Raoul Blankertz (2011). *Zerlegung von Polynomen (Decomposition of Polynomials)*. Diplomarbeit, Universität Bonn. Modified version available at http://arxiv.org/abs/1107.0687.

DECOMPOSITION OF POLYNOMIALS

Raoul Blankertz

July 4, 2011

Abstract. This diploma thesis is concerned with functional decomposition $f = g \circ h$ of polynomials. First an algorithm is described which computes decompositions in polynomial time. This algorithm was originally proposed by Zippel (1991). A bound for the number of minimal collisions is derived. Finally a proof of a conjecture in von zur Gathen, Giesbrecht & Ziegler (2010) is given, which states a classification for a special class of decomposable polynomials.

Note. *This is a modified version of the author's diploma thesis. The main changes concern notation and rephrasing of some results.*

Danksagung

Ich möchte mich an dieser Stelle bei den Personen bedanken, die mir bei der Erarbeitung dieser Diplomarbeit geholfen haben.

Mein Dank gilt Herrn Prof. Dr. Joachim von zur Gathen für die Betreuung dieser Diplomarbeit. Besonders zu schätzen weiß ich seine kritischen Anmerkungen und die anregenden Gespräche mit ihm. Außerdem bedanke ich mich bei Herrn Konstantin Ziegler für seine Betreuung und für die zahlreichen fachlichen Diskussionen, Ratschläge und Hilfestellungen. Bei Herrn Prof. Dr. Nitin Saxena bedanke ich mich für die Übernahme der Zweitkorrektur.

Für die sprachliche Durchsicht bedanke ich mich bei David Steimle und Jens Humrich.

Nicht zuletzt möchte ich mich bei meiner Familie bedanken. Bei meinen Eltern und Großeltern bedanke ich mich für die Unterstützung während des Studiums. Ganz besonders bedanke ich mich bei meiner Freundin Ellen für ihre Geduld und ihren Beistand.

Contents

1. Introduction

1.1. Deutsche Einleitung. Die Komposition zweier Polynome ist ebenfalls ein Polynom. Umgekehrt kann man sich fragen, unter welchen Umständen ein Polynom die Komposition zweier anderer Polynome ist – oder in anderen Worten: Wann ist ein Polynom funktional zerlegbar? Ritt (1922) beschäftigte sich mit dieser Frage, wobei er als Grundkörper $\mathbb{C}$ voraussetzte. Im Gegensatz zur multiplikativen Zerlegung ist die funktionale Zerlegung nicht eindeutig. Ritts erstes Theorem besagt, dass die Grade der Komponenten zweier verschiedener vollständigen Zerlegungen (das heißt die Komponenten sind unzerlegbar) modulo einer Permutation gleich sind. Ritts zweites Theorem gibt eine Klassifikation der Lösungen von $g_1 \circ h_1 = g_2 \circ h_2$ mit $\deg(g_1) = \deg(h_2)$ an. Diese beiden Theoreme konnten auf Körper der Charakteristik Null verallgemeinert werden (siehe Dorey & Whaples (1974)). Allerdings gibt es Gegenbeispiele für beide Theoreme über Körper mit positiver Charakteristik. In solchen Körpern treten so genannte gleichgradige Kollisionen auf. Diese Kollisionen machen es schwer, die Anzahl der zerlegbaren Polynome über einem endlichen Körper anzugeben oder auch nur zu schätzen, siehe von zur Gathen (2009). Auch algorithmisch ist das Problem im Falle positiver Charakteristik schwieriger, siehe von zur Gathen (1990a) und von zur Gathen (1990b).

Zuerst werden in Kapitel 2 grundlegende Konzepte eingeführt. In Kapitel 3 wird dann ein Algorithmus von Zippel (1991) besprochen. Der von Zippel vorgestellte Algorithmus berechnet Zerlegungen von rationalen Funktionen in polynomieller Zeit. Dabei zitiert Zippel einige Resultate aus Landau & Miller (1985), auf denen sein Algorithmus basiert. Diese Resultate wurden allerdings dort nur für den Grundkörper $\mathbb{Q}$ bewiesen und nicht wie benötigt über einem Funktionenkörper $F(t)$ von beliebiger Charakteristik. Eine komplette Beschreibung dieses Algorithmus für die Zerlegung von Polynomen und ein Beweis seiner Korrektheit wird in dieser Arbeit gegeben. Dabei wird zunächst eine Beziehung zwischen der funktionalen Zerlegung eines Polynoms und den Blöcken einer bestimmten Permutationsgruppe hergestellt. Dann wird gezeigt wie minimale Blöcke effizient berechnet werden können und wie man daraus die entsprechende Zerlegung gewinnt. Am Ende von Kapitel 3 wird eine obere Schranke für die Anzahl von minimalen Zerlegungen eines Polynoms hergeleitet.

Danach widmet sich Kapitel 4 der Klassifikation von Polynomen vom Grad p^2 mit mindestens zwei verschiedenen Zerlegungen über einem Körper der Charakteristik p. Diese Klassifikation wurde in von zur Gathen, Giesbrecht & Ziegler (2010) vorgeschlagen und soll nun in dieser Arbeit bewiesen werden. Der Beweis orientiert sich an den Beweisen von Ritts zweitem Theorem in den

Arbeiten von Dorey & Whaples (1974) und Zannier (1993).

1.2. English introduction. The functional composition of two polynomials is a polynomial itself. Conversely one could ask, when is a given polynomial the composition of two others polynomials – or in other words: When is a polynomial functionally decomposable?

At first in Section 2 basic notions and concepts will be introduced. An algorithm for computing decompositions, which was originally proposed in Zippel (1991) [1], will be discussed in Section 3. The proof relies on a generalization of results of Landau & Miller (1985). But this generalization lacked a proper foundation. A proof of correctness and a runtime estimation is provided (20 years later) in this paper. In the end of Section 3 an upper bound for the number of minimal decompositions of a polynomial will be deduced.

In von zur Gathen, Giesbrecht & Ziegler (2010) a classification for decomposable polynomials of degree p^2 over a field of characteristic p was proposed. This conjecture will be stated and proven in Section 4.

2. Decompositions

Let F be an arbitrary field. In the runtime considerations of the algorithm in Section 3 we restrict F to a field in which one can compute efficiently and in Section 4 we restrict F to a field of positive characteristic. One can think of F being a finite field, which is the most interesting case.

Definition 2.1. *A polynomial f in $F[x]$ is* decomposable *if there are g and h in $F[x]$, both of degrees at least two, such that $f = g \circ h$. The pair (g, h) is called a* decomposition *of f. A polynomial is* indecomposable *if it is not decomposable.*

We call a polynomial original *if its graph passes though the origin, or, equivalently, it has a root at zero. A polynomial is* normal *if it is monic and original. We call a decomposition (g, h)* normal *if g and h are normal and we call it* minimal *if h is normal and indecomposable.*

In a decomposition (g, h), g is uniquely determined by f and h, since the ring homomorphism $F[x] \to F[x]$ with $g \mapsto g \circ h$ is injective. Furthermore, g is easy to compute by the generalized Taylor expansion, see von zur Gathen & Gerhard (1999).

[1]There is a related (unpublished) paper by Zippel from 1996, which was not known to the author until the submission of the thesis. A subsequent publication of the author's work will refer to it.

Definition 2.2. *A linear left composition (linear right composition, linear composition) of f is the polynomial $\ell \circ f$ ($f \circ \ell$, $\ell \circ f \circ \hat{\ell}$, respectively) for some linear polynomials ℓ and $\hat{\ell}$.*

The conjugate *of a normal polynomial f by a linear polynomial $x + w$ is the normal polynomial $(x - f(w)) \circ f \circ (x + w)$.*

Each non-constant polynomial has a unique linear left composition which is normal. Namely, if a is the leading coefficient of f and b is its constant term, then $(a^{-1}x - a^{-1}b) \circ f$ is normal. The functional inverse of a linear polynomial $\ell = ax + b$ is $\ell^{-1} = a^{-1}x - a^{-1}b$. If a normal polynomial f has a decomposition (g, h) and l is a linear polynomial such that $\ell \circ h$ is normal, then $g \circ \ell^{-1}$ is normal. This is because the leading coefficient and constant term of $(g \circ \ell^{-1}) \circ (\ell \circ h) = f$ equal the leading coefficient and the constant term, respectively, of $g \circ \ell^{-1}$.

Functional decomposition is related to intermediate fields of certain field extensions in the following way. Let $F(t)$ be the function field in t over F. Then for a given non-constant polynomial $f \in F[x]$ let φ be the polynomial $f - t$ in $F(t)[x]$. Then φ is irreducible by the Eisenstein criterion.

If we assume that the derivative f' of f is not zero then the derivative of φ with respect to x is not zero and thus φ is separable. In this case, for a root α of φ, $F(t)[\alpha] = F(\alpha)$ is a separable field extension of $F(t)$.

In characteristic 0 we have $f' \neq 0$. If the characteristic of F is p and $f' = 0$ then there exists $\tilde{f}$ such that $f = \tilde{f}(x^{p^r})$ and $\tilde{f}' \neq 0$. If F is finite or is the algebraic closure of a finite field then the Frobenius endomorphism $x \mapsto x^p$ is an automorphism of F. In this case, by knowing all decompositions of $\tilde{f}$ one knows all decompositions of f. In general the Frobenius endomorphism is not an automorphism (for example on function fields), but we will anyway assume that $f' \neq 0$. This assumption excludes some cases in general, but we lose no generality if F is a finite field.

Now the following theorem states a correspondence between decompositions of f and intermediate fields of $F(\alpha) \mid F(t)$. A proof of it can be found in Fried & MacRae (1969).

Theorem 2.3. *Let f be a polynomial over F with $f' \neq 0$ and let α be a root of $f - t \in F(t)[x]$. Let $L = \{h \in F[x] \colon h \text{ is normal and } \exists g \in F[x] \colon f = g \circ h\}$ and let M be the set of intermediate fields between $F(\alpha)$ and $F(t)$. Then the map $L \to M$ with $h \mapsto F(h(\alpha))$ is bijective.*

The minimal polynomial of α over $F(h(\alpha))$ is $h(x) - h(\alpha)$. Thus we have $[F(\alpha) \colon F(h(\alpha))] = \deg(h)$.

The set M is a lattice with the inclusion as order (the intersection and the composition of fields are the meet and the join). It is clear that if $h = g \circ h^*$, then $F(h(\alpha)) \subseteq F(h^*(\alpha))$. Thus, if we take $h^* \leq h$ to mean that $h = g \circ h^*$ for some $g \in F[x]$, then the bijection in Theorem 2.3 is an order-reversing bijection of partially ordered sets. Thus $(L, \leq)$ is a lattice, which we call the *lattice of decompositions* of f.

Lemma 2.4. *Let $f \in F[x]$ with $f' \neq 0$ and ℓ_1, $\ell_2 \in F[x]$ be linear. Let α be a root of $f - t$ and β a root of $\ell_1 \circ f \circ \ell_2 - t$. Then the field extensions $F(\alpha) \mid F(t)$ and $F(\beta) \mid F(t)$ are isomorphic.*

Proof. Let $\Phi\colon F[\alpha] \to F[\beta]$ with $\alpha \mapsto \ell_2(\beta)$ be the evaluation homomorphism. Since β is transcendental over F this homomorphism is injective and extends to a field homomorphism $\Phi\colon F(\alpha) \to F(\beta)$. From $\Phi(t) = \Phi(f(\alpha)) = f(\ell_2(\beta)) = \ell_1^{-1}(t)$ follows that $F(t)$ is mapped to $F(t)$ under Φ. The degrees of the extensions are equal, hence Φ is surjective. □

Corollary 2.5. *If $\hat{f}$ is a linear composition of f, then the lattice of decompositions of $\hat{f}$ is isomorphic to the lattice of decompositions of f.*

Proof. Let α and β be roots of $f - t$ and $\hat{f} - t$, respectively. Then the lattice of decompositions of $\hat{f}$ is isomorphic to the lattice of intermediate fields between $F(\beta)$ and $F(t)$. By the previous lemma, this lattice is isomorphic to the lattice of intermediate fields between $F(\alpha)$ and $F(t)$, which is in turn isomorphic to the lattice of decompositions of f. □

Thus, one needs only to consider normal polynomials and normal decompositions. Furthermore the lattice of decompositions of a normal polynomial is invariant under conjugation.

