On Ritt’s decomposition theorem in the case of finite fields

Jaime Gutierrez\textsuperscript{a,*,1}, David Sevilla\textsuperscript{b,1}

\textsuperscript{a}Departamento de Matemáticas, Estadística y Computación, Universidad de Cantabria, E-39071 Santander, Spain

\textsuperscript{b}Department of Computer Science, Concordia University, 1455 de Maisonneuve W., Montreal, Canada H3G 1M8

Received 7 July 2004; revised 18 July 2005
Communicated by Gary L. Mullen

Abstract

A classical theorem by Ritt states that all the complete decomposition chains of a univariate polynomial satisfying a certain tameness condition have the same length. In this paper we present our conclusions about the generalization of these theorem in the case of finite coefficient fields when the tameness condition is dropped.

© 2005 Published by Elsevier Inc.

Keywords: ■; ■; ■

1. Introduction

Our starting point is the decomposition of polynomials and rational functions in one variable. First, we define the basic concepts of this topic.
Definition 1. Let $\mathbb{K}$ be any field, $x$ a transcendental over $\mathbb{K}$ and $\mathbb{K}(x)$ the field of rational functions in the variable $x$ with coefficients in $\mathbb{K}$. In the set $T = \mathbb{K}(x) \setminus \mathbb{K}$ we define the binary operation of composition as

$$ g(x) \circ h(x) = g(h(x)) = g(h). $$

We have that $(T, \circ)$ is a semigroup, the element $x$ being its neutral element.

If $f = g \circ h$, we call this a decomposition of $f$ and say that $g$ is a component on the left of $f$ and $h$ is a component on the right of $f$. We call a decomposition trivial if any of the components is a unit with respect to decomposition.

Given two decompositions $f = g_1 \circ h_1 = g_2 \circ h_2$ of a rational function, we call them equivalent if there exists a unit $u$ such that

$$ h_1 = u \circ h_2 \quad (\text{thus}, g_1 = g_2 \circ u^{-1}), $$

where the inverse is taken with respect to composition.

Given $f \in T$, we say that it is indecomposable if it is not a unit and all its decompositions are trivial.

We define a complete decomposition of $f \in \mathbb{K}(x)$ to be $f = g_1 \circ \cdots \circ g_r$ where every $g_i$ is indecomposable. The notion of equivalent complete decompositions is straightforward from the previous concepts.

Definition 2. Given a non-constant rational function $f(x) \in \mathbb{K}(x)$ where $f(x) = f_N(x)/f_D(x)$ with $f_N, f_D \in \mathbb{K}[x]$ and $(f_N, f_D) = 1$, we define the degree of $f$ as

$$ \deg f = \max\{\deg f_N, \deg f_D\}. $$

We also define $\deg a = 0$ when $a \in \mathbb{K}$.

From now on, we will use the previous notation when we refer to the numerator and denominator of a rational function. Unless explicitly stated, we will take the numerator to be monic, even though multiplication by constants will not be relevant.

Now, we can properly state the problem of decomposition of univariate rational functions, although this will not be our main object of study.

Problem 3. Given a univariate rational function, decide if it is decomposable, and in the affirmative case compute a non-trivial decomposition of the function.

It is clear that the solution of this problem provides the computability of a complete decomposition of a function if it exists.

Next, we introduce some basic results about univariate decomposition, see [1] for more details.
Lemma 4. (i) For every \( f \in T \), \( \deg f = [\mathbb{K}(x) : \mathbb{K}(f)] \).

(ii) \( \deg(g \circ h) = \deg g \cdot \deg h \).

(iii) \( f(x) \) is a unit with respect to composition if and only if \( \deg f = 1 \), that is, \( f(x) = \frac{ax + b}{cx + d} \) with \( a, b, c, d \in \mathbb{K} \) and \( ad - bc \neq 0 \).

(iv) Every non-constant element of \( \mathbb{K}(x) \) is cancelable on the right with respect to composition. In other words, if \( f(x), h(x) \in T \) are such that \( f(x) = g(h(x)) \) then \( g(x) \) is uniquely determined by \( f(x) \) and \( h(x) \).

