CHARACTERIZATION OF POLYNOMIAL PRIME BIDECOMPOSITIONS A SIMPLIFIED PROOF

FRANZ BINDER

ABSTRACT. Bidecompositions, i.e., solutions to $r \circ p = s \circ q$, play a central rôle in the study of uniqueness properties of complete decompositions with respect to functional composition. In [Rit22] all bidecompositions using polynomials over the complex number field have been characterized. Later the result was generalized to more general fields. All proofs tend to be rather long and involved. The object of this paper is to develop a version that is simpler than the existing ones, while keeping completely elementary, thus making it accessible to a wider community.

1. BIDECOMPOSITIONS

In the whole paper we will deal with polynomials over a field, which is usually denoted by $\mathbf{k}$. The indeterminant, or identity polynomial, will be denoted by x. Thus, whenever we just say *polynomial*, elements of $\mathbf{k}[x]$ are intended, if not specified differently. Their functional composition will be denoted by $\circ$. Thus for polynomials r and p we use the notations

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

interchangably. The degree of a polynomial p is denoted by $[p]$. As usual when dealing with composition, we use the convention $[0] = 0$. Thus the degree function is a homomorphism of the monoid $(\mathbf{k}[x], \circ)$ onto the monoid $(\mathbf{N}_0, \cdot)$. The units of $(\mathbf{k}[x], \circ)$ are the polynomials of degree 1.

We have to clarify some basic notions.

1.1. Definition.

1. Two polynomials p and q are called *associated*, in symbols $p \sim q$, whenever there exist units a and b such that
$$p = a \circ q \circ b.$$
2. A polynomial f is said to be *decomposable* iff it has a decomposition $f = r \circ p$ into two non-units r and p. Otherwise it is called

indecomposable. $f = f_n \circ \ldots \circ f_1$ is called a *prime decomposition* of f iff all its componets f_i are indecomposable.

3. A *bidecomposition* consists of two decompositions $f = r \circ p = s \circ q$ that are not associated. A *prime bidecomposition* consists of prime decompositions.
4. If $r \circ p = s \circ q$ is a bidecomposition and a, b, c, and d are units, then
$$(a \circ r \circ b) \circ (b^{-1} \circ p \circ c) = (a \circ s \circ d) \circ (d^{-1} \circ q \circ c)$$
is a bidecomposition that is *associated* to the original one.

1.2. *Remark.* Two polynomials over $\mathbf{k}$ that are not associated may be associated when viewed as polynomials over an extension field. So we must be careful with this notion.

Suppose that f has a prime decomposition of the form
$$f = \ldots \circ r \circ p \circ \ldots ,$$
and that $r \circ p = s \circ q$ is a prime bidecomposition. Then we can replace r and p by s and q to get another prime decomposition
$$f = \ldots \circ s \circ q \circ \ldots .$$
If two prime decompositions can be joined by a sequence of such replacements, then they are called *related*. The importance of bidecomposition comes from the following theorem, that goes back essentially to [Ritt 22]. First a definition.

1.3. Definition. A polynomial over $\mathbf{k}$ is called *tame* iff its degree is not a multiple of the characteristic of $\mathbf{k}$.

1.4. *Remark.* If $\operatorname{char} \mathbf{k} = 0$, then tame just means non-constant.

1.5. Theorem (Ritt)**.** *All prime decompositions of a tame polynomial f are related. In particular, the number of its indecomposable components and their degrees, but not nescessarily their order, are uniquely determined by f.*

Other proofs and generalizations of this theorem as well as related results are e.g. in [Eng41], [LN73], [DW74], [Sch82], and [Bin95].

Some other facts from the above references that usually are proved in this context should be mentioned.

1.6. Corollary. *In a tame nontrivial bidecomposition $r \circ p = s \circ q$ the outer resp. the inner polynomials have the same degrees, i.e.,*
$$[r] = [q] , \quad [p] = [s] .$$
Additionally, $[p]$ and $[q]$ are relatively prime.

1.7. Theorem (Lüroth)**.** *All intermediate fields of the extension* $\mathbf{k}(x) : \mathbf{k}$ *are simple. In particular, if polynomials* p *and* q *have no common right component, then there is a rational function* f *in two variables such that*

$$x = f(p, q).$$

Decompositions of tame polynomials do not depend on the ground field:

1.8. Proposition. *A tame polynomial is indecomposable over* $\mathbf{k}$ *iff it is indecomposable over some extension fields.* □

This allows us to pass over to the algebraic closure of $\mathbf{k}$ without losing indecomposablility. But bidecompositions associated over an extension field need not be associated over $\mathbf{k}$. For these reason we do the proof for arbitrary fields. algebraicall

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

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

This can be generalized a bit to

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

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

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

1.10. Definition. A bidecomposition associated to (1) is called *exponential*, one associated to (2) is called *trigonometric*.