3. Finding minimal decompositions

An algorithm that computes functional decompositions of rational functions was proposed in Zippel (1991). In this paper Zippel cites results of Landau & Miller (1985), on which the algorithm relies. But these results were only proven for the ground field $\mathbb{Q}$ – instead of $F(t)$, which would be needed. In the 20 years since then, nobody seems to have undertaken the somewhat ungrateful task of verifying whether Zippel's claims are actually true. A complete description of the algorithm for polynomial decomposition and a proof of its correctness will be given in this section.

The main idea for the algorithm is to relate decompositions of f to certain partitions of the set of roots of $\varphi = f - t$ and to find an efficient way to compute these partitions. To specify this idea, let (g, h) be a decomposition of f. For each root λ of $g - t$ the roots of $h - \lambda$ form a subset of the roots of φ. Furthermore two different roots of $g - t$ yield two disjoint subsets. In this way one can partition the set of roots of φ with respect to a decomposition of f. For getting a better understanding of the nature of these partitions we consider the notion of blocks.

3.1. Blocks of imprimitivity. We introduce the notion of blocks of imprimitivity and its relation to decompositions. For this propose consider a finite permutation group G on a finite set Z (that is a subgroup $G \subseteq S(Z)$, where $S(Z)$ is the symmetric group on Z). The following facts are mainly taken from Wielandt (1964).

DEFINITION 3.1. *A subset B of Z is a* block *of G if for all σ in G, the set $\sigma(B) \cap B$ is empty or equals B.*

Equivalently, B is a block of G if for all σ in G the sets B and $\sigma(B)$ are disjoint or equal. If B is a block, then any $\sigma(B)$ is a block. If G is transitive and $B \neq \emptyset$ then $\{\sigma(B)\}_{\sigma \in G}$ is a partition of Z and is called a *complete block system.*

DEFINITION 3.2. *For a subgroup $U \subseteq G$ and $\alpha \in Z$ the* orbit *of α under U is the subset $U(\alpha) = \{\sigma(\alpha) \mid \sigma \in U\}$. For a subset $S \subseteq Z$ the* stabilizer *of S is the subgroup $G_S = \{\sigma \mid \sigma(S) = S\}$. We write G_α for $G_{\{\alpha\}}$. A permutation group G on Z is called* regular *if G_α is trivial for all α in Z.*

For $\sigma \in G$ we have $\sigma G_\alpha \sigma^{-1} = G_{\sigma(\alpha)}$. In particular, if G is transitive, all stabilizers have the same cardinality, and G is regular if and only if G_α is trivial for some $\alpha \in Z$.

LEMMA 3.3. *If B and C are blocks then $B \cap C$ is a block.*

PROOF. Let σ be in G. Then $\sigma(B \cap C) \cap (B \cap C) = (\sigma B \cap B) \cap (\sigma C \cap C)$ and this is empty if and only if $\sigma B \cap B$ or $\sigma C \cap C$ is empty. If both are nonempty we get $\sigma(B \cap C) \cap (B \cap C) = B \cap C$, since B and C are blocks. □

Definition 3.4. *The blocks* $\emptyset$, Z, *and* $\{\gamma\}$, *for* $\gamma \in Z$, *are called* trivial blocks. *A nontrivial block is called* block of imprimitivity. *A permutation group* G *on* Z *is called* primitive *if there are only trivial blocks. It is called* imprimitive *otherwise.*

Example 3.5. The alternating group $\mathbb{A}_n$ on $\{1,\ldots,n\}$ is primitive: Let without lose of generality $n > 2$ and assume B is a block with at least two distinct elements, say $\alpha \neq \beta \in B$. Let γ be an arbitrary element in $\{1,\ldots,n\}$ distinct from α and β. Then for $\sigma = (\alpha\ \beta\ \gamma) \in \mathbb{A}_n$ we have $\sigma(\alpha) = \beta \in \sigma B \cap B$. Therefore $\sigma B = B$, and since $\sigma(\beta) = \gamma \in \sigma B$ we get $\gamma \in B$. This proves $B = \{1,\ldots,n\}$.

The same holds for $\mathbb{S}_n$. ♢

Example 3.6. Let the dihedral group $D_6 = \langle \sigma, \tau \rangle$ act on $\{1,\ldots,6\}$ by $\sigma = (1\ 3\ 5)(2\ 4\ 6)$ and $\tau = (1\ 4)(2\ 3)(5\ 6)$. Then D_6 is imprimitive, for example, $\{1,3,5\}$ and $\{1,2\}$ are nontrivial blocks. ♢

The following theorem is essential for the link between the decomposition of polynomials and the theory of blocks.

Theorem 3.7. *Let* G *be a finite transitive permutation group on a finite set* Z *and let* $\alpha \in Z$. *Then the lattice of subgroups between* G_α *and* G *is isomorphic to the lattice of blocks containing* α.

Proof. For a block B with $\alpha \in B$ define $\Phi\colon B \mapsto G_B$ and for a subgroup $G_\alpha \subseteq U \subseteq G$ define $\Psi\colon U \mapsto U(\alpha)$. To prove that both maps are well defined we first show that G_B contains G_α. Let σ be in G_α. Then $\sigma(\alpha) = \alpha \in B$. Thus $\sigma(B) \cap B \neq \emptyset$ and therefore $B = \sigma(B)$.

To prove that $U(\alpha)$ is a block, let $\sigma \in G$ and assume γ is in $\sigma(U(\alpha)) \cap U(\alpha)$. Then there are τ and τ' in U such that $\sigma\tau'(\alpha) = \gamma = \tau(\alpha)$. Thus $\tau^{-1}\sigma\tau'(\alpha) = \alpha$ and therefore $\tau^{-1}\sigma\tau' \in G_\alpha \subseteq U$. This implies that σ is in U and we have $\sigma(U(\alpha)) = U(\alpha)$.

Clearly $\Phi \circ \Psi(U) = \{\sigma \in G \mid \sigma(U(\alpha)) = U(\alpha)\} \supseteq U$. For the reverse inclusion consider $\sigma \in G$ such that $\sigma(U(\alpha)) = U(\alpha)$. Thus there is $\tau \in U$ such that $\sigma\tau(\alpha) = \alpha$. Then $\sigma\tau \in G_\alpha \subseteq U$ and therefore $\sigma \in U$. Thus we have proven that $\Phi \circ \Psi = id$. For the other direction, one finds that $\Psi \circ \Phi(B) = \{\sigma(\alpha) \mid \sigma(B) = B\} \subseteq B$. Let $\beta \in B$. Since G acts transitively there is $\sigma \in G$ such that $\sigma(\alpha) = \beta$. Then $\beta \in B \cap \sigma B$ and thus $\sigma(B) = B$. Therefore $\beta = \sigma(\alpha) \in \Psi \circ \Phi(B)$. Thus $\Psi \circ \Phi = id$ and we have proven that Ψ

and Φ are bijective. It is now sufficient to show that Φ is order preserving. Let $B \subseteq B'$ and $\sigma \in G_B$. Then $B = \sigma(B) \cap B \subseteq \sigma(B') \cap B'$ and thus $\sigma(B') \cap B'$ is nonempty. Hence σ is in $G_{B'}$. □

We fix the following notation. Let f be a polynomial in $F[x]$ of degree n with $f' \neq 0$. As before define $\varphi = f - t \in F(t)[x]$ and let α be a root of φ. Furthermore let L be the splitting field of φ over $F(t)$ and let G be its Galois group. Then G acts transitively on the set Z of roots of φ. We consider G as permutation group on Z.

COROLLARY 3.8. *(i) The lattice of decompositions of f and the lattice of blocks of G containing α are isomorphic.*

(ii) Let h be the right component of a normal decomposition of f and let B be the block corresponding to h. Then $\deg(h) = |B|$.

PROOF. The lattice of decompositions of f is isomorphic to the lattice of intermediate fields of $F(\alpha) \mid F(t)$. This in turn is by Galois theory isomorphic to the lattice of subgroups between G_α and G. Thus, by the previous theorem one achieves an isomorphism between the lattice of decompositions of f and the lattice of blocks containing α.

Let U be the subgroup corresponding to h and B be the corresponding block (that is, $F(h(\alpha)) = L^U$ and $U(\alpha) = B$). Then $\deg(h) = [F(\alpha) \colon F(h(\alpha))] = [L^{G_\alpha} \colon L^U] = (U \colon G_\alpha)$. On the other hand, we have $B = U(\alpha) = \{\sigma(\alpha) \mid \sigma \in U/G_\alpha\}$. Thus $|B| = (U \colon G_\alpha) = \deg(h)$. □

3.2. Finding minimal blocks. In this and in the next section we will discuss an algorithm that computes minimal blocks of the Galois group G. This algorithm and all intermediate results were introduced in Landau & Miller (1985) for the ground field $\mathbb{Q}$. In our case we have the ground field $F(t)$, but the proofs are essentially based on Landau & Miller (1985).

From now on we consider only blocks containing α. We call a nontrivial block B minimal if all blocks $B' \subseteq B$ are either trivial or equal to B. If B is minimal, then the corresponding decomposition is minimal.

LEMMA 3.9. *The set $B_\alpha = \{\beta \in Z \mid \forall \sigma \in G_\alpha \colon \sigma\beta = \beta\}$ is a block of G.*

PROOF. Let $\beta \in B_\alpha$. Then each $\sigma \in G_\alpha$ fixes β, hence $G_\alpha \subseteq G_\beta$. Since G is transitive we have $|G_\alpha| = |G_\beta|$ and thus $G_\alpha = G_\beta$.

Now let τ be in G and assume $\tau B_\alpha \cap B_\alpha$ is not empty. Then there are β, $\beta' \in B_\alpha$ such that $\tau(\beta) = \beta'$. We have $G_\alpha = G_\beta = G_{\beta'}$. Let $\gamma \in B_\alpha$. Then

$\tau^{-1}G_\alpha\tau = \tau^{-1}G_{\beta'}\tau = G_\beta = G_\gamma$. Thus for all $\sigma \in G_\alpha$ we have $\sigma\tau(\gamma) = \tau(\gamma)$. Hence $\tau(\gamma) \in B_\alpha$. □

Now factor φ over $F(\alpha)$ into irreducible factors ψ_i such that

$$\varphi = \prod_{i=1}^{s}(x - \alpha_i) \cdot \psi_{s+1} \cdot \ldots \cdot \psi_r, \tag{3.10}$$

with $\alpha = \alpha_1$, $\alpha_i \in F(\alpha)$, and $\psi_i = x - \alpha_i$ for $1 \le i \le s$, and $\deg \psi_i \ge 2$ for $s < i \le r$. Since $\alpha_i \in F(\alpha)$ for $1 \le i \le s$ there are rational functions ℓ_i such that $\alpha_i = \ell_i(\alpha_1)$. Since α is transcendental over F, from the equation $f(\alpha) = t = f(\ell_i(\alpha))$ follows that ℓ_i must be a linear polynomial. Clearly α_i is in B_α for all $1 \le i \le s$. Let $\beta \in B_\alpha$. Then $\beta \in L^{G_\alpha} = F(\alpha)$. Thus $\beta = \ell(\alpha)$ for some linear polynomial ℓ. We have proven that $B_\alpha = \{\alpha_i \mid 1 \le i \le s\}$.

