ON RITT'S POLYNOMIAL DECOMPOSITION THEOREMS

MICHAEL E. ZIEVE AND PETER MÜLLER

Abstract. Ritt studied the functional decomposition of a univariate complex polynomial f into prime (indecomposable) polynomials, $f = u_1 \circ u_2 \circ \cdots \circ u_r$. His main achievement was a procedure for obtaining any decomposition of f from any other by repeatedly applying certain transformations. However, Ritt's results provide no control on the number of times one must apply the basic transformations, which makes his procedure unsuitable for many theoretical and algorithmic applications. We solve this problem by giving a new description of the collection of all decompositions of a polynomial. One consequence is as follows: if f has degree $n > 1$ but f is not conjugate by a linear polynomial to either X^n or $\pm T_n$ (with T_n the Chebychev polynomial), and if the composition $a \circ b$ of polynomials a, b is the k^{th} iterate of f for some $k > \log_2(n+2)$, then either $a = f \circ c$ or $b = c \circ f$ for some polynomial c. This result has been used by Ghioca, Tucker and Zieve to describe the polynomials f, g having orbits with infinite intersection; our results have also been used by Medevedev and Scanlon to describe the affine varieties invariant under a coordinatewise polynomial action. Ritt also proved that the sequence $(\deg(u_1), \ldots, \deg(u_r))$ is uniquely determined by f, up to permutation. We show that in fact, up to permutation, the sequence of permutation groups $(G(u_1), \ldots, G(u_r))$ is uniquely determined by f, where $G(u) = \mathrm{Gal}(u(X)-t, \mathbb{C}(t))$. This generalizes both Ritt's invariant and an invariant discovered by Beardon and Ng, which turns out to be equivalent to the subsequence of cyclic groups among the $G(u_i)$.

1. Introduction

Around 1920, Fatou, Julia and Ritt made profound investigations of functional equations. For instance, each of them wrote at length on commuting polynomials, namely $f, g \in \mathbb{C}[X]$ such that $f \circ g = g \circ f$. This is a particular instance of the general functional equation $F = f_1 \circ \cdots \circ f_r = g_1 \circ \cdots \circ g_s$ with $f_i, g_j \in \mathbb{C}[X] \setminus \mathbb{C}$, which Ritt studied intensively [32]. Ritt's strategy was to write each nonlinear f_i and g_j as a composition of minimal-degree nonlinear polynomials, thereby obtaining two expressions of F as a composition of such 'prime' polynomials. This led him to study the extent of nonuniqueness of the 'prime factorization' of a polynomial under the operation of composition.

The 'primes' under this operation are the *indecomposable* polynomials, namely those $u \in \mathbb{C}[X]$ with $\deg(u) > 1$ which cannot be written as the composition of polynomials of strictly lower degrees. Given $f \in \mathbb{C}[X]$ with

Date: July 23, 2009.

$\deg(f) > 1$, a *complete decomposition* of f is a finite sequence $(u_1, \dots, u_r)$ of indecomposable polynomials $u_i \in \mathbb{C}[X]$ such that $f = u_1 \circ \cdots \circ u_r$. Clearly such a complete decomposition always exists if $\deg(f) > 1$; however, it need not be unique.

Ritt gave a procedure for obtaining all complete decompositions of f from a single such decomposition. Specifically, he showed that any complete decomposition of f can be obtained from any other via a finite sequence of steps, each of which involves replacing two adjacent indecomposables by two others which have the same composition. He then solved the equation $a \circ b = c \circ d$ in indecomposable $a, b, c, d \in \mathbb{C}[X]$. Up to composing with linears, the only solutions are the trivial $a \circ b = a \circ b$ and the nontrivial

$$X^n \circ X^s h(X^n) = X^s h(X)^n \circ X^n \tag{1.1}$$

$$T_n \circ T_m = T_m \circ T_n, \tag{1.2}$$

where $h \in \mathbb{C}[X]$ and n, s, m are positive integers. The polynomial T_n in (1.2) is the Chebychev polynomial, whose definition and basic properties are recalled in Section 3. We may view (1.1) as the least common generalization of the fact that $X^n \circ X^s = X^s \circ X^n$ and the fact that the square of an odd polynomial is even.

Ritt's results are analogous to the classical result in knot theory that any two knot diagrams belonging to the same knot can be obtained from one another by a sequence of certain basic transformations known as Reidemeister moves. Since in general there is no known bound on the number of Reidemeister moves required, this result does not resolve the problem of determining whether two knot diagrams belong to the same knot. However, the result has been useful in the study of invariants of knots, since any quantity which is unchanged by Reidemeister moves is necessarily a knot invariant. Likewise, Ritt's reults do not yield any bound on the number of Ritt moves required to pass between two complete decompositions. On the other hand, Ritt's results can be used to determine decomposition invariants. For instance, by inspecting the solutions of $a \circ b = c \circ d$ in indecomposable $a, b, c, d \in \mathbb{C}[X]$, we see that the degrees of a and b are the same as those of c and d, although possibly in reversed order. It follows from Ritt's procedure that the sequence of degrees of the indecomposables in a complete decomposition of f is uniquely determined (up to permutation) by f. Beardon and Ng [5] used the same method to exhibit another invariant: given any complete decomposition $f = u_1 \circ \cdots \circ u_r$, they showed that the sequence $(\#\Gamma_0(u_1), \dots, \#\Gamma_0(u_r))$ is uniquely determined (up to permutation) by f, where $\Gamma_0(u)$ is the set of linear $\ell \in \mathbb{C}[X]$ such that $u \circ \ell = u$. Our first result presents a new invariant which simultaneously generalizes Ritt's degree invariant and the Beardon–Ng invariant.

Definition. For $u \in \mathbb{C}[X] \setminus \mathbb{C}$, the *monodromy group* $\mathrm{Mon}(u)$ is the Galois group of $u(X) - t$ over $\mathbb{C}(t)$, viewed as a group of permutations of the roots of $u(X) - t$.

Theorem 1.3. *Pick $f \in \mathbb{C}[X]$ with $\deg(f) > 1$. Let $(u_1, \ldots, u_r)$ and $(v_1, \ldots, v_s)$ be two complete decompositions of f. Then $r = s$, and there is a permutation χ of the set $\{1, 2, \ldots, r\}$ such that $\mathrm{Mon}(u_i)$ and $\mathrm{Mon}(v_{\chi(i)})$ are isomorphic permutation groups for all $1 \le i \le r$.*

We will show that $\#\Gamma_0(u_i) = 1$ unless $\mathrm{Mon}(u_i)$ is cyclic, in which case $\#\Gamma_0(u_i) = \#\,\mathrm{Mon}(u_i)$; thus, the Beardon–Ng invariant is equivalent to the subsequence of cyclic groups in the sequence $(\mathrm{Mon}(u_1), \ldots, \mathrm{Mon}(u_r))$.

Ritt generalized his solution of $a \circ b = c \circ d$ in indecomposable a, b, c, d to give a similar description of all solutions to this equation which satisfy $\deg(a) = \deg(d)$ and $\gcd(\deg(a), \deg(c)) = 1$ (but without assuming indecomposability). This result has been applied to a variety of topics, for instance:

- The classification of all $f, g \in \mathbb{Z}[X]$ for which the Diophantine equation $f(X) = g(Y)$ has infinitely many integer solutions [7];
- The classification of $f, g \in \mathbb{C}[X]$ such that $f^{-1}(A) = g^{-1}(B)$ for some infinite compact sets $A, B \subset \mathbb{C}$ [30];
- The description of $K[f] \cap K[g]$ and $K(f) \cap K(g)$ for arbitrary $f, g \in K[X]$, where K is a field of characteristic zero [3];
- A proof that, for $f \in \mathbb{C}((X)) \setminus \mathbb{C}(X)$, the set of positive integers m for which $f(X^m) \in \mathbb{C}(X)[f]$ consists of the powers of a single integer [40].

However, to date there have been no applications of Ritt's procedure for passing from one complete decomposition to another (except for the derivation of the invariants mentioned above). Our main results transform Ritt's procedure into an applicable form. We give a new method for describing all complete decompositions of a polynomial. Unlike Ritt's procedure, in our procedure one can determine in advance exactly how many steps one must perform.

Our method is as follows. We first write the polynomial f as the composition of polynomials of two types, which we call blocks: either indecomposable polynomials which cannot be transformed into X^n or T_n by composing with linears, or (possibly decomposable) polynomials which can be so transformed. Then, when possible, we combine adjacent blocks of the form $\ell_1 \circ X^n \circ \ell_2$ (with the ℓ_i linear), so long as their composition again has this form; and we combine Chebychev blocks similarly. There can be many different decompositions of X^n, since it is the composition (in any order) of the various X^p where p runs through the prime factors of n counted with multiplicities; similar remarks apply to T_n. We obtain complete decompositions of f by inserting all such complete decompositions of each X^n or T_n block. These typically comprise all complete decompositions of f. There are only two ways to obtain further complete decompositions: first, if an X^n block is adjacent to a T_m block, and if the linears between X^n and T_m have appropriate composition, then we can move a degree-2 factor from one block to the other (since X^2 is the composition of T_2 with linears); however, we will show that after one degree-2 factor has been moved, no further degree-2 factors can be moved in the same direction. And second, if an X^n block is adjacent to an indecomposable of a special form, then we can use (1.1) to effectively move an X^k sub-block to the other side of the indecomposable; typically this will change the form of the indecomposable, but we will show that if k is chosen maximally then no further sub-block of the remaining $X^{n/k}$ can switch sides with the transformed indecomposable. We will give a detailed exposition of our procedure in Section 4.

One application of our results is to the decomposition of iterates of a polynomial. Here we write $f^{\langle e\rangle}$ for the e^{th} iterate of f, or in other words the e^{th} power of f under the operation of composition. By convention $f^{\langle 0\rangle} = X$, and for a linear $\ell \in \mathbb{C}[X]$ we write $\ell^{\langle -1\rangle}$ for the inverse of ℓ, which is again a linear polynomial.

Theorem 1.4. *Pick $f \in \mathbb{C}[X]$ of degree $n > 1$, and suppose there is no linear $\ell \in \mathbb{C}[X]$ for which $\ell \circ f \circ \ell^{\langle -1\rangle}$ is either X^n or T_n or $-T_n$. If $a, b \in \mathbb{C}[X]$ satisfy $a \circ b = f^{\langle e\rangle}$ for some $e \geq 1$, then*

$$a = f^{\langle i\rangle} \circ \hat{a} \quad \text{and} \quad b = \hat{b} \circ f^{\langle j\rangle} \quad \text{and} \quad \hat{a} \circ \hat{b} = f^{\langle k\rangle}$$

for some $\hat{a}, \hat{b} \in \mathbb{C}[X]$ and $i, j, k \geq 0$ with $k \leq \log_2(n + 2)$.

This result says that if $e > \log_2(n+2)$ then every decomposition $a \circ b = f^{\langle e\rangle}$ can be obtained from some decomposition of $f^{\langle \lfloor \log_2(n+2)\rfloor\rangle}$ by composing on the outside with several copies of f. The bound on k can be improved to $k \leq \lfloor \log_2(n)\rfloor$ if $n \neq 6$, but in Example 4.9 we will show that the bound cannot be improved further if $n = 2^m + 2$ with $m \geq 3$. We will prove a refined version of Theorem 1.4 in Section 4, as a consequence of the stronger Theorem 4.7.