We can relate decomposition and Field Theory by means of the following classical result:

Theorem 5 (Lüroth’s Theorem). Let \( \mathbb{F} \) be a field such that \( \mathbb{K} \subset \mathbb{F} \subset \mathbb{K}(x) \). Then there exists \( f \in \mathbb{K}(x) \) such that \( \mathbb{F} = \mathbb{K}(f) \). Also, if \( \mathbb{F} \) contains a polynomial, \( f \) can be chosen to be a polynomial.

Proof. See for example [9] for a proof in the case \( \mathbb{K} = \mathbb{C} \), [15] for one in the general case and [16] for an elementary one. Constructive proofs can be found in [10,13,1]. \( \square \)

Now, we state one of the classical Ritt’s theorems (see [11]) about the relations among the complete decompositions of a polynomial that satisfies a certain condition. First, we have to define that condition.

Definition 6. A polynomial \( f \in \mathbb{K}[x] \) is tame when \( \text{char} \mathbb{K} \) does not divide \( \deg f \).

Ritt’s theorem essentially proves that all the decompositions have the same length and are related in a rather direct way.

Definition 7. A bidecomposition is a 4-tuple of polynomials \( f_1, g_1, f_2, g_2 \) such that \( f_1 \circ g_1 = f_2 \circ g_2 \), \( (\deg f_1, \deg g_1) = 1 \) and \( \deg f_1 = \deg g_2 \).

Theorem 8 (Ritt’s Theorem). Let \( f \in \mathbb{K}[x] \) be tame and let \( f = g_1 \circ \cdots \circ g_r = h_1 \circ \cdots \circ h_s \) be two complete decompositions of \( f \). Then \( r = s \), and the sequences \( (\deg g_1, \ldots, \deg g_r) \) and \( (\deg h_1, \ldots, \deg h_s) \) are permutations of each other. Moreover, there exists a finite chain of complete decompositions

\[ f = f_1^{(j)} \circ \cdots \circ f_r^{(j)}, \quad j \in \{1, \ldots, k\}, \]

such that

\[ f_i^{(1)} = g_i, \quad f_i^{(k)} = h_i, \quad i = 1, \ldots, r, \]
and for each \( j < k \), there exists \( i_j \) such that the \( j \)th and \((j+1)\)th decomposition differ only in one of these aspects:

1. \( f_{i_j}^{(j)} \circ f_{i_{j+1}}^{(j)} \) and \( f_{i_j}^{(j+1)} \circ f_{i_{j+1}}^{(j+1)} \) are equivalent.
2. \( f_{i_j}^{(j)} \circ f_{i_{j+1}}^{(j)} = f_{i_j}^{(j+1)} \circ f_{i_{j+1}}^{(j+1)} \) is a bidecomposition.


In this paper, we will study the generalization of this result to polynomials with coefficients in finite fields. To that end, we will also analyze the structure of intermediate fields between \( K(f) \) and \( K(x) \). It is already known that Ritt’s theorem is false when the tameness condition is dropped, see [4] for a counterexample.

Let \( f = g(h) \). Then \( f \in K(h) \), thus \( K(f) \subset K(h) \). Also, \( K(f) = K(h) \) if and only if \( f = u \circ h \) for some unit \( u \). This allows the following bijection among decompositions of a function \( f \) and fields between \( K(f) \) and \( K(x) \):

**Theorem 9.** Let \( f \in K(x) \). In the set of decompositions of \( f \) we have the equivalence relation given by the definition of equivalence of decompositions. If we denote as \( [(g, h)] \) the class of the decomposition \( f = g(h) \), the we have then the bijection:

\[
\{ [(g, h)] : f = g(h) \} \longleftrightarrow \{ F : K(f) \subset F \subset K(x) \} \quad \longleftrightarrow \quad F = K(h).
\]

Thanks to the Primitive Element Theorem (see for example [7]), we know that for each non-constant \( f \in K(x) \) there exist finitely many fields between \( K(f) \) and \( K(x) \). Due to the second part of Lüroth’s Theorem, every rational decomposition of a polynomial is equivalent to a decomposition whose components are polynomials. Therefore, it suffices to care about polynomial decomposition in this case.