Claim 3.11. *$H = (\{\ell_i \mid 1 \le i \le s\}, \circ)$ is a group.*

The neutral element in H is $\ell_1 = x$. Since α is transcendental over F, the equation $f \circ \ell_i(\alpha) = t = f(\alpha)$ implies $f \circ \ell_i = f$. Then from $f(\ell_i \circ \ell_j(\alpha)) = f(\ell_j(\alpha)) = t$ follows that $\ell_i \circ \ell_j(\alpha)$ is a root of φ in $F(\alpha)$. Thus there exists k such that $\ell_i \circ \ell_j = \ell_k$. In the same way by $f(\ell_i^{-1}(\alpha)) = f \circ \ell_i \circ \ell_i^{-1}(\alpha) = t$ we conclude the existence of the inverse of ℓ_i in H.

The following lemma presents us with the opportunity to lay hands on the Galois group from a computational point of view.

Lemma 3.12. *The mapping $\Phi\colon G_{B_\alpha} \to H, \sigma \mapsto \ell_i^{-1}$ for $\sigma(\alpha) = \alpha_i = \ell_i(\alpha_1)$, is a surjective homomorphism with kernel G_α.*

Proof. Let σ, $\tau \in G_{B_\alpha}$ with $\Phi(\sigma) = \ell_i^{-1}$ and $\Phi(\tau) = \ell_j^{-1}$. Then $\sigma \circ \tau(\alpha) = \sigma(\ell_j(\alpha)) = \ell_j(\sigma\alpha) = (\ell_j \circ \ell_i)(\alpha)$. Thus $\Phi(\sigma\tau) = (\ell_j \circ \ell_i)^{-1} = \ell_i^{-1} \circ \ell_j^{-1} = \Phi(\sigma)\Phi(\tau)$. Each ℓ_i defines an $F(t)$-automorphism on $F(\alpha)$, extends to L and maps B_α to B_α. Thus Φ is surjective. Finally, σ is in the kernel if and only if $\sigma(\alpha) = \alpha$. This is if and only if σ is in G_α. □

Now let $s = n$. Then we have $B_\alpha = Z$ and $L = F(\alpha) = L^{G_\alpha}$. Thus $G_{B_\alpha} = G$ and $G_\alpha = 1$. In this case Φ is an isomorphism between G and H and we can compute all minimal blocks by the algorithm of Atkinson (1975) in polynomial time. Before we consider the other cases we will see how to compute the appropriate h from a block B. Instead of following Zippel (1991), we use the following new result.

Lemma 3.13. *Let B be a block and h be the right component of a decomposition of f corresponding to B. Then $h(x) - h(\alpha) = \prod_{\gamma \in B}(x - \gamma)$.*

Proof. The block B corresponds to the intermediate field L^{G_B} and by Theorem 2.3 there is a decomposition of f with right component h such that $L^{G_B} = F(h(\alpha))$. Set $\lambda = h(\alpha)$. Then the minimal polynomial of α over $F(\lambda)$ is $h - \lambda$. Since α is in B and both polynomials have the same degree, it is sufficient to show that $\prod_{\gamma \in B}(x - \gamma)$ is in $F(\lambda)[x]$. Let σ be in G_B. Then $\sigma(B) = B$ and therefore $\sigma(\prod_{\gamma \in B}(x - \gamma)) = \prod_{\gamma \in B}(x - \sigma\gamma) = \prod_{\gamma \in B}(x - \gamma)$. Since $F(\lambda) = L^{G_B}$ this proves that $\prod_{\gamma \in B}(x - \gamma)$ is in $F(\lambda)[x]$. □

Note that $\prod_{\gamma \in B}(-\gamma)$ is the constant term of $h(x) - h(\alpha)$. Since h is normal, we get $h = \prod_{\gamma \in B}(x - \gamma) - \prod_{\gamma \in B}(-\gamma)$, as explicit formula.

Example 3.14. Let p be an odd prime and F be a finite field of characteristic p. Let $f = x^2 \circ (x^p - x)$, a be an element of the prime field $\mathbb{F}_p$ of F and ζ be either 1 or -1. Then $f(\zeta x + a) = (\zeta^p x^p + a^p - \zeta x - a)^2 = f(x)$. Thus, for a root α of $f - t$ also $\zeta\alpha + a$ is a root of f. Thus we have $2p$ roots of $f - t$ in $F(\alpha)$ and therefore $F(\alpha) \mid F(t)$ is Galois. Its Galois group is isomorphic to $\{(\zeta x + a) \mid \zeta \in \{-1, 1\}, a \in \mathbb{F}_p\} \cong \mathbb{F}_p \rtimes \mathbb{Z}/2\mathbb{Z} \cong D_{2p}$. The dihedral group D_{2p} has one subgroup of order p and p subgroups of order two. Hence f has $p + 1$ decompositions. A block with two elements is of the form $\{\alpha, -\alpha + a\}$. Then $h(x) - h(\alpha) = (x - \alpha)(x - (-\alpha + a)) = x^2 - ax - (\alpha^2 - a\alpha)$ and we have found the right component of a decomposition of f, namely $h = x^2 - ax$. ◊

Let $1 < s < n$. Then the induced action of G_{B_α} on B_α is determined by the action of H on B_α, since G_α acts trivial on B_α. If there are minimal blocks of G_{B_α} containing α, one can find all of them in polynomial time (by the above mentioned algorithm of Atkinson).

Lemma 3.15. *If Λ is a minimal block of G_{B_α}, then Λ is a minimal block of G.*

Proof. Assume $\sigma(\Lambda) \cap \Lambda \neq \emptyset$ for some $\sigma \in G$. Since $\Lambda \subseteq B_\alpha$ and $\sigma(\Lambda) \subseteq \sigma(B_\alpha)$, we get $\sigma(\Lambda) \cap \Lambda \subseteq \sigma(B_\alpha) \cap B_\alpha \neq \emptyset$. Thus $\sigma(B_\alpha) = B_\alpha$, which means that σ is in G_{B_α}. Since Λ is a block of G_{B_α}, we have $\sigma(\Lambda) = \Lambda$. If $B \subseteq \Lambda$ is a nontrivial block of G, then B is a block of G_{B_α} and thus $B = \Lambda$. □

Thus, we can easily compute all minimal blocks that are contained in B_α. Note that if there is no nontrivial Block of G_{B_α}, then B_α is a minimal block of G.

Example 3.16. Let $p = 3$ and $f = x^9 - x$ over $\mathbb{F}_3$. Let α be a root of $f - t$. One checks that

$$\begin{aligned} f(x) - f(\alpha) &= (x-\alpha)(x-\alpha+1)(x-\alpha-1) \\ &\quad (x^2+\alpha x+\alpha^2+1)(x^2+(\alpha+1)x+\alpha^2-\alpha-1) \\ &\quad (x^2+(\alpha-1)x+\alpha^2+\alpha-1) \end{aligned}$$

is the factorization of $f - t$ into irreducible polynomials over $\mathbb{F}_3(\alpha)$. As shown above $\{\alpha, \alpha-1, \alpha+1\}$ forms a block. Thus for $h(x) - h(\alpha) = (x-\alpha)(x-\alpha+1)(x-\alpha-1) = x^3 - x - (\alpha^3 - \alpha)$ we have that $h = x^3 - x$ is the right component of a decomposition of f. One calculates that the corresponding left component is $g = x^3 + x$. ♢

If $s = 1$ then $B_\alpha = \{\alpha\}$ is trivial. Thus the method above does not apply. But also if $1 < s < n$, there could be minimal blocks Λ with $\Lambda \cap B_\alpha = \{\alpha\}$. Thus, let $s < n$. Then $F(\alpha)$ is not Galois and we have $G_\alpha \neq 1$. Hence G is not regular and the following theorem applies.

Theorem 3.17. *Let G be a finite permutation group on Z, which is transitive and not regular. Then G is primitive if and only if for all distinct α and β in Z we have $\langle G_\alpha, G_\beta \rangle = G$.*

Proof. Let G be imprimitive and Λ be a nontrivial block with α, $\beta \in \Lambda$ and $\alpha \neq \beta$. Then G_α, $G_\beta \subseteq G_\Lambda$. Thus by Theorem 3.7, we get $\langle G_\alpha, G_\beta \rangle \subseteq G_\Lambda \neq G$, since $\Lambda \neq Z$.

Now let $\langle G_\alpha, G_\beta \rangle \neq G$ for some $\alpha \neq \beta$. Then $\Lambda = \langle G_\alpha, G_\beta \rangle(\alpha)$ is a block $\neq Z$ as shown in the proof of Theorem 3.7. If Λ is nontrivial, then we are done. Thus assume $\Lambda = \{\alpha\}$. Then we have $\sigma(\alpha) = \alpha$ for all $\sigma \in G_\beta$. Thus $G_\beta \subseteq G_\alpha$. Since $|G_\alpha| = |G_\beta|$ we have $G_\alpha = G_\beta$. But then α, $\beta \in B_\alpha$ and B_α is trivial only if $B_\alpha = Z$. Since G is not regular there is $1 \neq \sigma \in G_\alpha$ and γ such that $\sigma(\gamma) \neq \gamma$. Thus $\gamma \notin B_\alpha$. Hence $B_\alpha \neq Z$, which is thus a nontrivial block. □

Proposition 3.18. *Let Λ be a minimal block of G with $\alpha \in \Lambda$ and $\Lambda \cap B_\alpha = \{\alpha\}$. Then for all $\beta \in \Lambda$ distinct from α the orbit $\langle G_\alpha, G_\beta \rangle(\alpha)$ equals Λ.*

Proof. Let $\beta \neq \alpha$ be in Λ. Clearly $\langle G_\alpha, G_\beta \rangle(\alpha) = \{\sigma(\alpha) \mid \sigma \in \langle G_\alpha, G_\beta \rangle\} \subseteq \Lambda$. Now if $G_\alpha = G_\beta$ then β would be fixed by G_α and thus $\beta \in B_\alpha$, which is a contradiction to the assumption. Thus we have $G_\alpha \neq G_\beta$ and therefore $|\langle G_\alpha, G_\beta \rangle(\alpha)| > 1$. Thus $\langle G_\alpha, G_\beta \rangle(\alpha)$ is a nontrivial block contained in Λ, which implies equality by the minimality of Λ. □

LEMMA 3.19. *Let $\varphi = \prod_{i=1}^{r} \psi_i$ be a factorization of φ into irreducible factors ψ_i over $F(\alpha)$. Let $\beta \in Z$ and j be such that $\psi_j(\beta) = 0$. Then $G_\alpha(\beta) = \{\gamma \in Z \colon \psi_j(\gamma) = 0\}$.*

PROOF. If $\psi_j = (x - \beta)$ then β is in $F(\alpha)$ and is thus fixed by G_α. Hence the claim holds. Thus let $\deg(\psi_j) \geq 2$. Each $\sigma \in G_\alpha$ acts trivial on $F(\alpha)$. Thus $\sigma(\psi_j) = \psi_j$ and $\psi_j(\sigma(\beta)) = \sigma(\psi_j(\beta)) = 0$. Hence $G_\alpha(\beta) \subseteq \{\gamma \in Z \colon \psi_j(\gamma) = 0\}$.

