# POLYNOMIAL DECOMPOSITION

## THEORETICAL RESULTS AND ALGORITHMS

**DIPLOMARBEIT**

ZUR ERLANGUNG DES AKADEMISCHEN GRADES EINES

**DIPLOMINGENIEURS**

IN DER STUDIENRICHTUNG TECHNISCHE MATHEMATIK

EINGEREICHT VON

**FRANZ BINDER**

ANGEFERTIGT AM INSTITUT FÜR MATHEMATIK

DER TECHNISCH-NATURWISSENSCHAFTLICHEN FAKULTÄT

DER JOHANNES KEPLER UNIVERSITÄT LINZ

EINGEREICHT BEI:

A. UNIV.-PROF. DR. GÜNTER F. PILZ

LINZ, JUNI 1995

Reprint with minor corrections, February 1996

# Zusammenfassung

*De todos es conocido que los polinomios constituyen el núcleo del álgebra.*

*Jaime Gutierrez*

*Polynome und Polynomfunktionen erfreuen sich von jeher großer Beliebtheit.*

*Erhard Aichinger*

Die Theorie der Zerlegung von Polynomen bezüglich des Einsetzens fand ihren ersten Höhepunkt durch die Arbeit von J. F. Ritt, welchem es 1922 gelang, durch konsequente Anwendung der damals schon voll entwickelten Theorie der Riemannschen Flächen, alle Zerlegungen von Polynomen mit komplexen Koeffizienten im Prinzip zu beschreiben. Sein Resultat teilt sich in zwei Teile. Der erste besagt, daß alle Primzerlegungen bis auf Bidekompositionen und lineare Transformationen eindeutig sind, der zweite charakterisiert dann alle Bidekompositionen. Ritt's Resultat wurde in mehreren Schritten verallgemeinert.

Motiviert durch die Fortschritte der Computeralgebra entstand in den letzten Jahren eine Neubelebung dieser Thematik. Dabei stand die Entwicklung leistungsfähiger Algorithmen zur Berechnung von Primzerlegungen im Vordergrund. Das überraschendste Ergebnis in diesem Bereich ist wohl, daß das Auffinden von Primzerlegungen, zumindest für Polynome über Körpern der Charakteristik 0, bedeutend schneller ist als die Faktorisierung.

In dieser Arbeit wird konsequent versucht, die beiden oben genannten Richtungen durch eine gemeinsame Behandlung zu vereinen. Dabei stellte sich einerseits heraus, daß die modernen Zerlegungsalgorithmen bereits implizit in den alten Eindeutigkeitsbeweisen enthalten sind, und andererseits, daß aus eben diesen Algorithmen zahlreiche theoretische Ergebnisse einfacher abgeleitet werden können. Daraus ergaben sich zahlreiche neue Aspekte, durchschaubarere Beweise der bekannten Resultate sowie verbesserte Algorithmen.

Das erste Kapitel befaßt sich hauptsächlich mit dem Beweis des Rittschen Eindeutigkeitssatzes und den dabei auf natürliche Weise auftretenden Algorithmen. Dabei wurden die wesentlichsten Schritte und Ideen etwas breiter ausgeführt.

Im ersten Abschnitt wird eine nichtkommutative Teilbarkeitstheorie versucht. Sie ist möglichst allgemein formuliert, die Anwendung auf Polynome bleibt dabei jedoch stets im Hinterkopf. Obwohl dieser Abschnitt eigentlich nur der klaren Begriffsbildung und Festlegung der Notation dient, so enthält er doch die, zumindest in dieser Form, bislang unbekannte Verallgemeinerung des Rittschen Eindeutigkeitssatzes auf Monoide mit Rechtskürzungsregel und semimodularem Teilbarkeitsverband. Das weitere Ziel des ersten Kapitels ist es dann, eben diese Bedingung für Polynome zu zeigen.

Überraschend wirkt hier die Tatsache, daß es nicht gelingt, die Semimodularität direkt zu zeigen, sondern nur über ein wesentlich stärkeres Resultat, nämlich daß der Teilerverband eines (zahmen) Polynoms durch die Gradfunktion in den Teilerverband seines Grades eingebettet ist. Dies ist das eigentliche theoretische Resultat dieses Kapitels.

Im zweiten Abschnitt wird aus Engstöms elementarem Beweis, daß die Teilbarkeitsstruktur eines Polynoms tatsächlich einen Verband bildet, ein allgemeiner Algorithmus zur Berechnung des größten gemeinsamen Rechtsteilers abgleitet. Es wird gezeigt, daß durch geschickte Ausnutzung der vorhandenen Wahlmöglichkeiten ein sehr schneller Algorithmus entsteht.

Stärkere Resultate können nur für den sogenannten *zahmen Fall* bewiesen werden, z. B. wenn die Charakteristik des Grundkörpers 0 ist. Dabei stellt sich der Begriff der $n$-ten (Näherungs)wurzel eines Polynoms als äußerst nützlich heraus. Ihm ist daher der dritte Abschnitt gewidmet. Durch dieses Hilfsmittel können unter anderem einerseits der schnellste bekannte Algorithmus zur Primzerlegung und ein Algorithmus zur Berechnung des kleinsten gemeinsamen Linksvielfachen hergeleitet werden, und andererseits, die Rationalität von Primzerlegungen sowie erste bedeutende Eindeutigkeitsaussagen bewiesen werden.

Um das Resultat über die Einbettung in den Teilerverband des Grades endgültig zu zeigen, ist noch ein Ausflug zu den Zerlegungen rationaler Funktionen notwendig, um etwas elementare Körpertheorie, vor allem den Satz von Lüroth, einsetzen zu können. Dabei erhalten wir das schöne Ergebnis, daß der Teilerverband der Polynome ein konvexer Unterverband des Teilerverbandes der rationalen Funktionen ist.

Das zweite Kapitel beinhaltet einen vereinfachten elementaren Beweis des Rittschen Satzes über die Charakterisiertung von Bidekompositionen. Die Vereifachungen sind vor allem in den beiden Abschnitten über exponentielle und trigonometrische Lösungen beinhaltet. Durch Vergleich vorhandener Beweise und Entwirrung logischer Verflechtungen, aber auch durch konsequennte Ausnutzung der Unzerlegbarkeit konnte die übliche Behandlung der Extrapunke zur Gänze eliminiert werden. Der Abschnitt über Verzweigungen mit seinen ausführlichen Beispielen sollte ebenfalls zur Klarheit beitragen. Außerdem beinhaltet er ein einfaches Verfahren um die Verzweigungsstruktur eines Polynoms zu bestimmen.

Die beschriebenen Algorithmen wurden größtenteils implementiert und teilweise getestet. Entsprechende Programmpakete für *Mathematica* und `Maple` sind in Vorbereitung.

Die Beweise für die theoretischen Resultate scheinen noch weiter verbesserungsfähig zu sei, vor allem die systematische Verwendung des Verzweigungspolynoms zusammen mit dem Resultantenkalkül anstatt des doch eher unduchsichtigen Studiums der Nullstellen ist sehr vielversprechend.

Weitere Verbesserungen sind zu erwarten durch Einbeziehung weiterer verwandter Themen, wie Zerlegung rationaler Funktionen, algebraischer Funktionen und Potenzreihen, sowie die Entwicklung von Zusammenhängen mit der Faktorisierung und der Gruppentheorie. Außerdem wäre es sehr aufschlußreich, mehr über die auftretenden algebraischen Strukturen, des Fastringes $(\Bbbk[x], +, \circ)$, des Kompositionsringes $(\Bbbk[x], +, \cdot, \circ)$ sowie besonders des Fastringes $(\Bbbk(x)_1, \cdot, \circ)$ der rationalen Funktionen mit $f(1) = 1$ zu wissen.

# Abstract

*De todos es conocido que los polinomios*
*constituyen el núcleo del álgebra.*

*Jaime Gutierrez*

*Polynome und Polynomfunktionen erfreuen*
*sich von jeher großer Beliebtheit.*

*Erhard Aichinger*

The theory of polynomial decompositions, with respect to substitution, owes most of its ideas to the work of J. F. Ritt in the years around 1922. Using Riemann surface theory, he could characterize virtually all decompositions of polynomials over the complex number field. His main result on this topic consists of two parts. First, he proved that all prime decompositions are unique up to linear transformations and bidecompositions. Second, he characterized all bidecompositions. Ritt's result was improved in several steps.

Recently, motivated by the rapid development of computational algebra, there was a renaissance of these topics. Now the development of efficient algorithms for the computation of prime decompositions became dominant. In this area it is mostly surprising that, at least in the *tame* case, decomposing is much more efficient than factoring.

This thesis is an approach to combine these two disciplines. It presents many results in a different light, e.g. the modern decomposition algorithms are already contained implicitly in old uniqueness proofs, and conversely, just these algorithms provide an easier derivation of numerous theoretical results.

The first chapter mainly contains a proof of Ritt's uniqueness theorem, together with some algorithms that appear naturally in this context. The most important steps are presented with some digression.

The first section is a first approach to a noncommutative divisibility theory. In spite of its abstract formulation, the application to polynomials always remains in the background. Its main purpose is to fix a consistent set of notations and terminologies, but it also contains a generalization of Ritt's uniqueness theorem to right cancellation monoids with semimodular component lattice. The main goal of the remaining sections of Chapter I then is to establish just this condition for polynomials.

It's somewhat surprising that semimodularity cannot be shown directly. We first have to prove the considerably stronger result that the degree function embeds the component lattice of a polynomial into the divisor lattice of its degree. So the last property should be considered as the main result of the first chapter.

Starting from Engström's direct proof that the component structure of a polynomial is in fact a lattice, the second section contains a general method to compute greatest common right components. Attached with a good heuristics, this provides a very efficient algorithm.

For the *tame* case (e.g., over a field of characteristic 0), a lot of even more interesting results can be proved. In this context, the notion of the $n$-th (approximate) root of a polynomial turns out to be most useful. In the third section, this tool is used to derive the fastest known decomposition algorithm and an algorithm for the computation of least common left multiples, as well as to prove the rationality of prime decompositions and remarkable uniqueness properties.

In order to complete the proof for the embedding into the divisor lattice a discourse to rational function decomposition is necessary. This allows us to use some elementary field theory, particularly Lüroth's theorem is needed. We get the nice result, that the component lattice of polynomials is a convex sublattice of that of rational functions.

The second chapter contains a somewhat simplified proof of Ritt's theorem on the characterization of prime bidecompositions. The improvements are contained mainly in the two sections on exponential and trigonometric solutions, respectively. Comparing previous proofs and doing some logical simplifications, but also by consequent use of the primality of the components of a prime decomposition, the usual treatment of extra points could be completely eliminated. The extensive example in the section on ramification should make clear what is actually going on. Additionally we have obtained an efficient method to compute the ramification structure of any polynomial.

Most of the algorithms discussed in this thesis have been implemented and partially tested. Well designed program packages for both *Mathematica* and `Maple` are being developed.

The proofs of some of the theoretical results seem to be open for further improvements, in particular, a more systematic use of ramification polynomials, together with the resultant calculus, might improve the detailed analysis of zeros.

A comparison with methods used for related topics such as decomposition of rational functions, algebraic functions, or power series, as well as the development of relations to factorization and group theory might be quite enlightening. Additionally, a more detailed knowledge of the appearing algebraic structures like the near-ring $(\Bbbk[x], +, \circ)$, the composition ring $(\Bbbk[x], +, \cdot, \circ)$, and particularly of the near-ring $(\Bbbk(x)_1, \cdot, \circ)$ of rational functions satisfying $f(1) = 1$, is supposed to provide some more insight.

# Dank

An dieser Stelle sei besonders Herrn Prof. Dr. Günter Pilz gedankt, ohne dessen Förderung und persönlichen Einsatz so manches nicht zustandekäme, so auch mein Interesse für Algebra im allgemeinen und diese Arbeit im besonderen.

Weiters gilt mein Dank allen ordentlichen und außerordentlichen Mitgliedern der PKS für die zahlreichen anregenden Diskussionen über Mathematik und PK's.

Nicht zuletzt danke ich meiner Familie und allen meinen Freunden für deren Unterstützung und Vertrauen in allen Lebenslagen.

*Im Gedenken an meinen Vater*

# Contents

CHAPTER I

# Uniqueness Results

## § 1. Divisibility in Noncommutative Monoids

Besides proving Theorem 1.10 in a very general setting, this section contains a first attempt to define a consistent set of notations and terminology for a noncommutative divisibility theory.

Let $M$ be a monoid, written multiplicatively, with neutral element 1. The invertible elements of $M$ are called its *units*.

**1.1. Definition.** Let $f \in M$.

(i) A sequence of elements $f_i \in M$ such that $f = f_n \dots f_1$ is called a *decomposition* of $f$ of *length* $n$ into the *components* $f_i$, and we write $f_i \subseteq f$, for all $i$.

(ii) If $f = rp$ then we call $p$ a *right component* and $r$ a *left component* of $f$. Equivalently, we say that $f$ is a *left multiple* of $p$ and a *right multiple* of $r$. We use the notations $p \mathrel{\subseteq\!\!\!\!\rightarrow} f$ and $f \mathrel{\leftarrow\!\!\!\!\supseteq} p$ for right components and left multiples, respectively, and, symmetrically, $r \mathrel{\subseteq\!\!\!\!\leftarrow} f$ and $f \mathrel{\rightarrow\!\!\!\!\supseteq} r$ for left components and right multiples.

(iii) $p$ and $q$ are called *(left,right) associated* iff they are (right, left) components (or (left, right) multiples) of each other. We use the symbols $\overset{\uparrow}{\cong}$ and $\overset{\downarrow}{\cong}$ to denote left and right association, respectively. Thus $p \overset{\uparrow}{\cong} q$ iff $p \mathrel{\subseteq\!\!\!\!\rightarrow} q$ and $p \mathrel{\leftarrow\!\!\!\!\supseteq} q$, and $p \overset{\downarrow}{\cong} q$ iff $p \mathrel{\subseteq\!\!\!\!\leftarrow} q$ and $p \mathrel{\rightarrow\!\!\!\!\supseteq} q$. Moreover $p \cong q$ iff $p \subseteq q$ and $p \supseteq q$, i.e., iff they are simply associated.

(iv) A (right, left) component of $f$ is a called *proper* iff it is not (left, right) associated to $f$. Equivalently, $f$ then is called a proper (left, right) multiple. We use the symbols $\mathrel{\subset\!\!\!\!\rightarrow}$, $\mathrel{\leftarrow\!\!\!\!\supset}$, $\mathrel{\subset\!\!\!\!\leftarrow}$, $\mathrel{\rightarrow\!\!\!\!\supset}$, $\subset$, $\supset$ with the obvious meanings to denote proper (right, left) components (multiples).

(v) $f$ is called *decomposable* iff there is $p \in M$ such that $1 \mathrel{\subset\!\!\!\!\rightarrow} p \mathrel{\subset\!\!\!\!\rightarrow} f$. Otherwise it is *indecomposable* or *prime*.

(vi) A *prime decomposition* is one that contains only prime components.

**1.2. Remark.** Note that all symbols derived from $\mathrel{\subseteq\!\!\!\!\rightarrow}$ have the arrow running out of the symbol, whereas those derived from $\mathrel{\subseteq\!\!\!\!\leftarrow}$ have it running inside.

The definition for *decomposable* is somewhat unsatisfactory as is not symmetric with respect to *left* and *right*. For the applications used in this thesis, however, this does not matter.

From now on, we mainly consider only the notions derived from $\mathrel{\subseteq\!\!\!\!\rightarrow}$, i.e., right components, left multiples, and left associates. By symmetry, there are always similar definitions and results for left components, right multiples, and right associates.

The theory of this section was developed in order to abstract some notions and properties of the monoid of polynomials with composition, as introduced in the next section, and which is dealt with in almost all parts of this thesis. But it is also used for the monoid of rational functions and might be of general interest.

**1.3. Proposition.**

(i) $\mathrel{\leftarrow\!\!\!\supseteq}$ *and* $\overset{\uparrow}{\triangleq}$ *are right compatible with multiplication, i.e.,*

$$f \mathrel{\leftarrow\!\!\!\supseteq} p \implies fq \mathrel{\leftarrow\!\!\!\supseteq} pq \quad \textit{and} \quad f \overset{\uparrow}{\triangleq} g \implies fp \overset{\uparrow}{\triangleq} gp.$$

(ii) *The left associates of the identity are just its right components, thus the elements that have a left inverse.*

(iii) $\overset{\uparrow}{\triangleq}$ *is an equivalence relation, and* $\mathrel{\leftarrow\!\!\!\supseteq}$ *is a reflexive and transitive relation on $M$, but antisymmetric only if* $\overset{\uparrow}{\triangleq}$ *is equality.*

(iv) $(M/\overset{\uparrow}{\triangleq}, \mathrel{\leftarrow\!\!\!\supseteq})$ *is an ordered set. Its global minimum is the equivalence class of the neutral element* 1. *Its atoms are just the (equivalence classes of the) prime elements.*

*Proof.* If $f = rp$ then $fq = rpq$, hence $pq \mathrel{\subseteq\!\!\!\rightarrow} fq$, showing right compatibility.

Because $1 \mathrel{\subseteq\!\!\!\rightarrow} a$ is always true, $a \overset{\uparrow}{\triangleq} 1$ is the same as $a \mathrel{\subseteq\!\!\!\rightarrow} 1$, which means that there is $b \in M$ such that $ba = 1$, i.e., a left inverse. The rest is trivial. □

As usual in contexts like this, we now stop the pedantic distinction between elements of $M$ and their equivalence classes, whenever no confusion can arise.

**1.4. Definition.**

(i) To every decomposition $f = p_n \dots p_1$ there is a corresponding *right composition series*

$$f = p_n \dots p_1 \mathrel{\leftarrow\!\!\!\supseteq} p_{n-1} \dots p_1 \mathrel{\leftarrow\!\!\!\supseteq} \cdots \mathrel{\leftarrow\!\!\!\supseteq} p_2 p_1 \mathrel{\leftarrow\!\!\!\supseteq} p_1 \mathrel{\leftarrow\!\!\!\supseteq} 1,$$

i.e., a chain in $(M/\overset{\uparrow}{\triangleq}, \mathrel{\leftarrow\!\!\!\supseteq})$. It is *proper* iff all $\mathrel{\leftarrow\!\!\!\supseteq}$'s in this chain are in fact $\mathrel{\leftarrow\!\!\!\supset}$'s.