1.11. Definition. A tame polynomial f is called *completely tame* iff for each multiplicity μ of a zero of f', $m+1$ is not a multiple of $\operatorname{char} \mathbf{k}$.

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

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

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

1.14. Corollary.

1. *Over a field of characteritic 0 all bidecompositions are either exponential or trigonometric.*
2. *If* $\operatorname{char}\mathbf{k} \neq 0$ *then all bidecompositions using polynomials of degrees* $< \operatorname{char}\mathbf{k}$ *are either exponential or trigonometric.*

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

The proof given in this paper is completely elementary, in the sense that the basic facts about field extensions are the most advanced results used. Nevertheless it is not longer, quite on the contrary, some simplifications, just in the most involved passages, were possible. Our schedule will be as follows.

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

2. Dickson Polynomials

As the (prime-degree) Dickson polynomials constitute bidecompositions, a closer look at their properties will be useful.

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

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

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

Note. The classical Chebyshev polynomials t_n, usually defined by $\cos nx = t_n(\cos x)$, are conjugate to our Dickson polynomials by $t_n(x) = \frac{1}{2}D_n(2x,1)$. Using Dickson polynomials instead of Chebyshev ones has the advantage that they are normed, whenever n is odd. And there is an additional parameter.

Almost directly from the definition we get

2.2. Proposition. *The Dickson polynomials satisfy*

1. $D_n(\lambda x, \lambda^2) = \lambda^n D_n(x,1)$.
2. $D_n(x,a) \circ (x + ax^{-1}) = (x + a^n x^{-1}) \circ x^n$.
3. $D_m(x,a) \circ D_n(x,a) == D_{nm}(x,a) = D_n(x,a) \circ D_m(x,a)$ *for arbitrary constants* a, λ. □

This result among many other ones can be found e.g. in [LMT93].

2.3. *Remark.* Part 1 of his proposition suggests that the second parameter is superfluous because all Dickson polynomials of degree n are associated to D_n. Suppose, however, that $\lambda \notin \mathbf{k}$, but $a := \lambda^2 \in \mathbf{k}$. Then

$$D_n(x, a) = \lambda^n x \circ D_n(x, 1) \circ \frac{1}{\lambda} x.$$

Thus $D_n(x, a)$ is associated to $D_n(x, 1)$ as polynomial over $\mathbf{k}(\lambda)$, but not necessarily over $\mathbf{k}$. Using this extra parameter we can avoid the usage of extension fields in such cases.

Note. Using proposition 2.2 it is easy to prove the well known differential equation for Dickson polynomials

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

Conversely, the Dickson polynomials D_n and their negatives constitute all polynomial solutions to this differential equation. This is proved e.g. in [LN73] or [Sch82]. The essential idea in the later reference is used in the proof of the next lemma, which will be enough for our purposes.

2.4. Lemma. *Let K be a field. Suppose that* $\operatorname{char} K \neq 2$*; If a polynomial f over K of degree n satisfies*

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

for some polynomials g_-, g_+ and $\lambda \in K$, then

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

Proof. Let $a = \lambda^2$. We substitute $x + ax^{-1}$ and multiply by x^n; thus obtain

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

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

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

Substracting these two equations we get

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

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

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

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

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

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

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

thus

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

which is the characteristic equation for a Dickson polynomial (2). □

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

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

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

for some constants $\xi_1, \xi_2 \in K$, polynomials g_1, g_2 over K, and e_1, $e_2 \in K$ that are two different solutions of some quadratic equation over $\mathbf{k}$, then $f \sim D(x, a)$ (as a polynomials over $\mathbf{k}$) for some $a \in \mathbf{k}$.

Proof. Being the solution of a quadratic equation, the e_i have the form

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

for some $e \in \mathbf{k}$ and $\lambda \in K$ such that $\lambda^2 \in \mathbf{k}$.

In particular, f has its coefficients in $\mathbf{k}[\lambda]$, and so has $x - \xi_i$, as this is a factor of the square-free factorization. Thus the ξ_i have the form

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

for some $\xi, c \in \mathbf{k}$. Let $n = [f]$; it is odd, directly from the assumption. After multipliying with $2\lambda^{n-1}$ ($\in \mathbf{k}$, as n is odd) the equations look like

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

Now $\tilde{f} = (2\lambda^{n-1} - e) \circ f \circ (\frac{c}{2}x + \xi) \sim f$ satisfies

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

which is the form required to use the lemma. □

3. The Tschirnhaus Transform

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

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

In other words, we obtain the Tschirnhaus by transforming the zeros of q by p. As it is a symmetric function of the zeros of q, it is clear that its value is always in $\mathbf{k}$. In fact, the Tschirnhaus can easily be expressed without any reference to an extension field as a resultant:

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

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

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

$$\mathrm{res}_x(p(x) - y, q(x)) = \prod_i (p(\xi_i) - y) =\neq {}^p q j$$ □

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

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

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

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

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

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

which means, that p maps the zeros of $q-b$ injectively to the zeros of $r-s(b)$. As $[p]=[q]$, this is even a bijection, and the proof is complete for transcendental b.