For the other direction let γ be such that $\psi_j(\gamma) = 0$. Let M be the splitting field of ψ_j over $F(\alpha)$. Then there is σ in $\mathrm{Gal}(M \mid F(\alpha))$ such that $\sigma(\beta) = \gamma$. Since $F(\alpha) \subseteq M \subseteq L$ we have that σ extends to an automorphism in $\mathrm{Gal}(L \mid F(\alpha)) = G_\alpha$. Thus γ is in $G_\alpha(\beta)$. □

Fix $\nu > s$ and β such that β is a root of ψ_ν. Note that $\beta \notin B_\alpha$ and thus $\langle G_\alpha, G_\beta \rangle(\alpha)$ is a block, which is minimal if there is a minimal block containing α and β. Let σ be in G such that $\sigma(\alpha) = \beta$ and set $\psi_i^* = \sigma(\psi_i)$ for all $1 \leq i \leq r$. Then the polynomials $\psi_i^* \in F(\beta)[x]$ are the polynomials ψ_i with β substituted for α and the irreducible factors of φ over $F(\beta)$ are precisely the polynomials ψ_i^*. Note that if γ is a root of ψ_j^* we have $G_\beta(\gamma) = \{\gamma' \in Z \colon \psi_j^*(\gamma') = 0\}$ by the previous lemma.

PROPOSITION 3.20. *Consider the bipartite graph Γ_β with the set of vertices consisting of ψ_i and ψ_i^* for $1 \leq i \leq r$ and with an undirected edge between ψ_i and ψ_j^* if $\gcd(\psi_i, \psi_j^*) \neq 1$. Let C_β be the the set of roots of those ψ_i that are connected to ψ_1. Then $\langle G_\alpha, G_\beta \rangle(\alpha) = C_\beta$.*

PROOF. Each element γ of $\langle G_\alpha, G_\gamma \rangle(\alpha)$ is of the form $\sigma_u \ldots \sigma_2 \sigma_1(\alpha)$ with σ_i in G_α or in G_β. We prove by induction on u that γ is in C_β. The induction basis is the fact that $\alpha \in C_\beta$. For the induction step let $\gamma = \sigma_{u-1} \ldots \sigma_2 \sigma_1(\alpha)$ be in C_β. Then there is some i such that $\psi_i(\gamma) = 0$ and ψ_i is connected to ψ_1. We distinguish two cases: First we have $\sigma_u \in G_\alpha$. Since $\sigma_u(\psi_i) = \psi_i$, we have $0 = \sigma_u(\psi_i(\gamma)) = \psi_i(\sigma_u \gamma)$. Thus also $\sigma_u(\gamma) \in C_\beta$. In the second case we have $\sigma_u \in G_\beta$. Let j such that $\psi_j^*(\gamma) = 0$. Then there is an edge between ψ_j^* and ψ_i and thus ψ_j^* is connected to ψ_1. Since $\sigma_u(\gamma) \in G_\beta(\gamma)$ we have that also $\sigma_u(\gamma)$ is a roots of ψ_j^*. Hence if $\sigma_u(\gamma)$ is a root of ψ_k we have $\gcd(\psi_k, \psi_j^*) \neq 1$ and thus $\sigma_u(\gamma) \in C_\beta$.

For the other direction, let γ be in C_β and i such that $\psi_i(\gamma) = 0$. Then there is a path P form ψ_1 to ψ_i, say $P = (\psi_1, \psi_\ell^*), \ldots, (\psi_k, \psi_j^*)(\psi_j^*, \psi_i)$. By an induction argument one can assume that the roots of ψ_k are already in $\langle G_\alpha, G_\gamma \rangle(\alpha)$. Since $\gcd(\psi_j^*, \psi_i) \neq 1$ there is β' such that $\psi_j^*(\beta') = 0$ and $\psi_i(\beta') = 0$. Thus, by Lemma 3.19 we get $\gamma \in G_\alpha(\beta')$ and $\beta' \in G_\beta(\alpha')$ where α' is a common root of ψ_k and ψ_j^*. Then $\alpha' \in \langle G_\alpha, G_\beta \rangle(\alpha)$ and $\gamma = \sigma_1\sigma_2(\alpha')$ for $\sigma_1 \in G_\beta$ and $\sigma_2 \in G_\alpha$. Hence γ is in $\langle G_\alpha, G_\beta \rangle(\alpha)$. □

Now let Λ be a minimal block containing α and β. In case $\Lambda \subseteq B_\alpha$ we saw that one can calculate Λ directly and therefore calculate h by Lemma 3.13. Otherwise by Proposition 3.18 we have $\Lambda = \langle G_\alpha, G_\beta \rangle(\alpha)$. Then as seen in Proposition 3.20 one can calculate h by $h(x) - h(\alpha) = \Pi_{\gamma\in\Lambda}(x - \gamma) = \Pi\psi_i$, where the last product is taken over all ψ_i that are connected to ψ_1 in Γ_β.

Example 3.16 continued. Let us continue to find decompositions of $f = x^9 - x$. Let

$$\begin{aligned}\psi_1 &= x^2 + \alpha x + \alpha^2 + 1,\\ \psi_2 &= x^2 + \alpha x + x + \alpha^2 - \alpha - 1 \text{ and}\\ \psi_3 &= x^2 + \alpha x - x + \alpha^2 + \alpha - 1.\end{aligned}$$

Then as before we have $f - t = (x - \alpha)(x - \alpha + 1)(x - \alpha - 1)\psi_1\psi_2\psi_3$. Now let ζ be a root of $x^2 + x - 1$ in $\mathbb{F}_9$ and note that we have then

$$\begin{aligned}\psi_1 &= (x - (\alpha + \zeta + 1))(x - (\alpha - \zeta - 1)),\\ \psi_2 &= (x - (\alpha + \zeta - 1))(x - (\alpha - \zeta)) \text{ and}\\ \psi_3 &= (x - (\alpha + \zeta))(x - (\alpha - \zeta + 1)).\end{aligned}$$

Let $\beta_1 = \alpha + \zeta + 1$. Then we have that ψ_1 with β_1 substituted for α is $\psi_1^* = (x - \alpha)(x - (\alpha - \zeta - 1))$ and thus $C_{\beta_1} = \{\alpha, \alpha + \zeta + 1, \alpha - \zeta - 1\}$ is a minimal block. We get $h(x) - h(\alpha) = (x - \alpha)\psi_1 = x^3 + x - (\alpha^3 + \alpha)$ and thus the corresponding decomposition has right component $x^3 + x$. Then the corresponding left component is $x^3 - x$.

Now for $\beta_2 = \alpha + \zeta - 1$ we get $\psi_3^* = (x - \alpha)(x - (\alpha - \zeta - 1))$. Thus $C_{\beta_2} = C_{\beta_1}$. In the same way $\alpha - \zeta + 1$ does not yield any further block. Therefore all in all f has exactly two decompositions over $\mathbb{F}_3$.

Note that going to the extension $\mathbb{F}_9$ of $\mathbb{F}_3$ unveils more structure. Actually, f has four decompositions over $\mathbb{F}_9$ as we will see in Example 3.27. ◊

3.3. The algorithm. Zippel (1991) describes loosely an algorithm that computes decompositions of rational functions. The following is a concrete description of an algorithm that computes all minimal decompositions of a polynomial, whose derivative does not vanish. It is mainly based on Zippel (1991) and on Landau & Miller (1985). The runtime estimation in Theorem 3.23 is new.

Algorithm 3.21 calls a subroutine `Atkinson`(G, Z, α) which returns a list of all minimal blocks of the permutation group G on Z that are containing α. If G is primitive this list consists of Z only.

ALGORITHM 3.21. Computing minimal decompositions.

Input: A monic polynomial $f \in F[x]$ of degree n with $f' \neq 0$.

Output: A list of decompositions (g, h) of f. This list is empty if f is indecomposable.

1. Set $List = \{\}$ and let $F(\alpha)$ be the rational function field in α.
2. Factor $f(x) - f(\alpha)$ in $F(\alpha)[x]$ into $\prod_{i=1}^{s}(x - \alpha_i) \cdot \psi_{s+1} \cdot \ldots \cdot \psi_r$ as in (3.10).
3. If $s > 1$ then
4.   Set $B_\alpha = \{\alpha_i \mid 1 \leq i \leq s\}$ and $H = \{\ell_i \mid 1 \leq i \leq s\}$ where $\alpha_i = \ell_i(\alpha)$.
5.   Set $AtkinsonBlocks = \texttt{Atkinson}(H, B_\alpha, \alpha)$.
6.   For $\Lambda \in AtkinsonBlocks$ with $|\Lambda| < n$ do 7–9
7.   Compute $h(x) = \prod_{\gamma \in \Lambda}(x - \gamma) - \prod_{\gamma \in \Lambda}(-\gamma)$.
8.   Compute g such that $f = g \circ h$.
9.   Attach (g, h) to $List$.
10. For $\nu \in \{s + 1, \ldots, r\}$ do 11–17
11.   Let β be a root of ψ_ν and let ψ_i^* be ψ_i with β substituted for α, for all $1 \leq i \leq r$.
12.   Compute the graph Γ_β as in Proposition 3.20.
13.   Compute $I_\nu = \{i \colon \psi_i \text{ is connected to } \psi_1 \text{ in } \Gamma_\beta\}$.
14.   If $I_\nu \neq \{1, \cdots, r\}$ then
15.   Compute $h(x) - h(\alpha) = \prod_{i \in I_\nu} \psi_i$, where $h(x)$ is in $F[x]$ normal.
16.   Compute g such that $f = g \circ h$.
17.   Attach (g, h) to $List$.
18. Return $List$.

THEOREM 3.22. *Algorithm 3.21 correctly computes all minimal decompositions of f.*

PROOF. Let (g, h) be a minimal decomposition and let Λ be the corresponding block. Then either $\Lambda \subseteq B_\alpha$ or $\Lambda \cap B_\alpha = \{\alpha\}$. In the first case Λ is computed in Step 5, by Lemma 3.15. Then h is recovered from Λ in Step 7, by Lemma 3.13. In the second case let $\beta \in \Lambda \setminus \{\alpha\}$ and ν such that $\psi_\nu(\beta) = 0$. By Proposition 3.20 we have $\Lambda = C_\beta = \{\gamma \colon \exists i \in I_\nu \colon \psi_i(\gamma) = 0\}$, where I_ν is computed in Step 13. Then in Step 15 we have $\prod_{i \in I_\nu} \psi_i = \prod_{\gamma \in \Lambda}(x - \gamma) = h(x) - h(\alpha)$, from which we can recover h. □

Note that C_β is a block even if there is no minimal block containing α and β. Then either C_β is minimal and $\beta \notin C_\beta$ or C_β is not minimal and contains minimal blocks. An algorithm that computes *only* minimal decompositions (and outputs each decomposition only once) should keep track of this.

For the runtime consideration let F be a field over which one can factor bivariate polynomials in polynomial time (these include, for example, finite fields). The gcd computation in Step 12 can be done by computing the resultant, which can be done in polynomial time. Thus the algorithm can be implemented with a polynomial runtime. This was already remarked by Zippel (1991). The following runtime estimation for finite fields is new.