(ii) Two decompositions are called *right associated* iff they lead to the same right composition series.

(iii) A decomposition is called *right proper* iff its right composition series is proper.

(iv) If elements $p, q \in M/\overset{\uparrow}{\triangleq}$ have a least upper bound $f$, then it is called the *least common left multiple* of $p$ and $q$, and we write $f = p \overset{\uparrow}{\cup} q$. Similarly a greatest common lower bound is called the *greatest common right component*, and is denoted by $p \underset{\downarrow}{\cap} q$.

(v) The set $[1, f] = \{\, p : 1 \mathrel{\subseteq\!\!\!\rightarrow} p \mathrel{\subseteq\!\!\!\rightarrow} f \,\}$ together with $\mathrel{\leftarrow\!\!\!\supseteq}$ is called the *(right) component structure* of $f$. If it is a lattice it is also called the *component lattice* of $f$.

(vi) $(M/\overset{\uparrow}{\triangleq}, \mathrel{\leftarrow\!\!\!\supseteq})$ is called the component structure of the monoid $M$. If is a lattice, it is also called the component lattice of the monoid $M$.

Note that the arrows in the symbols $\underset{\downarrow}{\cap}$ and $\overset{\uparrow}{\cup}$ run outside, as they are derived from $\mathrel{\subseteq\!\!\!\rightarrow}$.

**1.5. Definition.**

(i) An element $p \in M$ is *right cancellable* iff

$$fp = gp \implies f = g,$$

for all $f, g \in M$.

(ii) A *right cancellation monoid* is one in which all elements are right cancellable.

Note that the right cancellable elements always constitute a right cancellation monoid.

**1.6. Proposition.** *Let $M$ be a right cancellation monoid.*

(i) *The left invertible elements of $M$ are just the units.*

(ii) *$p, q \in M$ are left associates iff there is a unit $a \in M$ such that $p = aq$.*

(iii) *The decompositions left associated to $f = p_n \dots p_1$ are exactly those of the form*

$$a_n f = (a_n p_n a_{n-1}^{-1})(a_{n-1} p_{n-1} a_{n-2}^{-1}) \dots (a_2 p_2 a_1^{-1})(a_1 p_1)$$

*for some units $a_i$.*

(iv) *The maximal chains in $(M/\overset{\leftarrow}{=}, \leftarrow\!\!\!\supseteq)$ are in one-to-one correspondence to exactly one class of left associated prime decompositions.*

*Proof.*

(i) Suppose $ba = 1$. Then $aba = a$, and hence $ab = 1$, by right cancellation.

(ii) Suppose $p = aq$ and $q = bp$. Then $p = abp$ and $q = baq$. Now, by right cancellation, this implies $ab = 1$ and $ba = 1$, thus $a$ and $b$ are inverses. The converse is trivial.

(iii) Obviously, such a decomposition is a left associate. For the converse, let $p_n \dots p_1$ and $q_n \dots q_1$ be two left associated decompositions. By definition, $p_1 \overset{\leftarrow}{=} q_1$, thus $q_1 = a_1 p_1$ for some unit $a_1$. By right cancellation, the decompositions

$$q_n \dots q_3 q_2 \quad \text{and} \quad p_n \dots p_3 (p_2 a_1^{-1})$$

are left associated, thus, by induction

$$q_i = a_i p_i a_{i-1}^{-1},$$

for all $1 < i \leq n$.

(iv) We have to show that a decomposition is prime iff the corresponding composition series is maximal. By right cancellation, this means to show that $f$ is prime iff $1 \subset\!\!\!\rightarrow f$ is maximal. But this is true by definition. □

**1.7. Definition.**

(i) A *bidecomposition* $rp = sq$ is a set of two decompositions that are not left associated. *Prime bidecompositions* consist of two *prime* decompositions.

(ii) If $rp = sq$ is a prime bidecomposition, and $f \in M$ has a prime decomposition of the form $\cdots rp \cdots$, then we get another prime decomposition $\cdots sq \cdots$ from replacing $r$ and $p$ by $s$ and $q$, respectively. All (prime) decompositions that can be obtained in a finite number of steps by using bidecompositions this way, are called *related* to the original one.

(iii) Suppose that a lattice $(L, \supseteq)$ contains two incomparable elements $p$ and $q$, such that both $p \cap q \subset p \subset p \cup q$ and $p \cap q \subset q \subset p \cup q$ are maximal chains. Then the four-element sublattice $p \cap q \subset p, q \subset p \cup q$ is called a *unit rhomb* of $L$.

(iv) If $p \cap q \subset p, q \subset p \cup q$ is a unit rhomb of the lattice $(L, \supseteq)$, and a maximal chain contains the maximal subchain $p \cap q \subset p \subset p \cup q$, then we can replace it by $p \cap q \subset q \subset p \cup q$ to get another maximal chain. All (maximal) chains that can be obtained in a finite number of steps by using unit rhombs this way are called *related* to the original one.

**1.8. Proposition.** *Let $M$ be a right cancellation monoid with component lattice $(M/\overset{\uparrow}{\cong}, \overset{\uparrow}{\cup}, \underset{\downarrow}{\cap})$, then two prime decompositions are related iff their right composition series are.*

*Proof.* Immediately from the definitions. □

**1.9. Definition.**

(i) An element $f$ of a lattice is said to *cover* $p$ iff $f \supset p$ and no elements are between $p$ and $f$.

(ii) A lattice is *semimodular* iff whenever both $p$ and $q$ cover $p \cap q$, then $p \cup q$ covers both $p$ and $q$.

(iii) A monoid $M$ is *semimodular* iff $(M/\overset{\uparrow}{\cong}, \overset{\leftarrow}{\supseteq})$ is a semimodular lattice. An element $f \in M$ is *semimodular* iff its component structure, $([1, f], \overset{\leftarrow}{\supseteq})$, is a semimodular lattice.

**1.10. Theorem.**

(i) *If a semimodular lattice contains a finite maximal chain, then all maximal chains are related, in particular, their length is invariant.*

(ii) *If a semimodular element of a right cancellation monoid has at least one prime decomposition, then all its prime decompositions are related, in particular, the number of components is invariant.*

(iii) *All prime decompositions of an element in a semimodular monoid that has no infinite $\overset{\leftarrow}{\supseteq}$-chains are related.*

*Proof.* We need to proof only the first part. Let $\mathcal{A}$ and $\mathcal{B}$ be maximal chains. We may assume that $\mathcal{A}$ is finite. If one of $\mathcal{A}$ and $\mathcal{B}$ has length 0, i.e., contains only one element, then by maximality, the lattice contains only one element, too, so $\mathcal{A} = \mathcal{B}$, and we are through in this case. We proceed by induction on the length of $\mathcal{A}$. Thus assume that

$$\mathcal{A} = (1 \subset p \subset \mathcal{A}') \quad \text{and} \quad \mathcal{B} = (1 \subset q \subset \mathcal{B}').$$

for some (possibly) empty maximal chains $\mathcal{A}'$ and $\mathcal{B}'$.

*Case 1:* If $p = q$, then $p \subset \mathcal{A}'$ and $p \subset \mathcal{B}'$ are maximal chains in the lattice $[p, \infty] := \{\, q : q \supseteq p \,\}$, thus related by induction, so $\mathcal{A}$ is related to $\mathcal{B}$.

*Case 2:* If $p \neq q$, let $\mathcal{C}$ be a maximal chain in the lattice $[p \cup q, \infty]$. By semimodularity both

$$1 \subset p \subset \mathcal{C} \quad \text{and} \quad 1 \subset q \subset \mathcal{C}$$

are maximal chains; they are (directly) related, and each is related to $\mathcal{A}$ or $\mathcal{B}$ respectively, using case 1. □

## § 2. Decompositions of Polynomials

From now on we deal with the monoid $(\Bbbk[x], \circ)$ of polynomials in $x$ over a field of *constants* $\Bbbk$, together with (functional) *composition* defined by

$$(r \circ p)(x) := r(p(x))$$

If in the sequel we just say polynomial, elements of this set are intended.

When discussing algorithms, we generally assume that $\Bbbk$ has computable field operations and decidable equality.

**2.1. Notation.** Note that an expression like $p(x-1)$ is ambiguous because it either means that $(x-1)$ is substituted into $p$, just as in $p(x)$, or that $p$ and $(x-1)$ should be multiplied, as in $(x+1)(x-1)$. Therefore we denote multiplication of polynomials by a dot, e.g., we write $(x+1)\cdot(x-1)$, at least whenever the correct meaning is not obvious. Additionally, the correct meaning of the notation $p^n$ is not clear now. Therefore we reserve it to powers arising from multiplication, whereas $p^{\circ n}$ denotes an $n$-fold composition.

**2.2. Example.** The following examples of trivial formulas should eliminate any doubt about the notation.

$$\begin{aligned}(x+1)\cdot(x-1) &= x^2-1\\ (x+1)\circ(x-1) &= x\end{aligned}$$

$$\begin{aligned}(x+1)^2 &= (x+1)\cdot(x+1) = x^2+2x+1\\ (x+1)^{\circ 2} &= (x+1)\circ(x+1) = x+2\end{aligned}$$

$$\begin{aligned}(x+1)^{-1} &= \frac{1}{x+1}\\ (x+1)^{\circ -1} &= x-1\end{aligned}$$

$$\begin{aligned}2p &= 2\cdot p\\ p(2) &= p\circ 2\\ p(t) &= p\circ t\end{aligned}$$

$$\begin{aligned}xp &= x\cdot p\\ p(x) &= p\circ x = p\end{aligned}$$

We should mention for later reference the trivial

**2.3. Proposition.** $(\Bbbk[x],+,\cdot,\circ)$ *is a composition ring, i.e., we have the right distributive laws*

$$\begin{aligned}(r+s)\circ p &= r\circ p + s\circ p,\\ (r\cdot s)\circ p &= (r\circ p)\cdot(s\circ p).\end{aligned}$$

*for all polynomials* $p$, $r$, $s$. □

Note, however, that the corresponding left distributive laws are not generally satisfied. [Pil83] contains a description of the structure of this composition ring and of the near-ring $(\Bbbk[x],+,\circ)$.

## *Degree of Polynomials*

A very nice property of polynomials is that they have a *degree*, which we denote by square brackets ($[p]$). When dealing with composition, the convention $[0]=0$ is useful. Our first result is trivial, but crucial for all the subsequent theory. Let $(\mathbb{N}_0,\cdot)$ denote the multiplicative monoid of the positive integers including zero. As this monoid is commutative we omit the arrows in symbols like $\subseteq\!\!\!\!\!\rightarrow$, $\cap\!\!\!\!\!\downarrow$.

**2.4. Proposition.** *The degree function maps*

(i) *the monoid* $(\mathbb{k}[x], \circ)$ *homomorphically onto* $(\mathbb{N}_0, \cdot)$*, thus*

$$[r \circ p] = [r]\,[p]$$

(ii) *the ordered set* $(\mathbb{k}[x]/\stackrel{\leftrightarrows}{=}, \overset{\leftarrow}{\supseteq})$ *monotonically onto* $(\mathbb{N}_0, \supseteq)$*, thus*

$$\begin{aligned} p \overset{\rightarrow}{\subseteq} q &\implies [p] \subseteq [q]\,, \\ p \stackrel{\leftrightarrows}{=} q &\implies [p] = [q]\,. \end{aligned}$$

*Proof.* Let $r = b_0x^n + b_1x^{n-1} + \cdots$ and $r = a_0x^m + a_1x^{m-1} + \cdots$ with $a_0 \neq 0$ and $b_0 \neq 0$. Then

$$\begin{aligned} r \circ p &= b_0p^n + b_1p^{n-1} + \cdots \\ &= b_0(a_0x^m + a_1x^{m-1} + \cdots)^n + \cdots \\ &= b_0a_0x^{nm} + \cdots\,, \end{aligned}$$

and $b_0a_0 \neq 0$, thus $[r \circ p] = nm$. Of course $[x] = 1$. The second part is a consequence of the first. As there are polynomials of arbitrary degree, surjectivity is trivial.

The second part is a trivial consequence of the first. □

**2.5. Proposition.** *Consider the monoid* $(\mathbb{k}[x], \circ)$*.*

(i) *The decompositions of a constant are exactly those that contain at least one constant component.*

(ii) *The units are exactly the polynomials of degree* 1*.*

(iii) *Every polynomial of prime degree is prime.*

(iv) *The non-constant polynomials are exactly the right cancellable ones.*

(v) *A (right, left) component of a non-constant polynomial is proper iff it has a smaller degree.*

(vi) *The component structure of any non-constant polynomial contains no infinite* $\overset{\leftarrow}{\supseteq}$*-chains.*

(vii) *Every non-constant polynomial has a prime decomposition.*

*Proof.*

(i) $[f_n]\cdots[f_1] = 0$ iff at least one of the $[f_i] = 0$, i.e., $f_i$ is constant.

(ii) The inverse of $ax + b$ with $a \neq 0$ is given by $\frac{1}{a}x - \frac{b}{a}$. Polynomials of degree $\neq 1$ cannot be invertible because their degree is not.

(iii) If $f = r \circ p$ has prime degree, then $[f] = [r]\,[p]$, thus either $r$ or $p$ must have degree 1.

(iv) Of course, constants are not right cancellable, as different polynomials can have a common zero. For the converse assume that $r \circ p = s \circ p$ for some nonconstant polynomial $p$. Then, by right distributivity,

$$0 = [0] = [r \circ p - s \circ p] = [(r - s) \circ p] = [r - s]\,[p]\,.$$

Because $[p] \neq 0$, $r - s$ is constant. But

$$r - s = (r - s) \circ p = r \circ p - s \circ p = 0,$$

so $r = s$.

(v) Suppose $f = r \circ p$. Then $[f] = [p]$ is equivalent to $[r]$=1, i.e., that $f \stackrel{\leftrightarrows}{=} r$. Similarly for $\overset{\leftarrow}{\subseteq}$ and $\subseteq$.

(vi) If $f_1 \mathrel{\leftarrow\!\!\!\!\supset} f_2 \mathrel{\leftarrow\!\!\!\!\supset} f_3 \mathrel{\leftarrow\!\!\!\!\supset} \cdots$ is a chain, then $[f_1] \supset [f_2] \supset [f_3] \supset \cdots$ is a chain of positive integers. But no positive integer has an infinite number of divisors. So $[f_1] \neq 0$ implies that the chain cannot be infinite.

(vii) By the previous part, together with Proposition 1.6. □

Thus the discussion about existence of prime decompositions has been finished. (If, however, $\Bbbk$ is not a field, or at least a unique factorization domain, prime decompositions need not exist and the question becomes more interesting.) A more difficult problem is to develop algorithms for computing prime decompositions and to find interesting uniqueness properties. It turns out that very similar methods solve these two problems, so we treat them at once.

### *Polynomial Decomposition Algorithms*

Let us ask whether our existence proof contains any method to find a prime decomposition of a polynomial $f$. In fact it does, though not explicitly. The multiplicativity of the degree function shows that there is only a finite number of possible degrees for the components, one for each divisor of $[f]$. So by an approach with indetermined coefficients, we can test for nontrivial decompositions.

**2.6. Algorithm.** *The following method determines whether a given polynomial $f$ has a proper decomposition over some algebraic extension field of $\Bbbk$ and computes it in the affirmative case.*

```
For each non-trivial divisor $n$ of $[f]$
repeat take $p = \sum_{i=0}^{n} a_i x^i$, $r = \sum_{j=0}^{[f]/n} b_j x^j$
          with indetermined coefficients $a_i$ and $b_j$;
          Compute $r \circ p$ and compare its coefficients
                  to the corresponding ones of $f$;
          Test whether the resulting system of algebraic equations
                              has a solution for the $a_i$ and $b_i$;
If one of the systems has a solution,
   then $r \circ p$ with this solution is a decomposition,
   else $f$ is indecomposable.
```

**2.7. Remark.** Note that any system of algebraic equations can be solved, e.g., by computing the Gröbner basis (cf. e.g. [BL82]). The algorithm presented there either determines that no solution exists or transforms it into a triangular system, i.e., an equation for the first variable, one for the second, but using the first, one for the third, using the first two, and so on. So it is easy to find out, whether there are solutions in $\Bbbk$, and which field extension are necessary to obtain all solution. There is also an easy criterion to detect whether the system has a finite or an infinite number of solutions.

This way, the polynomial decomposition problem is solved, in principle. But, except for polynomials of very small degree, the system that must be solved has too many variables occurring with too high degrees for being tractable. Hence this method has not been studied in detail. On the other hand, there are much more equations than variables, and the Gröbner bases computation has a lot of choice, that can make it fast in particular situations. It is an open problem, whether one can do so for the polynomial decomposition problem.

[Zip91] contains another general algorithm, which has a polynomial computation time. But it uses polynomial factorization in two variables over an algebraic extension field, therefore is mainly of theoretical interest, as even exponential-time algorithms are usually faster in practice.

In the next section an algorithm that is very fast for the special but very important *tame* case will be developed.

But first we discuss some more properties and algorithms valid in the general case.

### *Taylor expansion*

**2.8. Proposition.** *Let $f, p$ be polynomials, $p$ non-constant. Then there are unique polynomials $r_i$ with $[r_i] < [p]$ such that*

$$f = \sum_i r_i \cdot p^i.$$

*Proof.* If $[f] < [p]$, the statement is clear, using $r_0 = f$. We do induction on the degree of $f$. If $[f] \geq [p]$, take any nonconstant left multiple $\tilde{p} = u \circ p$ of $p$ such that $[\tilde{p}] \leq [f]$, e.g., $p$ itself, or some of its powers $p^i$, but not too big. We use Euclidean division to get unique polynomials $q$ and $r$ such that

$$f = q \cdot \tilde{p} + r.$$

By induction, $q = \sum q_i \cdot p^i$, $r = \sum r_i \cdot p^i$, and by definition, $\tilde{p} = \sum a_i p^i$, for some unique polynomials $q_i, r_i$ of degrees $< [p]$ and constants $a_i$. Thus

$$f = \sum q_i \cdot p^i \cdot \sum a_i p^i + \sum r_i \cdot p^i,$$

which, after expansion, has the requested form. □

Of course, the step involving $\tilde{p}$ is unnecessary in order to prove the proposition. But it provides us with a more general construction, which we are going to use to obtain a more efficient algorithm.