Theorem 1.4 is one of the key ingredients in the companion paper [17], in which the following is proved:

Theorem 1.5. *For $x_0, y_0 \in \mathbb{C}$, if $f, g \in \mathbb{C}[X]$ are nonlinear and the orbits $\{x_0, f(x_0), f(f(x_0)), \ldots\}$ and $\{y_0, g(y_0), g(g(y_0)), \ldots\}$ have infinite intersection, then f and g have a common iterate.*

This question can be translated into a decomposition problem as follows. Supposing for simplicity that $x_0, y_0 \in \mathbb{Z}$ and $f, g \in \mathbb{Z}[X]$, the hypothesis implies that for any $i, j > 0$ the equation $f^{\langle i\rangle}(X) = g^{\langle j\rangle}(Y)$ has infinitely many solutions in integers X, Y. By Siegel's theorem, it follows that $f^{\langle i\rangle} \circ a = g^{\langle j\rangle} \circ b$ for some nonconstant $a, b \in \mathbb{C}(X)$ which are Laurent polynomials (i.e., rational functions whose denominator is a power of X). Since Ritt's results have been generalized to the setting of Laurent polynomials [7, 31, 41], this gives information about decompositions of $f^{\langle i\rangle}$ and $g^{\langle j\rangle}$, which leads to the application of Theorem 1.4. In an earlier paper [16], Theorem 1.5 was proved in case $\deg(f) = \deg(g)$; in this special case, the polynomial decomposition arguments simplify dramatically (essentially because of Corollary 2.9). However, the full strength of the results of the present paper seems to be needed to prove Theorem 1.5 in general.

Another application of the results of this paper was found by Medvedev and Scanlon [28]: combining our results with a model-theoretic result of Chatzidakis and Hrushovski, they described the subvarieties of $\mathbb{A}^n$ preserved by a coordinatewise polynomial map $(x_1, \dots, x_n) \mapsto (f_1(x_1), \dots, f_n(x_n))$ with $f_i \in \mathbb{C}[X]$.

Ritt's results are not well understood: in many treatments the statements of Ritt's results are either false [5, 15, 24] or weaker than the original versions [8, 9, 10, 13, 14, 25, 26, 29]. In light of this, we have included simplified accounts of Ritt's proofs (in modern language) in the present paper.

Ritt's proofs have two distinct flavors. His solution of $a \circ b = c \circ d$ uses that the curve $a(X) = c(Y)$ has genus zero; by expressing this genus in terms of the ramification in the covers $\mathbb{P}^1 \to \mathbb{P}^1$ corresponding to b and d, one obtains a system of equations satisfied by the ramification indices, and the main work is to solve this system. See the appendix for a simplified version of this argument. Ritt's proof of his iterative procedure uses Galois theory to translate the problem to a question about cyclic groups. We give an account of this in the next section, and by extending the method we prove Theorem 1.3 and other results. In Section 3 we give various properties of the special polynomials occurring in (1.1) and (1.2). We prove our main results in Section 4. Then in the final section we briefly survey related topics, including decomposition of rational functions, decomposition of polynomials over arbitrary fields, decomposition algorithms, and monodromy groups of indecomposable polynomials.

Acknowledgements: The first author thanks Dragos Ghioca and Tom Tucker for a stimulating collaboration on the paper [16], which led to a conjectural version of Theorem 1.4; proving this conjecture was the initial motivation for the research presented in this paper. The authors thank Avi Wigderson for suggesting the analogy with knot theory.

2. Monodromy groups and Ritt's first theorem

In this section we present a Galois-theoretic framework which enables us to translate many questions about polynomial decomposition into questions about subgroups of cyclic groups. In particular, we prove Ritt's result that one can pass from any complete decomposition of f to any other via finitely many changes of the following form:

Definition. If $(u_1, \dots, u_r)$ and $(v_1, \dots, v_r)$ are complete decompositions of a polynomial $f \in \mathbb{C}[X]$, then we say they are *Ritt neighbors* if there exists i with $1 \le i < r$ such that

- $u_j = v_j$ for $j \notin \{i, i+1\}$, and
- $u_i \circ u_{i+1} = v_i \circ v_{i+1}$.

Theorem 2.1. *Pick $f \in \mathbb{C}[X]$ with $\deg(f) > 1$. If $\mathcal{U}$ and $\mathcal{V}$ are complete decompositions of f, then there is a finite sequence $\mathcal{S}$ of complete decompositions of f such that $\mathcal{U}, \mathcal{V} \in \mathcal{S}$ and every pair of consecutive decompositions in $\mathcal{S}$ are Ritt neighbors.*

We use the following notation in this section.

- f is a nonconstant polynomial in $\mathbb{C}[X]$
- S is the set of pairs $(a, b) \in \mathbb{C}[X]^2$ such that $a \circ b = f$
- t is transcendental over $\mathbb{C}$
- L is the splitting field of $f(X) - t$ over $\mathbb{C}(t)$
- x is a root of $f(X) - t$ in L
- G is the monodromy group $\mathrm{Mon}(f) = \mathrm{Gal}(L/\mathbb{C}(t))$
- H is stabilizer of x in G, namely $H = \mathrm{Gal}(L/\mathbb{C}(x))$

2.1. **General formalism.** We begin by reviewing the Galois-theoretic framework developed by Ritt [32] for addressing polynomial decomposition problems. Our presentation is a modernized and simplified version of Ritt's.

Lemma 2.2. *The map $\rho\colon (a, b) \mapsto \mathbb{C}(b(x))$ is a surjection from S onto the set of fields between $\mathbb{C}(x)$ and $\mathbb{C}(t)$. For $d \in \mathbb{C}[X]$, we have $\rho((a, b)) = \mathbb{C}(d(x))$ if and only if there is a linear $\ell \in \mathbb{C}[X]$ such that $d = \ell \circ b$, in which case $f = (a \circ \ell^{\langle -1 \rangle}) \circ d$. Moreover, $[\mathbb{C}(x) : \mathbb{C}(b(x))] = \deg(b)$ and $[\mathbb{C}(b(x)) : \mathbb{C}(t)] = \deg(a)$.*

Proof. Let E be a field between $\mathbb{C}(x)$ and $\mathbb{C}(t)$. By Lüroth's theorem, $E = \mathbb{C}(b(x))$ for some $b \in \mathbb{C}(X)$. Since E is unchanged if we replace b by $\ell \circ b$ where $\ell \in \mathbb{C}(X)$ has degree one, we may assume $b(\infty) = \infty$. Since $t = f(x)$ lies in $\mathbb{C}(b(x))$, we have $f(X) = a(b(X))$ for some $a \in \mathbb{C}(X)$. Now $X = \infty$ is the unique preimage of ∞ under f, and $b(\infty) = \infty$, so $X = \infty$ is the unique preimage of ∞ under each of $a(X)$ and $b(X)$. Thus $a(X)$ and $b(X)$ are polynomials, so ρ is surjective.

By Gauss's lemma, $f(X) - t$ is irreducible over $\mathbb{C}(t)$, so $[\mathbb{C}(x) : \mathbb{C}(t)] = \deg(f)$. This argument implies the final statement of the result. Moreover, for $d \in \mathbb{C}[X]$ and $(a, b) \in S$, we have $\mathbb{C}(b(x)) = \mathbb{C}(d(x))$ if and only if $d = \ell \circ b$ and $b = \hat{\ell} \circ d$ for some $\ell, \hat{\ell} \in \mathbb{C}(X)$; then ℓ and $\hat{\ell}$ have degree one, and since b and d are polynomials it follows that ℓ is linear. $\square$

This result enables us to translate questions about decompositions of f into questions about intermediate fields between $\mathbb{C}(x)$ and $\mathbb{C}(t)$. Here we define a decomposition of f to be a sequence $(a_1, \dots, a_r)$ where $f = a_1 \circ \cdots \circ a_r$ and each $a_i \in \mathbb{C}[X]$ satisfies $\deg(a_i) > 1$ (we do not require the a_i to be indecomposable). Such a decomposition corresponds to the chain of fields $C(x) \supset \mathbb{C}(a_r(x)) \supset \mathbb{C}(a_{r-1} \circ a_r(x)) \supset \cdots \supset \mathbb{C}(a_1 \circ \cdots \circ a_r(x))$. Letting θ denote this map from decompositions of f to decreasing chains of fields from $\mathbb{C}(x)$ to $\mathbb{C}(t)$, we now describe the decompositions which map to the same chain of fields.

Definition. For $f \in \mathbb{C}[X]$, we say two decompositions $(a_1, \ldots, a_r)$ and $(b_1, \ldots, b_s)$ of f are *equivalent* if $r = s$ and there are linear $\ell_0, \ldots, \ell_r \in \mathbb{C}[X]$, with $\ell_0 = \ell_r = X$, such that $b_i = \ell_{i-1} \circ a_i \circ \ell_i^{\langle -1 \rangle}$ for $1 \le i \le r$.

This is an instance of the category-theoretic notion of equivalence of two factorizations of an arrow. Our next result follows from Lemma 2.2.

Corollary 2.3. *The map θ induces a bijection between equivalence classes of decompositions of f and decreasing chains of fields from $\mathbb{C}(x)$ to $\mathbb{C}(t)$. If the decomposition $(a_1, \ldots, a_r)$ corresponds to the chain of fields $\mathbb{C}(x) = E_r > E_{r-1} > \cdots > E_0 = \mathbb{C}(t)$, then $[E_i : E_{i-1}] = \deg(a_i)$ for $1 \le i \le r$.*

We have reduced the study of decompositions of f to the study of decreasing chains of fields from $\mathbb{C}(x)$ to $\mathbb{C}(t)$. As usual, the latter is equivalent to the study of increasing chains of groups from H to G. Concretely, the map $W \mapsto \mathbb{C}(x)^W$ is a bijection from the set of groups between H and G to the set of fields between $\mathbb{C}(x)$ and $\mathbb{C}(t)$, and $|W_1 : W_2| = [\mathbb{C}(x)^{W_2} : \mathbb{C}(x)^{W_1}]$ for any groups W_1, W_2 with $H < W_2 < W_1 < G$. Since there are only finitely many groups between H and G, this implies the following.

Corollary 2.4. *There are only finitely many equivalence classes of decompositions of f.*

We make one further reduction. Let I be the inertia group at a place of L lying over $t = \infty$, so I is a cyclic subgroup of G, and moreover I is transitive (since $t = \infty$ is totally ramified in $\mathbb{C}(x)/\mathbb{C}(t)$). Alternately, we could define I to be the Galois group of $f(X) - t$ over $\mathbb{C}((1/t))$, so I is cyclic because any finite extension of $\mathbb{C}((1/t))$ is cyclic, and I is transitive because the monic polynomial whose roots are the reciprocals of the roots of $f(X) - t$ is Eisenstein over $\mathbb{C}[[1/t]]$ and hence irreducible over $\mathbb{C}((1/t))$.

The following simple lemma reduces the study of decompositions of f to the study of increasing chains of groups from 1 to I.

Lemma 2.5. *Let I be a cyclic subgroup of the finite group G, and let H be a subgroup of G such that $G = HI$ and $H \cap I = 1$. For any group W between H and G, we have $W = HJ$ where $J = W \cap I$. Conversely, for any subgroup J of I, the set HJ is a group if and only if $HJ = JH$, in which case $|HJ : H| = |J|$.*

Corollary 2.6. *For groups W_1 and W_2 between H and G, write $J_i := W_i \cap I$; then*

(2.6.1) $<W_1, W_2> = HJ_1J_2$.
(2.6.2) $|<W_1, W_2> : H| = \operatorname{lcm}(|W_1 : H|, |W_2 : H|)$.
(2.6.3) $|G : <W_1, W_2>| = \gcd(|G : W_1|, |G : W_2|)$.
(2.6.4) $W_1 \cap W_2 = H(J_1 \cap J_2)$.
(2.6.5) $|(W_1 \cap W_2) : H| = \gcd(|W_1 : H|, |W_2 : H|)$.
(2.6.6) *If* $|W_1| = |W_2|$, *then* $W_1 = W_2$.
(2.6.7) $N_G(H) \le N_G(W_1)$.