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

4. Ramification

4.1. Definition. We say that e is a *ramification point* of some polynomial r iff $r-e$ and r' have a common zero. The degree of $\gcd(r-e,r')$ is called the *(ramification) index* of r at e and is denoted by $\operatorname{ind}_e r$.

4.2. *Remark.* As $r-e$ and r' have a common zero iff $\operatorname{res}(r-e,r')=0$, the ramification points of r are just the zeros of ${}^r r'$.

4.3. Proposition. *Let r be a tame polynomial. Then*

$$\sum_e \operatorname{ind}_e r = [r]-1$$

Proof. As r is tame, $[r']=[r]-1$, and if ξ is a zero of r' of multiplicity κ, then $(x-\xi)^\kappa$ divides $r-e$ for exactly one e. Thus the sum of the gcd's is $[r']$. □

The next very important proposition needs a stronger hypothesis (remember definition 1.11).

4.4. Proposition. *Suppose that r is completely tame. If we have the canonical factorization $r-e=c\prod(x-a_i)^{\alpha_i}$, then*

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

Proof. As r was assumed to be completely tame, all $\alpha_i \neq 0 \pmod{\operatorname{char}\mathbf{k}}$. Thus the multiplicity of a_i in r' is α_i-1, which proves the result. □

4.5. *Remark.* If a polynomial p has only one ramification point e, then $\operatorname{ind}_e p=[p']$, thus p is associated to $x^{[p]}$. Such polynomials are also called *exponential*.

4.6. *Remark.* For the rest of this section and the following two ones we fix a non-trivial completely tame prime bidecomposition

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

with $n = [p] = [s]$ and $m = [q] = [r]$. Whenever we want, we can assume all polynomials to be monic. For every point e we use the canonical factorizations over the algebraic closure of $\mathbf{k}$

$$r - e = \prod_i (x - a_i)^{\alpha_i}$$
$$s - e = \prod_j (x - b_j)^{\beta_j}.$$

Then

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

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

$$(p - a_i)^{\alpha_i} = \prod_j \prod_{\kappa=1}^{\gamma_{ij}} (x - \xi_{ij\kappa})^{\epsilon_{ij\kappa}}$$
$$(q - b_j)^{\beta_j} = \prod_i \prod_{\kappa=1}^{\gamma_{ij}} (x - \xi_{ij\kappa})^{\epsilon_{ij\kappa}}.$$

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

4.7. Lemma. *For i, j we have*

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

In particular $\epsilon_{ij\kappa} \leq \alpha_i \beta_j$ for all i, j, κ.

Proof. Taking the Tschirnhaus transform we get

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

But on the other hand, using proposition 3.3,

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

and this factorization must coincide with that obtained before. □

4.8. *Notation.* We will write $n \subseteq m$ or $m \supseteq n$ iff the integer n divides m. Note that $(\mathbf{N}_0, \supseteq)$ is a lattice. Thus it makes sense to use the symbols $\cap$ and $\cup$ to denote the greatest common divisor resp. the least common multiple of integers. This intuitive notation will simplify many of our formulas considerably.

4.9. Lemma. *For all i, j, κ we have*

$$\epsilon_{ij\kappa} \supseteq \alpha_i \cup \beta_j \tag{3}$$

$$\gamma_{ij} \subseteq \alpha_i \cap \beta_j. \tag{4}$$

Proof. Note that $\epsilon_{ij\kappa}$, the multiplicity of $\xi_{ij\kappa}$ in $f-e$ equals α_i times its multiplicity in $p-a_i$. Thus $\epsilon_{ij\kappa} \supseteq \alpha_i$. Similarly $\epsilon_{ij\kappa} \supseteq \beta_j$. Thus the first inequality is clear. From this, together with the previous lemma

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

Then the second equation follows from dividing by $\alpha_i \cup \beta_j$. □

4.10. Lemma. *For all i we have*

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

Proof. Using proposition 4.4 we get

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

The inequality then follows from the previous lemma. □

5. Exponential Solutions

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

5.1. Lemma. *Suppose that the positive integers α_i have no common divisor, i.e., $\bigcap_i \alpha_i = 1$. Then for all positive integers β*

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

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

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

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

The following result now has got a direct and considerably shorter proof. It continues in the style of the previous proofs.