Note that for $p = x - a$, $a \in \Bbbk$, the proposition just says that $f$ has a (finite) Taylor expansion around the point $a$. Therefore we define

**2.9. Definition.** The unique representation of Proposition 2.8 is called the *Taylor expansion* of $f$ around $p$ with *coefficients* $r_i$.

**2.10. Example.** Let us compute the Taylor expansion of

$$f = x^{12} + 12x^{11} + 66x^{10} + 223x^9 + 522x^8 + 900x^7 + 1179x^6 + 1188x^5 + 918x^4 + 533x^3 + 222x^2 + 60x$$

around $p = x^2 + 2x$. First we use the choice $\tilde{p} = p$. Dividing $f$ by $p$ we get the remainder $r_0 = 4x$, whose degree is $< 2$, and the quotient

$$q_1 = x^{10} + 10x^9 + 46x^8 + 131x^7 + 260x^6 + 380x^5 + 419x^4 + 350x^3 + 218x^2 + 97x + 28.$$

Continuing with this quotient as in the proof, we get the sequence

$$\begin{aligned}
r_1 &= 13x + 28\\
q_2 &= x^8 + 8x^7 + 30x^6 + 71x^5 + 118x^4 + 144x^3 + 131x^2 + 88x + 42\\
r_2 &= 18x + 42\\
q_3 &= x^6 + 6x^5 + 18x^4 + 35x^3 + 48x^2 + 48x + 35\\
r_3 &= 12x + 35\\
q_4 &= x^4 + 4x^3 + 10x^2 + 15x + 18\\
r_4 &= 3x + 18\\
q_5 &= x^2 + 2x + 6\\
r_5 &= 6\\
q_6 &= 1\\
r_6 &= 1.
\end{aligned}$$

Thus the Taylor expansion is

$$f = p^6 + 6p^5 + (3x + 18) \cdot p^4 + (12x + 35) \cdot p^3 + (18x + 42) \cdot p^2 + (13x + 28) \cdot p + 4x.$$

Note that, by the choice $\tilde{p} = p$, the remainder always had degree $< [p]$; therefore we just needed to continue with the quotient. This means that the problem of (Taylor) expanding $f$ is reduced to expanding a polynomial of degree $[f] - [p]$. Though this is quite practical for polynomials of low degree, for high degree polynomials a more balanced version seems to be better. If $[\tilde{p}] = \frac{[f]+1}{2}$, then the problem of (Taylor) expanding one $n$-th degree polynomials is reduced to expanding two polynomials of degree $\frac{[f]-1}{2}$. This approach is explained in the following algorithm.

**2.11. Algorithm.** *Given polynomials $f$ and non-constant $p$, then the following method computes the Taylor expansion of $f$ around $p$:*

```
if $[f] < [p]$ then return $f$
else set $i =$ power of 2 closest to $\frac{[f]+1}{2[p]}$;
    thus $[p^i] \approx \frac{[f]+1}{2}$
    use Euclidean division of $f$ by $p^i$;
    set $q$ to the quotient, $r$ to the remainder;
    return (expansion of $q$) $\cdot p^i +$ expansion of $r$.
```

**2.12. Remark.** Let $M(n)$ be the number of field operations necessary for multiplying two polynomials of degree $n$.

Suppose $[f] = n$. In every step, $i \leq n$. Thus computing all necessary powers $p^i$ by successively computing squares, takes at most $O(M(n) \cdot \log n)$ field operations. In the first step, the Euclidean division uses at most $O(M(n))$ field operations, i.e., has a bound $c \cdot M(n)$, for some constant $c$. Next, two problems of size $\approx \frac{n}{2}$ must be solved. So we have to do two Euclidean divisions, but of polynomials with degree bound $\frac{n}{2}$, so it takes at most $2c \cdot M(\frac{n}{2})$ field operations. Similarly, in the third step, we have the bound $4c \cdot M(\frac{n}{4})$, in general $2^i c \cdot M(\frac{n}{2^i})$. As $2^i \leq n$ the total cost of the algorithm is bounded by

$$O(M(n) \cdot \log n) + O(\sum_{i<\log n} 2^i c \cdot M(\frac{n}{2^i})) \leq O(M(n) \log n) + O(c \cdot \sum_{i<\log n} M(2^i \frac{n}{2^i}))$$

$$= O(M(n) \log n) + O(M(n) \log n).$$

So the algorithm uses at most $O(M(n) \log n)$ field operations.

**2.13. Example.** Let us illustrate the algorithm with the computation of the Taylor expansion of $f$ as in the last example

$$f = x^{12}+12x^{11}+66x^{10}+223x^9+522x^8+900x^7+1179x^6+1188x^5+918x^4+533x^3+222x^2+60x$$

but around $p = x^3 + 3x^2 + 3x$. According to the algorithm we must choose $i$ near $\frac{12+1}{2 \cdot 3} = \frac{13}{6}$, thus $i = 2$ and

$$\tilde{p} = p^2 = x^6 + 6x^5 + 15x^4 + 18x^3 + 9x^2.$$

We divide $f$ by $\tilde{p}$ to obtain

$$r_1 = 20x^3 + 60x^2 + 60x$$

as the remainder and

$$q_1 = x^6 + 6x^5 + 15x^4 + 25x^3 + 30x^2 + 21x + 18$$

as the quotient. Now we continue recursively, computing the Taylor expansion of both $r_1$ and $q_1$. As $[r_1] = 3 = [p]$, the choice $i = 1$ is the only possible. We get

$$r_2 = 0$$
$$q_2 = 20.$$

Thus $r_1 = 20p$. $[q_1] = 6$, hence $i = 1$ is appropriate, i.e., we divide $q_1$ by $p$, with the result

$$r_3 = 18$$
$$q_3 = x^3 + 3x^2 + 3x + 7 = p + 7.$$

So $q_1 = (p + 7) \cdot p + 18$, and we get

$$f = q_1 \cdot p^2 + r_1 = (p^2 + 7p + 18) \cdot p^2 + 20p = p^4 + 7p^3 + 18p^2 + 20p.$$

In this example, all Taylor coefficients of $f$ around $p$ happen to be constant. Of course, this is quite an incidence, and means that $p \mathrel{\underline{\subset\!\!\!\rightarrow}} f$, in fact

$$f = (x^4 + 7x^3 + 18x^2 + 20x) \circ p.$$

**2.14. Definition.** Let $p \mathrel{\underline{\subset\!\!\!\rightarrow}} f$. The unique polynomial $r$ such that $f = r \circ p$ is denoted by $f \div p$, and we call this operation *Taylor division.*

We have called this *Taylor* division, because it is a special case of Taylor expansion.

That $f \div p$ really is uniquely determined follows directly from the right cancellation law. This does not, however, equip us directly with a method to compute this operation. Additionally there could be the possibility that there exists $r$ such that $f = r \circ p$ only if it is allowed to have coefficients in some extension field of $\Bbbk$. But we can use Taylor expansion.

**2.15. Proposition.** *Let $f, p \in \Bbbk[x]$, $p$ non-constant.*

(i) *$p \mathrel{\underline{\subset\!\!\!\rightarrow}} f$ iff all coefficients of the Taylor expansion of $f$ around $p$ are constant.*

(ii) *The relation $p \mathrel{\underline{\subset\!\!\!\rightarrow}} f$ is independent of the ground field.*

(iii) *If $p \mathrel{\underline{\subset\!\!\!\!\rightarrow}} f$, then the coefficients of $f \div p$ are rational functions of those of $p$ and $f$.*

*Proof.* The first part is immediate from the definition. But the algorithm for the computation of the Taylor expansion uses only rational operations involving the coefficients of $f$ and $p$. Thus the remaining parts are also obvious. □

Remember that $[p] \subseteq [f]$ is a necessary condition for $p \mathrel{\underline{\subset\!\!\!\!\rightarrow}} f$.

**2.16. Algorithm.** *Given polynomials $f$ and non-constant $p$, then we can decide whether $p \mathrel{\underline{\subset\!\!\!\!\rightarrow}} f$ and compute $f \div p$ in the affirmative case just by computing the Taylor expansion as in algorithm 2.11, and aborting as soon as it computes a polynomial whose degree is not a multiple of $[p]$, because in this case $p$ cannot be a (right) component of $f$.*

Note that, if the algorithm is not aborted, then all Taylor coefficients are constant.

### *Component Lattice*

The following lemma is very surprising and seems to be rather unknown, though it appears implicitly in [Eng41].

**2.17. Lemma.** *Let $f$ and $q$ be polynomials, and*

$$f = q \cdot p + r; \quad [r] < [p].$$

*Then a polynomial $t$ is a common right component of $f$ and $p$ iff it is one of $p$, $q$ and $r$.*

*Proof.* The if-part is trivial. Therefore assume that $t$ is a common component of $f$ and $p$, thus there exist polynomials $\hat{f}$ and $\hat{p}$ such that

$$f = \hat{f}(t), \; p = \hat{p}(t).$$

Then, by Euclidean division, there exist polynomials $\hat{q}$, $\hat{r}$ such that

$$\hat{f} = \hat{q} \cdot \hat{p} + \hat{r}; \quad [\hat{r}] < [\hat{p}],$$

and by substituting $t$ into this equation

$$\hat{f}(t) = \hat{q}(t) \cdot \hat{p}(t) + \hat{r}(t); \quad [\hat{r}(t)] < [\hat{p}(t)],$$

i.e.,

$$f = \hat{q}(t) \cdot p + \hat{r}(t); \quad [\hat{r}(t)] < [p].$$

As the quotient and remainder are uniquely determined, it follows that $q = \hat{q}(t)$, $r = \hat{r}(t)$, thus $t$ is a component of both $q$ and $r$. □

**2.18. Proposition.** *Any finite set $F$ of polynomials has a greatest common right component.*

*Proof.* Because every polynomial is a right component of any constant, the constants can be removed from $F$ without changing the result. If $F = \varnothing$, then any constant is a greatest common right component. If $F = \{p\}$, then $p$ is the result. So assume that $F$ has at least two elements but no constants. If the greatest common divisor of the degrees of all polynomials in $F$ is 1, then $x$ is the greatest common right component, because its degree must divide 1. Otherwise choose polynomials

$f \in F$ and $p$ in $\Bbbk[F \setminus \{f\}]$ with $[f] \geq [p] > 0$. Thus $p$ can be any element of $F$ different from $f$, but can as well be formed by adding and multiplying any such elements. Using Euclidean division, we get a quotient $q$ and remainder $r$. Then, by the lemma, the sets $F$ and $(F \setminus \{f\}) \cup \{q, r\}$ have the same right components. Let $\tilde{F}$ be the later set with constants omitted. Note that $[q] + [r] \leq [f] - 1$, so the sum of the degrees in $\tilde{F}$ is smaller than that in $F$. We proceed inductively. Because the constants are omitted, we eventually must get a singleton set. Its element then has the same right components as the original set, i.e., it is the greatest common right component. □

Now we can state our lemma in a simpler form.

**2.19. Proposition.** *Let $f$ and $q$ be polynomials, and*

$$f = q \cdot p + r; \quad [r] < [p] .$$

*Then*

$$f \mathbin{\underset{\downarrow}{\cap}} p = p \mathbin{\underset{\downarrow}{\cap}} q \mathbin{\underset{\downarrow}{\cap}} r.$$

The proof of the Proposition 2.18 contains a new algorithm for the computation of $\underset{\downarrow}{\cap}$, which is both simpler and more general than previous ones. It works like a kind of *compositional Euclidean algorithm.*

**2.20. Algorithm.** *The following program computes the greatest common right component of a finite set $F$ of polynomials:*

```
Remove all constants from the set $F$;
**if** $F = \varnothing$ **then return** 0;
**while** $F$ contains at least two elements
        and the gcd of their degrees is $> 1$
**repeat** choose polynomials $f \in F$ and $p \in \Bbbk[F \setminus \{f\}]$ with $[f] \geq [p] > 0$;
         use Euclidean division of $f$ by $p$, giving $q$ and $r$;
         remove $f$ from $F$;
         add $q$ and $r$ instead, if nonconstant;
**return** the single element of $F$
         resp. $x$, if we terminated because the gcd was $= 1$.
```

**2.21. Example.** Let us compute the greatest common right component of

$f = x^{12}+12x^{11}+66x^{10}+223x^9+522x^8+900x^7+1179x^6+1188x^5+918x^4+533x^3+222x^2+60x$

and

$$\begin{aligned} g = x^{27} &+ 27x^{26} + 351x^{25} + 2924x^{24} + 17526x^{23} + 80454x^{22} + 293985x^{21} + 877383x^{20} \\ &+ 2177361x^{19} + 4550901x^{18} + 8084232x^{17} + 12282381x^{16} + 16023713x^{15} + 17986719x^{14} \\ &+ 17374647x^{13} + 14417930x^{12} + 10238064x^{11} + 6178974x^{10} + 3134779x^9 + 1313667x^8 \\ &+ 442593x^7 + 115013x^6 + 21450x^5 + 2487x^4 + 307x^3 + 183x^2 + 3x. \end{aligned}$$

At the first step we divide $g$ by $f$ and get the quotient

$$\begin{aligned} q_1 = x^{15} &+ 15x^{14} + 105x^{13} + 451x^{12} + 1317x^{11} + 2739x^{10} + 4133x^9 \\ &+ 4527x^8 + 3555x^7 + 1970x^6 + 777x^5 + 228x^4 + 37x^3 - 6x^2 + 3x \end{aligned}$$

and the remainder

$$r_1 = x^3 + 3x^2 + 3x$$

Now $F = \{f, q_1, r_1\}$, with degrees $\{12, 15, 3\}$. In the next step we have a lot of choice. Let us get rid of the biggest polynomial, $q_1$, which has degree 15. If we divide it by $f$, we get a quotient of degree 3 and a remainder of degree $< 12$. If we divide by $r_1$, then the quotient has degree 12 and the remainder $< 3$. If, however, we divide by

$$r_1^3 = x^9 + 9x^8 + 36x^7 + 81x^6 + 108x^5 + 81x^4 + 27x^3,$$

then the quotient will have degree 6 and the remainder $< 9$, which situation is more balanced. In fact,

$$\begin{aligned} q_2 &= x^6 + 6x^5 + 15x^4 + 19x^3 + 12x^2 + 3x + 2, \\ r_2 &= -x^6 - 6x^5 - 15x^4 - 17x^3 - 6x^2 + 3x. \end{aligned}$$

We observe that these two polynomials are left associated ($r_2 = (2-x) \circ q_2$). Because we are computing in $(\mathbb{k}[x]/\overset{\leftarrow}{\triangleq}, \mathrel{\overleftarrow{\supseteq}})$, i.e., modulo $\overset{\leftarrow}{\triangleq}$, we can omit one of them. Thus $F = \{f, q_2, r_1\}$, with degrees $\{12, 6, 3\}$. The next choice is rather straightforward: we get rid of $f$ using division by $q_2$, thus

$$\begin{aligned} q_3 &= x^6 + 6x^5 + 15x^4 + 24x^3 + 27x^2 + 18x + 10, \\ r_3 &= -2x^3 - 6x^2 - 6x - 20. \end{aligned}$$

Both polynomials happen to be left associated to some already in $F$, so $f$ can be eliminated from $F$ without compensation. $F = \{q_2, r_1\}$ now, and in the last step we have to divide these two elements, with the result

$$\begin{aligned} q_4 &= x^3 + 3x^2 + 3x + 1, \\ r_4 &= 2. \end{aligned}$$

Again, $q_4$ is left associated to $r_1$, and $r_4$ can be omitted as it is constant. So, finally, $F$ is a singleton and its element $r_1$ (or any of its left associates) is the greatest common right component.

**2.22. Remark.** There is a lot of arbitrariness in this algorithm, involved by the word *choose*, which can affect the efficiency of the algorithm. In the example we have used the strategy to replace the polynomial of highest degree by two ones that have about the same degree. Choosing the second polynomial $p$ of the algorithm appropriately in $\mathbb{k}[F \setminus \{f\}]$, not just in $F$, it can always be accomplished that $[p] \approx \frac{[f]}{2}$. Suppose that the biggest polynomial in $F$ has degree $n$. If all polynomials have degree $\approx n$, then we get rather small polynomials. Thus polynomials of any order of magnitude between the smallest and the biggest can be computed with $O(n \log n)$ field operations. Dividing the $n$-th degree polynomial by one of degree $\approx \frac{n}{2}$ replaces it by two ones of degree at most $\approx \frac{n}{2}$. Thus, with this strategy, we get the same complexity bound as for our algorithm for Taylor expansion, which was $O(n \log n)$. In fact, these two algorithms are not very different and e.g. both can be used to decide whether $p \mathrel{\overrightarrow{\subseteq}} f$.

**2.23. Theorem.** *$(\mathbb{k}[x]/\overset{\leftarrow}{\triangleq}, \mathrel{\overleftarrow{\supseteq}})$ is a bounded lattice with minimum $x$ and maximum $0$.*

*Proof.* We have already shown that any two elements have an infimum. 0 is always a common left multiple, thus a least common left multiple must exist, because there are no infinite $\mathrel{\overleftarrow{\supseteq}}$-chains. □

**2.24. Remark.** The degree function maps the lattice $(\Bbbk[x]/\triangleq, \mathbin{\cup\!\!\!\!\uparrow}, \mathbin{\cap\!\!\!\!\downarrow})$ monotonically onto the lattice $(\mathbb{N}, \cup, \cap)$. But, in general, this is not a lattice homomorphism. We just get, as an immediate consequence of monotonicity (Proposition 2.4), the considerably weaker facts

$$[p \mathbin{\cap\!\!\!\!\downarrow} q] \subseteq [p] \cap [q] , \quad [p \mathbin{\cup\!\!\!\!\uparrow} q] \supseteq [p] \cup [q] .$$

There is, however, a very important *local* replacement, stated in theorem 5.4. We will be concerned with its proof in the next sections.

**2.25. Example.** Here is an easy counterexample. Obviously

$$x^2 \mathbin{\cap\!\!\!\!\downarrow} (x^2 + x) = x.$$

Later(Proposition 5.2) we will show that this implies that

$$x^2 \mathbin{\cup\!\!\!\!\uparrow} (x^2 + x) = 0,$$

at least if $\operatorname{char} \Bbbk = 0$.