THEOREM 3.23. *Let F be a finite field, $c \geq 3$ be a natural number and n be the degree of the input polynomial f. Denote the complexity of multiplying two polynomials over F of degree at most n by $M(n)$ and let q be the size of F. Then there is an implementation of Algorithm 3.21 that takes an expected number of $\mathcal{O}^{\sim}(cn^4 M(n)^2 \log(q))$ operations in F with an error probability of at most $n^3(4n)^{-c}$.*

PROOF. The factorization in Step 2 can be done in $\mathcal{O}^{\sim}(n^{\omega+1})$, where $2 \leq \omega \leq 3$ is the matrix multiplication exponent, see Bostan *et al.* (2004) and Lecerf (2008). Atkinson's algorithm runs in $\mathcal{O}(n^3)$, see Atkinson (1975). Butler (1992) improved the runtime of Atkinson's algorithm to $\mathcal{O}(n^2 \log n)$. In Step 8 and 16 for each right component h the appropriate left component g can be computed in $\mathcal{O}(M(n) \log n)$ by the generalized Taylor expansion, see von zur Gathen & Gerhard (1999). Since there are at most s minimal blocks computed by the algorithm of Atkinson, Step 8 is called at most s times. Step 16 is called at most $r - s$ times. Thus we get $\mathcal{O}(rM(n) \log n)$ for this part.

To compute the graph in Step 12 we need at most r^2 gcd computations. We have to compute $r - s$ such graphs. Thus in total we have at most $(r-s)r^2 \leq n^3$ such gcd computations. Since the field arithmetic of $F(\alpha, \beta)$ is quite costly, one should use a modular algorithm that checks if two polynomials in $F(\alpha, \beta)[x]$ are coprime. For example one could use Algorithm 3.24 below. We will prove that it has expected runtime $\mathcal{O}^{\sim}(nM(n)^2 \log(q))$ and an error probability of at most $(4n)^{-1}$. If we repeat this coprimality check c times we get for all n^3 computations an expected runtime of $\mathcal{O}^{\sim}(cn^4 M(n)^2 \log(q))$, which dominates the runtime of the other computations, and an error probability of at most $1 - (1 - (4n)^{-c})^{n^3} \leq 1 - (1 - n^3(4n)^{-c}) = n^3(4n)^{-c}$, by the Bernoulli inequality. This finishes the proof of Theorem 3.23. □

LEMMA 3.25. *Let F be a finite field of size q. If the total degree of all input polynomials is bounded by n, then Algorithm 3.24 takes an expected number of $\mathcal{O}^{\sim}(M(n)^2 n \log(q))$ operations in F. It returns True only if g and h are*

ALGORITHM 3.24. Coprimality in $F(\alpha,\beta)[x]$.

Input: An irreducible polynomial $G \in F[x,y]$ of total degree at most n, that defines $F(\alpha,\beta)$ by $G(\alpha,\beta)=0$, and two polynomials g, $h \in F[\alpha,\beta,x]$ that are monic in x.

Output: *True* / *False*.

1. Let $K' \mid F$ be a field extension of F with $[K' \colon \mathbb{F}_p] \geq 4\log(16n)$.
2. Randomly choose a in K'.
3. Compute a root b of $G(a,y)$ in an extension K of K'.
4. Compute $r = \operatorname{res}(g(a,b,x), h(a,b,x))$.
5. Return *True* if $r \neq 0$ and *False* else.

coprime. If g and h are coprime the algorithm returns $False$ with probability at most $(4n)^{-1}$.

PROOF. First note that for $(a,b) \in K^2$ with $G(a,b)=0$ the maximal ideal $(\alpha - a, \beta - b)$ in $K[\alpha,\beta]$ gives rise to a place P in $K(\alpha,\beta)$. Substituting a and b for α and β, respectively, is the same as reducing modulo P. Thus the degree of P is one. If on the other hand P is a place in $K(\alpha,\beta)$ of degree one such that it is neither a pole of α nor of β, then there exists a, $b \in K$ such that $\alpha \equiv a \mod P$ and $\beta \equiv b \mod P$. Then we have $P \cap K[\alpha,\beta] = (\alpha - a, \beta - b)$. Denote such a places by $P_{a,b}$.

Since the leading coefficient of g in x is one, and therefore not zero modulo P, we have $\operatorname{res}(g(a,b,x),h(a,b,x)) = 0$ if and only if $\operatorname{res}(g,h) \equiv 0 \mod P$ (see Lemma 6.25 in von zur Gathen & Gerhard (1999)). Assume $\rho = \operatorname{res}(g,h) \in K[\alpha,\beta]$ is not zero. Let $A = |\{(a,b) \in K' \times K \colon G(a,b)=0\}|$ and $B = |\{(a,b) \in A \colon \rho \equiv 0 \mod P_{a,b}\}|$. Then the probability of ρ being zero modulo P is B/A.

We have $B \leq |\{(a,b) \in K^2 \mid G(a,b) = 0 = R(a,b)\}|$, where R is a representative of ρ in $F[x,y]$ of degree less then $2n^2$. If G would divide R, then ρ would be zero. Since G is irreducible we get $\gcd(G,R) = 1$ and thus by Bézout's Theorem we have $B \leq \deg(G)\deg(R) \leq 2n^3$.

Since α and β can have at most n poles we have $A \geq N - 2n$, where N is the number of places in $K'(\alpha,\beta)$ of degree one. By the Hasse–Weil Bound (see Theorem V.2.3 in Stichtenoth (1993)) we get $N \geq q' + 1 - 2gq'^{1/2}$, where q' is the size of K' and g is the genus of $K'(\alpha,\beta)$. By the degree assumption on K' we have that $q' \geq p^{4\log(16n)} \geq 16n^4$. By the Riemann Inequality (see Corollary III.10.4 in Stichtenoth (1993)) the genus is bounded by $([K'(\alpha,\beta) \colon K'(\alpha)] - 1)([K'(\alpha,\beta) \colon K'(\beta)] - 1) \leq (n-1)^2$. Thus we have $A \geq N - 2n \geq q' + 1 -$

$2gq'^{1/2} - 2n \geq q'^{1/2}(q'^{1/2} - 2g) - 2n \geq 4n^2(4n^2 - 2(n-1)^2) - 2n \geq 8n^4$. Hence $B/A \leq (2n^3)/(8n^4) = 1/(4n)$, which gives us the claimed error bound.

The resultant computation takes $\mathcal{O}((M(n)+n)\log n)$ operations in K (see Corollary 11.16 in von zur Gathen & Gerhard (1999)). Finding a root of $G(a, y)$ in K takes an expected number of $\mathcal{O}(M(n)\log n \log(n\tilde{q}))$ operations in K, where $\tilde{q}$ is the size of K (see Corollary 14.16 in von zur Gathen & Gerhard (1999)). We have $\log(\tilde{q}) \leq [K\colon \mathbb{F}_p] = [K\colon F][F\colon \mathbb{F}_p]$. Unfortunately the degree of K over K' can only be bounded by $\deg_y(G(a, y)) \leq n$. Thus $[K\colon F]$ is in $\mathcal{O}(n \log n)$ and finding a root takes $\mathcal{O}^{\sim}(M(n)n\log(q))$ operations in K. Arithmetic in K costs us $\mathcal{O}(M(n\log n)\log(n\log n))$ operations in F. Hence with omitting the log factor the expected runtime is in $\mathcal{O}^{\sim}(M(n)^2 n \log(q))$. □

3.4. An upper bound. Now we will deduce two sharp upper bounds on the number of minimal decompositions of a polynomial. These bounds coincide partly with results in von zur Gathen *et al.* (2010).

Let B and B' be minimal blocks. By Lemma 3.3 their intersection is a block and therefore trivial. Hence, the minimal blocks minus $\{\alpha\}$ are distinct sets in $Z \setminus \{\alpha\}$. Therefore the sum of the cardinality of all minimal blocks minus $\{\alpha\}$ is less than $n-1$. Since the cardinality of a block equals the degree of the right component of the corresponding normal decomposition, we get the following results.

COROLLARY 3.26. *Let f be a decomposable polynomial of degree n with $f' \neq 0$.*

(i) Let d divide n. Then there are at most $(n-1)/(d-1)$ minimal decompositions (g, h) of f with $\deg(h) = d$.

(ii) Let q be the smallest prime divisor of n. Then there are at most $(n-1)/(q-1)$ minimal decompositions of f.

EXAMPLE 3.27. Let p be the characteristic of F and let f be a separable additive polynomial of degree p^r with $r \geq 2$, that is f is of the form $\sum_{i=0}^{r} a_i x^{p^i}$ with $a_0 \neq 0$. Furthermore assume that f splits completely over F. Then the roots of f form a group $G \subseteq F$ which is isomorphic to $(\mathbb{Z}/p\mathbb{Z})^r$. If α is a root of $\varphi = f - t$ then so is $\alpha + a$ for all roots a of f. Thus φ is Galois and its Galois group is isomorphic to G. But G has exactly $(p^r-1)/(p-1)$ subgroups of order p. Thus f has exactly $(p^r-1)/(p-1)$ minimal decompositions. This shows that both bounds are sharp. ◊

Another class of examples, which also shows that these bounds are sharp, is discussed in the next section.

4. A taxonomy

Zannier (1993) proved Ritt's second Theorem over arbitrary fields with the assumption that the derivatives of the right components are nonzero and that their degrees are relatively prime. Unfortunately this result is not applicable when the degrees are equal – these are so called equal-degree collisions. A classification for a special case of equal-degree collision was proposed in von zur Gathen *et al.* (2010), Conjecture 6.7. The main result of this section (Theorem 4.5) is a proof of this conjecture. The idea is to adapt the proofs of Ritt's second Theorem from Dorey & Whaples (1974) and Zannier (1993). In both papers the classification is obtained from studying ramification in a rational function field.

4.1. Preliminaries. First we will state some facts from the ramification theory of function fields. For more details see, for example, Stichtenoth (1993). Let F be a field of characteristic p and K be its algebraic closure. Let f be a normal polynomial over F of degree n such that $f' \neq 0$ and $p \mid n$. As in Section 2 let t be transcendental over K. Then $\varphi = f - t$ is irreducible and separable over $K(t)$. Let α be a root of φ. Then $K(\alpha)$ is an extension of $K(t)$ of degree n. Both function fields are rational (that is, of genus 0). Thus, each finite place P in $K(t)$ corresponds to a monic and irreducible polynomial in $K[t]$ (see Section I.2 in Stichtenoth (1993)). This polynomial is linear, since K is algebraically closed, say of the form $t - c$ with $c \in K$. In $K(\alpha)$ we have $t - c = f(\alpha) - c = \prod g_i^{e_i}(\alpha)$, where $\prod g_i^{e_i}$ is a factorization of $f - c$ into irreducible factors $K[x]$. The g_i are linear and correspond to places S_i in $K(\alpha)$. Then $S_i^{e_i}$ divides P. Since $\sum e_i = \deg f = [K(\alpha) \colon K(t)]$ we obtain a decomposition $P = \prod S_i^{e_i}$. Thus the multiplicities of $f - c$ correspond to the ramification indices of P, that is $e_i = e(S_i \mid P)$.

Later in this section we wish to have certain multiple roots at the right "place". One can achieve this by conjugation of f (see Definition 2.2). If w has multiplicity m in $f - f(w)$ then the conjugate $(x - f(w)) \circ f \circ (x + w)$ has a root at 0 with multiplicity m.

We will make use of the notion of the different exponent $d(S \mid P)$ of a place $S \mid P$. Mainly we need the following facts about the different exponent (for a definition and further facts see Section III.4 in Stichtenoth (1993)): A place S is unramified over P if and only if $d(S \mid P) = 0$. If S is tamely ramified over P, then $d(S \mid P) = e(S \mid P) - 1$. Since K is algebraically closed, the relative degree of $S \mid P$ equals one. Thus, if P is tamely ramified, we get $\sum_{S|P} d(S \mid P) = n - \rho$, where ρ is the number of places in $K(\alpha)$ lying over P.

