + 5)$
2. $(q, r) =$ **PolyDivide**$(x^2 - 1, 5x + 5) = ((x - 1)/5, 0)$
3. $s \leftarrow qs = (-x^2 + 4x - 3)/5$
4. $t \leftarrow qt = (x^3 - 7x^2 + 16x - 10)/5$

So we get the following solution:

$$\left(\frac{-x^2 + 4x - 3}{5}\right) a + \left(\frac{x^3 - 7x^2 + 16x - 10}{5}\right) b = x^2 - 1\,. \tag{1.6}$$

```
**HalfExtendedEuclidean**$(a, b, c)$
(* Half extended Euclidean algorithm – diophantine version *)

    (* Given a Euclidean domain $D$ and $a, b, c \in D$ with $c \in (a, b)$, return
    $s \in D$ such that $sa \equiv c \pmod{b}$ and either $s = 0$ or $\nu(s) < \nu(b)$. *)

    $(s, g) \leftarrow$ HalfExtendedEuclidean$(a, b)$           $(* \ sa \equiv g \pmod{b} \ *)$
    $(q, r) \leftarrow$ EuclideanDivision$(c, g)$               $(* \ c = gq + r \ *)$
    if $r \neq 0$ then error "$c$ is not in the ideal generated by $a$ and $b$"
    $s \leftarrow qs$
    if $s \neq 0$ and $\nu(s) \geq \nu(b)$ then
        $(q, r) \leftarrow$ EuclideanDivision$(s, b)$           $(* \ s = bq + r \ *)$
        $s \leftarrow r$
    return $s$
```

As earlier, the "half" variant yields a more efficient alternative to the extended diophantine version, since the second coefficient can be obtained via

$$t = \frac{c - sa}{b}$$

where the division is always exact.

```
ExtendedEuclidean$(a, b, c)$
(* Extended Euclidean algorithm – "half/full"diophantine version *)

    (* Given a Euclidean domain $D$ and $a, b, c \in D$ with $c \in (a, b)$, return
    $s, t \in D$ such that $sa + tb = c$ and either $s = 0$ or $\nu(s) < \nu(b)$. *)

    $s \leftarrow$ HalfExtendedEuclidean$(a, b, c)$              (* $sa \equiv c \pmod{b}$ *)
    $(t, r) \leftarrow$ EuclideanDivision$(c - sa, b)$             (* $r$ must be 0 *)
    return$(s, t)$
```

Example 1.3.5. Solving $sa + tb = x^2 - 1$ in $\mathbb{Q}[x]$ with the a and b of example 1.3.1, we get

1. $s =$ **HalfExtendedEuclidean**$(a, b, x^2 - 1) = (-x^2 + 4x - 3)/5$
2. $c - sa = x^2 - 1 - sa = (x^6 - 6x^5 + 5x^4 + 30x^3 - 46x^2 - 24x + 40)/5$
3. $(t, r) =$ **PolyDivide**$(c - sa, b) = ((x^3 - 7x^2 + 16x - 10)/5, 0)$

so we recover (1.6).

Since the extended Euclidean algorithm can be used to solve diophantine equations, it is also useful for computing partial fraction decompositions. Let $d \in D \setminus \{0\}$ and let $d = d_1 \cdots d_n$ be any factorization of d (not necessarily into irreducibles) where $\gcd(d_i, d_j) = 1$ for $i \neq j$. Then, for any $a \in D \setminus \{0\}$, there are $a_0, a_1, \ldots, a_n$ in D such that either $a_i = 0$ or $\nu(a_i) < \nu(d_i)$ for $i \geq 1$, and

$$\frac{a}{d} = \frac{a}{\prod_{i=1}^n d_i} = a_0 + \sum_{i=1}^n \frac{a_i}{d_i}\,.$$