Proof. We have $HJ_1J_2 = J_1HJ_2 = J_1J_2H$, so $<W_1, W_2> = HJ_1J_2$, which implies (2.6.1), (2.6.2) and (2.6.3). Since $W_1 \cap W_2 \cap I = J_1 \cap J_2$, we obtain

(2.6.4) and (2.6.5). For (2.6.6), note that I has at most one subgroup of a given order. Finally, for $\tau \in N_G(H)$ we have $H \le W_1^\tau \le G$, so (2.6.7) follows from (2.6.6). $\square$

Remark 2.7. Corollary 2.3 is implicit in [32, §2] and explicit in [26, §3]. Corollary 2.4 is due to Ritt [32, p. 55].

2.2. **Greatest common divisors and Ritt's first theorem.** In this subsection we prove Ritt's result (Theorem 2.1) describing how to obtain any complete decomposition of f from any other. We then deduce that the sequence of monodromy groups of the indecomposables in a complete decomposition of f is uniquely determined (up to permutation) by f. Our first result describes the left and right greatest common divisors of two decompositions.

Lemma 2.8. *If $a, b, c, d \in \mathbb{C}[X] \setminus \mathbb{C}$ satisfy $a \circ b = c \circ d$, then there exist $\hat{a}, \hat{b}, \hat{c}, \hat{d}, g, h \in \mathbb{C}[X]$ such that*

- $g \circ \hat{a} = a,\ g \circ \hat{c} = c,\ \deg(g) = \gcd(\deg(a), \deg(c))$*;*
- $\hat{b} \circ h = b,\ \hat{d} \circ h = d,\ \deg(h) = \gcd(\deg(b), \deg(d))$*; and*
- $\hat{a} \circ \hat{b} = \hat{c} \circ \hat{d}$*.*

Proof. Let $a, b, c, d \in \mathbb{C}[X] \setminus \mathbb{C}$ satisfy $a \circ b = c \circ d = f$. Let W_1 and W_2 be the subgroups of G fixing $b(x)$ and $d(x)$, respectively, so $H \le W_1, W_2 \le G$. Putting $W := <W_1, W_2>$, Corollary 2.3 implies that the chain of groups $H \le W_1 \cap W_2 \le W_1 \le W \le G$ corresponds to the chain of fields $\mathbb{C}(x) \ge \mathbb{C}(h(x)) \ge \mathbb{C}(b(x)) \ge \mathbb{C}(\hat{a}(b(x))) \ge \mathbb{C}(f(x))$ with $\hat{a}, h \in \mathbb{C}[X]$, and by Lemma 2.2 we have $b = \hat{b} \circ h$ and $a = g \circ \hat{a}$ for some $\hat{b}, g \in \mathbb{C}[X]$. Likewise, the chain of groups $H \le W_1 \cap W_2 \le W_2 \le W \le G$ corresponds to the chain of fields $\mathbb{C}(x) \ge \mathbb{C}(h(x)) \ge \mathbb{C}(d(x)) \ge \mathbb{C}(\hat{a}(b(x)) \ge \mathbb{C}(f(x))$, so $d = \hat{d} \circ h$ and $\hat{a} \circ b = \hat{c} \circ d$ with $\hat{c}, \hat{d} \in \mathbb{C}[X]$, whence $c = g \circ \hat{c}$ and $\hat{a} \circ \hat{b} = \hat{c} \circ \hat{d}$. Finally, the statements about degrees follow from (2.6.3) and (2.6.5). $\square$

Corollary 2.9. *Suppose $a, b, c, d \in \mathbb{C}[X] \setminus \mathbb{C}$ satisfy $a \circ b = c \circ d$.*

(2.9.1) *If $\deg(c) \mid \deg(a)$, then $a = c \circ \hat{a}$ for some $\hat{a} \in \mathbb{C}[X]$.*
(2.9.2) *If $\deg(d) \mid \deg(b)$, then $b = \hat{b} \circ d$ for some $\hat{b} \in \mathbb{C}[X]$.*
(2.9.3) *If $\deg(a) = \deg(c)$, then there is a linear $\ell \in \mathbb{C}[X]$ such that $a = c \circ \ell$ and $b = \ell^{\langle -1 \rangle} \circ d$.*

Assertion (2.9.3) implies that, up to the insertion of linears and their inverses between consecutive indecomposables, a complete decomposition is uniquely determined by the sequence of degrees of the involved indecomposables. This yields a refinement of Corollary 2.4.

We now prove Theorem 2.1.

Proof of Theorem 2.1. By Corollary 2.3 and Lemma 2.5, the result is a consequence of the following lemma about chains of subgroups of I. $\square$

Lemma 2.10. *Let $\mathcal{J}$ be a set of subgroups of a finite cyclic group I, and assume that $1, I \in \mathcal{J}$ and $\mathcal{J}$ is closed under intersections and products. Let $1 = A_0 < A_1 < \cdots < A_r = I$ and $1 = B_0 < B_1 < \cdots < B_s = I$ be two maximal increasing chains of groups in $\mathcal{J}$. Then one can pass from the first chain to the second via finitely many steps, each of which involves replacing a chain $1 = C_0 < C_1 < \cdots < C_r = I$ by a chain $1 = D_0 < D_1 < \cdots < D_r = I$ where $D_i = C_i$ for all i not equal to a single value j (with $0 < j < r$).*

Proof. We proceed by induction on $|I|$. So suppose the result holds for any cyclic group of order less than $|I|$. Let $\mathcal{A} = (A_0, A_1, \ldots, A_r)$ and $\mathcal{B} = (B_0, \ldots, B_s)$ be maximal chains as prescribed. By the inductive hypothesis, the conclusion holds for any two chains containing A_{r-1}. So suppose $A_{r-1} \neq B_{s-1}$; maximality of the chains implies $A_{r-1}B_{s-1} = I$, so there is no group in $\mathcal{J}$ properly between $A_{r-1} \cap B_{s-1}$ and A_{r-1} (since $A_{r-1} \cap B_{s-1} < J < A_{r-1}$ implies $|JB_{s-1} : B_{s-1}| = |J : J \cap B_{s-1}| = |J : A_{r-1} \cap B_{s-1}|$). Let $1 = U_0 < U_1 < \cdots < U_k = A_{r-1} \cap B_{s-1}$ be a maximal increasing chain of groups in $\mathcal{J}$ contained in $A_{r-1} \cap B_{s-1}$; then $\mathcal{U} := (U_0, U_1, \ldots, U_k, A_{r-1}, I)$ is a maximal chain in $\mathcal{J}$. By inductive hypothesis, we can pass from $\mathcal{A}$ to $\mathcal{U}$ by steps of the required type. In one more such step we replace $\mathcal{U}$ by $\mathcal{V} := (U_0, U_1, \ldots, U_k, B_{s-1}, I)$. Finally, by inductive hypothesis we can pass from $\mathcal{V}$ to $\mathcal{B}$ by steps of the required type, and the result follows. $\square$

In light of Theorem 2.1, invariants of pairs of Ritt neighboring complete decompositions of f are invariants of any pair of complete decompositions of f. Lemma 2.8 implies the following result about the degrees of the indecomposables in a pair of Ritt neighbors.

Corollary 2.11. *Suppose indecomposable $a, b, c, d \in \mathbb{C}[X]$ satisfy $a \circ b = c \circ d$. Then either there is a linear ℓ such that $a = c \circ \ell$ and $b = \ell^{\langle -1 \rangle} \circ d$, or $\gcd(\deg(a), \deg(c)) = \gcd(\deg(b), \deg(d)) = 1$ (in which case $\deg(a) = \deg(d)$ and $\deg(b) = \deg(c)$).*

In combination with Theorem 2.1, this result shows that the sequence of degrees of the indecomposables in a complete decomposition of f is uniquely determined (up to permutation) by f:

Corollary 2.12. *Pick $f \in \mathbb{C}[X]$ with $\deg(f) > 1$. Let $(u_1, \ldots, u_r)$ and $(v_1, \ldots, v_s)$ be complete decompositions of f. Then $r = s$, and there is a permutation χ of the set $\{1, 2, \ldots, r\}$ such that u_i and $v_{\chi(i)}$ have the same degree for all $1 \leq i \leq r$.*

We now show that χ can be chosen so that u_i and $v_{\chi(i)}$ share a finer invariant than the degree: we can require them to have the same monodromy group. Note that the monodromy group is a permutation group whose degree equals the degree of the polynomial, so this result refines Corollary 2.12.

Definition. Let G and $\tilde{G}$ be permutation groups acting on sets Ω and $\tilde{\Omega}$, respectively. We say that G and $\tilde{G}$ are isomorphic as permutation groups if there is a group isomorphism $\phi \colon G \to \tilde{G}$ and a bijection $\psi \colon \Omega \to \tilde{\Omega}$ such that $\psi(\omega^\tau) = \psi(\omega)^{\phi(\tau)}$ for each $\omega \in \Omega$ and $\tau \in G$.

Theorem 2.13. *Suppose* $a, b, c, d \in \mathbb{C}[X] \setminus \mathbb{C}$ *satisfy* $a \circ b = c \circ d$ *and* $\gcd(\deg(a), \deg(c)) = 1 = \gcd(\deg(b), \deg(d))$*. Then* $\mathrm{Mon}(a)$ *and* $\mathrm{Mon}(d)$ *are isomorphic permutation groups, and so are* $\mathrm{Mon}(b)$ *and* $\mathrm{Mon}(c)$*.*

Proof. Let x be transcendental over $\mathbb{C}$, let $t = a(b(x))$, and let L be a normal closure of $\mathbb{C}(x)/\mathbb{C}(t)$. Set $G = \mathrm{Gal}(L/\mathbb{C}(t))$. Let U, V, and H be the stabilizers in G of $b(x)$, $d(x)$, and x, respectively.

Let $N := \bigcap_{\tau\in G} U^\tau$ be the core of U in G; then N is the kernel of the action of G on the set G/U of right cosets of U in G. Thus $\mathrm{Mon}(a)$ is isomorphic to G/N with respect to this action. Let $C := \bigcap_{v\in V} H^v$ be the core of H in V; then V/C, in its action on the coset space V/H, is isomorphic to $\mathrm{Mon}(d)$.

Recall that $G = HI$ with I cyclic. Since $|U:H| = \deg(b)$ is coprime to $|V:H| = \deg(d)$, we have $U \cap V = H$. Then $|G:U| = \deg(a) = \deg(d) = |V:H|$ implies $G = UV$. Since $G = UI$, we have $N = \bigcap_{\tau\in G} U^\tau = \bigcap_{\tau\in I} U^\tau \geq \bigcap_{\tau\in I} (U\cap I)^\tau = U \cap I$. From $U \cap I \leq N$ we get $U = H(U\cap I) \leq HN$, whence $U = HN$ and $VN = VHN = VU = G$.

Since $H = U \cap V$ and $G = UV$, we have

$$C = \bigcap_{v\in V} H^v = \bigcap_{v\in V} (U\cap V)^v = (\bigcap_{v\in V} U^v) \cap V = (\bigcap_{v\in UV} U^v) \cap V = N \cap V.$$

Hence the natural map $V \to G/N$ is surjective with kernel $N \cap V = C$, and thus induces a natural isomorphism $V/C \to G/N$. This isomorphism maps H/C to $HN/N = U/N$, so V/C and G/N are isomorphic permutation groups with respect to their actions on the coset spaces V/H and G/U, respectively. Thus $\mathrm{Mon}(d)$ and $\mathrm{Mon}(a)$ are isomorphic as permutation groups.

The isomorphy of $\mathrm{Mon}(b)$ and $\mathrm{Mon}(c)$ follows by symmetry. □

Theorem 1.3 follows from Theorem 2.1 and the previous result.