The following results tell us more about the ramification in rational function fields over K. (These results are true for arbitrary fields, but they are needed here only for the algebraic closure of F.)

Proposition 4.1. *The place at infinity in $K(t)$ is totally ramified in $K(\alpha)$.*

Proof. See Proposition 3.2. in Fried & MacRae (1969). □

Proposition 4.2. *Let $E \mid K(t)$ be a finite separable extension. Let P be a place in $K(t)$ and S be a place in E which is totally ramified over P. Let π be a prime element of S and ψ its minimal polynomial over $K(t)$. Then $d(S \mid P) = v_S(\psi'(\pi))$, where v_S is the valuation at S.*

Proof. See Proposition III.5.12 in Stichtenoth (1993). □

Lemma 4.3. *Let P_∞ be the infinite place of $K(t)$ and S_∞ be the place in $K(\alpha)$ over P_∞. Then $d(S_\infty \mid P_\infty) = 2n - 2 - \deg(f')$ and*

$$\sum_{S \text{ finite}} d(S \mid P) = \deg(f').$$

Proof. Since S_∞ is totally ramified we can apply Proposition 4.2. We have that α^{-1} is a primitive element of S_∞. Let ψ be the minimal polynomial of α^{-1}. We have $0 = \alpha^{-n}(f(\alpha) - t) = \hat{f}(\alpha^{-1}) - t\alpha^{-n}$, with $\hat{f}$ being the reversal of f. Since f is original we have $\deg(\hat{f}) < n$. Then $x^n - t^{-1}\hat{f}(x)$ is a monic polynomial, and since $[K(\alpha^{-1})\colon K(t)] = n$, we get $\psi = x^n - t^{-1}\hat{f}(x)$. Thus we have $\psi' = -t^{-1}\hat{f}'(x)$ and Proposition 4.2 yields $d(S_\infty \mid P_\infty) = v_\infty(\psi'(\alpha^{-1})) = v_\infty(-t^{-1}\hat{f}'(\alpha^{-1})) = v_\infty(-t^{-1}) + v_\infty(\hat{f}'(\alpha^{-1}))$. Since t^{-1} is a primitive element of P_∞ we have $v_\infty(-t^{-1}) = n$. Let $\hat{a}_j$ be the coefficients of $\hat{f}$. Then by the strong triangle inequality we get $v_\infty(\hat{f}'(\alpha^{-1})) \geq \min\{v_\infty(j\hat{a}_j\alpha^{-(j-1)}) \mid j\hat{a}_j \neq 0\}$ and equality since we have $v_\infty(j\hat{a}_j\alpha^{-(j-1)}) = j + 1 \neq i + 1 = v_\infty(i\hat{a}_i\alpha^{-(i-1)})$ for all $i \neq j$. The $(j-1)$-th coefficient of $\hat{f}'$ is nonzero if $p \nmid j$ and $\hat{a}_j \neq 0$. But since $p \mid n$ this is the case if and only if $p \nmid (n-j)$ and the $(n-j)$-th coefficient of f is nonzero. Thus, the last nonzero coefficient in $\hat{f}'$ is the first nonzero coefficient in f'. Hence $v_\infty(\hat{f}'(\alpha^{-1})) = n - (\deg(f') + 1) - 1$ and therefore $d(S_\infty \mid P_\infty) = 2n - 2 - \deg(f')$.

By the Hurwitz Genus Formula we have $2g' - 2 = [K(\alpha)\colon K(t)](2g-2) + \sum_S d(S \mid P)$, where g and g' is the genus of $K(t)$ and $K(\alpha)$, respectively (see Theorem III.4.12 in Stichtenoth (1993)). In our case we have $g, g' = 0$ and thus obtain $\sum_S d(S \mid P) = 2[K(\alpha)\colon K(t)] - 2 = 2n - 2$. By subtracting $d(S_\infty \mid P_\infty)$ we get $\sum_{S \text{ finite}} d(S \mid P) = 2n - 2 - (2n - 2 - \deg(f')) = \deg(f')$. □

There is also an elementary proof for the last equation of Lemma 4.3 if there is no finite wildly ramified place; see the second proof of Lemma 2 in Dorey & Whaples (1974).

LEMMA 4.4. *Let M and N be two intermediate fields of $K(\alpha) \mid K(t)$ such that $MN = K(\alpha)$ and let Q and R be finite places in M and N, respectively, over a place P in $K(t)$. Let the ramification indices $e = e(Q \mid P)$ and $\tilde{e} = e(R \mid P)$ be not divisible by the characteristic of K. Then there are $\gcd(e, \tilde{e})$ places S in $K(\alpha)$ which lie over Q and over R. Moreover for such a place we have $e(S \mid P) = \operatorname{lcm}(e, \tilde{e})$.*

This lemma is proven in Dorey & Whaples (1974) with the assumption that the characteristic of K is zero. The following proof is only slightly different, but does not use this assumption.

PROOF. At first we apply Abhyankar's Lemma (see Proposition III.8.9 in Stichtenoth (1993)) and find that for a place S in $K(\alpha)$, which lies over Q and over R, the ramification index $e(S \mid P)$ equals $\operatorname{lcm}(e, \tilde{e})$.

Then we proceed as in Dorey & Whaples (1974). Let $\widehat{K(t)}$, $\widehat{M}^Q$, and $\widehat{N}^R$ be the completions of $K(t)$, M, and N with respect to P, Q, and R, respectively. For readability set $E = K(t)$ and $E^* = \widehat{K(t)}$. Note that $K(\alpha) \otimes_E \widehat{K(t)}$ is the direct product of the completions of $K(\alpha)$ with respect to the places over P in $K(\alpha)$ (see Proposition II.8.3 in Neukirch (2007)). Since $N \otimes_E M \cong NM = K(\alpha)$ we get $K(\alpha) \otimes_M \widehat{M}^Q \cong N \otimes_E M \otimes_M \widehat{M}^Q = N \otimes_E \widehat{M}^Q = N \otimes_E (\widehat{K(t)} \otimes_{E^*} \widehat{M}^Q) \cong (N \otimes_E \widehat{K(t)}) \otimes_{E^*} \widehat{M}^Q \cong \bigoplus_R \widehat{N}^R \otimes_{E^*} \widehat{M}^Q$. Thus $\widehat{N}^R \otimes_{E^*} \widehat{M}^Q$ is the direct product of the completions of $K(\alpha)$ with respect to the places that lie over Q and R. These fields are of degree $\operatorname{lcm}(e, \tilde{e})$ and the E^*-dimension of $\widehat{N}^R \otimes_{E^*} \widehat{M}^Q$ is $e \cdot \tilde{e}$. Thus there are $e \cdot \tilde{e} / \operatorname{lcm}(e, \tilde{e}) = \gcd(e, \tilde{e})$ places over Q and R. □

4.2. Decompositions of polynomials of degree p^2. We have already seen an example of decomposable polynomials of degree p^2: additive polynomials of degree p^2 have $p+1$ decompositions over a sufficiently large field (see Example 3.27). A classification of the polynomials of degree p^2 with at least two normal decompositions is given in the next theorem.

THEOREM 4.5. *Let F be a field of characteristic $p > 0$. Let f be a normal polynomial in $F[x]$ of degree p^2 with $f' \neq 0$ and at least two normal decompositions. Then exactly one of the following statements holds:*

(i) There is $w \in F$ and a divisor m of $p-1$ such that for each normal decomposition (g, h) of $(x - f(w)) \circ f \circ (x + w)$ there are a and b in $F^\times$ such that $g = x(x^\ell - a)^m$ and $h = x(x^\ell - b)^m$ with $\ell = (p-1)/m$.

(ii) There is $w \in F$ and an integer $1 < m < p - 1$ such that for each normal decomposition (g, h) of $(x - f(w)) \circ f \circ (x + w)$ there is $r \in \{m, p - m\}$, and $a, b \in F^\times$ such that $g = x^r(x - a)^{p-r}$, $h = x^{p-r}q$, and $h - a = (x - b)^r \tilde{q}$, where q and $\tilde{q}$ are squarefree polynomials of degree r and $p - r$, respectively.

We note that m depends only on f. Since w is a root of $f(x) - f(w)$, the polynomial $(x - f(w)) \circ f \circ (x + w)$ is a conjugate of f. We will see that w is unique in Case (i) and there are two alternative values for w in Case (ii). Additive polynomials fall into Case (i) with $m = 1$. The proof of this theorem will take the rest of this section.

Let (g_1, h_1) and (g_2, h_2) be two normal decompositions of f. Then there are two intermediate fields M and N of $K(\alpha) \mid K(t)$ that correspond to (g_1, h_1) and (g_2, h_2), respectively. Throughout this section let Q, R, and S denote places in M, N, and $K(\alpha)$, respectively. We have that $M = K(h_1(\alpha))$ and $g_1 - t$ is the minimal polynomial of $h_1(\alpha)$ over $F(t)$. Thus $\sum_{Q \text{ finite}} d(Q \mid P) = \deg(g_1')$, by Lemma 4.3. Since $h_1 - h_1(\alpha)$ is the minimal polynomial of α over M we have $\sum_{S \text{ finite}} d(S \mid Q) = \deg(h_1')$. The analog holds for N. Figure 4.1 illustrates the relation between this fields.

First we will show that we are in the situation in which we can apply Lemma 4.4. Then Corollary 4.10, Lemma 4.12 (which are similar to results in Dorey & Whaples (1974)) and Lemma 4.11 (which is due to Zannier (1993)) will tell us more about the ramification indices in M and N. From this we can make a case distinction, whether there is an unramified place in N over a certain place in $K(t)$ or not. This will lead to the two cases in the theorem.

CLAIM 4.6. *$MN = K(\alpha)$.*

Clearly $M \subseteq MN \subseteq K(\alpha)$. If $MN = M$ then $N \subseteq M$, which can not be since $h_1 \neq h_2$. But since $[K(\alpha)\colon MN] \mid [K(\alpha)\colon M] = p$ we have $[K(\alpha)\colon MN] = 1$.

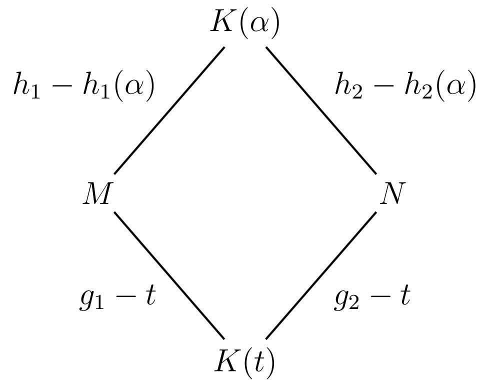

Figure 4.1

Claim 4.7. *There is no finite place in $K(t)$ that is wildly ramified in $K(\alpha)$.*

Assume for contradiction that P is wildly ramified, that is $p \mid e(S \mid P) = e(S \mid Q)e(Q \mid P)$ for $Q = S \cap M$. But then $p \mid e(S \mid Q)$ or $p \mid e(Q \mid P)$. Therefore we have $h_1 - b = (X - a)^p$ or $g_1 - b = (X - a)^p$, which is a contradiction to the assumption $f' \neq 0$.

Now we can apply Lemma 4.4. As in Dorey & Whaples (1974) we need the notion of extra places.