Whereas we have got a general and very efficient method for the computation of the $\mathbin{\cap\!\!\!\!\downarrow}$-operation, no general method for computing the $\mathbin{\cup\!\!\!\!\uparrow}$-operation is known. The reason is that the existence proof for least common left multiples was not constructive. But we can test whether there is a common left multiple of a specified degree, because this leads to a system of linear equations. ([Alo94])

**2.26. Algorithm.** *We can test whether two polynomials $p$, $q$ have a common left multiple of degree $n$, and compute it in the affirmative case.*

```
if $n \not\supseteq [p] \cup [q]$ then return no common multiple;
try polynomials $r, s$ of degrees $\frac{n}{[p]}$ and $\frac{n}{[q]}$, respectively,
                    with indetermined coefficients;
Find a solution satisfying $r \circ p = s \circ q$
    this is a linear system of $[p]\,[q]$ equations with $[p] + [q]$ variables;
if it has a solution
  then return $r \circ p = s \circ q$ (for that solution)
  else return no common multiple.
```

**2.27. Algorithm.** *The following semialgorithm computes $p \mathbin{\cup\!\!\!\!\uparrow} q$, if it is not a constant, and never stops otherwise.*

```
$n := [p] \cup [q]$;
for $k \in \mathbb{N}$
repeat if there is a common left multiple $f$ of degree $kn$
         then return $f$
         else continue.
```

Thus the general case is not very satisfactory. Because we cannot wait until the end of time to see that the algorithm did not stop, we hope to obtain a bound for $[p \mathbin{\cup\!\!\!\!\uparrow} q]$. The general case is unsolved, but for $\operatorname{char} \Bbbk = 0$, there is a very satisfactory answer, given in the next section. But in this case we can use the characterization of bidecompositions in chapter II to obtain an even more efficient algorithm.

### *Normed Polynomials*

The component lattice contains *equivalence classes* of polynomials. This is sometimes inconvenient, in particular, if we try to compose these equivalence classes, because $\triangleq$ is not a congruence with respect to composition. This prohibits having both lattice operations and composition in *one* structure. But we can choose a good system of canonical representatives to achieve this.

**2.28. Definition.** Let $p = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0$ be a polynomial of degree $n$.

(i) $p$ is called *zerosymmetric* iff $a_0 = 0$.
(ii) $p$ is called *monic* iff $a_n = 1$.
(iii) $p$ is called *normed* iff it is monic and zerosymmetric.
(iv) a *decomposition* is *normed* iff all its components are normed.

**2.29. Proposition.**

(i) *The (normed, zerosymmetric, monic) polynomials form a submonoid of the monoid* $(\mathbb{k}[x], \circ)$. *They are also closed under multiplication.*
(ii) *Each non-constant polynomial* $q$ *has exactly one decomposition*

$$q = u \circ p,$$

*such that* $u$ *is linear and* $p$ *normed.*
(iii) *Each polynomial has exactly one normed left associate.*
(iv) *Each decomposition of a normed non-constant polynomial is left associated to exactly one normed one.*

*Proof.*

(i) Trivial.
(ii) Let $q = d_n x^n + \cdots + d_0$, $u = ax + b$, and $p = 1x^n + c_{n-1}x^{n-1} + \cdots + 0$. We compare the coefficients of $q$ and $u \circ p$; thus obtain

$$\begin{aligned} d_n &= a \\ d_i &= ac_i, \quad \text{for } 0 < i < n \\ d_0 &= b, \end{aligned}$$

which, if $q$ is given, has a unique solution for $a$, $b$, and all $c_i$.
(iii) For non-constant polynomials this is clear by the previous part. Note that 0 is the only normed constant, and for each constant $c$ we have $c = (x + c) \circ 0$.
(iv) Let $f = q_n \circ \cdots \circ q_2 \circ q_1$ be a decomposition. Each left associate of this decomposition looks like

$$f = (u_n \circ q_n \circ u_{n-1}^{\circ -1}) \circ \cdots \circ (u_2 \circ q_2 \circ u_1^{\circ -1}) \circ (u_1 \circ q_1).$$

We have to choose the $u_i$ appropriately to make all components normed. For the rightmost component, part (ii) shows that there is exactly one choice. But then $u_2$ must make $q_2 \circ u_1^{\circ -1}$ normed, and by the same argument, we get exactly one solution. This way we continue until $u_n$ is determined. □

## § 3. Roots of Tame Polynomials

**3.1. Definition.** Let $f$ be a normed polynomial of degree $nm$. A normed polynomial $p$ such that

$$[f - p^n] \le nm - m$$

is called an $n$-th (approximate) *root* of $f$. We will use the notation $p = \sqrt[n]{f}$, if it exists uniquely.

**3.2. Proposition.** *Let* $r$ *and* $p$ *be nonconstant normed polynomials of degrees* $n$ *and* $m$. *Then* $p$ *is an* $n$*−th root of* $r \circ p$.

*Proof.* Let $r = x^n + b \cdot x^{n-1} + \cdots$. Then

$$r \circ p = p^n + b \cdot p^{n-1} + \cdots .$$

But $\left[b \cdot p^{n-1}\right] = (n-1)m = nm - m$. All further terms are even smaller. □

**3.3. Example.** If $\Bbbk$ has characteristic 2, then the polynomial $x^4 + x^3$ has no $2^{\text{nd}}$ root, because $(x^2 + ax)^2 = x^4 + a^2x^2$, for each $a \in \Bbbk$, which has no term for $x^3$. On the other hand, each polynomial of this form is a 2nd root of $x^4$. We want to avoid such *wild* behavior.

**3.4. Definition.** An integer is called *tame* (relatively to $\Bbbk$) iff it is has an inverse in $\Bbbk$, i.e., if it is not a multiple of the characteristic. A polynomial is *tame* iff its degree is tame.

**3.5. Remark.** If $\operatorname{char} \Bbbk = 0$, a polynomial is tame iff it is non-constant.

**3.6. Lemma.** *Suppose that the nonconstant normed polynomials $f$ and $p$ with $[f] = nm$, $[p] = m$ and tame $n$ satisfy*

$$[f - p^n] \le nm - k$$

*for some $1 \le k < m$. Then*

(i) *for each polynomial $q$ of degree at most $m - k$, we again have*

$$[f - (p+q)^n] \le nm - k.$$

(ii) *with $\tilde{k} := \min(2k, m)$, there is exactly one zerosymmetric polynomial $q$ of degree at most $\tilde{k} - k$ such that*

$$\left[f - (p + q \cdot x^{m-\tilde{k}})^n\right] \le nm - \tilde{k}.$$

*Proof.*

(i) $[f - (p+q)^n] = \left[f - p^n - np^{n-1} \cdot q - \cdots\right] \le nm-k$, because $\left[np^{n-i} \cdot q^i\right] \le (n-i)m + i(m-k) = nm - ik \le nm - k$ for $i \ge 1$.

(ii) The condition on $q$ is

$$\left[f - (p + q \cdot x^{m-\tilde{k}})^n\right] = \left[f - p^n - np^{n-1} \cdot q \cdot x^{m-\tilde{k}} - \cdots\right] \le nm - \tilde{k}.$$

The omitted terms have degrees $\le (n-i)m + i(\tilde{k} - k) + i(m - \tilde{k}) = nm - ik \le nm - \tilde{k}$ for $i \ge 2$. By the assumption,

$$\left[f - p^n - \hat{f} \cdot x^{nm-\tilde{k}}\right] \le nm - \tilde{k},$$

for (exactly) one zerosymmetric polynomial $\hat{f}$ of degree at most $\tilde{k} - k$. Thus the condition turns into

$$\left[\hat{f} \cdot x^{nm-\tilde{k}} - np^{n-1} \cdot x^{n-\tilde{k}} \cdot q\right] \le nm - \tilde{k}.$$

Because $\left[p^{n-1} \cdot x^{m-\tilde{k}}\right] = nm - \tilde{k}$, and $n$ is tame, we see, after dividing by $x$, that $\frac{q}{x}$ is the unique Euclidean quotient of $\hat{f} \cdot x^{nm-\tilde{k}-1}$ by $np^{n-1}$. □

**3.7. Proposition.** *Let $f$ be a normed polynomial with $[f] = nm$. If $n$ is tame, then there exists exactly one normed $n$-th root of $f$.*

*Proof.* There is exactly one polynomial $p$ with only one term such that $[f - p^n] \leq nm - 1$, namely $x^m$. Thus we can apply the lemma with $p = x^m$ and $k = 1$, and subsequently with $k = 2,4,8,\dots,2^i < m$ until we get an $n$-th root. Because the additional coefficients that we get at each step are unique, $\sqrt[n]{f}$ is also. □

**3.8. Remark.** Note that the proof of the lemma always deals only with the leading $k$, or $\tilde{k}$ coefficients of the occurring polynomials. In the following algorithm the notation $c_1^{(k)}(f)$ denotes the polynomial of degree $< k$ constructed from the leading $k$ coefficients, thus $\left[f - c_1^{(k)}(f) \cdot x^{n-k+1}\right] \leq n - k$. Similarly $c_2^{(k)}(f)$ denotes the next $k$ coefficients.

**3.9. Algorithm.** *The $\sqrt[n]{f}$ can be computed according to the proof of the lemma in the following way:*

```
$m := \frac{[f]}{n}$;
$\tilde{f} := c_1^{(m)}(f)$; we forget the remaining coefficients!
$k := 1$; $p_1 := x^m$;
while $k < n$
repeat $\tilde{k} := \min(2k, m)$;
        $f_1 := c_1^{k}(\tilde{f})$;
        $f_2 := c_2^{\tilde{k}-k}(\tilde{f})$;
        $p_2 := c_2^{k}(p_1^m)$;
        $q_1 := \mathrm{EuclideanQuotient}(\frac{p_1 \cdot (f_2 - p_2)}{n f_1})$;
        $p_1 := p_1 x^{\tilde{k}-k} + q_1$;
        $k := \tilde{k}$
return $p \cdot x$.
```

**3.10. Example.** Let us compute the second root of the example in the previous section

$$f = x^{12} + 12x^{11} + 66x^{10} + 223x^9 + 522x^8 + 900x^7 + \cdots .$$

Here, we do not even want to know what the remaining coefficients are. In the notation of the Algorithm we have $n = 2$, $m = 6$, and

$$\tilde{f} = x^5 + 12x^4 + 66x^3 + 223x^2 + 522x + 900.$$

frees us from the superfluous coefficients.

We start with $k := 1$ and $p_1 := 1$ (according to the first approximation $x^m$). In the first step we want to obtain the first $\tilde{k} = 2k = 2$ coefficients. We see immediately

$$\begin{aligned} f_1 &:= 1 \\ f_2 &:= 12 \\ p_2 &:= (\text{2nd coefficient of } p_1^2) = 0. \end{aligned}$$

Now we obtain the 2nd coefficient of the root

$$q_1 = p_1(f_2 - p_2) : nf_1 = 1(12 - 0) : 2 = 6,$$

thus we enter with

$$p_1 := p_1 x + q_1 = x + 6$$

and $k := 2$ into the second step, to obtain the first $\tilde{k} = 2k = 4$ coefficients We read the next coefficients from $\tilde{f}$:

$$\begin{aligned} f_1 &:= x + 12 \\ f_2 &:= 66x + 223. \end{aligned}$$

and compute $p_1^2$:

$$(x+6)^2 = x^2 + 12x + 36,$$

The first 2 coefficients *must* coincide with that of $f$, and

$$p_2 := 36x.$$

Now we get the next *two* coefficients by

$$q_1 = p_1(f_2 - p_2) : nf_1 = (x+6)((66x+223) - (36x)) : 2(x+12) = 15x + \frac{43}{2},$$

thus we enter with

$$p_1 := p_1 x^2 + q_1 = x^3 + 6x^2 + 15x + \frac{43}{2}$$

and $k := 4$ into the third step, to obtain the first $\tilde{k} = \min(2k, m) = \min(8, 6) = 6$ coefficients, i.e., the complete root. We read the next coefficients from $\tilde{f}$:

$$\begin{aligned} f_1 &:= x^3 + 12x + 66x + 223 \\ f_2 &:= 522x + 900. \end{aligned}$$

This time $f_2$ has smaller degree, because there are no more coefficients. We compute $p_1^2$:

$$(x^3 + 6x^2 + 15x + \frac{43}{2})^2 = x^6 + 12x^5 + 66x^4 + 223x^3 + 483x^2 + 645x + \cdots.$$

The last coefficient will not be needed. The first 4 coefficients again *must* coincide with that of $f$, and from the next $\tilde{k} - k = 2$ ones we get

$$p_2 := 483x + 645.$$

Now we get the remaining coefficients by

$$\begin{aligned} q_1 &= p_1(f_2 - p_2) : nf_1 \\ &= (x^3 + 6x^2 + 15x + \frac{43}{2})((522x + 900) - (483x + 645)) : 2(x^3 + 12x^2 + 66x + 223) \\ &= \frac{39}{2}x + \frac{21}{2}. \end{aligned}$$

Thus

$$p_1 := p_1 x^2 + q_1 = x^5 + 6x^4 + 15x^3 + \frac{43}{2}x^2 + \frac{39}{2}x + \frac{21}{2},$$

gives all coefficients of the root and $\sqrt[2]{f} = x \cdot p_1$ is normed and of degree 6. In fact, squaring this polynomial gives

$$(\sqrt[2]{f})^2 = x^{12} + 12x^{11} + 66x^{10} = 223x^9 + 522x^8 + 900x^7 + \frac{4693}{4}x^6 + \frac{2307}{2}x^5 + \frac{3327}{4}x^4 + \frac{819}{2}x^3 + \frac{441}{4},$$

and we check that its first 6 coefficients coincide with that of $f$.

**3.11. Remark.** Because at each step in the iteration the number of coefficients of $\sqrt[n]{f}$ already computed is doubled, our algorithm needs only $O(\log n)$ iterations. The most expensive part in the $i$-th iteration is the computation of the first $2^i$ coefficients of the $n$-th power of a polynomial of degree $2^{i-1}$. This can be accomplished by the usual method of successive squaring with $O(\log n \cdot M(2^i))$ field operations. Again $M(k)$ denotes the number of steps used for multiplying polynomials of degree $k$. According to [SS71] we can choose $M(k) = k \log k$. For practical purposes, however, $M(k) = k^{1.5}$ is more appropriate (Karatsuba method). In any case, we have $M(2k) \geq 2M(k)$, thus the total cost of our algorithm is dominated by the cost of the last step, which is $O(M(m)\log n)$. This is a very good bound, at least if $k$ is a finite field. For infinite fields, the growth of the size of the coefficients becomes essential. A good polynomial bound is obtained in [vzG90].

**3.12. Theorem.** *Let $f$ be a tame normed polynomial.*

(i) *For each divisor $n$ of $[f]$ there is exactly one normed root $\sqrt[n]{f}$; its coefficients are rational functions of the first $\frac{[f]}{n}$ coefficients of $f$.*

(ii) *For each divisor $m$ of $[f]$ there is at most one normed right component $p \mathrel{\subseteq\mkern-14mu\rightarrow} f$ of degree $m$, and, in the affirmative case, $p = \sqrt[n]{f}$, with $n = \frac{[f]}{m}$.*

(iii) *For each finite sequence $n_k, \ldots, n_1 \in \mathbb{N}$ there is at most one normed decomposition $f = p_k \circ \cdots \circ p_1$ such that $[p_i] = n_i$.*

(iv) *One gets no more normed decompositions of $f$ when components are allowed to have coefficients in some algebraic extension field of $\mathbb{k}$. In particular, a polynomial is prime over an extension field iff it is prime over $\mathbb{k}$.*

*Proof.*

(i) Proposition 3.7 proves uniqueness, and from the algorithm we see that only elementary field operations are used in its computation.

(ii) Each right component of degree $m$ must be an $n$-th root by proposition 3.2. Thus it is the unique one.

(iii) The rightmost component is unique by the previous part. Then we use Taylor division to see that $f_k \circ \cdots \circ f_2$ is also uniquely determined. Applying the same argument recursively, we see that all components are determined.

(iv) As both root computation and Taylor division use only rational operations, this is clear from the construction in the previous part. □

This theorem and its proof also show that we have got a fast method to compute a prime decomposition of a tame polynomial $f$. We just compute roots of $f$ for each divisor $n$ and get a good (the only possible) *candidate* for being a right component of the appropriate degree. We can test this using Taylor division and continue by decomposing $f \div \sqrt[n]{f}$.

**3.13. Algorithm.** *Let $f$ be a normed tame polynomial. Its prime decomposition can be computed in the following way:*

```
For each proper divisor $m$ of $[f]$ (smallest first);
         compute the candidate $p$ of degree $m$
                 as an appropriate root.
         test whether this is a right component using Taylor division;
         in the affirmative case continue recursively with $f \div p$,
         otherwise test the next divisor.
If all divisors are exhausted, without finding a right component,
                         then $f$ is prime.
```

In fact, this algorithm finds the the *first* prime decomposition, i.e., that with smallest components on the right. If we try all proper divisors, (a variant of) this algorithm even finds all prime decompositions.

**3.14. Example.** Let us now decompose our polynomial

$f = x^{12}+12x^{11}+66x^{10}+223x^9+522x^8+900x^7+1179x^6+1188x^5+918x^4+533x^3+222x^2+60x.$

into prime components. According to the algorithm, we first look for the candidate of degree 2; it must be $\sqrt[6]{f}$. Algorithm 3.9 finds, with only one iteration,

$$\sqrt[6]{f} = x^2 + 2x,$$

but using Taylor division we have already seen in Example 3.10 that this is not a right component. Note that, for this purpose, we need not do all the computations in that example, because we obtain a linear Taylor coefficient already at the first step. So let us compute the candidate of degree 3; we get

$$p := \sqrt[4]{f} = x^3 + 3x^2 + 3x.$$

We already know this polynomial from Example 2.10, where it was shown, using Taylor division, that this is in fact a right component, and

$$r := f \div p = x^4 + 7x^3 + 18x^2 + 20x.$$

As $r$ has degree 4 it could be decomposable. But $\sqrt[2]{r} = x^2 + \frac{x}{2}$, and, using Taylor division we see that this is not a right component. So $r$ is prime and we have found the prime decomposition $f = r \circ p$.