Remark 2.14. Letting $a := X^i h(X)^n$ and $d := X^i h(X^n)$ with $\gcd(i, n) = 1$, we have $a \circ X^n = X^n \circ d$, so Theorem 2.1 implies that a is indecomposable if and only if d is indecomposable. This has been observed previously, and has been regarded as mysterious (cf. [25, p. 140] or [5, p. 128]). It is explained by Theorem 2.13, since a polynomial is indecomposable precisely when its monodromy group is primitive.

Remark 2.15. Beardon and Ng [5] studied the set $\Gamma_0(u)$ of Euclidean isometries of a polynomial $u \in \mathbb{C}[X] \setminus \mathbb{C}$, defined as the set of linear $\ell \in \mathbb{C}[X]$ for which $u \circ \ell = f$. Writing $\gamma(u) := |\Gamma_0(u)|$, they showed that if $(u_1, \dots, u_r)$ is a complete decomposition of f then $(\gamma(u_1), \dots, \gamma(u_r))$ is uniquely determined (up to permutation) by f. We now deduce this from Theorem 1.3. Each element of $\Gamma_0(f)$ is an automorphism of $\mathbb{C}(x)$ which fixes $\mathbb{C}(f(x))$; conversely, any such automorphism is a degree-one rational function fixing the unique preimage $X = \infty$ of $f = \infty$, and so lies in $\Gamma_0(f)$. Thus $\Gamma_0(f) \cong N_G(H)/H \cong N_G(H) \cap I$ is cyclic. If f is indecomposable and $\Gamma_0(f) \neq \{X\}$ then $N_G(H) = G$; since H contains no nontrivial normal subgroups of G (because L is the normal closure of L^H/L^G), we must have $H = 1$ so G is cyclic of order $\gamma(f)$. Thus the Beardon–Ng invariant amounts to the subsequence of cyclic groups among the $\mathrm{Mon}(u_i)$. Moreover, it is easy to see (cf. Lemma 3.6) that G is cyclic of order n precisely when $f = \ell_1 \circ X^n \circ \ell_2$ with ℓ_1, ℓ_2 linear; this yields all but one of the new results in [5]. The remaining result is [5, Thm. 1.2], which says $\gamma(a \circ b) \mid \gamma(a)\gamma(b)$; the above interpretation (and Corollary 2.6) implies the refinements $\gcd(\deg(b), \gamma(a\circ b)) = \gamma(b)$ and $\mathrm{lcm}(\deg(b), \gamma(a\circ b)) \mid \gamma(a)\deg(b)$.

Remark 2.16. The crux of Lemma 2.8 is implicit in [32, p. 57]; a preliminary explicit version is [10, Thm. 2.2 and Thm. 3.1]. Our version first appeared in [38, p. 334]. Assertion (2.9.3) is due to Ritt [32, p. 56]; subsequently, Levi [26, §2] proved it by comparing coefficients (an argument also anticipated by Ritt [34, p. 221]), and this proof later led to fast decomposition algorithms [15, 24]. Corollary 2.11 occurs in [32, p. 57]. In case a, b, c, d are indecomposable, Theorem 2.13 is shown in the proof of [29, Thm. R.2].

2.3. **Ritt's second theorem and Ritt moves.** Ritt's second theorem determines all Ritt neighbors, by solving the equation $a \circ b = c \circ d$ in indecomposable $a, b, c, d \in \mathbb{C}[X]$. This equation has the trivial solution $a = c \circ \ell$ and $b = \ell^{\langle -1\rangle} \circ d$ with $\ell \in \mathbb{C}[X]$ linear; by Corollary 2.11, any other solution satisfies $\gcd(\deg(a), \deg(c)) = \gcd(\deg(b), \deg(d)) = 1$. Ritt solved the functional equation assuming only this constraint on the degrees (and not assuming indecomposability):

Theorem 2.17 (Ritt)**.** *Suppose $a, b, c, d \in \mathbb{C}[X] \setminus \mathbb{C}$ satisfy $a \circ b = c \circ d$ and $\gcd(\deg(a), \deg(c)) = \gcd(\deg(b), \deg(d)) = 1$. Then there are linear $\ell_j \in \mathbb{C}[X]$ such that (after perhaps switching (a, b) and (c, d)) the quadruple $(\ell_1 \circ a \circ \ell_2,\ \ell_2^{\langle -1\rangle} \circ b \circ \ell_3,\ \ell_1 \circ c \circ \ell_4,\ \ell_4^{\langle -1\rangle} \circ d \circ \ell_3)$ has one of the forms*

$$(T_n,\ T_m,\ T_m,\ T_n) \quad \textit{or} \tag{2.17.1}$$

$$(X^n,\ X^s h(X^n),\ X^s h(X)^n,\ X^n), \tag{2.17.2}$$

where $m, n > 0$ are coprime, $s \geq 0$ is coprime to n, and $h \in \mathbb{C}[X] \setminus X\mathbb{C}[X]$.

We will prove Theorem 2.17 in the appendix.

For applications, it is often useful to combine Theorem 2.17 with Lemma 2.8 in the following manner:

Corollary 2.18. *For $a, b, c, d \in \mathbb{C}[X] \setminus \mathbb{C}$ with $\deg(a) \leq \deg(c)$, we have $a \circ b = c \circ d$ if and only if there exist $\hat{a}, \hat{b}, \hat{c}, \hat{d}, g, h, \ell_1, \ell_2 \in \mathbb{C}[X]$ such that ℓ_i is linear and the following three conditions hold:*

- *$g \circ \hat{a} = a,\ g \circ \hat{c} = c,\ \deg(g) = \gcd(\deg(a), \deg(c))$;*
- *$\hat{b} \circ h = b,\ \hat{d} \circ h = d,\ \deg(h) = \gcd(\deg(b), \deg(d))$; and*
- *the tuple $(\hat{a} \circ \ell_1,\ \ell_1^{\langle -1\rangle} \circ \hat{b},\ \hat{c} \circ \ell_2,\ \ell_2^{\langle -1\rangle} \circ \hat{d})$ has the form of either (2.17.1) or (2.17.2).*

As noted above, if indecomposable $a, b, c, d \in \mathbb{C}[X]$ satisfy $a\circ b = c\circ d$, and if there is no linear $\ell \in \mathbb{C}[X]$ for which $(a, b) = (c \circ \ell,\ \ell^{\langle -1\rangle} \circ d)$, then a, b, c, d satisfy the hypotheses of Theorem 2.17. In this situation, we refer to the replacement of $a \circ b$ by $c \circ d$ as a *Ritt move*. Thus, Theorem 2.1 says that one can pass from any complete decomposition to any other by a sequence of steps, each of which is either a Ritt move or is the replacement of consecutive indecomposables a and b by $a \circ \ell$ and $\ell^{\langle -1 \rangle} \circ b$ for some linear $\ell \in \mathbb{C}[X]$. Note that the insertion of ℓ and $\ell^{\langle -1 \rangle}$ does not affect the sequence of degrees of the indecomposables in a complete decomposition, and in a Ritt move two consecutive coprime degrees in this sequence are interchanged. Recall that a complete decomposition is uniquely determined by the sequence of degrees of the involved indecomposables, up to the insertion of pairs of inverse linears between adjacent indecomposables. Now pick $f \in \mathbb{C}[X]$ with $\deg(f) > 1$, and let $(u_1, \dots, u_r)$ and $(v_1, \dots, v_r)$ be two complete decompositions of f. Then the sequence $(\deg(v_1), \dots, \deg(v_r))$ can be obtained from the sequence $(\deg(u_1), \dots, \deg(u_r))$ via finitely many steps, each of which involves interchanging two consecutive coprime entries. We note that there are examples in which every permutation of $(\deg(u_1), \dots, \deg(u_r))$ occurs – namely, if f is X^n or T_n. However, it turns out that such examples are quite special, and in general there are further constraints on which permutations can occur. We will deduce these constraints in Section 4; naturally, they depend on the form of the polynomials u_i rather than merely their degrees.

3. The polynomials involved in Ritt moves

The difficulty in applying Ritt's results is that, after applying a Ritt move to an adjacent pair of indecomposables in a complete decomposition, it may happen that one of the resulting indecomposables can be involved in another Ritt move, and so on. In this section we prove various results about the special polynomials involved in Ritt moves, which will allow us to control all subsequent Ritt moves involving the resulting polynomials. We also give useful characterizations of these special polynomials.

We will use the following terminology.

Definition. We say $f, g \in \mathbb{C}[X]$ are *equivalent* if there are linear $\ell_1, \ell_2 \in \mathbb{C}[X]$ such that $f = \ell_1 \circ g \circ \ell_2$.

Definition. For $f \in \mathbb{C}[X]$, we say f is *cyclic* if it is equivalent to X^n for some $n > 1$, and we say f is *dihedral* if it is equivalent to T_n for some $n > 2$.

Here the (normalized) Chebychev polynomial T_n is defined by the functional equation $T_n(Y + 1/Y) = Y^n + 1/Y^n$; the classical Chebychev polynomial $C_n(X)$ defined by $C_n(\theta) = \cos(n \arccos \theta)$ satisfies $T_n(2X) = 2C_n(X)$. Thus $T_0 = 2$ and $T_1 = X$, and in general $T_n = XT_{n-1} - T_{n-2}$, so T_n is a degree-n polynomial and for $n > 1$ the two highest-degree terms of T_n are X^n and $-nX^{n-2}$. Also, $T_n \circ (-X) = (-1)^n T_n$ and $T_n \circ T_m = T_m \circ T_n$.

3.1. **Ramification.**

We will need some properties of the ramification in the cover $\pi_f \colon \mathbb{P}^1_x \to \mathbb{P}^1_t$ corresponding to $f \in \mathbb{C}[X]$, where x is transcendental over $\mathbb{C}$ and $t = f(x)$ (and $\mathbb{P}^1_x$ denotes the projective line with coordinate x). We use the standard notions of ramification indices, ramification points, and branch points for the cover π_f. We also refer to a point of $\mathbb{P}^1_x$ as a 'special point' if it is unramified in π_f but its image is a branch point. In our concrete setting these notions have the following explicit definitions:

Definition. Pick $f \in \mathbb{C}[X] \setminus \mathbb{C}$. For $x_0 \in \mathbb{C}$, the *ramification index* of x_0 in f, denoted $e_f(x_0)$, is the multiplicity of x_0 as a root of $f(X) - f(x_0)$. The *finite ramification points* of f are the values $x_0 \in \mathbb{C}$ for which $e_f(x_0) > 1$. The *finite branch points* of f are the values $f(x_0)$, where x_0 is a finite ramification point. The *special points* of f are the values $x_0 \in \mathbb{C}$ which are not finite ramification points, but for which $f(x_0)$ is a finite branch point.

We briefly record some standard ramification facts in polynomial language. If $e_1, e_2, \ldots, e_k$ are the multiplicities of the roots of $f(X) - x_0$, then $\mathrm{Mon}(f)$ contains an element having cycle lengths $e_1, \ldots, e_k$ (but this fact is not used in this paper). Ramification indices are multiplicative in towers: for $f, g \in \mathbb{C}[X] \setminus \mathbb{C}$ and $x_0 \in \mathbb{C}$, we have $e_{f \circ g}(x_0) = e_f(g(x_0)) \cdot e_g(x_0)$. The Riemann–Hurwitz formula for π_f says

$$\deg(f) - 1 = \sum_{x_0 \in \mathbb{C}} (e_f(x_0) - 1);$$

since the finite ramification points of f are precisely the roots of the derivative $f'(X)$, this amounts to writing the degree of $f'(X)$ as the sum of the multiplicities of its roots. We will also use the Riemann–Hurwitz formula for the Galois closure of the cover π_f, as well as the following variant of Abhyankar's lemma:

Lemma 3.1. *Let F_1, F_2 be finite extensions of $\mathbb{C}(x)$ whose compositum is E. Let Q be a place of $F := F_1 \cap F_2$, let P_i be a place of F_i lying over Q, and let e_i denote the ramification index of P_i/Q. Then for each place P of E lying over both P_1 and P_2, the ramification index of P/Q is* $\mathrm{lcm}(e_1, e_2)$. *If $[F_1 : F]$ and $[F_2 : F]$ are coprime, then there are precisely* $\gcd(e_1, e_2)$ *such places P.*

Proof. Let L be the Galois closure of E/F, let $G := \mathrm{Gal}(L/F)$, and let H_1, H_2, and H be the stabilizers in G of F_1, F_2, and E, respectively. Let R be a place of L lying over P_1 and P_2, let I be the inertia group of R/Q, and let P be the place of E lying under R. Then the inertia groups of R/P and R/P_i are $I \cap H$ and $I \cap I_i$; since $H = H_1 \cap H_2$, it follows that $I \cap H = (I \cap H_1) \cap (I \cap H_2)$. Cyclicity of I implies $|I \cap H| = \gcd(I \cap H_1, I \cap H_2)$, so the ramification index of P/Q is $|I : I \cap H| = \mathrm{lcm}(|I : I \cap H_1|, |I : I \cap H_2|) = \mathrm{lcm}(e_1, e_2)$.