In Section 2, we introduce several elementary results about univariate function fields that arise from Galois theory. In Section 3, we present a function that is fixed by all the automorphisms of a univariate function field over a finite field and several results related to it. In particular, we provide an essentially new counterexample of Ritt’s theorem for finite coefficient fields.

### 2. The fixing group and the fixed field

In this section, we introduce several simple notions from the classical Galois theory. Let \( \Gamma(K) = \text{Aut}_K K(x) \) (we will write simply \( \Gamma \) if there can be no confusion about the field). The elements of \( \Gamma(K) \) can be identified with the images of \( x \) under the automorphisms, that is, with Möbius transformations (non-constant rational functions...
of the form \((ax + b)/(cx + d) \in \mathbb{K}(x)\), which are also the units of \(\mathbb{K}(x)\) under composition.

**Definition 10.**
- Let \(f \in \mathbb{K}(x)\). We define \(G(f) = \{u \in \Gamma(\mathbb{K}) : f \circ u = f\}\).
- Let \(H < \Gamma(\mathbb{K})\). We define \(\text{Fix}(H) = \{f \in \mathbb{K}(x) : f \circ u = f \ \forall u \in H\}\).

This definitions correspond to the classical Galois correspondences (not bijective in general) between the intermediate fields of an extension and the subgroups of its automorphism group, as the following diagram shows:

\[
\begin{array}{c}
\mathbb{K}(x) \leftrightarrow \{\text{id}\} \\
\mathbb{K}(f) \rightarrow G(f) \\
\text{Fix}(H) \leftrightarrow H \\
\mathbb{K} \leftrightarrow \Gamma
\end{array}
\]

**Remark 11.** As \(\mathbb{K}(f) = \mathbb{K}(f')\) if and only if \(f = u \circ f'\) for some unit \(u\), we have that the application \(\mathbb{K}(f) \mapsto G(f)\) is well-defined.

We are interested in the computability of these elements, the following results solves one of the two parts of this question.

**Theorem 12.** Let \(H = \{h_1, \ldots, h_m\} \subset \mathbb{K}(x)\) be a finite subgroup of \(\Gamma\). Let \(P(T) = \prod_{i=1}^{m} (T - h_i) \in \mathbb{K}(x)[T]\). Then any non-constant coefficient of \(P(T)\) generates \(\text{Fix}(H)\).

**Sketch of proof.** It can be shown that \(P(T)\) is the minimal polynomial of \(x\) over \(\text{Fix}(H) \subset \mathbb{K}(x)\). Then, a known proof of Lüroth’s theorem (see [10]) gives the desired result. □

The previous theorem obviously provides an algorithm to compute the fixed field for a given finite subgroup of \(\Gamma\): compute the symmetric elementary functions in \(h_1, \ldots, h_m\) until a non-constant one is found.

About the computation of the fixing group, an elementary but inefficient algorithm is given by the resolution of the equations given by

\[
f(x) - f\left(\frac{ax + b}{cx + d}\right) = 0
\]

in terms of \(a, b, c, d\). Another algorithm (see [14]) combines this idea with certain normalization of the rational function, which simplifies the equations substantially.
Next, we state several interesting properties of the fixed field and the fixing group, see [14] for details.

**Theorem 13.** Let $H < \Gamma$.

- $H$ is infinite $\Rightarrow$ Fix$(H) = K$.
- $H$ is finite $\Rightarrow$ $K \subseteq$ Fix$(H)$, Fix$(H) \subset K(x)$ is a normal extension, and in particular $\text{Fix}(H) = K(f)$ with $\deg f = |H|$.

**Theorem 14.**

(i) Given a non-constant $f \in K(x)$, $|G(f)|$ divides $\deg f$. Moreover, for any field $K$ there is always a function $f \in K(x)$ such that $1 < |G(f)| < \deg f$.