We ask whether there are any more prime decompositions. Thus compute the candidate of degree 4:

$$q := \sqrt[3]{f} = x^4 + 4x^3 + 6x^2 + 5x.$$

Now Taylor division shows that $q$ is in fact another right component with

$$s := f \div q = x^3 + 6x^2 + 12x.$$

We know already that $q$ is prime, because $f$ has no right component of degree 2. Thus $f$ has the two essentially different (i.e., not associated) prime decompositions $f = r \circ p = s \circ q$. In fact, we have got a *prime bidecomposition.* To obtain all prime decompositions of $f$ we can now test the candidate of degree 6, i.e., $\sqrt[2]{f}$. But in Remark 5.6 we will see that this is in fact not necessary.

**3.15. Remark.** Though the notion of root for polynomials (in this sense) as well as its systematic use is new, a proof of proposition 3.7 is already contained implicitly in [Eng41]. [LN73] contains a similar proof. Additionally, the algorithms for polynomial decomposition in [Gut88] and [KL89] use very similar constructions.

Our proof is not more complicated than the ones mentioned above, and has the advantage that it almost directly leads to the fastest known algorithms. Whereas the above methods essentially compare the first coefficients, one by one, our proof and algorithm compare the coefficients in a second order manner, thus doubling the accuracy at each step.

The similarity with Newton's iteration method is not incidental: Every polynomial $f = \sum a_i x^{n-i}$ can be identified with the Laurant series $\sum a_i \left(\frac{1}{x}\right)^{i-n}$ around $\infty$, which has only negative terms. Thus, if we consider only the leading $k$ coefficients of the polynomials, we are doing essentially power series arithmetic up to order $O(x^{n-k})$. It is well known that the class of power series with leading coefficient 1 has unique roots. The paper [BK78] outlines how these roots can be computed efficiently using Newton's iteration method, and [vzG90] proposes this for polynomial decomposition. So our Algorithm 3.9 does essentially the same as that in [vzG90].

The proofs in [Rit22] and [DW74] do not contain any version of proposition 3.7, but use Riemann surfaces resp. valuation theory instead, which essentially reduce to the use of Laurant series. Using roots for polynomials directly, we can avoid the discourse to infinite structures completely.

Roots have proved very useful in developing good algorithms for decomposition as well as some interesting uniqueness results. One can get even more.

**3.16. Remark.** Yet another way to express part (ii) of theorem 3.12 is that the degree function injectively maps the component lattice into the divisor lattice of $[f]$. Though it is trivially monotone, we do not yet know that it is a lattice homomorphism. The next proposition proves one half of this, the rest must be postponed.

**3.17. Lemma.** *Let $r$ be a normed polynomial and let $n$ be a tame divisor of $[r]$; then for all normed polynomials $p$*

$$\sqrt[n]{r \circ p} = \sqrt[n]{r} \circ p.$$

*Proof.* We have to prove that $\sqrt[n]{r} \circ p$ satisfies the characteristic property of an $n$-th root of $r \circ p$. Thus we estimate

$$\begin{aligned}
\left[r \circ p - (\sqrt[n]{r} \circ p)^n\right] &= \left[r \circ p - x^n \circ \sqrt[n]{r} \circ p\right] \\
&= \left[r - x^n \circ \sqrt[n]{r}\right][p] \\
\text{(by definition of } \sqrt[n]{r}\text{)} \qquad &\leq ([r] - \frac{[r]}{n})\,[p] \\
&= [r \circ p] - \frac{[r \circ p]}{n},
\end{aligned}$$

which is what we wanted. □

**3.18. Proposition.** *If $p$ and $q$ have any tame common left multiple, then*

$$[p \overset{\uparrow}{\cup} q] = [p] \cup [q]\,.$$

*Proof.* Let $f = r \circ p = s \circ q$ be tame, then $[f] \supseteq [p] \cup [q]$, and with

$$n := \frac{[f]}{[p] \cup [q]} = [r] \cap [s]$$

we have

$$\sqrt[n]{f} = \sqrt[n]{r} \circ p = \sqrt[n]{s} \circ q$$

as another common left multiple, and this one has the appropriate degree $[p] \cup [q]$. □

The proof of the corresponding result for $\underset{\downarrow}{\cap}$ (5.2) is completely different and surprisingly needs a discourse to rational function decomposition.

That the greatest common right components are independent of the ground field was not surprising, as this is so for greatest common divisors, too. But for complete factorizations the ground field is essential. Thus, prime decompositions have a considerably simpler structure in this respect, at least in the tame case. On the other hand, every polynomial can be factored into linear ones over its splitting field. There is no (known) compositional replacement for this. One could expect that every polynomial can be decomposed into ones of prime degree, which are trivially prime, just like the linear polynomials are trivially irreducible. But this, by far, is not true, as most polynomials are prime. In fact, If $f = r \circ p$ and $g$ is some polynomial with $[g] \leq [f] - [p]$, then $[(f + g) - p^n] \leq [f] - [p]$, thus $p = \sqrt[n]{f} = \sqrt[n]{f + g}$. Suppose $f + g = \hat{r} \circ p$; As $f = r \circ p$, $g = (r - \hat{r}) \circ p$. Thus, for e.g. every polynomial $g$ such that its degree is not a multiple of $[p]$, $f + g$ is indecomposable. So for each decomposable polynomial we get a whole bunch of prime polynomials of any degree. Another way to see this is looking at the number of coefficients: $r \circ p$ is computed from $[r] + [p]$ coefficients, whereas a general polynomial of the some degree has $[r] \cdot [p]$ coefficients.

In this context it would be particularly interesting to know what happens when decomposing into algebraic functions.

Another interesting question is whether there is a compositional replacement for squarefree factorizations.

## § 4. Rational Function Decomposition

**4.1. Notation.** The elements of the field $\Bbbk(x)$ will be called rational functions, as it consists of all rational expression involving $x$. It is the quotient field of the integral domain of polynomials $\Bbbk[x]$, thus the elements can be represented in the form $\frac{p}{q}$, where $p$ and $q \neq 0$ are polynomials. $\frac{p}{q}$ is said to be in *prime form* iff $p$ and $q$ are relatively prime. Of course, every rational function has a prime form which is unique up to constant factors.

If $f$ and $g$ are rational functions, then we can substitute $g$ for the $x$ in $f$ to to get another rational function $g(f)$. We get problems, however, if $f$ is constant and $g$ has a pole at $f$. In this case we assign a new constant value $\infty$ to $g(f)$. Note, in particular, that $\frac{1}{x} \circ 0 = \frac{-1}{x} \circ 0 = \infty$, thus $\infty = -\infty$. Consistently, we assign $g(\frac{1}{x})(0)$ to $g(\infty)$, and we define $\infty \circ f = \infty$. So we can view rational functions as functions from $\Bbbk(x)_\infty := \Bbbk(x) \cup \{\infty\}$ onto itself. Note that $x$ then is viewed as the identity function, and that $g(x) = g$, so different rational functions give rise to different functions. This justifies the name rational *function*.

As the rational functions are really functions, they can be composed, and we have

$$f \circ g = f \circ g \circ x = (f \circ g)(x) = f(g(x)) = f(g),$$

thus extending composition of polynomials to rational functions.

We have done this rather pedantic introduction of composition to be sure that associativity is preserved even if constants are involved. But now the following is immediate.

**4.2. Proposition.** $(\Bbbk(x)_\infty, \circ)$ *is a monoid with identity $x$. It contains the submonoid of polynomials.* □

**4.3. Remark.** A rational functions $f$ does not necessarily give rise to a function of $\Bbbk$ into itself, as it can have poles. But $\bar{f} : a \mapsto f \circ a$ is a function of $\Bbbk_\infty$ into itself. Note, however, that, if $\Bbbk$ is finite, $\bar{f}$ may vanish, without $f$ being zero. For example, $x^2 + x$ corresponds to the zero function of $\mathbb{Z}_2$ into itself.

**4.4. Notation.** Because the rational functions form a monoid, we can use the theory of §1. In particular, we speak of right components of a rational function, its component structure, decompositions, and so on, just like for polynomials. However, we have to be careful here, because a polynomial, indecomposable as an element of $(\Bbbk[x], \circ)$, could have a non-trivial decomposition into rational functions. We will prove at the end of this section that this cannot happen and that no ambiguity is possible here. Until then, the rational function meaning is used exclusively.

## *Rational Function Fields*

Though rational functions have a more complicated structure than polynomials, there is one advantage: $\mathbb{k}(x)$ is a field, thus we can use the well developed theory of field extensions. We establish some important facts in this area, mainly along the lines of [vdW66, §73].

**4.5. Notation.** Let $k$ and $K$ be arbitrary fields. If $k \subseteq K$, i.e., if $k$ is a subfield of $K$, then $K$ is called an *extension* of $k$, and we denote it by $K : k$. Its *degree*, i.e., the dimension of $K$ as a vector space over $k$, is denoted by $[K : k]$. Fields between $k$ and $K$ are called the *intermediate* fields of $K : k$. Extensions of the form $k(f) : k$, are called *simple*. If $L$ is another extension of $k$, then a homomorphism from $K : k$ to $L : k$ is one from $K$ to $L$ that fixes $k$. It is also called a $k$-homomorphism.

**4.6. Remark.** In particular, for each $f \in \mathbb{k}(x)$, $\mathbb{k}(x)$ is a (simple) extension of $\mathbb{k}(f)$. In general, the intermediate fields of $\mathbb{k}(x) : \mathbb{k}$ are called the *rational function fields.*

**4.7. Proposition.** *Let $f$ be a non-constant rational function. The mapping*

$$\begin{aligned} \circ f : \mathbb{k}(x) &\to \mathbb{k}(f) \\ g &\mapsto g \circ f \end{aligned}$$

*defines an isomorphism of the extension fields $\mathbb{k}(x) : \mathbb{k}$ and $\mathbb{k}(f) : \mathbb{k}$.*

*Proof.* We have to prove the distributive laws

$$\begin{aligned} (g + h) \circ f &= g \circ f + h \circ f \\ (g \cdot h) \circ f &= g \circ f \cdot h \circ f \\ g^{-1} \circ f &= (g \circ f)^{-1}, \end{aligned}$$

but these are satisfied by the definition as substitution. Also $1 \circ f = 1$. Being a homomorphism of fields, the map is automatically injective, it is onto by the definition of $\mathbb{k}(f)$. Obviously, the constants are fixed. □

**4.8. Remark.** The distributive laws are also satisfied if $f$ is constant, as long as $\infty$ is not involved. More exactly, an indeterminate expression like $\infty + \infty$ or $\frac{0}{0}$ must not occur. For example, $(\frac{1}{x} + \frac{1}{x}) \circ 0 = \frac{2}{x} \circ 0 = \frac{2}{0} = \infty$, but $\frac{1}{x} \circ 0 + \frac{1}{x} \circ 0 = \frac{1}{0} + \frac{1}{0} = \infty + \infty$; or $\frac{x}{x} \circ 0 = 1 \circ 0 = 1$, but $\frac{x \circ 0}{x \circ 0} = \frac{0}{0}$.

**4.9. Proposition.** *Let $f, h \in \mathbb{k}(x)$ then*

$$\begin{aligned} f \mathrel{\subseteq\!\!\!\rightarrow} h &\iff \mathbb{k}(f) \supseteq \mathbb{k}(h) \\ f \mathrel{\overset{\uparrow}{\cong}} h &\iff \mathbb{k}(f) = \mathbb{k}(h) \end{aligned}$$

*Proof.* $f \mathrel{\subseteq\!\!\!\rightarrow} h$ just means that $h \in \mathbb{k}(f)$. But $\mathbb{k}(h)$ is the smallest field containing $\mathbb{k}$ and $h$, so $\mathbb{k}(f) \supseteq \mathbb{k}(h)$. Conversely, from $h \in \mathbb{k}(f)$, we have $h \mathrel{\leftarrow\!\!\!\supseteq} f$. The second assertion is a trivial consequence of the first. □

**4.10. Remark.** This means that the component structure of rational functions, $(\Bbbk(x)/\overset{\uparrow}{\simeq}, \mathrel{\leftarrow\!\!\supseteq})$ can be embedded into the lattice of intermediate fields of $\Bbbk(x) : \Bbbk$, ordered by $\subseteq$. Note the reversion of the symbol.

**4.11. Definition.** We extend the notion of *degree* to rational functions by defining

$$[f] := \max([p], [q]),$$

where $f = \frac{p}{q}$ is in prime form.

Note that $f = \frac{p}{q}$ must be in prime form to make this well-defined.

**4.12. Notation.** As $\Bbbk(x)$ is a field, it will be convenient to consider polynomials over $\Bbbk(x)$. For this reason, we choose a new variable $y$ to denote the indeterminate of such polynomials. Thus polynomials over a rational function field are understood to be elements of $\Bbbk(x)[y]$.

We cite one form of Gauß's lemma ([Coh77] or [vdW66, §30]). Note that a polynomial over a ring is called *primitive* iff its coefficients are coprime.

**4.13. Lemma** (Gauß). *A polynomial over $\Bbbk[x]$ is irreducible iff it is primitive and irreducible over $\Bbbk(x)$.*

**4.14. Proposition.** *Let $f = \frac{p}{q} \in \Bbbk(x)$ be a non-constant rational function in prime form. Then $\Bbbk(x) : \Bbbk(f)$ is a finite field extension. The minimal polynomial of $x$ over $\Bbbk(f)$ is given by*

$$m(y) = p(y) - f \cdot q(y),$$

*thus $[\Bbbk(x) : \Bbbk(f)] = [f]$.*

*Proof.* Obviously $m(y) \in \Bbbk(f)[y]$, and it satisfies $m(x) = p(x) - f \cdot q(x) = p - \frac{p}{q} \cdot q = 0$. Thus $x$ is algebraic over $\Bbbk(f)$. $m(y)$ has degree $\max([p], [q]) = [f]$ (in $y$), So, if we can show that $m(y)$ is irreducible over $\Bbbk(f)$, all the remaining assertions are also clear.

Note that the field $\Bbbk(f)$ is isomorphic to $\Bbbk(x)$, thus we can treat $f$ as an independent variable. Because $m(y) \in \Bbbk[f][y]$, and $\Bbbk[f][y] = \Bbbk[y][f]$, we can also view $m(y)$ as a polynomial in $f$ over $\Bbbk(y)$. As such, it is linear, hence irreducible, and primitive because $p$ and $q$ are coprime. Thus, by Gauß's lemma, $m(y)$ is also irreducible in $\Bbbk[y][f] = \Bbbk[f][y]$. Hence, again by Gauß's lemma, irreducible over $\Bbbk(f)$. □

**4.15. Theorem** (Lüroth). *All rational function fields are simple, i.e., of the form $\Bbbk(f)$ for some rational function $f$.*

*Proof.* E.g. [vdW67] or [Coh77] contain elementary proofs. They make essential use of Proposition 4.14 and Lemma 4.13. □

**4.16. Corollary.** *The component structure of rational functions, $(\Bbbk(x)_\infty/\overset{\uparrow}{\simeq}, \mathrel{\leftarrow\!\!\supseteq})$, is isomorphic to the lattice of intermediate fields of $\Bbbk(x) : \Bbbk$, ordered by $\subseteq$.*

*Proof.* We have already remarked (4.10) that $(\Bbbk(x)_\infty//\overset{\uparrow}{\simeq}, \mathrel{\leftarrow\!\!\supseteq})$ can be embedded into the lattice of intermediate fields. But Lüroth's theorem ensures that this embedding is surjective. □

### *Component Lattice*

**4.17. Proposition.**

(i) *The degree function is a homomorphism from* $(\Bbbk(x)_\infty, \circ)$ *onto* $(\mathbb{N}_0, \cdot)$, *i.e.,*

$$[g \circ f] = [g]\,[f]$$

*for all rational functions* $f$ *and* $g$.

(ii) *The units of* $(\Bbbk(x)_\infty, \circ)$ *are those of degree 1.*

(iii) $f$ *is a right cancellable element of* $(\Bbbk(x)_\infty, \circ)$ *iff it is not constant.*

*Proof.*

(i) If one of $f$ and $g$ is constant so is $g \circ f$, and the result is immediate. Thus assume that both are non-constant. Then both $\Bbbk(g\circ f) : \Bbbk(f)$ and $\Bbbk(f) : \Bbbk(x)$ are finite field extensions, thus

$$\begin{aligned}[g \circ f] = [\Bbbk(x) : \Bbbk(g \circ f)] &= [\Bbbk(f) : \Bbbk(g \circ f)][\Bbbk(x) : \Bbbk(f)]] \\ &= [\Bbbk(x) : \Bbbk(g)][\Bbbk(x) : \Bbbk(f)] = [g]\,[f]\,.\end{aligned}$$

(ii) One can use the degree function, just as for polynomials (Proposition 2.5). Here is another possibility: Let $f$ be non-constant. Using the injectivity of the isomorphism in proposition 4.7, $g \circ f = 0$ implies $g = 0$. So, by the distributive law, $f$ is right cancellable. Conversely, constants are not cancellable: $g(c) = h(c)$ just means that $g$ and $h$ have the same value on $c$.

(iii) By the first part, every unit must have degree 1. On the other hand $[\Bbbk(x) : \Bbbk(u)] = 1$ whenever $[u] = 1$. Thus $u$ induces an automorphism of $\Bbbk(x)$ (cf. 4.7), mapping some element $v$ to $x$, i.e., $v \circ u = x$. □

The multiplicativity of the degree is particularly good news. For example, it allows us to compute complete decompositions of rational functions, by an approach with indetermined coefficients, just like in the polynomial case (2.6). The polynomial time algorithm in [Zip91] also works for rational functions, in fact, was designed for this case. [AGR] contains an algorithm that has exponential worst case complexity, but is faster in practice.