Since G (resp. H_i) acts transitively on the places of R lying over Q (resp. P_i), and I is the stabilizer of R, for $g \in G$ the place gR lies over P_i if and only if $g \in H_i I$. Thus the places of E lying over P_1 and P_2 are the restrictions to E of places gR with $g \in H_1 I \cap H_2 I$; since the restrictions to E of $g_1 R$ and $g_2 R$ are the same precisely when $g_2^{-1} g_1 \in HI$, the number of places of E lying over P_1 and P_2 is $|H_1 I \cap H_2 I|/|HI|$. Note that $|HI| = |H||I : I \cap H| = |H| \,\mathrm{lcm}(e_1, e_2)$, so we must show that $|H_1 I \cap H_2 I| = |H| e_1 e_2$. Assume $[F_1 : F]$ and $[F_2 : F]$ are coprime, or equivalently $|G : H_1|$

and $|G:H_2|$ are coprime. Then $|G:H_1|$ divides $|G:H| = |G:H_2||H_2:H|$, and so divides $|H_2:H|$, so $|G| \le |H_1||H_2|/|H| = |H_1H_2|$, whence $G = H_1H_2$. Thus the set of right-cosets $H\backslash G$ has the same cardinality as $H_1\backslash G \times H_2\backslash G$; since $gH \mapsto (gH_1, gH_2)$ defines an injection $\rho\colon H\backslash G \to H_1\backslash G \times H_2\backslash G$, it follows that ρ is bijective. Finally, $e_i = |I : I \cap H_i| = |H_iI|/|H_i|$, so $e_1e_2 = |\rho^{-1}(H_1I, H_2I)|$, whence indeed $e_1e_2|H| = |H_1I \cap H_2I|$ as desired. □

We now characterize cyclic and dihedral polynomials in terms of their ramification.

Lemma 3.2. *Pick $f \in \mathbb{C}[X]$ with $\deg(f) > 1$. Then f is cyclic if and only if f has a unique finite branch point (or equivalently, f has a unique finite ramification point). Likewise, f is dihedral if and only if f has precisely two finite branch points and every finite ramification point has ramification index 2; these conditions imply there are precisely two special points.*

Proof. If f has a unique finite branch point or a unique finite ramification point, then by Riemann–Hurwitz it has both a unique finite branch point α and a unique finite ramification point β. Thus $f(X+\beta) - \alpha$ has no nonzero roots, and so equals $\gamma X^{\deg(f)}$.

Now suppose that f has precisely two finite branch points, and further that every finite ramification point has ramification index 2. Letting L denote the Galois closure of the extension $\mathbb{C}(x)/\mathbb{C}(f(x))$, Lemma 3.1 implies that $L/\mathbb{C}(f(x))$ is ramified over precisely two finite places of $\mathbb{C}(f(x))$ (both with ramification index 2) and over the infinite place (with ramification index n). By Riemann–Hurwitz we compute $[L:\mathbb{C}(x)] = 2$, so Lemma 3.3 implies f is dihedral.

Finally, if f is cyclic or dihedral then it is well-known (and easy to verify) that the ramification of π_f is as described. □

Lemma 3.3. *Pick $f \in \mathbb{C}[X]$ with $\deg(f) > 1$, let x be transcendental over $\mathbb{C}$, and let L be the Galois closure of $\mathbb{C}(x)/\mathbb{C}(f(x))$. Then $L = \mathbb{C}(x)$ if and only if f is cyclic, and $[L:\mathbb{C}(x)] = 2$ if and only if f is dihedral.*

Proof. If $L = \mathbb{C}(x)$ then all points of L lying over the same point of $\mathbb{C}(f(x))$ are in a single orbit of $\mathrm{Gal}(L/\mathbb{C}(f(x)))$, and so have the same ramification index. By Riemann–Hurwitz, it follows that f has a unique finite ramification point, so f is cyclic. Conversely, if f is cyclic then visibly $\mathbb{C}(x)/\mathbb{C}(f(x))$ is Galois, and likewise if f is dihedral then $[L:\mathbb{C}(f(x))] = 2$.

Henceforth assume $[L:\mathbb{C}(f(x))] = 2$. Then each root of $f(X) - f(x)$ has degree at most 2 over $\mathbb{C}(x)$, so $f(X) - f(x)$ is the product of irreducibles $\Phi_i(X, x) \in \mathbb{C}[X, x]$ each of which has X-degree at most 2. By symmetry, also the x-degree of each $\Phi_i(X, x)$ is at most 2. The leading coefficient of $\Phi_i(X, x)$ (viewed as a polynomial in X with coefficients in $\mathbb{C}[x]$) divides the corresponding leading coefficient of $f(X) - f(x)$, and hence lies in $\mathbb{C}^*$; likewise the same property holds if we interchange x and X, so $\Phi_i(X, x)$ has total degree at most 2. Since $L \neq \mathbb{C}(x)$, some Φ_i has X-degree 2. Let $x_i \in L$

satisfy $\Phi_i(x_i, x) = 0$, so $L = \mathbb{C}(x, x_i)$. Since Φ_i has total degree 2, the genus of L is zero, so $L = \mathbb{C}(z)$ for some z.

Set $x = b(z)$ with $b \in \mathbb{C}(X)$ of degree $[\mathbb{C}(z) : \mathbb{C}(x)] = 2$. The infinite place of $\mathbb{C}(t)$ is totally ramified in each conjugate of $\mathbb{C}(x)$, so by Lemma 3.1 the infinite place of $\mathbb{C}(x)$ is unramified in $\mathbb{C}(z)$. Thus, after a linear fractional change of z and a linear change of x (and f), we have $x = z + 1/z$.

For each $\mathbb{C}$-automorphism σ of $\mathbb{C}(z)$, the image z^σ of z is a linear fractional change of z. If σ fixes $t := f(x)$, then σ fixes the set of values of z which map to $t = \infty$, namely $\{0, \infty\}$, so $z^\sigma = \alpha z^\epsilon$ with $\alpha \in \mathbb{C}^*$ and $\epsilon \in \{1, -1\}$. Let τ generate $\mathrm{Gal}(\mathbb{C}(z)/\mathbb{C}(x))$, so $z^\tau = 1/z$. Then $\mathrm{Gal}(\mathbb{C}(z)/\mathbb{C}(t)) = C{<}\tau{>}$, where C consists of the maps $z \mapsto \alpha z$ with $\alpha^n = 1$. The fixed field of C is $\mathbb{C}(z^n)$, and the fixed field of $C{<}\tau{>}$ is $\mathbb{C}(t) = \mathbb{C}(z^n + 1/z^n)$. Thus $t = \ell(z^n + 1/z^n)$ for some degree-one rational function ℓ. Since $\ell(\infty) = \infty$, in fact ℓ is a polynomial, so a linear change to t makes $t = z^n + 1/z^n$. But $f(z + 1/z) = z^n + 1/z^n$ implies $f = T_n$, and the result follows. □

For $n > 1$, the unique finite branch point of X^n is 0, which is also the unique finite ramification point. For $n > 2$, the special points of T_n are 2 and -2, which are also the finite branch points of T_n.

The analogous ramification characterization of $X^s h(X)^n$ is immediate:

Lemma 3.4. *Pick $f \in \mathbb{C}[X] \setminus \mathbb{C}$ and integers $n > 0$ and $s \geq 0$. Then f is equivalent to $X^s h(X)^n$ for some $h \in \mathbb{C}[X] \setminus X\mathbb{C}[X]$ if and only if there exists $x_0 \in \mathbb{C}$ such that $e_f(x_0) = s$ and $n \mid e_f(x_1)$ for every $x_1 \neq x_0$ satisfying $f(x_0) = f(x_1)$.*

We do not give a ramification characterization of $X^s h(X^n)$. Instead we characterize these polynomials in a different way in Lemma 3.17.

Remark 3.5. Variants of Lemma 3.1 are classical, but we know no reference for this version; for instance, a special case is proved in [32]. We know of three other proofs of Lemma 3.2; here we only discuss the most difficult part, where we assume that f has precisely two finite branch points and every finite ramification point has ramification index 2. Ritt [32, p. 65] argued as follows: since $\mathrm{Mon}(f)$ has an n-cycle, one can show there is only one possibility for the permutation representations induced by generators of the inertia groups in the Galois closure of π_f, so by topological considerations there is just one equivalence class of such polynomials f, whence f is dihedral since T_n has the prescribed ramification. Versions of this argument appear in [12, Lemma 9] and [38, Prop. 4]. Levi [26, §13] proves this result by observing that, after composing with linears, we have $n^2(f^2-4) = (X^2-4)f'(X)^2$; but $\pm T_n$ solve this differential equation, so one can deduce that f is dihedral by showing there are at most two solutions in degree-n polynomials. Beginning with $f^2 - 4 = (X^2 - 4)h^2$, Dorey and Whaples [9, p. 97] factor $f - 2$ and $f + 2$; upon substituting $X = Y + Y^{-1}$ and subtracting the expression for $f - 2$ from that for $f + 2$, they find that $4Y^n$ is the difference between the squares of two degree-n polynomials, which determines the polynomials and consequently the form of f. The advantages of our proof are that it uses similar methods to the rest of this paper, and also that Lemma 3.3 provides additional information which does not follow from these other proofs.

3.2. **Monodromy groups.** We now show that X^n and T_n are uniquely determined (up to equivalence) by their monodromy groups.

Lemma 3.6. *Pick $f \in \mathbb{C}[X]$ of degree $n > 1$, and put $G := \mathrm{Mon}(f)$. Then G is cyclic if and only if f is cyclic, in which case $|G| = n$. Likewise, if $n > 2$ then G is dihedral if and only if f is dihedral, in which case $|G| = 2n$.*

Proof. Since G contains an n-cycle, if G is cyclic then $|G| = n$, and if G is dihedral and $n > 2$ then $|G| = 2n$. The result now follows from Lemma 3.3. □

Remark 3.7. A different proof is given in [6, Thm. 3.8], using the fact that if the multiplicities of the roots of $f(X) - x_0$ are $e_1, \dots, e_k$ then G has an element whose cycle lengths are $e_1, \dots, e_k$. There are only a few possibilities for the cycle structure of an element of a dihedral group, and in combination with the Riemann–Hurwitz formula for π_f this implies that if G is dihedral (and $n \neq 4$) then f has precisely two finite branch points and every finite ramification point has ramification index 2.