Such a decomposition is called the *partial fraction decomposition of a/d with respect to the factorization $d = \prod_{i=1}^n d_i$*, and computing it reduces to solving equations of the form (1.5), so to the extended Euclidean algorithm. Indeed, write first $a = da_0 + r$ by the Euclidean division, where either $r = 0$ or $\nu(r) < \nu(d)$. If $n = 1$, then $a/d = a_0 + r/d$ is already in the desired form. Otherwise, since $\gcd(d_i, d_j) = 1$ for $i \neq j$, we have $\gcd(d_1, d_2 \cdots d_n) = 1$, so by the extended Euclidean algorithm, we can find a_1 and b in D such that

$$r = a_1 (d_2 \cdots d_n) + bd_1 \tag{1.7}$$

and either $a_1 = 0$ or $\nu(a_1) < \nu(d_1)$. We can recursively find $b_0, a_2, \dots, a_n \in D$ such that either $a_i = 0$ or $\nu(a_i) < \nu(d_i)$, and

$$\frac{b}{d_2 \cdots d_n} = b_0 + \sum_{i=2}^{n} \frac{a_i}{d_i}\,.$$

Dividing (1.7) by d and adding a_0, we get

$$\frac{a}{d} = a_0 + \frac{r}{d} = a_0 + \frac{a_1}{d_1} + \frac{b}{d_2 \cdots d_n} = (a_0 + b_0) + \sum_{i=1}^{n} \frac{a_i}{d_i}\,.$$

We note that in the case of polynomial rings, since $\deg(r) < \deg(d) = \deg(d_1) + \deg(d_2 \cdots d_n)$ and $\deg(a_1) < \deg(d_1)$ in (1.7), then $\deg(b) < \deg(d_2 \cdots d_n)$, so $b_0 = 0$.

```
PartialFraction($a, d_1, \dots, d_n$)        (* Partial fraction decomposition *)

    (* Given a Euclidean domain $D$, a positive integer $n$ and $a, d_1, \dots, d_n \in$
    $D \setminus \{0\}$ with $\gcd(d_i, d_j) = 1$ for $i \neq j$, return $a_0, a_1, \dots, a_n \in D$ such that

        $\frac{a}{d_1 \cdots d_n} = a_0 + \sum_{i=1}^{n} \frac{a_i}{d_i}$

    and either $a_i = 0$ or $\nu(a_i) < \nu(d_i)$ for $i \geq 1$. *)

    $(a_0, r) \leftarrow$ EuclideanDivision($a, d_1 \cdots d_n$)    (* $a = (d_1 \cdots d_n)a_0 + r$ *)
    if $n = 1$ then return($a_0, r$)
    $(a_1, t) \leftarrow$ ExtendedEuclidean($d_2 \cdots d_n, d_1, r$)      (* $\nu(a_1) < \nu(d_1)$ *)
    $(b_0, a_2, \dots, a_n) \leftarrow$ PartialFraction($t, d_2, \dots, d_n$)
    return($a_0 + b_0, a_1, a_2, \dots, a_n$)
```

Example 1.3.6. We compute the partial fraction decomposition of

$$f = \frac{a}{d} = \frac{x^2 + 3x}{x^3 - x^2 - x + 1} \in \mathbb{Q}(x)$$

with respect to the factorization $d = (x+1)(x^2 - 2x + 1) = d_1 d_2$. Applying **PartialFraction** to a, d_1 and d_2, we get:

1. $(a_0, r) =$ **PolyDivide**$(a, d) = (0, x^2 + 3x)$
2.

$$\begin{aligned}(a_1, t) &= \textbf{ExtendedEuclidean}(x^2 - 2x + 1, x + 1, x^2 + 3x) \\ &= \left(-\frac{1}{2}, \frac{3x+1}{2}\right)\end{aligned}$$