**4.18. Proposition.** *Let both* $g = \frac{r}{s}$ *and* $f = \frac{p}{q}$ *be rational functions in prime form.*

(i) $r \circ p$ *and* $s \circ p$ *are relatively prime.*

(ii) *Let*

$$u := (r \circ f) \cdot q^{[g]},$$
$$v := (s \circ f) \cdot q^{[g]},$$

*i.e., with* $r = \sum_{i=0}^{n} r_i x^i$ *and* $s = \sum_{i=0}^{m} s_i x^i$

$$u := r_n p^n q^{[g]-n} + r_{n-1} p^{n-1} \cdot q^{[g]-n+1} + \cdots + r_0 q^{[g]}$$
$$v := s_m p^m q^{[g]-m} + s_{m-1} p^{m-1} \cdot q^{[g]-m+1} + \cdots + s_0 q^{[g]}$$

*Then* $g \circ f = \frac{u}{v}$ *is in prime form.*

(iii) *If* $g \circ f$ *is a non-constant polynomial and* $[p] > [q]$, *then both* $f$ *and* $g$ *are polynomials.*

*Proof.*

(i) As $r$ and $s$ are coprime, $a \cdot r + b \cdot s = 1$, for some polynomials $a, b$ (Bezout's relation). We substitute $p$, and get

$$a(p) \cdot r(p) + b(p) \cdot s(p) = 1,$$

thus $r(p)$ and $s(p)$ are coprime again.

(ii) As above, we substitute $f$ into Bezout's relation: $a(f) \cdot r(f) + b(f) \cdot s(f) = 1$. This time, however, rational functions are involved. To transform this into a relation involving only polynomials, we multiply by an appropriate power of $q$ to get an equation of the form

$$\tilde{a} \cdot u + \tilde{b} \cdot v = q^k,$$

such that both $\tilde{a}$ and $\tilde{b}$ are polynomials. So $\gcd(u, v)$ must divide $q^k$. But at least one of $u$ and $v$ has the form $r_n p^{[g]} + q \cdot (\dots)$ or $s_m p^{[g]} + q \cdot (\dots)$, respectively, so is coprime to $q$, as $p$ and $q$ are coprime. Thus $\gcd(u, v) = 1$.

(iii) If $g \circ f$ is a polynomial, then $[v] = 0$. But $[p] > [q]$ implies

$$0 = [v] = m\,[p] + ([g] - m)\,[q]\,.$$

As $[g] \geq m$ and $[p] > [q] \geq 0$, we conclude $m = 0$. But then $[g]\,[q] = 0$, so $[q] = 0$, as $[g] \neq 0$. □

Thus we have got the prime form of $g \circ f$ quite explicitly in terms of that of $f$ and $g$. No polynomial gcd-computation is necessary for its computation.

**4.19. Definition.** A sublattice $S$ of a lattice $L$ is *convex* iff for all $a, b \in S$ and $c \in L$, $a \leq c \leq b$ implies $c \in S$.

**4.20. Theorem.**

(i) *The component lattice of a polynomial is independent whether is considered in* $(\mathbb{k}(x)_\infty, \circ)$ *or in* $(\mathbb{k}[x], \circ)$.

(ii) *The component lattice of polynomials is a convex sublattice of the component lattice of rational functions.*

*Proof.* If $p$ and $q$ are left associated polynomials, then $p = u \circ q$ for some some fractional linear function $u$. Thus, with 4.18.(iii), $u$ is a linear *polynomial.* Hence we can identify $(\mathbb{k}[x]/\triangleq, \overset{\uparrow}{\cup}, \underset{\downarrow}{\cap})$ with a subset of $(\mathbb{k}(x)_\infty/\triangleq, \overset{\uparrow}{\cup}, \underset{\downarrow}{\cap})$.

Suppose that $g \circ f$ is a polynomial and $f = \frac{p}{q}$. If $[p] > [q]$, we can apply 4.18(iii) directly. If $[p] < [q]$, we apply it to the associated decomposition $(g \circ \frac{1}{x}) \circ \frac{q}{p}$. In the case $[p] = [q]$ the quotient of the Euclidean division of $p$ by $q$ is some constant, say $c$, thus $p = c \cdot q + r$, where $r$ is the remainder (so $[r] < [q]$), thus

$$\frac{p}{q} = c + \frac{r}{q} = (c + x) \circ \frac{r}{q},$$

and again we get an associated decomposition to which the proposition can be applied. In any case, the decomposition is left associated to one using only polynomials. This proves that the component lattices are the same.

For the second part, it remains to show that, for arbitrary polynomials $p$ and $q$, $p \overset{\uparrow}{\cup} q$ is (left associated to) a polynomial. If $p \overset{\uparrow}{\cup} q$ is constant, this is trivial. Otherwise write it in the form

$$p \overset{\uparrow}{\cup} q = \frac{r}{s} \circ p = \frac{\hat{r}}{\hat{s}} \circ q,$$

then, $\frac{r\circ p}{s\circ p} = \frac{\hat{r}\circ q}{\hat{s}\circ q}$, and, by the first part of proposition 4.18 both sides are in prime form. Thus, up to a constant factor, $r \circ p = \hat{r} \circ q$ and $s \circ p = \hat{s} \circ q$. The non-constant one is a polynomial common left multiple whose degree is $\leq [p \Cup q]$. □

Expressed less formally, this theorem says that we never have to take care whether notions like *right component, component lattice of $f$, least common left multiple, prime decomposition* are relative to the monoid of polynomials or that of rational functions.

## § 5. The Invariant Integers

We continue considering tame polynomials over the field $\Bbbk$.

**5.1. Lemma.** *Suppose the polynomials $p$ and $q$ have a tame common left multiple, but no nontrivial common right component, i.e., $p \Cap q = x$. Then their degrees are coprime, i.e., $[p] \cap [q] = 1$.*

*Proof.* Let $p \Cup q = r \circ p = s \circ q$. From proposition 3.18,

$$[p \Cup q] = [p] \cup [q] = [r]\,[p] = [s]\,[q]\,.$$

Thus $[r] \cap [s] = 1$, and we will prove $[p] = [s]$, $[q] = [r]$. Obviously, $[p] \supseteq [s]$ and $[q] \supseteq [r]$, and $[p] > 0$. We show $[p] \leq [s]$, then $[q] \leq [r]$ follows by symmetry, proving the proposition.

Define the polynomial $m(y) = s(y) - r \circ p$, thus $m(y) \in \Bbbk(p)[y]$, with degree $[s]$ in $y$. Then $m(q) = s(q) - r \circ p = 0$. This means (cf. Proposition 4.14),

$$[\Bbbk(p)(q) : \Bbbk(p)] \leq [s]\,.$$

But from corollary 4.16,

$$\Bbbk(p)(q) = \Bbbk(p, q) = \Bbbk(p \Cap q) = \Bbbk(x).$$

As $[\Bbbk(x) : \Bbbk(p)] = [p]$, $[p] \leq [s]$. □

**5.2. Proposition.** *If polynomials $p$ and $q$ have a tame common left multiple, then*

$$[p \Cap q] = [p] \cap [q]\,.$$

*Proof.* One simply gets rid of the common component using Taylor division and uses the lemma. In detail: Let $t = p \Cap q$. We already have $t \subseteq [p] \cap [q]$ (cf. Remark 2.24). By Taylor division, there are unique polynomials $\tilde{p}$ and $\tilde{q}$ such that $p = \tilde{p} \circ t$ and $q = \tilde{q} \circ t$. By the lemma, $[\tilde{p}] \cap [\tilde{q}] = 1$. But $[\tilde{p}]\,[t] = [p]$ and $[\tilde{q}]\,[t] = [q]$, so $[t] \supseteq [p] \cap [q]$. □

Somewhat strange, we need the existence of a nontrivial common *multiple* to prove this property of common *components*. Note that the corresponding equality for $\Cup$ has been proved completely differently, and was in fact used here.

**5.3. Corollary.** *Prime bidecompositions permute the degrees, i.e., in the prime bidecomposition $r \circ p = s \circ q$,*

$$[p] = [s] \ \textit{ and } \ [r] = [q]\,. \qquad \square$$

Now we are fine out and have got the essential result of this chapter:

**5.4. Theorem.** *The component lattice of a tame polynomial $f$ is isomorphic to a sublattice of the divisor lattice of $[f]$. The degree function provides the embedding.*

*Proof.* By the corollaries 3.18 and 5.2, the degree is a lattice homomorphism, and by theorem 3.12 it is injective. □

**5.5. Example.** Let $f = x^{12}$. Its right components are $x, x^2, x^3, x^4, x^6, x^{12}$. Thus, in this case, the right component lattice of $f$ is even isomorphic to the divisor lattice of 12. Of course, we have the same situation with all polynomials of the form $x^n$. The Dickson polynomials (described in Chapter II) provide a class of polynomials with the same property.

**5.6. Remark.** Of course, these polynomials are rather special. It is not surprising that most polynomials miss components of certain degrees. Conversely it is somewhat remarkable that, if a polynomial has right components of degrees e.g. 6 and 4, then it has also one of degree 2, because $2 = 6 \cap 4$, and the component lattice is a sub*lattice*. This can save us a lot of computations, if we want to know all prime decompositions of a given polynomial.

**5.7. Example.** Let us reconsider the polynomial $f$ from Example 3.14. It has right components of degrees 3 and 4, but not of degrees 2. Hence it cannot have one of degree 6, which frees us from testing the candidate of degree 6. Additionally, it was unnecessary in that example to test whether $r$ is prime, because that would imply a right component of degree 6.

Summarizing, the right component lattice of $f$ is isomorphic to the lattice $1 \subset 3, 4 \subset 12$, which is a proper sublattice of the divisor lattice of 12.

**5.8. Corollary.** *The component lattice of any tame polynomial is distributive.*

*Proof.* By the theorem, it is (homomorphic to) a sublattice of the distributive lattice $(\mathbb{N}, \supseteq)$. □

Note that every bounded sublattice of $(\Bbbk[x]_n, \mathrel{\leftarrow\!\!\!\!\supseteq})$, not containing 0, is a sublattice of the component lattice of some polynomial (namely the maximum).

**5.9. Definition.** A lattice is called to have some property *locally* iff it is true for every bounded sublattice that does not contain a global maximum.

With this notion we can express our local result in a *global* form:

**5.10. Corollary.** *Let* $\operatorname{char} \Bbbk = 0$. *Then the lattice* $(\Bbbk[x]_n, \mathrel{\leftarrow\!\!\!\!\supseteq})$ *is embedded locally into* $(\mathbb{N}, \supseteq)$ *by the degree function. Thus it is locally distributive.*

*Proof.* By the assumption about the characteristic, every bounded sublattice not containing 0 is the component lattice of a *tame* polynomial. Thus the assertion follows with the theorem and its corollary. □

Using our abstract theory of §1 we get the classical result on prime decompositions as a corollary to our Theorem 5.4

**5.11. Theorem** (Ritt). *Let $f$ be a tame polynomial.*

(i) *All prime decompositions of $f$ are related.*

(ii) *The number and the degrees of the components in a prime decomposition, but not necessarily their order, are invariant.*

*Proof.* The right component structure is distributive, thus modular, thus semimodular. Thus theorem 1.10 can be used. By corollary 5.3, prime bidecompositions just permute the degrees. □

Using theorem 4.7, there is another interesting consequence.

**5.12. Corollary.** *For every tame polynomial $f$ the lattice of intermediate fields of $\mathbb{k}(x):\mathbb{k}(f)$ is isomorphic to a sublattice of $[f]$, hence is distributive and all its maximal chains are related.* □

**5.13. Remark.** To proof that all prime decompositions are related one just needs that the component lattice is semimodular. No easier proof for semimodularity than that via distributivity via the embedding into the integers is known, nor handy conditions on a non-tame polynomial for having a semimodular component lattice.

The theorem leaves open the question, how many bidecompositions there are and how they look like. This is the topic of the next chapter.

CHAPTER II

# Characterization of Prime Bidecompositions

## § 1. Bidecompositions

This chapter contains a simplified proof of Ritt's characterization of all prime bidecompositions of the monoid $(\mathbb{k}[x], \circ)$.

**1.1. Example.** An easy example of bidecompositions is given by the powers, because they, trivially, satisfy

$$x^m \circ x^n = x^n \circ x^m.$$

This can be generalized a bit to

$$(x^m \cdot t(x)^n) \circ x^n = x^n \circ (x^m \cdot t(x^n)), \tag{1}$$

for an arbitrary polynomial $t$, as can be verified immediately. A second important class comes from the Dickson polynomials, as defined in the next section. They satisfy

$$D_m(x, a^n) \circ D_n(x, a) = D_n(x, a^m) \circ D_m(x, a), \tag{2}$$

for all constants $a$.

**1.2. Definition.** Let $r \circ p = s \circ q$ be a bidecomposition. For all units $a, b, c, d$ the bidecomposition

$$(a \circ r \circ b) \circ (b^{\circ -1} \circ p \circ c) = (a \circ s \circ d) \circ (d^{\circ -1} \circ q \circ c)$$

is called *associated* to the original one.

**1.3. Definition.** A bidecomposition associated to one of type (1) is called *exponential*, one associated to one of type (2), but not of type (1), is called *trigonometric.*

**1.4. Notation.** With $\mathbb{k}^{\mathrm{alg}}$ we denote the algebraic closure of $\mathbb{k}$.

We will need a stronger hypothesis than just *tame*:

**1.5. Definition.** A tame polynomial $f$ is called *completely tame* iff for all $e \in \mathbb{k}^{\mathrm{alg}}$, $f - e$ has no zero (in $\mathbb{k}^{\mathrm{alg}}$) whose multiplicity $\mu$ is a multiple of $\operatorname{char} \mathbb{k}$. A (bi)decomposition is completely tame iff all its components are.

**1.6. Remark.** Again, in the case of characteristic 0, completely tame just means non-constant. Otherwise a sufficient condition is $[f] < \operatorname{char} \mathbb{k}$.

Now we can express the theorem that we want to proof in the next five sections.

**1.7. Theorem** (Ritt). *All completely tame prime bidecompositions over a field not of characteristic* 2 *are either exponential or trigonometric.*

**1.8. Corollary.**

(i) *Over a field of characteristic 0 all prime bidecompositions are either exponential or trigonometric.*

(ii) *If* $\operatorname{char}\Bbbk \neq 0$ *then all prime bidecompositions using polynomials of degrees* $< \operatorname{char}\Bbbk$ *are either exponential or trigonometric.*

This theorem again goes back to [Rit22], with generalizations in [Lev42], [LN73], [DW74], [Sch82].

The proof given here is completely elementary, in the sense that, except for the results proved in chapter I, the basic theory of field extensions is the most advanced mathematics involved. Nevertheless it is not longer, quite on the contrary, some simplifications, just in the most involved passages, were possible. Our schedule will be as follows.

After discussing some not so widely known properties of Dickson (or Chebyshev) polynomials and of the Tschirnhaus transform, we will take a closer look at the ramification structure of the components in a bidecomposition. Then, in §5, we can give a condition for a bidecomposition to be exponential. The same is done in §6 for the trigonometric case. As exactly one of these two conditions is always satisfied the proof is complete then.

From Proposition I.3.12 we see that every bidecomposition that is prime over an extension field of $\Bbbk$ is also prime over $\Bbbk$. Nevertheless we cannot simply restrict us to algebraically closed fields, because the theorem says more: that every prime bidecomposition is *associated* to one of the specified types, and polynomials over $\Bbbk$ that are associated over an extension field need *not* be associated over $\Bbbk$. One may obtain this stronger result from that for algebraically closed fields by a careful analysis of the linear polynomials involved as in [Sch82]. As an alternative, we give the proof in a version that directly proves the characterization for general fields.

On the other hand, every bidecomposition is associated (even over the ground field $\Bbbk$) to one containing only monic polynomials. Therefore we can restrict ourselves to monic polynomials whenever we want.

## § 2. Dickson Polynomials

As the Dickson polynomials constitute bidecompositions, a closer look at their properties will be useful.

**2.1. Definition.** Let $a \in \Bbbk$. We define the *Dickson polynomials* $D_n(x,a)$ recursively as

$$D_{n+2}(x,a) = x \cdot D_{n+1}(x,a) - aD_n(x,a); \quad D_0(x,a) = 2,\ D_1(x,a) = x.$$

Instead of $D_n(x,1)$ we sometimes simply write $D_n$.

**Note.** The classical Chebyshev polynomials $t_n$, defined by $\cos nx = t_n(\cos x)$, are conjugate to our Dickson polynomials by $t_n(x) = \frac{1}{2}D_n(2x,1)$. One advantage of the usage of Dickson polynomials instead of Chebyshev ones is that they are monic. Using the additional parameter we sometimes can avoid extensions of the constant field. Confer the next remark and the discussion at the end of the first section.

[LMT93] contains a detailed treatment of such polynomials. For convenience we mention some well-known and easy to establish properties.

**2.2. Proposition.** *The Dickson polynomials satisfy*

(i) $D_n(\lambda x, \lambda^2) = \lambda^n D_n(x,1)$,

(ii) $D_n(x,a) \circ (x + ax^{-1}) = (x + a^n x^{-1}) \circ x^n$,
(iii) $D_m(x,a^n) \circ D_n(x,a) = D_{nm}(x,a) = D_n(x,a^m) \circ D_m(x,a)$,

*for arbitrary constants $a$ and $\lambda$.* □

**2.3. Remark.** Obviously $D_n(x,0) = x^n$, and part (i) of this proposition in particular says that for $\lambda \neq 0$

$$D_n(x,\lambda^2) \cong D_n.$$

Thus, if $\Bbbk$ is algebraically closed, or at least closed under the square root operation, the extra parameter is superfluous for the characterization of prime bidecompositions. But for the general case it is needed.

**Note.** Using Proposition 2.2 it is easy to prove a well known differential equation for Dickson polynomials

$$(D_n^2 - 4) \cdot n^2 = (x^2 - 4) \cdot {D'_n}^2.$$

Conversely, the Dickson polynomials $D_n$, together with their negatives $-D_n$, constitute all polynomial solutions to this differential equation. This is proved e.g. in [LN73] and, in an even stronger form, in [Sch82]. The idea in the latter reference is used in the proof of the next lemma, which will be enough for our purposes.