Definition 4.8. *Define*

$$i(P, N \mid K(t)) = \sum_{R|P} d(R \mid P)$$

and

$$i(P, K(\alpha) \mid M) = \sum_{S|P} d(S \mid S \cap M).$$

Call P extra *in N if $i(P, N \mid K(t)) > i(P, K(\alpha) \mid M)$.*

By Proposition 6.7 in von zur Gathen *et al.* (2010) we have $\deg_2(h_1) = \deg_2(g_2)$, where $\deg_2$ denotes the second degree (that is $\deg(f - ax^{\deg(f)})$ for a polynomial f with leading coefficient a). Since the degree of h_1 and g_2 is p we have that the second degree is the degree of the derivative plus one and thus $\deg(h_1') = \deg(g_2')$. Then we get $d(R_\infty \mid P_\infty) = 2p - 2 - \deg(g_2') = 2p - 2 - \deg(h_1') = d(S_\infty \mid Q_\infty)$, which proves that P_∞ cannot be extra in N.

Let P be a finite place in $K(t)$ and let Q and R be places over P in M and in N, respectively. Set $e = e(Q \mid P)$ and $\tilde{e} = e(R \mid P)$. For a place S over Q and R we have $e(S \mid P) = e(S \mid Q) \cdot e$ and $e(S \mid Q) = e(S \mid P)/e = \operatorname{lcm}(e, \tilde{e})/e$. Thus

$\sum_{S|R} d(S \mid Q) = \sum(\mathrm{lcm}(e,\tilde{e})/e-1) = \gcd(e,\tilde{e})\cdot(\mathrm{lcm}(e,\tilde{e})/e-1) = \tilde{e}-\gcd(e,\tilde{e})$. We define

$$c(Q,R) = \sum_{S|R} d(S \mid Q) = \tilde{e} - \gcd(e,\tilde{e})$$

and note that $i(P, K(\alpha) \mid M) = \sum_{Q,R} c(Q,R)$.

Lemma 4.9. *Let P be a finite place in $K(t)$. Then $\sum_{Q|P} c(Q,R) \geq e(R \mid P)-1$ for all places R over P.*

Proof. Let $P = \prod_{i=0}^{l} Q_i^{e_i}$ in M. For $d = \gcd(e_i \mid 0 \leq i \leq l)$ we have $d \mid \sum e_i = p$. If $d > 1$ then we would have $d = p$ and P would be wildly ramified in M, which cannot be. Thus $d = 1$.

Let R be a place over P with ramification index $\tilde{e} = e(R \mid P)$. Then as above we have $c(Q_i, R) = \tilde{e} - \gcd(e_i, \tilde{e})$. If $\tilde{e} = 1$ we have $\sum_i c(Q_i, R) = \sum_i(\tilde{e} - \gcd(e_i, \tilde{e})) = 0 = \tilde{e} - 1$. Thus assume $\tilde{e} > 1$. Then $\tilde{e}$ cannot divide e_i for all i, since their gcd is one. We distinguish two cases:

Case 1: $\tilde{e}$ divides all but one places Q over P in M. Then let Q_0 be the place such that $\tilde{e} \nmid e_0$. The gcd of $\tilde{e}$ and e_0 divides $\tilde{e}$ and thus divides all ramification indies of places over P in M. But their gcd is one. Thus the gcd of $\tilde{e}$ and e_0 is one and we have $\sum_{Q|P} c(Q,R) \geq c(Q_0, R) = \tilde{e} - 1$.

Case 2: There are at least two places, say Q_1 and Q_2, over P in M such that $\tilde{e} \nmid e_i$ for $i = 1, 2$. Then we have $\tilde{e}/\gcd(e_i, \tilde{e}) > 1$ since else we would have $\tilde{e} \mid e_i$, which is a contradiction. Thus $\gcd(e_i, \tilde{e}) \leq \tilde{e}/2$. Hence $\tilde{e} - \gcd(e_i, \tilde{e}) \geq \tilde{e}/2$ and thus $\sum_i c(Q_i, R) \geq c(Q_1, R) + c(Q_2, R) \geq \tilde{e} > \tilde{e} - 1$, as claimed. □

Corollary 4.10. *There is no finite place in $K(t)$ which is extra in N.*

Proof. Let P be a finite place. By Lemma 4.9 we have $\sum_{Q|P} c(Q,R) \geq e(R \mid P) - 1$ for all R. But then $i(P, K(\alpha) \mid M) = \sum_{Q,R} c(Q,R) \geq \sum_R (e(R \mid P) - 1) = i(P, N \mid K(t))$ which shows that P is not extra in N. □

As seen in Section 4.1 we have by the Hurwitz Genus Formula

$$\sum_P i(P, N \mid K(t)) = \sum_{R|P} d(R \mid P) = 2p - 2 = \sum_P i(P, K(\alpha) \mid M).$$

But since there are no extra places in N we get $i(P, N \mid K(t)) = i(P, K(\alpha) \mid M)$ for all places P.

LEMMA 4.11. *Let P be a finite place in $K(t)$. Then the following statements hold:*

(i) For each ramified place R over P in N the ramification index $e(R \mid P)$ divides $e(Q \mid P)$ for all but one place Q over P in M.

(ii) For each ramified place Q over P in M the ramification index $e(Q \mid P)$ divides $e(R \mid P)$ for all but one place R over P in N.

(iii) P is ramified in M if and only if it is ramified in N.

PROOF. To prove (i), we claim that the second case in the proof of Lemma 4.9 does not occur. For R falling into Case 2 we had seen that $\sum_{Q|P} c(Q,R) \geq e(R \mid P)$. Set $\varepsilon(R) = 1$ if R falls into this case and $\varepsilon(R) = 0$ else. Then we have in any case $\sum_{Q|P} c(Q,R) \geq e(R \mid P) - 1 + \varepsilon(R)$. Hence we get $i(P, K(\alpha) \mid M) = \sum_{Q,R} c(Q,R) \geq \sum_R e(R \mid P) - 1 + \varepsilon(R) = i(P, N \mid K(t)) + \sum_R \varepsilon(R)$. But since $i(P, N \mid K(t)) = i(P, K(\alpha) \mid M)$ we have $\sum_R \varepsilon(R) = 0$. This proves the claim.

The second statement can be proven analogously to the first one, by interchanging the role of M and N in the previous results.

Finally, if P is ramified in N then by (i) there is a place R with $1 < e(R \mid P) \mid e(Q \mid P)$ for some place Q in M. Thus P is ramified in M. The other direction follows in the same way from (ii). □

LEMMA 4.12. *There is at most one finite place in $K(t)$ that is ramified in M. Moreover if there is a place that is ramified in M then it has at most one unramified factor.*

PROOF. Let P be a finite place in $K(t)$, which is ramified in M. Assume there is a place Q such that $e(Q \mid P) = 1$. Then $\sum_{S|Q} d(S \mid Q) = \sum_R c(Q,R) \geq \sum_R e(R \mid P) - \gcd(1, e(R \mid P)) = \sum_R (e(R \mid P) - 1) = i(P, N \mid K(t))$. If there are two unramified places Q_1 and Q_2 then $i(P, K(\alpha) \mid M) \geq \sum_{S|Q_1} d(S \mid Q_1) + \sum_{S|Q_2} d(S \mid Q_2) \geq 2i(P, N \mid K(t))$. But since $i(P, K(\alpha) \mid M) = i(P, N \mid K(t))$ this can only be if $i(P, N \mid K(t)) = 0$. Hence P is unramified in N and by Lemma 4.11 unramified in M, in contradiction to our assumption. Thus there can be at most one unramified place over P. If ρ denotes the number of places in M over P we have $1 + 2(\rho - 1) \leq \sum e(Q \mid P) = p$ and thus $\rho \leq (p+1)/2$. Therefore $i(P, M \mid K(t)) = p - \rho \geq (p-1)/2$. But since $\sum_{P \text{ finite}} i(P, M \mid K(t)) = \deg(g_1') < p - 1$ there can be at most one such a place in $K(t)$. □

Setting $k = \deg_2(g_1)$, we have $k = \deg_2(g_i) = \deg_2(h_i)$ for $i = 1, 2$ (see Proposition 6.7 in von zur Gathen *et al.* (2010)). We have a special case when $k = 1$. Then $\sum_{P \text{ finite}} i(P, M \mid K(t)) = \deg(g_1') = k - 1 = 0$ and thus there are no ramified places. But since the second degree of the decompositions equals one, we have in this case that g_1 and h_1 are of the form $x^p - ax$ and $x^p - bx$, respectively. Thus f is an additive polynomial and we are in the Case (i) of Theorem 4.5.

We assume $k > 1$. Then we have $\sum_{P \text{ finite}} i(P, M \mid K(t)) = k - 1 > 0$ and thus there is a finite place P which is ramified in M. By Lemma 4.12 we have P is the only finite place which is ramified in M and thus $i(P, M \mid K(t)) = k - 1$. By Lemma 4.11 and by interchanging M and N in Lemma 4.12 we get that P is the only place that is ramified in N. Then we have $i(P, M \mid K(t)) = p - \rho_1 = k - 1 = p - \rho_2 = i(P, N \mid K(t))$, where ρ_1 and ρ_2 are the numbers of places over P in M and N, respectively. Thus $\rho_1 = \rho_2 = p - k + 1 = \ell + 1$ with $\ell = p - k$.

By the correspondence between ramification and multiplicities we get that there is exactly one c in K such that $f - c$ has multiple roots. Then for each automorphism σ of K that fixes F we have $f - \sigma(c) = \sigma(f - c)$ has multiple roots and thus $\sigma(c) = c$. This proves that c is in F. We will later see that there is also a root $w \in F$ of $f - c$, that is $f(w) = c$. Thus the conjugate $(x - f(w)) \circ f \circ (x + w)$ lies in $F[x]$ and has multiple roots.

Now assume that there is an unramified place R_0 over P in N. We will prove that f falls into Case (i) of Theorem 4.5.

Let R_i be the ramified places over P in N with ramification indices $\tilde{e}_i$, for $0 \leq i \leq \ell$. Then let e_0 denote the ramification index in M that is not divided by $\tilde{e}_1$. If $e_0 \neq 1$ then e_0 cannot divide $\tilde{e}_0 = 1$ and thus it must divide $\tilde{e}_1$. But this would imply that e_0 divides all ramification indices in M, which cannot be. Thus we get $e_0 = 1$. Then $\tilde{e}_i$ divides all ramification indices in M that are greater than 1, and the other way round. Thus all of this ramification indices equal, say we have $m = e_i = \tilde{e}_i$ for all $1 \leq i \leq \ell$.

Then we get that $g_1 - c$ is of the form $(x - \tilde{a})\tilde{g}^m$, where $\tilde{a}$ is in K and $\tilde{g}$ is a squarefree polynomial of degree ℓ. We have that $g_1 - c$ is a polynomial over F and the irreducible factors of $g_1 - c$ over F have only simple roots. Thus $\tilde{g}$ is defined over F and $\tilde{a}$ is in F.

Since $e(S \mid P) = \operatorname{lcm}(e(Q \mid P), e(R \mid P)) = m$ or $= 1$ we get $e(S \mid Q) = m$ if and only if $e(Q \mid P) = 1$ and $e(R \mid P) = m$. Thus only Q_0 is ramified in $K(\alpha) \mid M$ and has the same ramification like P in $N \mid K(t)$. Thus as above $h_1 - \tilde{a}$ is of the form $(x - w)\tilde{h}^m$, where $w \in F$ and $\tilde{h}$ is a squarefree polynomial of degree ℓ. Now we conjugate as follows: $(x - c) \circ f \circ (x + w) = (x - c) \circ g_1 \circ (x + \tilde{a}) \circ (x - \tilde{a}) \circ h_1 \circ (x + w) = x\hat{g}^m \circ x\hat{h}^m$, with $\hat{g} = \tilde{g} \circ (x + \tilde{a})$

and $\hat{h} = \tilde{h} \circ (x + w)$.