3. $(b_0, a_2) =$ **PartialFraction**$((3x+1)/2, x^2-2x+1) = (0, (3x+1)/2)$

so the partial fraction decomposition of f is

$$\frac{x^2+3x}{x^3-x^2-x+1} = \frac{-1/2}{x+1} + \frac{(3x+1)/2}{x^2-2x+1}\,.$$

We can combine this with the Euclidean division to get a refinement of the partial fraction decomposition: let $m \geq 1$ and $d \in D \setminus \{0\}$. Then, for any $a \in D \setminus \{0\}$, there are $a_0, a_1, \dots, a_m \in D$ such that either $a_j = 0$ or $\nu(a_j) < \nu(d)$ for $j \geq 1$, and

$$\frac{a}{d^m} = a_0 + \sum_{j=1}^{m} \frac{a_j}{d^j}\,.$$

Such a decomposition is called the *d-adic expansion of a/d^m*. Write $a = dq + a_m$ by the Euclidean division, where either $a_m = 0$ or $\nu(a_m) < \nu(d)$. Then,

$$\frac{a}{d^m} = \frac{dq + a_m}{d^m} = \frac{q}{d^{m-1}} + \frac{a_m}{d^m}\,.$$

If $m = 1$, then the above is in the desired form with $a_0 = q$. Otherwise, we recursively find $a_0, a_1, \dots, a_{m-1} \in D$ such that either $a_j = 0$ or $\nu(a_j) < \nu(d)$ for $j \geq 1$, and

$$\frac{q}{d^{m-1}} = a_0 + \sum_{j=1}^{m-1} \frac{a_j}{d^j}\,.$$

Thus,

$$\frac{a}{d^m} = \frac{q}{d^{m-1}} + \frac{a_m}{d^m} = a_0 + \sum_{j=1}^{m} \frac{a_j}{d^j}\,.$$

Let now $d \in D \setminus \{0\}$ and let $d = d_1^{e_1} \cdots d_n^{e_n}$ be any factorization of d, not necessarily into irreducibles, where $\gcd(d_i, d_j) = 1$ for $i \neq j$, and the e_i's are positive integers. Then, for any $a \in D \setminus \{0\}$, we can first compute the partial fraction decomposition of a/d with respect to $d = b_1 \cdots b_n$ where $b_i = d_i^{e_i}$:

$$\frac{a}{d} = a_0 + \sum_{i=1}^{n} \frac{a_i}{b_i} = a_0 + \sum_{i=1}^{n} \frac{a_i}{d_i^{e_i}}$$

and then compute the d_i-adic expansion of each summand to get

$$\frac{a}{d} = \frac{a}{\prod_{i=1}^{n} d_i^{e_i}} = \tilde{a} + \sum_{i=1}^{n} \sum_{j=1}^{e_i} \frac{a_{ij}}{d_i^j}$$

where $\tilde{a} \in D$ and either $a_{ij} = 0$ or $\nu(a_{ij}) < \nu(d_i)$ for each i and j. This decomposition is called the *complete partial fraction decomposition of a/d with respect to the factorization $d = \prod_{i=1}^{n} d_i^{e_i}$*, or simply the complete partial fraction decomposition of a/d when the factorization of d into irreducibles[1] is used.

```
PartialFraction($a, d_1, \dots, d_n, e_1, \dots, e_n$)
(* Complete partial fraction decomposition *)

    (* Given a Euclidean domain $D$, positive integers $n, e_1, \dots, e_n$ and
    $a, d_1, \dots, d_n \in D \setminus \{0\}$ with $\gcd(d_i, d_j) = 1$ for $i \neq j$, return
    $a_0, a_{1,1}, \dots, a_{1,e_1}, \dots, a_{n,1}, \dots, a_{n,e_n} \in D$ such that

        $\frac{a}{d_1^{e_1} \cdots d_n^{e_n}} = a_0 + \sum_{i=1}^{n} \sum_{j=1}^{e_i} \frac{a_{ij}}{d_i^j}$

    and either $a_{ij} = 0$ or $\nu(a_{ij}) < \nu(d_i)$. *)

    $(a_0, a_1, \dots, a_n) \leftarrow$ PartialFraction($a, d_1^{e_1}, \dots, d_n^{e_n}$)
    for $i \leftarrow 1$ to $n$ do
        for $j \leftarrow e_i$ to 1 step $-1$ do
            $(q, a_{ij}) \leftarrow$ EuclideanDivision($a_i, d_i$)        (* $a_i = d_i q + a_{ij}$ *)
            $a_i \leftarrow q$
        $a_0 \leftarrow a_0 + a_i$
    return($a_0, a_{1,1}, \dots, a_{1,e_1}, \dots, a_{n,1}, \dots, a_{n,e_n}$)
```