(ii) $|G(f)| = \deg f \Rightarrow K(f) \subseteq K(x)$ is normal. Moreover, if the extension $K(f) \subseteq K(x)$ is separable, then $K(f) \subseteq K(x)$ is normal $\Rightarrow |G(f)| = \deg f$.

(iii) Given a finite subgroup $H$ of $\Gamma$, there is a bijection between the subgroups of $H$ and the fields between $\text{Fix}(H)$ and $K(x)$. Also, if $\text{Fix}(H) = K(f)$, there is a bijection between the right components of $f$ (up to equivalence by units) and the subgroups of $H$.

**Proof.** For the first item, we take $f = x^2 (x - 1)^2$ gives $G(f) = \{x, 1 - x\}$. The other ones are straightforward. □

3. Finite fields

In this section, $K = \mathbb{F}_q$ where $q = p^m$ and $p = \text{char} \mathbb{F}_q$, see [8] for several useful results. As before, we will denote $\Gamma = \Gamma(\mathbb{F}_q)$.

**Definition 15.** For any $K$, $\Gamma_0 = \Gamma \cap K[x] = \{ax + b : a \in K^*, b \in K\}$.

**Theorem 16.** $K(x)$ is Galois over $K$ (that is, the only functions fixed by $\Gamma(K)$ are the constants) if and only if $K$ is infinite.

**Proof.** The “if” part is the first part of Theorem 13. The “only if” part is a consequence of Theorem 12, as $\Gamma(K)$ is finite whenever $K$ is finite. □

The interest of $\Gamma$ and $\Gamma_0$ in the case of finite fields lies in the fact that both groups provide non-trivial fixed fields.

**Theorem 17.** The fixed field for $\Gamma_0$ is generated by $(x^q - x)^{q-1}$.

**Proof.** According to Theorem 12 any non-constant coefficient of $Q(T) = \prod_{u \in \Gamma_0} (T - u)$ generates the field. But the constant term of $Q$ is precisely $\prod_{u \in \Gamma_0} u = (x^q - x)^{q-1}$. □
From now on, we will denote $P_q = (x^q - x)^{q-1}$.
As $\Gamma_0 \subset \Gamma$, if $f$ generates the fixed field for $\Gamma$ then $f = h(P_q)$ for some $h \in \mathbb{K}(x)$.
Moreover, $h$ has degree $[\Gamma : \Gamma_0] = q + 1$.

**Theorem 18.** Let

$$h_q = (x^{q+1} + x + 1)/x^q.$$  

Then the rational function $f_q = h_q(P_q)$ generates $\text{Fix}(\Gamma)$.

**Proof.** It is easy to prove that $\Gamma_0 \cup \{1/x\}$ generates $\Gamma$. As $f_q$ is a function of $P_q$ and its degree is equal to the order of the group, it suffices to show that $f_q(1/x) = f_q(x)$.

A simple computation shows that this is indeed the case: let $y = x^{q-1}$. Then $P_q(x) = y(y - 1)^{q-1}$ and $P_q(1/x) = (y - 1)^{q-1}/y^q$. Thus,

$$f_q(1/x) - f_q(x) = \frac{(y - 1)^{q-1}}{y^{q+q}} + \frac{(y - 1)^{q-1}}{y^q} + 1 - \frac{y^{q+1}(y - 1)^{q-1} + y(y - 1)^{q-1} + 1}{y^q(y - 1)^{q-2}}$$

$$= \frac{(y - 1)^{q-1} + y^q(y - 1)^{q-1} + y^{q+2} - y^{q+1}(y - 1)^{q-1} - y(y - 1)^{q-1}}{y^q(y - 1)^{q-2}}$$

$$= \frac{(y - 1)^{q-1}(1 - y^{q+1}) + (y - 1)^{q-1}(y^q - y) + y^{q+2} - 1}{y^q(y - 1)^{q-2}}$$