3.3. **Decompositions.** We now determine all decompositions of the special polynomials X^n, T_n, $X^s h(X^n)$, and $X^s h(X)^n$. First, $X^n = X^k \circ X^{n/k}$ and $T_n = T_k \circ T_{n/k}$ for any divisor k of n, and (2.9.3) implies these are the only decompositions of X^n and T_n up to equivalence:

Lemma 3.8. *If $a, b \in \mathbb{C}[X] \setminus \mathbb{C}$ satisfy $a \circ b = X^n$, then $a = X^k \circ \ell$ and $b = \ell^{\langle -1 \rangle} \circ X^{n/k}$ for some linear $\ell \in \mathbb{C}[X]$. If $a, b \in \mathbb{C}[X] \setminus \mathbb{C}$ satisfy $a \circ b = T_n$, then $a = T_k \circ \ell$ and $b = \ell^{\langle -1 \rangle} \circ T_{n/k}$ for some linear $\ell \in \mathbb{C}[X]$.*

Conversely, we now describe which compositions of cyclic polynomials are cyclic, and likewise for dihedral polynomials.

Lemma 3.9. *If a and b are cyclic, then $a \circ b$ is cyclic if and only if the finite ramification point of a equals the finite branch point of b. If a and b are dihedral, then $a \circ b$ is dihedral if and only if the special points of a coincide with the finite branch points of b.*

Proof. This is an immediate consequence of Lemma 3.2 and the multiplicativity of ramification indices. □

In practice, this result is often used in the following explicit form.

Corollary 3.10. *Pick integers m, n and linear $\ell, \ell_1, \ell_2 \in \mathbb{C}[X]$. If $m, n > 1$ and $X^m \circ \ell \circ X^n = \ell_1 \circ X^{mn} \circ \ell_2$, then $\ell = \alpha X$ for some $\alpha \in \mathbb{C}^*$. If $m, n > 2$ and $T_m \circ \ell \circ T_n = \ell_1 \circ T_{mn} \circ \ell_2$, then $\ell = \epsilon X$ for some $\epsilon \in \{1, -1\}$.*

Now we address the same question for polynomials of the forms $X^s h(X^n)$ or $X^s h(X)^n$. We first observe that polynomials of these forms behave well under composition: $X^s h(X^n) \circ X^{\hat{s}} \hat{h}(X^n) = X^{s\hat{s}} \tilde{h}(X^n)$ where $\tilde{h}(X) =$

$\hat{h}(X)^s h(X^{\hat{s}}\hat{h}(X)^n)$, and likewise $X^s h(X)^n \circ X^{\hat{s}}\hat{h}(X)^n = X^{s\hat{s}}\tilde{h}(X)^n$. Conversely, we now show that these are the only ways that polynomials of these forms can decompose.

Lemma 3.11. *Pick $a, b, h \in \mathbb{C}[X] \setminus \mathbb{C}$ and coprime positive integers s and n. If $a \circ b = X^s h(X)^n$ then $a = X^j \hat{h}(X)^n \circ \ell$ and $b = \ell^{\langle -1\rangle} \circ X^k \tilde{h}(X)^n$ for some $j, k > 0$ and some $\hat{h}, \tilde{h}, \ell \in \mathbb{C}[X]$ with ℓ linear. If $a \circ b = X^s h(X^n)$ then $a = X^j \hat{h}(X^n) \circ \ell$ and $b = \ell^{\langle -1\rangle} \circ X^k \tilde{h}(X^n)$ for some $j, k > 0$ and some $\hat{h}, \tilde{h}, \ell \in \mathbb{C}[X]$ with ℓ linear.*

Proof. Suppose $a \circ b = X^s h(X)^n$. After replacing a and b by $a \circ \ell^{\langle -1\rangle}$ and $\ell \circ b$ for some linear $\ell \in \mathbb{C}[X]$, we may assume $a(0) = b(0) = 0$ and a is monic. Write $a = X^j \prod_\beta (X - \beta)^{n_\beta}$, where β varies over the distinct nonzero roots of a. Since the various polynomials $b - \beta$ are coprime to one another and to b, it follows that b^j equals X^s times an n^{th} power. But j divides $e_{a\circ b}(0)$, which is coprime to n, so $\gcd(j, n) = 1$ and thus $b = X^k \tilde{h}(X)^n$ for some $\tilde{h} \in \mathbb{C}[X]$. Every b-preimage of β has ramification index divisible by $n/\gcd(n, n_\beta)$; if $n \nmid n_\beta$ then this yields too large a contribution to the Riemann–Hurwitz formula for b. Thus $a = X^j \hat{h}(X)^n$ for some $\hat{h} \in \mathbb{C}[X]$.

Now suppose $a \circ b = X^s h(X^n)$. For any primitive n^{th} root of unity ζ, we have $a(b(\zeta X)) = \zeta^s a(b(X))$, so (2.9.3) implies $a = \zeta^s a \circ \ell_3$ and $b(\zeta X) = \ell_3^{\langle -1\rangle} \circ b$ for some linear $\ell_3 \in \mathbb{C}[X]$. Thus $b = \beta + X^k \tilde{h}(X^n)$ for some $\beta \in \mathbb{C}$, so replacing a and b by $a(X + \beta)$ and $b - \beta$ implies $\ell_3 = \zeta^k X$. Since $\deg(a \circ b)$ is coprime to n, we have $\gcd(k, n) = 1$, so $\hat{\zeta} := \zeta^k$ is a primitive n^{th} root of unity, and we conclude from $a = \zeta^s a \circ \hat{\zeta} X$ that $a = X^j \hat{h}(X^n)$. □

Remark 3.12. The fact that odd polynomials only decompose into odd polynomials was proved in [23, Prop. 1]; the analogous fact for decompositions of $X^s h(X^n)$ with n prime is [22, Thm. 4.3].

3.4. **Equivalence.** We now determine all equivalences between polynomials of the forms X^n, T_n, $X^s h(X^n)$, and $X^s h(X)^n$. This enables us to describe all Ritt moves involving any prescribed polynomial.

Lemma 3.13. *If $n > 1$ and $\ell_1, \ell_2 \in \mathbb{C}[X]$ satisfy $\ell_1 \circ X^n \circ \ell_2 = X^n$, then $\ell_2 = \alpha X$ and $\ell_1 = X/\alpha^n$ for some $\alpha \in \mathbb{C}^*$. If $n > 2$ and $\ell_1, \ell_2 \in \mathbb{C}[X]$ satisfy $\ell_1 \circ T_n \circ \ell_2 = T_n$, then $\ell_2 = \epsilon X$ and $\ell_1 = \epsilon^n X$ for some $\epsilon \in \{1, -1\}$.*

Proof. In either case, comparing degrees gives $\deg(\ell_1) = \deg(\ell_2) = 1$, and comparing coefficients of X^{n-1} implies $\ell_2(0) = 0$, so $\ell_2 = \alpha X$ with $\alpha \in \mathbb{C}^*$. If $\ell_1 \circ X^n \circ \alpha X = X^n$ then $\ell_1 = X/\alpha^n$. Now suppose that $n > 2$ and $\ell_1 \circ T_n \circ \alpha X = T_n$. Since the ratio of the coefficients of X^n and X^{n-2} in $\ell_1 \circ T_n \circ \alpha X$ is α^2 times the corresponding ratio in T_n, we have $\alpha \in \{1, -1\}$. Since $T_n(-X) = (-1)^n T_n(X)$, this implies $\ell_1 = \alpha^n X$. □

Lemma 3.14. *The polynomials T_n and X^n are equivalent if and only if $n \leq 2$.*

Proof. This follows from Lemma 3.6, but we give an alternate proof. For $n \le 2$, there is a unique equivalence class of degree-n polynomials. Now suppose $n > 2$ and $T_n = \ell_1 \circ X^n \circ \ell_2$ with $\ell_1, \ell_2 \in \mathbb{C}[X]$ linear. Since the coefficient of X^{n-1} is zero in both T_n and X^n, we must have $\ell_2(0) = 0$. But then the coefficient of X^{n-2} in $\ell_1 \circ X^n \circ \ell_2$ is zero, yet the coefficient of X^{n-2} in T_n is nonzero, contradiction. $\square$

Lemma 3.15. *Pick $n, s > 0$ and $h \in \mathbb{C}[X]$, and let f be either $X^s h(X)^n$ or $X^s h(X^n)$. If f is cyclic and $n > 1$ then h is a monomial. If f is dihedral then $n \le 2$.*

Proof. If $f = X^s h(X)^n$ is equivalent to X^k with $k > 1$, then the unique finite branch point of f has just one f-preimage. If also $n > 1$ then each nonzero root of h is a ramification point of f having the same h-image as $X = 0$, a contradiction; thus h has no nonzero roots, so h is a monomial.

If $f = X^s h(X)^n$ is equivalent to T_k with $k > 2$, then each ramification point of f has ramification index 2; thus $s \le 2$, so h is non-constant, and each root α of h satisfies $e_f(\alpha) \ge n$ so $n \le 2$.

Suppose $\ell_1 \circ X^k \circ \ell_2 = X^s h(X^n)$ with $k > 1$ and the ℓ_i linear. If $n > 1$ then equating coefficients of X^{k-1} gives $\ell_2(0) = 0$, so evaluating at $X = 0$ gives $\ell_1(0) = 0$, whence h is a monomial.

Suppose $\ell_1 \circ T_k \circ \ell_2 = X^s h(X^n)$ with $k > 2$ and the ℓ_i linear. If $n > 2$ then the coefficients of X^{k-1} and X^{k-2} on the right side are zero, but it is not possible for the corresponding coefficients on the left side to both be zero. $\square$

We now describe the Ritt moves involving at least one dihedral polynomial. Here the crucial point is that if such a move has type (2.17.2) then it can be rewritten as a move of type (2.17.1).

Lemma 3.16. *Suppose $n > 2$ and $a \circ b = c \circ d$ where $a, b, c, d \in \mathbb{C}[X] \setminus \mathbb{C}$ satisfy $\gcd(\deg(a), \deg(c)) = \gcd(\deg(b), \deg(d)) = 1$. If $c = T_n$ then $d = \epsilon T_m \circ \ell$ and $a = \epsilon^n T_m \circ \hat{\ell}$ and $b = \hat{\ell}^{\langle -1 \rangle} \circ T_n \circ \ell$ where $\ell, \hat{\ell} \in \mathbb{C}[X]$ are linear and $\epsilon \in \{1, -1\}$. If $d = T_n$ then $a = \ell \circ T_n \circ \hat{\ell}$ and $b = \hat{\ell}^{\langle -1 \rangle} \circ \epsilon T_m$ and $c = \ell \circ \epsilon^n T_m$ where $\ell, \hat{\ell} \in \mathbb{C}[X]$ are linear and $\epsilon \in \{1, -1\}$.*

Proof. First suppose $c = T_n$. Since $n > 2$, Lemma 3.14 implies c is not cyclic. By Theorem 2.17, there are linear $\ell_j \in \mathbb{C}[X]$ for which the quadruple $Q := (\ell_1 \circ a \circ \ell_2,\ \ell_2^{\langle -1 \rangle} \circ b \circ \ell_3,\ \ell_1 \circ c \circ \ell_4,\ \ell_4^{\langle -1 \rangle} \circ d \circ \ell_3)$ has one of the forms (2.17.1) or (2.17.2). If it is (2.17.1), then Lemma 3.13 implies $\ell_4 = \epsilon X$ and $\ell_1 = \epsilon^n X$ for some $\epsilon \in \{1, -1\}$, and the result follows. So assume Q has the form (2.17.2); we will show that, after perhaps changing the ℓ_j's, we can also write Q in the form (2.17.1). Now $c = \ell_1^{\langle -1 \rangle} \circ X^s h(X)^N \circ \ell_4^{\langle -1 \rangle}$ where $h \in \mathbb{C}[X]$, $s \ge 0$, and $N := \deg(a)$. Lemma 3.15 implies $N \le 2$. If $N = 1$ then a and d are linear, so $a \circ b = c \circ d$ can be written in the form of (2.17.1) as $(T_1 \circ a) \circ (a^{\langle -1 \rangle} \circ T_n \circ d) = T_n \circ (T_1 \circ d)$. Now assume $N = 2$, so n is odd (since $\gcd(\deg(a), \deg(c)) = 1$). Since each ramification point of T_n has ramification index 2, we must have $s = 1$; thus $X = 0$ is a special point of $s^i h(X)^N$, so $\ell_4(0)$ is a special point of T_n and hence equals 2ϵ for some $\epsilon \in \{1, -1\}$. Thus $\ell_4 = \alpha X + 2\epsilon$ where $\alpha \in \mathbb{C}^*$. Now $\ell_4 \circ X^2 = \alpha X^2 + 2\epsilon = -\epsilon T_2(\gamma X)$ where $\gamma^2 = -\epsilon\alpha$, so $d = -\epsilon T_2 \circ \gamma \ell_3^{\langle -1 \rangle}$. Since $T_n(-\epsilon X) = -\epsilon T_n$, it follows that $c \circ d = -\epsilon T_{2n} \circ \gamma \ell_3^{\langle -1 \rangle} = (-\epsilon T_2) \circ (T_n \circ \gamma \ell_3^{\langle -1 \rangle})$. Since also $c \circ d = a \circ b$, by (2.9.3) we have $a = -\epsilon T_2 \circ \hat{\ell}$ and $b = \hat{\ell}^{\langle -1 \rangle} \circ T_n \circ \gamma \ell_3^{\langle -1 \rangle}$ for some linear $\hat{\ell}$. Combined with the expressions $c = -\epsilon T_n \circ (-\epsilon X)$ and $d = -\epsilon T_2 \circ \gamma \ell_3^{\langle -1 \rangle}$, this shows that (after perhaps changing the ℓ_j's) the quadruple Q can be written in the form (2.17.1).