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

Proof. Let e be the unique ramification point. Then $e \in \mathbf{k}$, and in our factorizations

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

where n must be prime by indecomposability. Hence some α_i is relatively prime to n, again by indecomposability. Thus let us assume $n \cap \alpha_1 = 1$. Now from lemma 4.10

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

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

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

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

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

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

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

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

6. Trigonometric Solutions

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

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

Proof. Because $r - e$ is not exponential, but indecomposable, $\bigcap \alpha_i = 1$. Thus we can apply the lemma for each β_j and, after summation, we get

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

Now we use lemma 4.10 to estimate

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

thus proving the $\geq$-part.

To see equality we consider the factorizations of remark 4.6 for various e's. Note that for $r - e_1$ and $r - e_2$ have no common zero whenever $e_1 \neq e_2$, thus all the elements $\alpha_i^{(e)}$ are distinct, so from summing up over all $e \in \mathbf{k}$ we get

$$m - 1 = \sum_e \operatorname{ind}_e p = \sum_e \sum_i \operatorname{ind} a_i p \geq \sum_e \operatorname{ind}_e s = m - 1,$$

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

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

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

Proof. By proposition 6.1 together with lemma 4.10

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

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

$$= \sum_j (\beta_j - 1) + \sum_{i \neq 1} \sum_j (\beta_j - \alpha_i \cap \beta_j) \tag{5}$$

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

6.3. *Remark.* If $r - e$ has no simple zero, then all its zeroes are at least double, hence their number is at most $\frac{m}{2}$, so $\mathrm{ind}_e\, r \geq \frac{m}{2}$. This cannot happen twice.

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

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

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

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

for some polynomials g_1, g_2. Because f has exactly two ramification points, e_1 and e_2 satisfy a quadratic equation over $\mathbf{k}$ (4.2). So we can apply proposition 2.4, and obtain:

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

7. Final Remarks

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

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

That we use the Tschirnhaus transform instead of the norm as previous elementary proofs is mainly a matter of taste. Note that ${}^pq \cdot p = \pm N_{\mathbf{k}(x):\mathbf{k}(p)}(q)$. The usage of the resultant is new in this context and may supply further improvements, when used more extensively. Our proof of proposition 3.3 serves as an alternative to the usage of norms and minimal polynomials; it seems to be more direct.

The section on ramification contains results mixed from the previous proofs. Lemma 4.9 has got an elementary proof.

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

These improvements essentially use that the components of prime bidecompostions are indecomposable. Thus they do not generalize as in [Sch82], partially characterizing bidecompositions that need not be prime. This raises the question, whether the two theorems of Ritt (1.5 and 1.13) can be used to give a general exact description of all possible decompositons. In particular, we may ask whether there is a canonical decomposition.

The decompositions of exponential polynomials (associated to x^n) may be considered to be trivial as they simply correspond to the factors of n. The same is true with Dickson polynomials. This suggest that a canonical decomposition could look like this: a composition of polynomials that are either Dickson or occur in some bidecomposition or have nothing to do with bidecompositions.

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

The assumption about $\operatorname{char} 2$ in the theorem was necessary because proposition 2.4 uses it, which again is needed in 6.6. It is not clear

whether we get any additional bidecompositions in case of characteristic 2.

The restriction to completely tame polynomials was necessary in proving 4.4, which is basic for all results about the index. It is not known how far this can be weakened, e.g. to tame polynomials..

Another open problem is, how far the assumptions can be weakened, or how the assertion should be changed for finite characteristic. One form of counterexample can be obtained as follows. Let $\chi = \operatorname{char} \mathbf{k}$; then for all indecomposable polynomials f

$$x^\chi \cdot f = f \cdot x^\chi$$

is a bidecompositions. Can all bidecompositions be reduced to trigonometric or exponential form using this ambiguity somehow?

The characterization of bidecompositions of rational functions seems to be more difficult. In this case, for example one gets the Redei functions as another class of permuting functions.

References

[Bin95] Franz Binder, *Polynomial decomposition*, Master's thesis, University of Linz, June 1995.

[DW74] F. Dorey and G. Whaples, *Prime and composite polynomials*, Journal of Algebra (1974), no. 28, 88–101.

[Eng41] H. T. Engström, *Polynomial substitutions*, American Journal of Mathematics (1941), no. 63, 249–255.

[Lev42] H. Levi, *Composite polynomials with coefficients in an arbitrary field of characteristic zero*, American Journal of Mathematics (1942), no. 23, 51–66.

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

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

[Rit22] J. F. Ritt, *Prime and composite polynomials*, Transactions of the American Mathematical Society (1922), no. 23, 51–66.

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

Department of Mathematics, Johannes Kepler University Linz, Austria

E-mail address: xbx@bruckner.stoch.uni-linz.ac.at