To see that we are in Case (i) of Theorem 4.5 we have to prove that $\hat{g}$ and $\hat{h}$ are of the appropriate form. For a polynomial $g = x\hat{g}^m$ the derivative of g is $g' = \hat{g}^{m-1}(\hat{g} + mx\hat{g}')$. If a_i are the coefficients of $\hat{g}$ then we have $\hat{g} + mx\hat{g}' = \sum_i (a_i + mia_i)x^i$. Let ℓ' be the degree of $\hat{g} + mx\hat{g}'$. Then $k - 1 = \deg(g') = (m-1)\ell + \ell' = m\ell - \ell + \ell' = p - \ell - 1 + \ell' = k - 1 + \ell'$ and thus $\ell' = 0$. But this means that we have $a_i + mia_i = 0$ for all $1 \leq i \leq \ell$. For $1 \leq i < \ell$ this is the case only if $a_i = 0$. Thus we get $\hat{g} = (x^\ell - a)$ and g is of the form as claimed.

Now we consider the other case, where there is no unramified place over P in N. Then there is neither one in M. To prove that we are actually in Case (ii), we first prove that $\ell = 1$. For this propose we translate Lemma 4.11 into the language of graphs. Let $V = A \cup B$ be the set of vertices, where $A = \{e_i : 0 \leq i \leq \ell\}$ and $B = \{\tilde{e}_i : 0 \leq i \leq \ell\}$. Let the set of edges E consist of $(e_i, \tilde{e}_j)$ if $e_i | \tilde{e}_j$ and of $(\tilde{e}_i, e_j)$ if $\tilde{e}_i | e_j$. Then this yields a directed bipartite graph with outdegree $\delta(v) = \ell$ for each $v \in V$. Note that if there is a vertex in A which is connected to all other vertices in A then we get that the gcd of all e_i is greater then one, which is a contradiction.

Lemma 4.13. *Let $G = (V, E)$ be a directed bipartite graph, with bipartition $V = A \cup B$. Assume A and B have the same cardinality $\ell + 1 > 2$ and the outdegree of each vertex equals ℓ. Then there is $a \in A$ such that a is connected to all vertices in A or there is $b \in B$ such that b is connected to all vertices in B.*

Proof. Assume that there are a_0 and a_1 in A such that there is no a_1-a_0 path (if such a_0 and a_1 would not exist we would be done). Since the outdegree of a_1 is ℓ there is b_0 in B such that $(a_1, b_0) \notin E$. Then for all b in $B' = B \setminus \{b_0\}$ there is a edge (a_1, b) and thus no edge (b, a_0). If now (b_0, a_1) is in G then b_0 is connected to all b in B' via a_1 and we would be done. Thus we assume that (b_0, a_1) is not in G.

We claim that $G' = G \setminus \{a_0, b_0\}$ is a complete bipartite graph. We have already seen that for each $b \in B'$ the edge (b, a_0) is not in G. Thus b has outdegree ℓ in G'. Let a be in $A' = A \setminus \{a_0\}$. Then a has outdegree ℓ in G and it would have less outdegree in G' only if there would be (a, b_0) in G. Then for b in B' we get that $(a_1, b), (b, a), (a, b_0), (b_0, a_0)$ is a a_1-a_0 path, which is a contradiction to our assumption. Thus G' is complete.

Since the outdegree of a_0 is $\ell > 1$ there is a b_1 in B' such that (a_0, b_1) is in G. Let b be any vertex in B'. Since G' is complete there is a b_1-b path p. But then $(b_0, a_0), (a_0, b_1), p$ is a b_0-b path and thus b_0 is connected to all vertices in B. □

By the previous discussion we know that $\ell = 1$ and therefore P splits exactly into two ramified places in both fields, M and N, say $P = Q_0^m Q_1^{p-m}$ in M and $P = R_0^m R_1^{p-m}$ in N, with $1 < m < p-1$. But this means that $g_1 - c$ is of the form $(x-a)^m (x-\tilde{a})^{p-m}$, for suitable $a, \tilde{a} \in F$.

Now there are $\gcd(m, m) = m$ places S over Q_0 and R_0. For such a place we have $e(S \mid Q_0) = \operatorname{lcm}(m, m)/m = 1$. Furthermore there is one place S over Q_0 and R_1 with $e(S \mid Q_0) = \operatorname{lcm}(m, p-m)/m = p-m$. Thus $h_1 - a$ is of the form $(x-b)^{p-m} q$, where q is a squarefree polynomial of degree m. Similarly we get $h_1 - \tilde{a} = (x - \tilde{b})^m \tilde{q}$. It is left to prove that a, $\tilde{a}$, b, and $\tilde{b}$ are in F. Assume a and $\tilde{a}$ would not be in F. Then $(x-a)(x-\tilde{a})$ must be an irreducible factor of $g_1 - c$. But then $m = p - m$ which is a contradiction (note that for $p = 2$ there is anyway no such m). Similarly we get that b and $\tilde{b}$ are in F. By conjugating f with $x + w$ for $w \in \{b, \tilde{b}\}$, we achieve the desired form.

Finally we note that the case, in which the polynomial falls, depends on whether there is an unramified place over P or not. Thus the two cases are distinct. This finishes the proof of Theorem 4.5.

4.3. Parametrization. The polynomials examined in Section 6 in von zur Gathen *et al.* (2010) fall into Case (i) of Theorem 4.5.

THEOREM 4.14. *For parameters $\varepsilon \in \{0, 1\}$, u, $s \in \mathbb{F}^\times$ and ℓ a positive divisor of $p-1$ let f be the polynomial*

$$f = x(x^{\ell(p+1)} - \varepsilon u s^r x^\ell + u s^{p+1})^m,$$

with $m = (p-1)/\ell$. Then f has for each root t of $x^{p+1} - \varepsilon u x + u$ in F a decomposition (g, h) with $g = x(x^\ell - u s^p t^{-1})^m$ and $h = x(x^\ell - st)^m$. These decompositions are pairwise distinct and there are no other possible decompositions of f.

PROOF. See Theorem 6.2 in von zur Gathen *et al.* (2010). □

COROLLARY 4.15. *All polynomials which fall into Case (i) of Theorem 4.5 can be parametrized as in Theorem 4.14.*

PROOF. We have $g = x(x^\ell - a)^m$ and $h = x(x^\ell - b)^m$ for suitable a and b in $F^\times$. Then we get $f = x(x^\ell - b)^m((x(x^\ell - b)^m)^l - a)^m = x(x^{\ell(p+1)} - (b^p + a)x^\ell + ab)^m$. Now define $\varepsilon = 0$ if $b^p + a = 0$ else define $\varepsilon = 1$. In case $\varepsilon = 0$ we set $s = 1$, $t = b$, and $u = ab$. If $\varepsilon = 1$ we set $s = ab(b^p + a)^{-1}$, $t = b/s$, and $u = ab/s^{p+1}$. In both cases we have that u, s, and t are in F and the equations $t^{p+1} - \varepsilon ut + u = 0$, $b = st$, and $a = us^p t^{-1}$ hold as claimed. □

Note that if the field F is large enough, the polynomial $x^{p+1} - \varepsilon ux + u$ has $p + 1$ roots and thus f has $p + 1$ (minimal) decompositions. This is another example, that shows that the bounds in Section 3.4 are sharp. Still open is a closer examination of the polynomials falling into Case (ii).

References

M. D. ATKINSON (1975). An algorithm for finding the blocks of a permutation group. *Mathematics of Computation* **29**, 911–913. ISSN 0378-4754.

A. BOSTAN, G. LECERF, B. SALVY, É. SCHOST & B. WIEBELT (2004). Complexity issues in bivariate polynomial factorization. In *ISSAC 2004: Proceedings of the 2004 International Symposium on Symbolic and Algebraic Computation*, RUDOLF FLEISCHER & GERHARD TRIPPEN, editors, 42–49. ACM. URL http://dx.doi.org/10.1145/1005285.1005294.

GREG BUTLER (1992). An analysis of Atkinson's algorithm. *SIGSAM Bull.* **26**, 1–9. ISSN 0163-5824. URL http://doi.acm.org/10.1145/130933.130935.

F. DOREY & G. WHAPLES (1974). Prime and Composite Polynomials. *Journal of Algebra* **28**, 88–101. URL http://dx.doi.org/10.1016/0021-8693(74)90023-4.

MICHAEL D. FRIED & R. E. MACRAE (1969). On the invariance of chains of Fields. *Illinois Journal of Mathematics* **13**, 165–171.

JOACHIM VON ZUR GATHEN (1990a). Functional Decomposition of Polynomials: the Tame Case. *Journal of Symbolic Computation* **9**, 281–299. URL http://dx.doi.org/10.1016/S0747-7171(08)80014-4.

JOACHIM VON ZUR GATHEN (1990b). Functional Decomposition of Polynomials: the Wild Case. *Journal of Symbolic Computation* **10**, 437–452. URL http://dx.doi.org/10.1016/S0747-7171(08)80054-5.

JOACHIM VON ZUR GATHEN (2009). The Number of Decomposable Univariate Polynomials — Extended Abstract. 359–366. ISBN 978-1-60558-609-0. Preprint (2008) at http://arxiv.org/abs/0901.0054.

JOACHIM VON ZUR GATHEN & JÜRGEN GERHARD (1999). *Modern Computer Algebra*. Cambridge University Press, Cambridge, UK, First edition. ISBN 0-521-64176-4. URL http://cosec.bit.uni-bonn.de/science/mca/. Other available editions: Second edition 2003, Chinese edition, Japanese translation.

JOACHIM VON ZUR GATHEN, MARK GIESBRECHT & KONSTANTIN ZIEGLER (2010). Composition collisions and projective polynomials. Statement of results. Preprint available at http://arxiv.org/abs/1005.1087.

S. LANDAU & G. L. MILLER (1985). Solvability by Radicals is in Polynomial Time. *Journal of Computer and System Sciences* **30**, 179–208.

G. Lecerf (2008). Fast Separable Factorization and Applications. *Applicable Algebra in Engineering, Communication and Computing* **19**(2), 135–160. URL `http://www.springerlink.com/content/75430606n6r52k67/fulltext.pdf`.

Jürgen Neukirch (2007). *Algebraische Zahlentheorie*. Springer-Verlag, Berlin. ISBN 3-540-37547-3.

J. F. Ritt (1922). Prime and Composite Polynomials. *Transactions of the American Mathematical Society* **23**, 51–66. URL `http://www.jstor.org/stable/1988911`.

Henning Stichtenoth (1993). *Algebraic Function Fields and Codes*. Springer-Verlag, 260 pages .

Helmut Wielandt (1964). *Finite permutation groups*. Academic Press, New York. ISBN 0-127-49656-4. Translated from the German by R. Bercov.

U. Zannier (1993). Ritt's Second Theorem in arbitrary characteristic. *Journal für die reine und angewandte Mathematik* **445**, 175–203. URL `http://www.digizeitschriften.de/index.php?id=loader&tx_jkDigiTools_pi1[IDDOC]=503382`.

Richard Zippel (1991). Rational Function Decomposition. In *Proceedings of the 1991 International Symposium on Symbolic and Algebraic Computation ISSAC '91*, Bonn, Germany, Stephen M. Watt, editor, 1–6. ACM Press, Bonn, Germany. ISBN 0-89791-437-6.

Raoul Blankertz
b-it cosec, University of Bonn
Dahlmannstr. 2
D-53113 Bonn
Germany
`blankertz@bit.uni-bonn.de`