One can use a similar (but easier) argument to prove the result when $d = T_n$; alternately, Theorem 2.13 and Lemma 3.6 imply a is dihedral, so the result follows from what was proved above. □

We now characterize the polynomials $X^s h(X^n)$ in terms of their self-equivalences.

Notation. For $f \in \mathbb{C}[X]$ with $\deg(f) > 1$, let $\Gamma(f)$ be the set of linear $\ell \in \mathbb{C}[X]$ for which there exists $\hat{\ell} \in \mathbb{C}[X]$ with $\hat{\ell} \circ f = f \circ \ell$.

Note that $\Gamma(f)$ is closed under composition and inversion, and hence is a group under composition. Further, $\Gamma(f)$ contains the group $\Gamma_0(f)$ defined in Remark 2.15.

Lemma 3.17. *Pick $f \in \mathbb{C}[X]$ with $\deg(f) = k > 1$. Then $\Gamma(f)$ is infinite if and only if there are linear $\ell_1, \ell_2 \in \mathbb{C}[X]$ for which $\ell_1 \circ f \circ \ell_2 = X^k$, in which case $\Gamma(f) = \{\ell_2 \circ \alpha \ell_2^{\langle -1 \rangle} : \alpha \in \mathbb{C}^*\}$. Also $|\Gamma(f)| = n > 1$ if and only if there are linear $\ell_1, \ell_2 \in \mathbb{C}[X]$ for which $\ell_1 \circ f \circ \ell_2 = X^s \hat{f}(X^n)$ where $s \geq 0$ and $\hat{f} \in \mathbb{C}[X]$ is neither a monomial nor a polynomial in X^j for any $j > 1$. In this case $\Gamma(f) = \{\ell_2 \circ \alpha \ell_2^{\langle -1 \rangle} : \alpha^n = 1\}$ is cyclic.*

Proof. Pick linear $\ell_1, \ell_2 \in \mathbb{C}[X]$ such that $g := \ell_1 \circ f \circ \ell_2$ is monic and has no terms of degrees $k - 1$ or 0. If $g \neq X^k$ then $\Gamma(g) = \{\alpha X : \alpha^n = 1\}$ where n is the greatest common divisor of the differences between degrees of terms of g, so $\Gamma(g)$ has order n where $g = X^s \hat{f}(X^n)$ with $s, \hat{f}$ as required. Since $\Gamma(X^k) = \{\alpha X : \alpha \in \mathbb{C}^*\}$ and $\Gamma(f) = \ell_2 \circ \Gamma(g) \circ \ell_2^{\langle -1 \rangle}$, the result follows. □

Much of Lemma 3.17 was proved in [2, §3]. This result allows us to determine all decompositions of even polynomials:

Corollary 3.18. *For $f, g \in \mathbb{C}[X] \setminus \mathbb{C}$ and $n > 1$, we have $f \circ g \in \mathbb{C}[X^n]$ if and only if $f = \hat{f}(X^{n/\gcd(n,s)}) \circ \ell^{\langle -1 \rangle}$ and $g = \ell \circ X^s \hat{g}(X^n)$ for some $r \geq 0$ and some $\hat{f}, \hat{g}, \ell \in \mathbb{C}[X]$ with ℓ linear.*

Proof. Let ζ be a primitive n^{th} root of unity. If $f \circ g \in \mathbb{C}[X^n]$ then $f \circ g = f \circ g(\zeta X)$, so by Corollary 2.9 we have $g = \tilde{\ell} \circ g(\zeta X)$ for some linear $\tilde{\ell}$. Thus $\zeta X \in \Gamma(g)$, so Lemma 3.17 implies $g = \ell \circ X^s \hat{g}(X^n) \circ \hat{\ell}$ where $s \geq 0$ and $\hat{g}, \ell, \hat{\ell} \in \mathbb{C}[X]$ with $\ell, \hat{\ell}$ linear; moreover, $\zeta X = \hat{\ell} \circ \hat{\zeta} \hat{\ell}^{\langle -1 \rangle}$ for some $\hat{\zeta} \in \mathbb{C}^*$, whence $\hat{\zeta} = \zeta$ and $\hat{\ell} = \gamma X$. Now $\ell^{\langle -1 \rangle} \circ g(\zeta X) = \zeta^s \ell^{\langle -1 \rangle} \circ g$, so $f\circ g = f\circ g(\zeta X) = f\circ \ell\circ \zeta^s \ell^{\langle -1\rangle}\circ g$. Thus $f = f\circ\ell\circ\zeta^s\ell^{\langle -1\rangle}$, so $f\circ\ell = (f\circ\ell)\circ\zeta^s X$, whence $f\circ\ell\in\mathbb{C}[X^{n/\gcd(n,s)}]$. $\square$

Remark 3.19. Corollary 3.18 was proved for $n = 2$ in [23, Prop. 1 and Thm. 1], and for prime n in [22, Thm. 4.3].

We now determine the equivalences between polynomials of the form $X^s h(X^n)$. Note that a polynomial can be written in this form with different values of s, n, and h; we now show that composing with linears does not introduce any essentially different expressions of this form.

Lemma 3.20. *Suppose $f := X^s h(X^n)$ and $g := X^r \hat{h}(X^m)$ satisfy $f = \ell_1\circ g\circ \ell_2$, where $h, \hat{h}\in\mathbb{C}[X]\setminus X\mathbb{C}[X]$, the ℓ_i are linear, and $m, n > 1$ and $r, s > 0$. Then $r = s$ and $h\in\mathbb{C}[X^{m/\gcd(m,n)}]$, and moreover if f is nonlinear then $\ell_1 = \gamma X$ and $\ell_2 = \alpha X$ with $\alpha,\gamma\in\mathbb{C}^*$.*

Proof. We may assume $\deg(f) > 1$, since otherwise the conclusion visibly holds. Since neither f nor g has a term of degree $(\deg(f) - 1)$ or 0, we must have $\ell_2 = \alpha X$ and $\ell_1 = \gamma X$ with $\alpha,\gamma\in\mathbb{C}^*$. Thus $X^s h(X^n) = \gamma\alpha^r X^r \hat{h}(\alpha^m X^m)$, so $r = s$ and $h(X^n)\in\mathbb{C}[X^n]\cap\mathbb{C}[X^m] = \mathbb{C}[X^{\operatorname{lcm}(m,n)}]$, which implies the result. $\square$

There can be nontrivial equivalences between polynomials of the form $X^s h(X)^n$: for instance, $X^2(X+1)^3 = X^3(X-1)^2\circ(X+1)$. We now give a presentation of a polynomial which displays all such equivalences.

Lemma 3.21. *Pick $h\in\mathbb{C}[X]\setminus X\mathbb{C}[X]$ and coprime $n > 1$ and $s > 0$, and suppose that $f := X^s h(X)^n$ is neither linear nor cyclic nor dihedral. Then $f = \tilde{h}(X)^{mq}\prod_{i=1}^k (X-\beta_i)^{mr_i q/q_i}$ where $\tilde{h}\in\mathbb{C}[X]$, $m, r_i > 0$, $q_i > 1$, $q = \prod_{i=1}^k q_i$, $\gcd(q_i, r_i q/q_i) = 1$, the $\beta_i\in\mathbb{C}$ are distinct, and $\tilde{h}(\beta_i)\neq 0$. Moreover, there is an expression of f in this form for which the following holds: for any linear $\ell_1,\ell_2\in\mathbb{C}[X]$ such that $\ell_1\circ f\circ\ell_2 = X^{\hat{s}}\hat{h}(X)^{\hat{n}}$ with $\hat{h}\in\mathbb{C}[X]\setminus X\mathbb{C}[X]$ and coprime $\hat{n} > 1$ and $\hat{s} > 0$, there exists i such that $\ell_1 = \gamma X$ and $\ell_2 = \alpha X + \beta_i$ with $\gamma,\alpha\in\mathbb{C}^*$, where $\hat{s} = mr_i q/q_i$ and $\hat{n}\mid q_i$.*

Proof. Let S be the set of roots of f, so S consists of 0 and the set of roots of h. Since f is neither linear nor cyclic, h is nonconstant, so since $h\notin X\mathbb{C}[X]$ it follows that h has nonzero roots. Each nonzero element of S is a ramification point of f with ramification index divisible by n; also, $e_f(0) = s$. Put $m := \gcd(e_f(\beta) : \beta\in S)$, and let $\beta_1 := 0, \beta_2, \dots, \beta_k$ be the elements of S for which $q_i := \gcd(e_f(\beta)/m : \beta\in S\setminus\{\beta_i\})$ satisfies $q_i > 1$. Write $R_i := e_f(\beta_i)/m$, so $\gcd(R_i, q_i) = 1$. Since $q_i \mid R_j$ for $i\neq j$, we must have $\gcd(q_i, q_j) = 1$. Putting $q := \prod_{i=1}^k q_i$, it follows that $f = \tilde{h}(X)^{mq}\prod_{i=1}^k (X-\beta_i)^{mr_i q/q_i}$ for some $\tilde{h}\in\mathbb{C}[X]$, where $r_i := R_i/\prod_{j\neq i} q_j$ is a positive integer coprime to q_i.

The roots of f contribute at least $(\deg(f)-1)/2$ to the Riemann–Hurwitz formula for the cover $\pi_f\colon\mathbb{P}^1\to\mathbb{P}^1$ corresponding to f, and if equality holds then $\deg(f)$ is odd and every root has multiplicity at most 2. Pick linear

$\ell_1, \ell_2 \in \mathbb{C}[X]$ such that $\ell_1 \circ f \circ \ell_2 = X^{\hat{s}}\hat{h}(X)^{\hat{n}}$, with $\hat{h} \in \mathbb{C}[X] \setminus X\mathbb{C}[X]$ and coprime $\hat{n} > 1$ and $\hat{s} > 0$. Then the preimages of $\ell_1^{\langle -1\rangle}(0)$ under f also contribute at least $(\deg(f) - 1)/2$ to the Riemann–Hurwitz formula for π_f. But the sum of the Riemann–Hurwitz contributions at all finite values is $\deg(f) - 1$, so if $\ell_1^{\langle -1\rangle}(0) \neq 0$ then f has precisely two finite branch points, and every finite ramification point has ramification index 2, whence f is dihedral (by Lemma 3.2), contradiction. Thus $\ell_1 = \gamma X$ for some $\gamma \in \mathbb{C}^*$. Next, $\hat{n} \mid e_f(\beta)$ for every $\beta \in S \setminus \{\ell_2(0)\}$, but $\hat{n}$ is coprime to $\hat{s} = e_f(\ell_2(0))$, so $\ell_2(0) = \beta_i$ and $\hat{n} \mid q_i$ for some i. Thus $\ell_2 = \alpha X + \beta_i$ for some $\alpha \in \mathbb{C}^*$, so $\gamma f(\alpha X + \beta_i) = X^{\hat{s}}\hat{h}(X)^{\hat{n}}$. Finally, equating the ramification indices of $X = 0$ on both sides gives $\hat{s} = mr_i q/q_i$. □

Finally, we determine equivalences between $X^s h(X)^n$ and $X^r\hat{h}(X^m)$.