$$= \frac{(y - 1)^{q-1}(1 - y^{q+1}) + (y - 1)^{q-1}((y - 1)^2 - (y - 1)) + y^{q+2} - 1}{y^q(y - 1)^{q-2}}$$

$$= \frac{(y - 1)^{q-1}(1 - y^{q+1}) + (y - 1)^{q-1} + y^{q+2} - 1}{y^q(y - 1)^{q-2}}$$

$$= \frac{(y - 1)^{q-1}(1 - y^{q+1} + y - 1) - (y - 1)^{q+1} + y^{q+2} - 1}{y^q(y - 1)^{q-2}}$$

$$= \frac{(y - 1)^{q-1}(1 - y^{q+1} + y^q - 1) - (y - 1)^q + (y^{q+1} - 1)^q}{y^q(y - 1)^{q-2}}$$

$$= \frac{-(y - 1)^q y^q - (y - 1)^q (1 + y + \cdots + y^q)}{y^q(y - 1)^{q-2}}$$

$$= \frac{-(y - 1)^q y^q + (y - 1)^q (y + \cdots + y^q)}{y^q(y - 1)^{q-2}}$$
Let \( f \in \mathbb{F}_q(x) \). Let \( C = \{ \mathbb{K} : \mathbb{F}_q \subseteq \mathbb{K} \subseteq \mathbb{F}_q(x) \} \) and
\[
\phi : C \longrightarrow C
\]
\[
\mathbb{F}_q(f) \rightarrow \text{Fix}(G(f)) = \mathbb{F}_q(f')
\]
which is a well-defined application. Then it is easy to check that \( f' \) is a (not necessarily proper) right-component of \( f \). Also, as \( G(f) \subset \Gamma \), \( f' \) is a right-component of \( f_q \). Thus, \( \mathbb{F}_q(f) \subseteq \mathbb{F}_q(f') \) and \( \mathbb{F}_q(f_q) \subseteq \mathbb{F}_q(f') \).

On the other hand, the polynomial \( P_q \) has at least two different decompositions:
\[
P_q = x^{q-1} \circ (x^q - x) = \left( x(x - 1)^{q-1} \right) \circ x^{q-1}.
\]
This gives at least two decompositions for \( h_q \), both involving the component \( \frac{x^{q+1} + x + 1}{x^q} \).

**Theorem 19.**

(i) \( \frac{x^{q+1} + x + 1}{x^q} \) is indecomposable.

(ii) \( x^q - x \) is decomposable iff \( q \) is composite, that is, \( q = p^m \) with \( m \geq 2 \).

(iii) \( x(x - 1)^{q-1} \) is indecomposable.

**Proof.** (i) We will prove that for certain units \( u, v \in \mathbb{F}_q(x) \), the function
\[
u \circ \frac{x^{q+1} + x + 1}{x^q} \circ v
\]
is indecomposable. In particular, let \( u = x + 1, v = 1/(x - 1) \). Then
\[
u \circ \frac{x^{q+1} + x + 1}{x^q} \circ v = \frac{x^{q+1}}{x - 1}.
\]
As the degree is multiplicative with respect to composition, and so is the difference in the degrees of numerator and denominator (see [14, Theorem 1.14 and Corollary 1.15]), there is no possible decomposition for this function and the original function is also indecomposable.
This allows to prove the following theorem.

\[25\]

Let \(q = p^m\). Let \(x(x - 1)^{q-1} = g(h)\) with \(g = x^{p^r} + g_0\), \(\deg g_0 \leq p^r - 1\) and \(h = x^{p^r} + h_0\), \(\deg h_0 \leq p^s - 1\). Then

\[g \circ h = h^{p^r} + g_0 \circ h = (x^{p^r} + h_0)^{p^r} + g_0 \circ h = x^q + h_0^{p^r} + g_0 \circ h\]

with \(\deg h_0^{p^r} \leq q - p^r\) and \(\deg g_0 \circ h \leq q - p^s\). But

\[x(x - 1)^{q-1} = x^q + x^{q-1} + \cdots + x^2 + x,\]

thus either \(r = 0\) or \(s = 0\) and the decomposition is trivial. \(\square\)

**Corollary 20.** If \(q\) is not prime, \(P_q\) has two complete decomposition chains of different lengths.

As there is a bijection between the subgroups of \(\Gamma_0\) and the components of \((x^q - x)^{q-1}\) on the right, we will study those subgroups in order to determine whether this polynomial has complete decompositions of different length when \(q\) is prime.