**2.4. Lemma.** *Let $K$ be any field not of characteristic 2. If a polynomial $f$ of degree $n$ over $K$ satisfies*

$$f - 2\lambda^n = (x - 2\lambda) \cdot g_-^2$$
$$f + 2\lambda^n = (x + 2\lambda) \cdot g_+^2$$

*for some polynomials $g_-, g_+ \in K[x]$ and $\lambda (\neq 0) \in K$, then*

$$f = D_n(x,\lambda^2).$$

*Proof.* Note that $n$ must be odd. Let $a = \lambda^2$. We substitute $x + ax^{-1}$ into the first equation and multiply by $x^n$; thus obtain

$$\begin{aligned}(f(x+ax^{-1}) - 2\lambda^n) \cdot x^n &= (x + ax^{-1} - 2\lambda) \cdot x \cdot g_-^2(x + ax^{-1}) \cdot x^{n-1}\\ &= (x-\lambda)^2 \cdot g_-^2(x+ax^{-1}) \cdot x^{n-1}\\ &= h_-^2\end{aligned}$$

for some polynomial $h_-$, because $[g_-] = \frac{n-1}{2}$. Similarly

$$(f(x+ax^{-1}) + 2\lambda^n) \cdot x^n = h_+^2.$$

Substracting these two equations we get

$$4\lambda^n x^n = h_+^2 - h_-^2 = (h_+ + h_-) \cdot (h_+ - h_-) \qquad (*)$$

But both $h_+$ and $h_-$ have degree $n$. As $\operatorname{char} K \neq 2$, we can choose the signs such that $[h_+ + h_-] = n$. But then $[h_+ - h_-] = 0$, thus $h_+ - h_- = c$ for some constant $c$. We substitute $\lambda$ for $x$ into equation $(*)$ to obtain

$$4a^n = (2h_-(\lambda) + c) \cdot c.$$

Using $h_-(\lambda) = 0$ we see $c^2 = 4a^n$, thus can assume $c = 2\lambda^n$. Now equation $(*)$ turns into

$$4\lambda^n x^n = (2h_+ - 2\lambda^n) \cdot 2\lambda^n,$$

from which it follows that $h_+ = x^n + \lambda^n$ and consequently $h_- = x^n - \lambda^n$. Therefore

$$\begin{aligned} f(x + ax^{-1}) + 2\lambda^n &= x^{-n} \cdot (x^n + \lambda^n)^2 \\ &= x^n + 2\lambda^n + a^n x^{-n}, \end{aligned}$$

thus

$$f(x + ax^{-1}) = x^n + ax^{-n},$$

which is the characteristic equation for a Dickson polynomial (2.2). □

The assumption in 2.4 was rather special. Using linear transformations we can make it more general.

**2.5. Corollary.** *Let $K \geq \Bbbk$ be an extension field of $k$. If a polynomial $f$ over $\Bbbk$ satisfies*

$$\begin{aligned} f - e_1 &= (x - \xi_1) \cdot g_1^2 \\ f - e_2 &= (x - \xi_2) \cdot g_2^2, \end{aligned}$$

*for some constants $\xi_1, \xi_2 \in K$, polynomials $g_1, g_2$ over $K$, and $e_1$, $e_2 \in K$ that are two different solutions of some quadratic equation over $\Bbbk$, then $f \cong D(x, a)$ (even as polynomials over $\Bbbk$) for some $a \in \Bbbk$.*

*Proof.* If $\Bbbk$ is algebraically closed this is rather trivial. The point is to show that no field extensions are necessary.

Being the solution of a quadratic equation, the $e_i$ have the form

$$e_{1,2} = e_0 \pm \lambda$$

for some $e_0 \in \Bbbk$ and $\lambda \in K$ such that $\lambda^2 \in \Bbbk$.

In particular, the $f - e_i$ are polynomials over $\Bbbk[\lambda]$, and so are $x - \xi_i$, as they are factors of a square-free factorization over $\Bbbk[\lambda]$. Thus the $\xi_i$ can be written as $\xi_0 \pm c_i\lambda$, with $\xi_0, c_1, c_2 \in \Bbbk$. But $\xi_1 + \xi_2 \in \Bbbk$, so $c_1 = -c_2$, and we have more precisely

$$\xi_{1,2} = \xi_0 \pm c\lambda$$

for some $\xi_0, c \in \Bbbk$. Let $n = [f]$; after multiplying with $2\lambda^{n-1}$ ($\in \Bbbk$, as $n$ is odd) the equations look like

$$2\lambda^{n-1}(f - e_0) \pm 2\lambda^n = (x - \xi_0 \pm c\lambda) 2\lambda^{n-1} g_{1,2}^2.$$

Thus $\tilde{f} := 2\lambda^{n-1}(x - e_0) \circ f \circ (\frac{c}{2}x + \xi_0) \cong f$ satisfies

$$\tilde{f} \pm 2\lambda^n = (x \pm 2\lambda) \cdot \lambda^{n-1} \big(g_{1,2}(\tfrac{c}{2}x + \xi_0)\big)^2,$$

which is in the form required to use the lemma, thus $\tilde{f} = D_n(x, \lambda^2)$ and $f \cong D_n(x, \lambda^2)$, even over $\Bbbk$. □

## § 3. The Tschirnhaus Transform

**3.1. Definition.** Let $p, q \in \Bbbk[x]$, $q \neq 0$ monic with canonical factorization $\prod_i (x - \xi_i)^{\nu_i}$ over its splitting field. Then the *Tschirnhaus transform* of $q$ by $p$, denoted by ${}^{p}q$ is defined by

$${}^{p}q := \prod_i (x - p(\xi_i))^{\nu_i}.$$

In other words, we obtain the Tschirnhaus by transforming the zeros of $q$ by $p$. As a symmetric function of the zeros of $q$, it is clear that it is a polynomial over $\Bbbk$. In fact, the Tschirnhaus can easily be expressed without any reference to an extension field as a resultant:

**3.2. Proposition.** *For any polynomials $p$, $q$, we have, up to the sign,*

$$^{p}q(y) = \operatorname{res}_x(p(x) - y, q(x)).$$

*Proof.* Let $q = \prod_i (x - \xi_i)^{\nu_i}$ as above. Then by an elementary property of the resultant

$$\operatorname{res}_x(p(x) - y, q(x)) = \prod_i (p(\xi_i) - y)^{\nu_i} = \pm {}^{p}q(y). \qquad \square$$

For bidecompositions the following property turns out to be most useful.

**3.3. Proposition.** *Let $f = r \circ p = s \circ q$ be a prime bidecomposition using monic polynomials; then*

$$^{p}(q - b) = r - s(b).$$

*Proof.* Let $q - b = \prod_i (x - \beta_i)$. Thus $^{p}(q - b) = \prod_i (x - p(\beta_i))$ and $p(\beta_i)$ is also a zero of $r(x) - s(b)$, because $r(p(\beta_i)) = s(q(\beta_i)) = s(b)$.

Assume that $b$ is transcendental. Then all the $\beta_i$ are distinct and transcendental, as $q$ is tame. Suppose $p(\beta_1) = p(\beta_2)$. As $p$ and $q$ have no common right component, Lüroth's Theorem (4.15), provides a rational function $f$ such that $f(p, q) = x$. Now

$$\beta_1 = f(p, q)(\beta_1) = f(p(\beta_1), q(\beta_1)) = f(p(\beta_2), q(\beta_2)) = \beta_2,$$

where the transcendency of the $\beta_i$ guarantees the validity of substitution here. But this means, that $p$ maps the zeros of $q - b$ injectively to the zeros of $r - s(b)$. As $[p] = [q]$, this is even a bijection, and the proof is complete for transcendental $b$.

For arbitrary $b$ we choose some new transcendental element, say $y$. Then $^{p}(q - y) = r - s(y)$. Proposition 3.2 allows us to substitute $b$ for $y$ here, thus providing the full assertion. □

## § 4. Ramification

**4.1. Definition.** Let $f$ be a polynomial. We say that $e$ is a *ramification point* of $f$ iff $f - e$ and $f'$ have a common zero. The degree of $\gcd(f - e, f')$ is called the *(ramification) index* of $f$ at $e$ and is denoted by $\operatorname{ind}_e f$. If $f' \neq 0$ has leading coefficient $c$, we call the Tschirnhaus transform $^{f}(\frac{1}{c} f')$ the *ramification polynomial* of $f$ (this name is justified by the next proposition).

**4.2. Proposition.** *Suppose that the ramification polynomial of $f$ has the canonical factorization over $\Bbbk^{\mathrm{alg}}$*

$$^{f}f' = \prod_i (x - e_i)^{\varepsilon_i},$$

*then the $e_i$ are just the ramification points of $f$ and*

$$\operatorname{ind}_{e_i} f = \varepsilon_i.$$

*Proof.* $\operatorname{ind}_e f$ counts the number of zeros $\xi$ of $f'$, with multiplicities, that fulfill $f(\xi) = e$. But the Tschirnhaus transforms exactly these zeros, together with their multiplicity, into the zero $e$ of the ramification polynomial. □

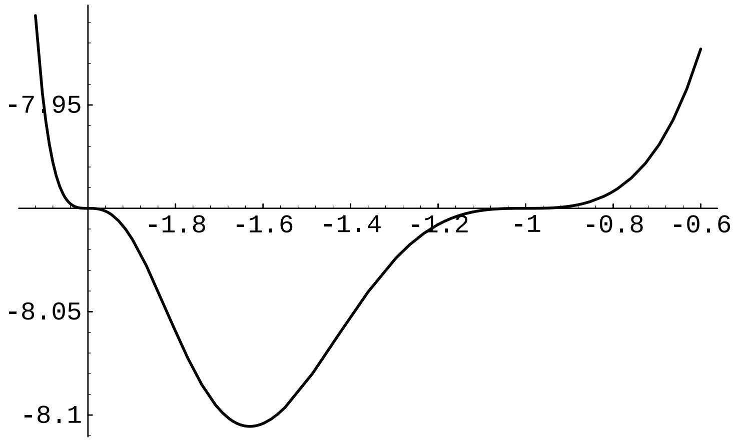
**4.3. Corollary.** *Let $f$ be a tame polynomial. Then*

$$\sum_{e \in \Bbbk^{\mathrm{alg}}} \operatorname{ind}_e f = [f] - 1.$$

*Proof.* Tameness guarantees that $[f'] = [f] - 1$. So this is a trivial consequence of the proposition. □

**4.4. Example.** Let us consider our standard example from chapter I:

$f = x^{12}+12x^{11}+66x^{10}+223x^9+522x^8+900x^7+1179x^6+1188x^5+918x^4+533x^3+222x^2+60x,$



Its ramification polynomial (computed from $\operatorname{res}_x(f - e, f')$)

$${}^{r}r'(e) = (e+8)^8 \cdot (e + \tfrac{2075}{256})^3$$

tells us that $-8$ and $-\frac{2075}{256} \approx -8.10547$ are its ramification points (obvious also from the picture) with indices 8 and 3, respectively (not obvious from the picture). Observe that we can read off the number of ramification points and their indices already from the squarefree factorization of the ramification polynomial.

The next proposition, which will be used frequently, unfortunately needs *completely* tame as hypothesis (Definition 1.5).

**4.5. Proposition.** *Suppose that $f$ is completely tame and $e \in \Bbbk^{\mathrm{alg}}$. If $f - e = \prod(x - a_i)^{\alpha_i}$ is the canonical factorization, then*

$$\operatorname{ind}_e f = \sum_i (\alpha_i - 1).$$

*Proof.* As $f$ was assumed to be completely tame, all $\alpha_i \not\equiv 0 \pmod{\operatorname{char} \Bbbk}$. Thus the multiplicity of $a_i$ in $f'$ is $\alpha_i - 1$, which proves the result. □

**4.6. Example.** We continue with *our* $f$. We have

$$f + 8 = (x+1)^3 \cdot (x+2)^3 \cdot (x^2+x+1)^3.$$

The first two zeros are clear from the graph, the remaining two are complex. All four zeros have multiplicity 3, thus $\operatorname{ind}_{-8} f = (3-1)+(3-1)+(3-1)+(3-1) = 8$, as we have already seen from the ramification polynomial. For the second ramification point we get the factorization

$$f + \tfrac{2075}{256} = (x^3 + 3x^2 + 3x + \tfrac{5}{4})^2 \cdot (x^6 + 6x^5 + 15x^4 + \tfrac{45}{2}x^3 + \tfrac{45}{2}x^2 + \tfrac{27}{2}x + \tfrac{83}{16}),$$

thus it has three double zeros, the remaining being simple. Again we verify that the index at this point is $(2-1)+(2-1)+(2-1)=3$.

**4.7. Remark.** The polynomials with only one ramification point are exactly those which are associated to some power $x^n$.

**4.8. Convention.** For the rest of this section and the following two ones we fix a completely tame prime bidecomposition

$$f = r \circ p = s \circ q$$

with $n = [p] = [s]$ and $m = [q] = [r]$. We assume all these polynomials to be monic. For every point $e \in \Bbbk^{\mathrm{alg}}$ we use the following canonical factorizations over $\Bbbk^{\mathrm{alg}}$

$$\begin{aligned} r - e &= \prod_{i=1}^{\mu} (x - a_i)^{\alpha_i} \\ s - e &= \prod_{j=1}^{\nu} (x - b_j)^{\beta_j}. \end{aligned}$$

Then

$$\begin{aligned} f - e &= \prod_{i=1}^{\mu} (p - a_i)^{\alpha_i} \\ &= \prod_{j=1}^{\nu} (q - b_j)^{\beta_j} \\ &= \prod_{i,j} \prod_{\kappa=1}^{\gamma_{ij}} (x - \xi_{ij\kappa})^{\varepsilon_{ij\kappa}}, \end{aligned}$$

where the $\xi_{ij\kappa}$ should be the zeros of $f - e$ classified according to $p(\xi_{ij\kappa}) = a_i$ and $q(\xi_{ij\kappa}) = b_j$; the $\varepsilon_{ij\kappa}$ denote their multiplicities and the $\gamma_{ij}$ the number of such zeros. Comparing the above factorizations we see that for all $i$ resp. $j$

$$(p - a_i)^{\alpha_i} = \prod_{j} \prod_{\kappa=1}^{\gamma_{ij}} (x - \xi_{ij\kappa})^{\varepsilon_{ij\kappa}} \tag{3}$$

$$(q - b_j)^{\beta_j} = \prod_{i} \prod_{\kappa=1}^{\gamma_{ij}} (x - \xi_{ij\kappa})^{\varepsilon_{ij\kappa}}. \tag{4}$$

All these notions depend on the point $e$. If it is necessary to indicate this dependence, we use upper indices: $a_i^{(e)}$, $\xi_{ij\kappa}^{(e)}$ and so on.

**4.9. Example.** As detected in chapter I, Example 3.14, *our* polynomial $f$ has the bidecomposition

$$\begin{aligned} &f = r \circ p = s \circ q \\ &\quad = (x^4 + 7x^3 + 18x^2 + 20x) \circ (x^3 + 3x^2 + 3x) = (x^3 + 6x^2 + 12x) \circ (x^4 + 4x^3 + 6x^2 + 5x). \end{aligned}$$

Let us compute the factorizations for this example. For the first ramification point, $-8$, we get

$$\begin{aligned} r+8 &= (x+1)\cdot(x+2)^3, \\ s+8 &= (x+2)^3. \end{aligned}$$

Thus $a_1=-1$, $\alpha_1=1$; $a_2=-2$, $\alpha_2=3$; $b_1=-2$, $\beta_1=3$. With these zeros we continue factoring

$$\begin{aligned} p+1 &= (x+1)^3, \\ p+2 &= (x+2)\cdot(x^2+x+1), \\ q+2 &= (x+1)\cdot(x+2)\cdot(x^2+x+1). \end{aligned}$$

Of course, we have got the zeros of $f$ again, now classified according to their values by $r$ and $s$, respectively:

$$\begin{aligned} &\gamma_{11}=1, & \varepsilon_{111}&=3, & \xi_{111}&=-1, \\ &\gamma_{21}=3, & \varepsilon_{211}=\varepsilon_{212}=\varepsilon_{213}&=3, & \xi_{211}&=-2, \end{aligned}$$

and $\xi_{212}$ and $\xi_{213}$ satisfy $x^2+x+1$.

We do the same for the second ramification point

$$\begin{aligned} r+\tfrac{2075}{256} &= (x+\tfrac{5}{4})\cdot(x^2+\tfrac{9}{8}x+\tfrac{83}{256}), \\ s+\tfrac{2075}{256} &= x^3+6x^2+12x+\tfrac{2075}{256}. \end{aligned}$$

Thus $a_1=-\frac{5}{4}$, with $\alpha_1=2$, and $a_2$ and $a_3$ are zeros of $(x^2+\frac{9}{8}x+\frac{83}{256})$, with $\alpha_2=\alpha_3=1$, whereas $b_1$, $b_2$, and $b_3$ all satisfy an irreducible polynomial of degree 3, with $\beta_1=\beta_2=\beta_3=1$. Because

$$p+\tfrac{5}{4}=x^3+3x^2+3x+\tfrac{5}{4},$$

we have got back one of the factors of $f-\frac{2075}{256}$. We continue computing factorizations

$$\begin{aligned} p^2+\tfrac{9}{8}p+\tfrac{83}{256} &= x^6+6x^5+15x^4+\tfrac{45}{2}x^3+\tfrac{45}{2}x^2+\tfrac{27}{2}x+\tfrac{83}{16}, \\ q^3+6q^2+12q+\tfrac{2075}{256} &= (x^3+3x^2+3x+\tfrac{5}{4})^2\cdot(x^6+6x^5+15x^4+\tfrac{45}{2}x^3+\tfrac{45}{2}x^2+\tfrac{27}{2}x+\tfrac{83}{16}), \end{aligned}$$

and obtain the remaining factors of $f-\frac{2075}{256}$.

**4.10. Lemma.** *For all $i,j$ we have*

$$\alpha_i\beta_j=\sum_{\kappa=1}^{\gamma_{ij}}\varepsilon_{ij\kappa}.$$

*In particular $\varepsilon_{ij\kappa}\le\alpha_i\beta_j$ for all $i,j,\kappa$.*

*Proof.* Using the Tschirnhaus transform we get for each $j$

$$\begin{aligned} {}^{p}(q-b_j)^{\beta_j} &= \prod_i\prod_{\kappa=1}^{\gamma_{ij}}{}^{p}(x-\xi_{ij\kappa})^{\varepsilon_{ij\kappa}} \\ &= \prod_i\prod_{\kappa=1}^{\gamma_{ij}}(x-p(\xi_{ij\kappa}))^{\varepsilon_{ij\kappa}}. \\ &= \prod_i\prod_{\kappa=1}^{\gamma_{ij}}(x-a_i)^{\varepsilon_{ij\kappa}}=\prod_i(x-a_i)^{\sum_\kappa\varepsilon_{ij\kappa}}. \end{aligned}$$

But on the other hand, using Proposition 3.3,

$$^{p}(q-b_j)^{\beta_j}=(r-s(b_j))^{\beta_j}=(r-e)^{\beta_j}=\prod_i(x-a_i)^{\alpha_i\beta_j},$$

and this canonical factorization must coincide with that obtained before. □

**4.11. Remark.** Remember the symbolism for the lattice $(\mathbb{N},\cup,\cap)$. We will frequently use the following simple properties:

$$n\subseteq m\implies n\le m,$$
$$n\subset m\implies n\le\frac{m}{2},$$

valid for all $n,m\in\mathbb{N}$.