Example 1.3.7. We compute the complete partial fraction fraction decomposition of

$$f = \frac{a}{d} = \frac{x^2 + 3x}{x^3 - x^2 - x + 1} \in \mathbb{Q}(x)$$

with respect to the factorization $d = (x+1)(x-1)^2 = d_1 d_2^2$. Applying **PartialFraction** to a, d_1, d_2, and the exponents 1 and 2, we get:

$$(a_0, a_1, \dots, a_n) = \textbf{PartialFraction}(x^2+3x, x+1, (x-1)^2) = (0, -\frac{1}{2}, \frac{3x+1}{2})$$

and then:

i	j	a_i	d_i	q	a_{ij}	a_0
1	1	$-1/2$	$x+1$	0	$-1/2$	0
2	2	$(3x+1)/2$	$x-1$	$3/2$	2	0
2	1	$3/2$	$x-1$	0	$3/2$	0

so the complete partial fraction decomposition of f is

$$\frac{x^2 + 3x}{x^3 - x^2 - x + 1} = \frac{-1/2}{x+1} + \frac{2}{(x-1)^2} + \frac{3/2}{x-1}.$$

[1] We show in Sect. 2.7 how to compute that decomposition for linear factors without factoring d.

The algorithm for computing partial fraction decompositions that we presented here dates back to Hermite in the 19^{th} century. There are alternative and faster approaches for rational functions that we do not detail here. See [1, 45] for other approaches and their complexities.

1.4 Resultants and Subresultants

We describe in this section the fundamental properties of the resultant of two polynomials. Although they originate from 19^{th}-century work on solving systems of nonlinear equations, resultants play a crucial role in integration. Throughout this section, let R be a commutative ring and x be an indeterminate over R.

Definition 1.4.1. *Let $A, B \in R[x] \setminus \{0\}$. Write $A = a_n x^n + \cdots + a_1 x + a_0$ and $B = b_m x^m + \cdots + b_1 x + b_0$ where $a_n \neq 0$, $b_m \neq 0$ and at least one of n or m is nonzero. The* Sylvester matrix of A and B *is the $n+m$ by $n+m$ matrix defined by*

$$S(A,B) = \left(\begin{array}{cccccccc} a_n & \cdots & \cdots & \cdots & a_1 & a_0 & & \\ & & & \ddots & & & & \\ & & a_n & \cdots & \cdots & \cdots & a_1 & a_0 \\ b_m & \cdots & b_1 & b_0 & & & & \\ & & \ddots & & & & & \\ & & & \ddots & & & & \\ & & & & \ddots & & & \\ & & & & b_m & \cdots & b_1 & b_0 \end{array}\right) \begin{array}{l} \left.\begin{array}{c} \\ \\ \\ \end{array}\right\} m \text{ rows} \\ \left.\begin{array}{c} \\ \\ \\ \\ \\ \end{array}\right\} n \text{ rows} \end{array}$$

where the A-rows are repeated m times and the B-rows are repeated n times. The resultant of A and B *is the determinant of $S(A,B)$.*

Example 1.4.1. Let $R = \mathbb{Z}[t]$, $A = 3tx^2 - t^3 - 4 \in R[x]$, and $B = x^2 + t^3 x - 9 \in R[x]$. The Sylvester matrix of A and B is

$$S(A,B) = \begin{pmatrix} 3t & 0 & -t^3-4 & 0 \\ 0 & 3t & 0 & -t^3-4 \\ 1 & t^3 & -9 & 0 \\ 0 & 1 & t^3 & -9 \end{pmatrix}$$

and the resultant of A and B is

$$\det(S(A,B)) = -3t^{10} - 12t^7 + t^6 - 54t^4 + 8t^3 + 729t^2 - 216t + 16\,.$$

The first useful property of the resultant of two polynomials is that it can be expressed in terms of their roots.