**Definition 21.** \(H_0 = \{x + b \mid b \in \mathbb{F}_q\}\).

**Lemma 22.** \(\Gamma_0\) is the semidirect product of \(H_0\) and \{\(ax \mid a \in \mathbb{F}_q^*_q\)\}.

Let \(G\) be a subgroup of \(\Gamma_0\). As \(H_0\) has prime order, we have two cases:

- \(G \cap H_0 = H_0\). Then \(H_0 \subseteq G\). If \(ax + b \in G\), then for every \(b' \in \mathbb{F}_q\) we have \(ax + b' \in G\). In particular, \(ax \in G\), and \(G_0 = \{a \in \mathbb{F}_q^*_q \mid ax \in G\} \triangleleft \mathbb{F}_q^*_q\). But \(\mathbb{F}_q^*_q\) is cyclic of order \(q - 1\), thus \(G_0\) is cyclic of order \(m \mid q - 1\). In this case, \(G = H_0 \rtimes G_0 \cong C_q \rtimes C_m\).

- \(G \cap H_0 = \{x\}\). Then for every \(a \in G_0\) there exists exactly one \(b \in \mathbb{F}_q\) such that \(ax + b \in G\), because \((ax + b) \circ (ax + b')^{-1} = x - b' + b\). As \(G_0\) is cyclic, we have that \(G\) is generated by some \(a_0x + b_0\) where \(a_0\) generates \(G_0\) and \(b_0 \in \mathbb{F}_q\).

This allows to prove the following theorem.

**Theorem 23.** If \(q\) is prime, then all the maximal chains of subgroups of \(\Gamma_0(\mathbb{F}_q)\) have the same length.

**Proof.** Let \(G_0 = \{x\} < G_1 < \cdots < G_n = \Gamma_0(\mathbb{F}_q)\) be a maximal chain. Let \(i \in \{1, \ldots, n\}\) be such that \(G_{i-1} \cap H_0 = \{x\}\) and for all \(j \geq i\), \(H_0 \subseteq G_j\). For each \(j \geq i\) there exists a cyclic group \(C_i\) of order \(m_i\) with \(m_i \mid q - 1\) such that \(G_i = H_0 \rtimes C_i\).

Thus, the numbers \(m_i, m_{i+1}, \ldots, m_n\) are a maximal chain of divisors of \(q - 1\) greater or equal than \(m_i\).
On the other hand, \( G_{i-1} \) must be a cyclic group of order \( m_i \), therefore, the orders of \( G_1, \ldots, G_{i-1} \) are a maximal chain of divisors of \( m_i \).

Therefore, the length of the chain \( G_0, \ldots, G_n \) is equal to the number of prime factors in a complete factorization of \( q - 1 \) plus two. □

**Corollary 24.** The polynomial \( (x^q - x)^{q-1} \in \mathbb{F}_q[x] \) has maximal decomposition chains of different lengths iff \( q \) is not prime.

**Remark 25.** It is possible to determine all the subgroups of \( \Gamma(\mathbb{F}_q) \) by finding all subgroups of \( GL(2, q) \). Then all chains of subgroups can be computed, finding out whether the function \( f \) has decompositions of different lengths.

### 4. Conclusions

The results in the last section show some new information about the structure of decompositions of rational functions in the finite case; it is our hope that more can be said about possible versions of Ritt’s theorems for finite fields. Also, the algorithms presented here indicate that fast decomposition algorithms in the finite case can be achievable, by using this structure.

### Acknowledgments

The authors are grateful to the anonymous referee for helpful comments and suggestions.

### References