**4.12. Lemma.** *For all $i,j,\kappa$ we have*

$$\varepsilon_{ij\kappa}\ge\alpha_i\cup\beta_j$$
$$\gamma_{ij}\le\alpha_i\cap\beta_j.$$

*Proof.* From the factorization (3) we see that $\varepsilon_{ij\kappa}\supseteq\alpha_i$. Similarly $\varepsilon_{ij\kappa}\supseteq\beta_j$. Thus $\varepsilon_{ij\kappa}\supseteq\alpha_i\cup\beta_j$ and the first inequality is clear. From this, together with Lemma 4.10,

$$\alpha_i\beta_j=\sum_{\kappa=1}^{\gamma_{ij}}\varepsilon_{ij\kappa}\ge\gamma_{ij}(\alpha_i\cup\beta_j).$$

We divide by $\alpha_i\cup\beta_j$ and obtain the second inequality. □

**4.13. Lemma.** *For all $i$ we have*

$$\operatorname{ind}_{a_i}p=\sum_j(\beta_j-\gamma_{ij})\ge\sum_j(\beta_j-\alpha_i\cap\beta_j).$$

*Proof.* Using Proposition 4.5 we get

$$\operatorname{ind}_{a_i}p=\sum_j\sum_{\kappa=1}^{\gamma_{ij}}\left(\frac{\varepsilon_{ij\kappa}}{\alpha_i}-1\right)=\sum_j\left(\sum_{\kappa=1}^{\gamma_{ij}}\frac{\varepsilon_{ij\kappa}}{\alpha_i}-\sum_{\kappa=1}^{\gamma_{ij}}1\right)=\sum_j(\beta_j-\gamma_{ij}).$$

The inequality then follows from the previous lemma. □

## § 5. Exponential Solutions

The following result now has got a direct and considerably shorter proof.

**5.1. Proposition.** *If $s$ has only one ramification point, then our bidecomposition is exponential.*

*Proof.* Let $e$ be the unique ramification point. Then $e\in\Bbbk$, and in our factorizations

$$\nu=1,\quad\beta_1=n,$$

where $n$ must be prime by the primality of $s$. Hence some $\alpha_i$ is relatively prime to $n$ as $p$ is prime. Thus let us assume $n\cap\alpha_1=1$. Now from Lemma 4.13

$$n-1\ge\operatorname{ind}_{a_1}p\ge\sum_{j=1}^{\nu}(\beta_j-\alpha_1\cap\beta_j)=n-\alpha_1\cap n=n-1.$$

Thus $a_1$ is the unique ramification point of $p$, and as such is in $\Bbbk$. For $i \neq 1$ we have

$$0 = \operatorname{ind}_{a_i} p \geq n - \alpha_i \cap n,$$

hence $\alpha_i \supseteq n$. So $r$ has the form

$$r - e = (x - a_1)^{\alpha_1} \cdot t^n$$

for some polynomial $t$. $a_1$ and the coefficients of $t$ are elements of $\Bbbk$ because they can be computed from the squarefree factorization. The form of $q$ is determined by the other three polynomials. □

**5.2. Example.** The bidecomposition of the examples in the previous section is exponential, because $s$ has the single ramification point $-8$, as $s + \frac{2075}{256}$ is squarefree, i.e., it splits into linear factors over $\Bbbk^{\mathrm{alg}}$. In fact, it is verified immediately that all polynomials have the specified forms.

Because the results in this section are symmetric in the sense that we can interchange the rôles of the two decompositions $r \circ p$ and $s \circ q$, we can summarize

**5.3. Proposition.** *If at least one of the two polynomials $r$ and $s$ has only one ramification point, then our bidecomposition is exponential.* □

## § 6. Trigonometric Solutions

The next proposition is very important for our simplifications. First we need a technical lemma.

**6.1. Lemma.** *Suppose that the $\alpha_i \in \mathbb{N}$ have no common divisor, i.e., $\bigcap_i \alpha_i = 1$. Then for all $\beta \in \mathbb{N}$*

$$\sum_i (\beta - \alpha_i \cap \beta) \geq \beta - 1.$$

*Proof.* Suppose that $\alpha_i$ is not a multiple of $\beta$. Then $\alpha_i \cap \beta \subset \beta$, thus $\leq \frac{\beta}{2}$, and the $i$-th summand is $\geq \frac{\beta}{2}$. If there are two such summands, then they sum up to $\beta$ and the lemma is proved. Thus consider the case that $\alpha_i \supseteq \beta$ for all but at most one $i$. Take $i = 1$ for the possible exception. Then

$$1 = \bigcap_i \alpha_i = \alpha_1 \cap \bigcap_{i \neq 1} \alpha_i \supseteq \alpha_1 \cap \beta,$$

thus $\alpha_1 \cap \beta = 1$, and we just have to look at the first summand $\beta - \alpha_1 \cap \beta = \beta - 1$ to prove the lemma also in this case. □

**6.2. Proposition.** *If $r$ has at least two ramification points, then*

$$\sum_i \operatorname{ind}_{a_i} p = \operatorname{ind}_e s.$$

*Proof.* Because $r - e$ is not associated to a power, but prime, $\bigcap \alpha_i = 1$. Thus we can apply the lemma for all $\beta_j$:

$$\sum_i (\beta_j - \alpha_i \cap \beta_j) \geq \beta_j - 1.$$

Now we take the sum over all $j$ and, together with Lemma 4.13, obtain the estimation

$$\sum_i \operatorname{ind}_{a_i} p \geq \sum_i \sum_j (\beta_j - \alpha_i \cap \beta_j) \geq \sum_j (\beta_j - 1) = \operatorname{ind}_e s,$$

thus proving the $\geq$-part.

To see equality we consider the factorizations of Convention 4.8 for various $e$'s. Note that $r - e_1$ and $r - e_2$ have no common zero whenever $e_1 \neq e_2$, thus all the elements $\alpha_i^{(e)}$ are distinct, so from summing up over all $e \in \mathbb{k}$ we get

$$m - 1 = \sum_e \operatorname{ind}_e p = \sum_e \sum_i \operatorname{ind}_{a_i^{(e)}} p \geq \sum_e \operatorname{ind}_e s = m - 1,$$

hence the $\geq$ here is an equality, and, by the part just proved, all summands are even equal. □

**6.3. Example.** Again we illustrate this with our bidecomposition from the previous sections. Let us check it for the ramification point $-8$: From the factorizations we obtain $\operatorname{ind}_{-8} s = 2$, $\operatorname{ind}_{a_1} p = 0$, $\operatorname{ind}_{a_2} p = 2$. As $2 = 0 + 2$ this is in accordance with the proposition. One can also check this for the second ramification point $e := -\frac{2075}{256}$, and gets the same observation, as must be the case, because $r$ has two ramification points. But $s$ has only one ramification point, as $s - e$ is squarefree, so the symmetric property

$$\sum_j \operatorname{ind}_{b_j} q = \operatorname{ind}_e r$$

might be false in this case. In fact, $\operatorname{ind}_e r = 1$, but $\operatorname{ind}_{a_i} = 0$ for all $i$, and $1 \neq 0 + 0 + 0$.

**6.4. Lemma.** *If $r$ has two ramification points and $r - e$ contains a simple zero, say $a_1$ (i.e. $\alpha_1 = 1$), then*

$$\alpha_i \supseteq \bigcup_j \beta_j, \qquad \textit{for all } i \neq 1.$$

*Proof.* By Proposition 6.2 together with Lemma 4.13

$$\sum_j (\beta_j - 1) = \operatorname{ind}_e s = \sum_i \operatorname{ind}_{a_i} p \geq \sum_i \sum_j (\beta_j - \alpha_i \cap \beta_j),$$

and using $\alpha_1 \cap \beta_j = 1$,

$$= \sum_j (\beta_j - 1) + \sum_{i \neq 1} \sum_j (\beta_j - \alpha_i \cap \beta_j)$$

Thus for $i \neq 1$ and all $j$ we have $\alpha_i \supseteq \beta_j$. □

**6.5. Remark.** If $r - e$ has no simple zero, then all its zeros are at least double, hence their number is at most $\frac{m}{2}$, so $\operatorname{ind}_e r \geq \frac{m}{2}$. By Proposition 4.3, this cannot happen twice.

**6.6. Proposition.** *If both $r$ and $s$ have at least two ramification points, then they have exactly two (common) ones. Let $e$ be one of them. Then both $r - e$ and $s - e$ have exactly one simple zero, the remaining ones being double.*

*Proof.* Suppose $e$ is a ramification point of $s$ such that $r-e$ has a simple zero, say $a_1$, thus $\alpha_1 = 1$. By the lemma, all the remaining $\alpha_i$ are multiples of all the $\beta_j$. But some $b_j > 1$, thus, in particular, $a_i \geq 2$ for all $i \neq 1$. Hence $e$ is also a ramification point of $r$ and $\operatorname{ind}_e r \geq \frac{m-1}{2}$ because $\mu \leq \frac{m-1}{2}$. If $e'$ is another ramification point of $r$, then its index is bounded by $\frac{m-1}{2}$, so $r-e'$ has a simple zero, too, and the whole story is equally true for this second ramification point. Thus $r$ has exactly the two ramification points $e$ and $e'$, both with index $\frac{m-1}{2}$, hence $\mu = \frac{m+1}{2}$. $r-e$ has one simple zero, the multiplicities of the remaining $\frac{m-1}{2}$ ones sum up to $m-1$, thus are double. The same is true for $e'$ and, by symmetry, for the ramification points of $s$. □

**6.7. Remark.** This means that $\alpha_1 = \beta_1 = 1$ and $\alpha_i = \beta_i = 2$ for all $i \neq 1$, for both ramification points. Thus $\gamma_{11} = 1$, $\varepsilon_{111} = 1$, and $\varepsilon_{ij\kappa} \supseteq 2$, if not $i = j = 1$. In particular, if $e_1, e_2$ are the two ramification points, then

$$f - e_1 = (x - \xi_1) \cdot g_1^2,$$
$$f - e_2 = (x - \xi_2) \cdot g_2^2$$

for some polynomials $g_1, g_2$. Because $f$ has exactly two ramification points, $e_1$ and $e_2$ satisfy a quadratic equation over $\Bbbk$ (4.2). So we can apply Corollary 2.5, and obtain:

**6.8. Corollary.** *If both $r$ and $s$ contain two ramification points, then our bidecomposition is trigonometric.* □

## § 7. Final Remarks

Now the proof of Ritt's bidecomposition theorem is complete. Let us outline where simplifications have been made, and which further improvements seem to be possible.

Previous proofs assume that the ground field $\Bbbk$ is algebraically closed. In [Sch82] the theorem for general fields is obtained as a corollary to that for algebraically closed ones. Our version proves the general form directly. There are only few points where we must take care of this, mainly in Corollary 2.5, whose nontrivial part says that the linear transformations can be chosen in the ground field.

That we use the Tschirnhaus transform instead of the norm as the previous proofs that avoid valuation theoretic or analytic methods is mainly a matter of taste. Note that ${}^{p}q \circ p = \pm N_{\Bbbk(x):\Bbbk(p)}(q)$. The usage of resultants is new in this context and may supply further improvements, when used more extensively. Our proof of Proposition 3.3 serves as an alternative to the usage of norms and minimal polynomials; it seems to be more direct.

The section on ramification contains results mixed from the previous proofs. Lemma 4.12 has got an elementary proof. [DW74] even prove equality for this statement, using valuation theoretic methods. This stronger form can also be obtained as a corollary to the characterization Theorem 1.7.

Our major simplifications are contained in sections 5 and 6. There is no discussion of extra points any more. We just make the distinction on the number of ramification points and rather quickly see, by analyzing the ramification structure, that we have the exponential or trigonometric case, respectively.

These improvements essentially use that the components of prime bidecompositions are prime. Thus they do not generalize as in [Sch82], partially characterizing bidecompositions that need not be prime. This raises the question, whether Ritt's Theorems (5.11 in chapter I and 1.7) can be used to give an even more explicit description of all possible decompositions. In particular, we may ask whether there is a canonical decomposition.

The decompositions of polynomials associated to $x^n$ may be considered to be trivial as they simply correspond to the divisors of $n$. The same is true with Dickson polynomials. This suggests that a canonical decomposition could look like this: a composition of polynomials that are either of exponential or of trigonometric type, or do not contain any of these.

As another further improvement it might be possible to extensively use the resultant calculus and square-free factorizations instead of the involved analysis of the zeros and their multiplicities in sections 4 to 6.

The assumption about char 2 in the theorem was necessary because Proposition 2.4 uses it, which in turn is needed in 6.8. It is not clear whether we get any additional bidecompositions in case of characteristic 2.

The restriction to completely tame polynomials was necessary in proving 4.5, which is basic for all results about the index. It is not known how far this can be weakened, e.g. to tame polynomials. No counterexample for this is known. The example in the note of [Cor90] does not work for this purpose as it is of exponential type.

A whole class of counterexamples using non-tame polynomials is given by

$$x^\chi \circ f = f \circ x^\chi,$$

where $\chi := \operatorname{char} \mathbb{k}$. Perhaps all prime bidecompositions can be reduced to a trigonometric or exponential form using this ambiguity somehow.

# Lebenslauf

*Sonntag, den 13. Jänner 1963, wurde ich als dritter und jüngster Sohn des Bergbauernehepaares Matthäus und Katharina Binder in Bad Ischl geboren. Dort besuchte ich die Volksschule und anschließend das Bundesrealgymnasium, an welchem ich im Juni 1981 die Reifeprüfung mit gutem Erfolg ablegte.*

*Im Oktober 1981 begann ich an der Johannes Kepler Universität in Linz mit dem Studium der Informatik, im darauffolgenden Semester zusätzlich der Technischen Mathematik, welches daraufhin allmählich zu meiner Hauptstudienrichtung wurde. Den ersten Studienabschnitt für den Studienzweig Informations- und Datenverarbeitung schloß ich im Dezember 1985 mit ausgezeichnetem Erfolg ab.*

*Den Präsenzdienst leistete ich von Oktober 1990 bis Mai 1991. In den Sommermonaten arbeitete ich zeitweise als Programmierer, wobei ich vor allem mit Compilerbau- und Datenbankproblemen befaßt war, aber auch mit zahlreichen verschiedenen Standard-Softwareprodukten.*

*Weiters war ich zeitweise als Universitätsinstruktor, Projektmitarbeiter, Studienassistent sowie Leiter von Tutorien am Institut für Mathematik beschäftigt.*

*Mein mathematisches Hauptinteresse gilt der Erforschung konstruktiver Methoden in der Algebra.*

# Bibliography

[Alo94] Cesar L. Alonso González, *Desarrollo, análisis e implementación de algoritmos para la manipulación de variedades paramétricas*, Ph.D. thesis, Universidad de Cantabria, Santander, 1994.

[AGR] T. Recio C. Alonso, J. Gutierrez, *A rational function decomposition algorithm by near-separated polynomials*, Journal of Symbolic Computation **to appear**.

[BK78] R. P. Brent and H. T. Kung, *Fast algorithms for manipulating formal power series*, Journal of the Association for Computing Machinery **25** (1978), no. 4, 581–595.

[BL82] B. Buchberger and R. Loos, *Algebraic simplification*, Computer Algebra (G. E. Collins B. Buchberger and R. Loos, eds.), Springer-Verlag, 1982, pp. 11–43.

[Coh77] P. M. Cohn, *Algebra*, vol. II, John Wiley & Sons, London, 1977.

[Cor90] Capi Corrales-Rodrigáñez, *A note on ritt's theorem on decomposition of polynomials*, Journal of Pure and Applied Algebra **95** (1990), 293–296.

[DW74] F. Dorey and G. Whaples, *Prime and composite polynomials*, Journal of Algebra **28** (1974), 88–101.

[Eng41] H. T. Engström, *Polynomial substitutions*, American Journal of Mathematics **63** (1941), 249–255.

[Gut88] Jaime Gutierrez, *Algunos aspectos de la teoría de casi-anilos de polinomios*, Ph.D. thesis, Universidad de Cantabria, Santander, 1988.

[KL89] D. Kozen and S. Landau, *Polynomial decomposition algorithms*, Journal of Symbolic Computation **7** (1989), 445–456.

[Lev42] H. Levi, *Composite polynomials with coefficients in an arbitrary field of characteristic zero*, American Journal of Mathematics **23** (1942), 51–66.

[LMT93] R. Lidl, G. L. Mullen, and G. Turnwald, *Dickson Polynomials*, Pitman Monographs and Surveys in Pure and Applied Mathematics, vol. 65, Longman Scientific & Technical, London, 1993.

[LN73] Hans Lausch and Wilfried Nöbauer, *Algebra of Polynomials*, North-Holland Mathematical Library, vol. 5, North Holland, Amsterdam, 1973.

[Pil83] Günter F. Pilz, *Near-Rings*, 2 ed., North-Holland, Amsterdam, 1983.

[Rit22] J. F. Ritt, *Prime and composite polynomials*, Transactions of the American Mathematical Society **23** (1922), 51–66.

[Sch82] A. Schinzel, *Seclected Topics on Polynomials*, Ann Arbor, University of Michigan press, 1982.

[SS71] A. Schönhage and V. Strassen, *Schnelle Multiplikation großer Zahlen*, Computing **7** (1971), 281–292.

[vdW66] B. L. van der Waerden, *Algebra*, 7 ed., vol. I, Springer-Verlag, Berlin Heidelberg New York, 1966.

[vdW67] B. L. van der Waerden, *Algebra*, 5 ed., vol. II, Springer-Verlag, Berlin Heidelberg New York, 1967.

[vzG90] Joachim von zur Gathen, *Functional decomposition of polynomials: the tame case*, Journal of Symbolic Computation **9** (1990), 281–299.

[Zip91] R. Zippel, *Rational function decomposition*, Proc. of ISSAC-91, ACM press (1991).

# Index