Lemma 3.22. *Suppose $f := X^s h(X^n)$ and $g := X^r \hat{h}(X)^m$ satisfy $g = \ell_1 \circ f \circ \ell_2$, where $h, \hat{h} \in \mathbb{C}[X] \setminus X\mathbb{C}[X]$, the $\ell_i \in \mathbb{C}[X]$ are linear, and $m, n > 1$ and $r, s > 0$ are such that $\gcd(r, m) = \gcd(s, n) = 1$. If f is not linear or dihedral then $r = s$ and $h = h_0^m$ for some $h_0 \in \mathbb{C}[X]$, and moreover $\ell_1 = \gamma X$ and $\ell_2 = \alpha X$ with $\alpha, \gamma \in \mathbb{C}^*$.*

Proof. Since $\Gamma(f)$ contains ζX where ζ is a primitive n^{th} root of unity, $\Gamma(g)$ contains $\ell := \ell_2^{\langle -1\rangle} \circ \zeta\ell_2$. Here $\ell = \zeta X + \delta$ with $\delta \in \mathbb{C}$, and there is a linear $\hat{\ell} \in \mathbb{C}[X]$ for which

$$X^r\hat{h}(X)^m \circ \ell = \hat{\ell} \circ X^r\hat{h}(X)^m.$$

If f is not cyclic then Lemma 3.21 implies $\hat{\ell}(0) = 0$; by equating the roots of multiplicity s on the two sides of the above equality, we obtain $\ell(0) = 0$. If f is cyclic then Lemma 3.15 implies $\hat{h} \in \mathbb{C}^*$, so by Lemma 3.13 we again have $\ell(0) = \hat{\ell}(0) = 0$. Thus, in either case, $\hat{h}(\zeta X)$ is a scalar times $\hat{h}(X)$. Since these polynomials have the same nonzero constant term, it follows that $\hat{h} = \tilde{h}(X^n)$ for some $\tilde{h} \in \mathbb{C}[X] \setminus X\mathbb{C}[X]$. Since $\deg(g) = \deg(f)$ is coprime to n, and $\deg(g) \equiv r \pmod{n}$, we must have $\gcd(r, n) = 1$. Now Lemma 3.20 implies that $\ell_1 = \gamma X$ and $\ell_2 = \alpha X$ for some $\alpha, \gamma \in \mathbb{C}^*$, and also $r = s$. Thus $h(X^n)$ is a scalar times $\tilde{h}(X^n)^m$, so h is an m^{th} power. □

Remark 3.23. If f and g satisfy the hypotheses of Lemma 3.22, and f is neither linear nor cyclic nor dihedral, then we must have $k = 1$ in Lemma 3.21. If f satisfies the hypotheses of Lemma 3.21, and f is not a nontrivial power of another polynomial, then $m = 1$ and $k \leq 2$; this applies in particular when f is indecomposable. The non-dihedral hypotheses in the previous two lemmas cannot be removed: if n is odd then $T_n = X\tilde{h}(X^2 - 2)$ for some squarefree $\tilde{h} \in \mathbb{C}[X]$, and consequently $T_n + 2 = (X + 2)\tilde{h}(X)^2$ and $T_n - 2 = (X - 2)\tilde{h}(-X)^2$. If f is linear or cyclic then the last assertion of Lemma 3.21 does not hold.

4. Combining multiple Ritt moves

Pick $f \in \mathbb{C}[X]$ with $\deg(f) > 1$, and let $\mathcal{U} = (u_1, \ldots, u_r)$ and $\mathcal{V} = (v_1, \ldots, v_s)$ be complete decompositions of f. By Corollary 2.12, $s = r$ and the sequence $(\deg(u_1), \ldots, \deg(u_r))$ is a permutation of $(\deg(v_1), \ldots, \deg(v_r))$. Thus there is a unique permutation $\sigma = \sigma_{\mathcal{U},\mathcal{V}}$ of $\{1, 2, \ldots, r\}$ such that both

- $\deg(u_i) = \deg(v_{\sigma(i)})$ for $1 \le i \le r$; and
- if $1 \le i < j \le r$ satisfy $\deg(u_i) = \deg(u_j)$ then $\sigma(i) < \sigma(j)$.

Here σ defines a bijection between $\mathcal{U}$ and $\mathcal{V}$, via $\sigma\colon u_i \mapsto v_{\sigma(i)}$.

In this section we use the permutation $\sigma_{\mathcal{U},\mathcal{V}}$ to obtain information about the shape of f. We begin with a simple observation:

Lemma 4.1. *If integers $1 \le i < j \le r$ satisfy $\sigma_{\mathcal{U},\mathcal{V}}(i) > \sigma_{\mathcal{U},\mathcal{V}}(j)$, then* $\gcd(\deg(u_i), \deg(u_j)) = 1$.

Proof. By Theorem 2.1, there is a finite sequence of complete decompositions $\mathcal{U} = \mathcal{U}_0, \mathcal{U}_1, \ldots, \mathcal{U}_m = \mathcal{V}$ such that $\mathcal{U}_{k+1}$ is obtained from $\mathcal{U}_k$ by replacing two consecutive indecomposables a, b by two others c, d such that $a \circ b = c \circ d$. In this situation, Corollary 2.11 implies that either $\deg(a) = \deg(c)$ (and $\deg(b) = \deg(d)$) or $\gcd(\deg(a), \deg(b)) = \gcd(\deg(c), \deg(d)) = 1$. Pick the minimal k for which $\sigma_{\mathcal{U},\mathcal{U}_k}(i) > \sigma_{\mathcal{U},\mathcal{U}_k}(j)$, so $\mathcal{U}_k$ is obtained from $\mathcal{U}_{k-1}$ by a Ritt move involving the indecomposables of $\mathcal{U}_{k-1}$ which correspond to u_i and u_j, whence these indecomposables have coprime degrees. $\square$

We need more notation to state our results. For $1 \le i, j \le r$, define

$$\begin{aligned}
\mathcal{LL}(\mathcal{U}, \mathcal{V}, i, j) &= \{k : 1 \le k < i,\ \sigma(k) < \sigma(j)\};\\
\mathcal{LR}(\mathcal{U}, \mathcal{V}, i, j) &= \{k : 1 \le k < i,\ \sigma(k) > \sigma(j)\};\\
\mathcal{RL}(\mathcal{U}, \mathcal{V}, i, j) &= \{k : i < k \le r,\ \sigma(k) < \sigma(j)\};\\
\mathcal{RR}(\mathcal{U}, \mathcal{V}, i, j) &= \{k : i < k \le r,\ \sigma(k) > \sigma(j)\}.
\end{aligned}$$

Thus, for instance, $\mathcal{LR}(\mathcal{U}, \mathcal{V}, i, j)$ is the set of positions of indecomposables in $\mathcal{U}$ which lie to the left of u_i, but which correspond to indecomposables in $\mathcal{V}$ lying to the right of the indecomposable corresponding to u_j. We also write

$$LL(\mathcal{U}, \mathcal{V}, i, j) = \prod_{k \in \mathcal{LL}(\mathcal{U}, \mathcal{V}, i, j)} \deg(u_k),$$

and define $LR(\mathcal{U}, \mathcal{V}, i, j)$, $RL(\mathcal{U}, \mathcal{V}, i, j)$, and $RR(\mathcal{U}, \mathcal{V}, i, j)$ analogously.

Proposition 4.2. *Pick two complete decompositions $\mathcal{U} = (u_1, \ldots, u_r)$ and $\mathcal{V}$ of some polynomial $f \in \mathbb{C}[X]$, and pick k with $1 \le k \le r$. Write $LL = LL(\mathcal{U}, \mathcal{V}, k, k)$, and define LR, RL, and RR analogously. Then LR, RL and*

$\deg(u_k)$ *are pairwise coprime, and there exist polynomials*

$$\begin{array}{rl} a & \textit{of degree } LL, \\ d & \textit{of degree } RR, \\ b, \hat{b}, \tilde{b}, \dot{b} & \textit{of degree } LR, \\ c, \tilde{c}, \overline{c}, \dot{c} & \textit{of degree } RL, \textit{ and} \\ \hat{u}, \tilde{u}, \overline{u} & \textit{indecomposable of the same degree as } u_k \end{array}$$

such that

(4.2.1) $u_1 \circ u_2 \circ \cdots \circ u_{k-1} = a \circ b$ *and* $u_{k+1} \circ \cdots \circ u_r = c \circ d$*;*
(4.2.2) $b \circ u_k = \hat{u} \circ \hat{b}$*;*
(4.2.3) $\hat{u} \circ \hat{b} \circ c = \tilde{c} \circ \tilde{u} \circ \tilde{b}$*;*
(4.2.4) $u_k \circ c = \overline{c} \circ \overline{u}$*; and*
(4.2.5) $b \circ \overline{c} \circ \overline{u} = \dot{c} \circ \tilde{u} \circ \dot{b}$*.*

Proof. The coprimality assertions follow from Lemma 4.1. Write $p = \deg(u_k)$. Put $g = u_1 \circ \cdots \circ u_{k-1}$ and $h = u_{k+1} \circ \cdots \circ u_r$. Then $f = g \circ u_k \circ h$ and $\deg(g) = LL \cdot LR$ and $\deg(h) = RL \cdot RR$. Likewise, letting $\tilde{u}$ denote the indecomposable in $\mathcal{V}$ corresponding to u_k, from $\mathcal{V}$ we get $f = \tilde{g} \circ \tilde{u} \circ \tilde{h}$ where $\tilde{g}, \tilde{h} \in \mathbb{C}[X]$ have degrees $LL \cdot RL$ and $LR \cdot RR$, respectively. By Lemma 2.8, there are $a, b, \hat{g} \in \mathbb{C}[X]$ such that $g = a \circ b$ and $\tilde{g} = a \circ \hat{g}$, where $\deg(a) = \gcd(\deg(g), \deg(\tilde{g}))$; since $\gcd(LR, RL) = 1$, this means $\deg(a) = LL$ (so $\deg(b) = LR$). Likewise, there are $c, d, \hat{h} \in \mathbb{C}[X]$ such that $h = c \circ d$ and $\tilde{h} = \hat{h} \circ d$, where $\deg(d) = RR$ and $\deg(c) = RL$. This proves (4.2.1).

Applying Lemma 2.8 to $(a \circ b \circ u_k) \circ (c \circ d) = (a \circ \hat{g} \circ \tilde{u}) \circ (\hat{h} \circ d)$, we obtain $a_0, \hat{b} \in \mathbb{C}[X]$ such that $a \circ b \circ u_k = a_0 \circ \hat{b}$ and $\deg(a_0) = \gcd(LL \cdot LR \cdot p, LL \cdot RL \cdot p) = LL \cdot p$, so $\deg(\hat{b}) = \deg(b)$. Applying Corollary 2.9 to $a \circ (b \circ u_k) = a_0 \circ \hat{b}$ gives $b \circ u_k = \hat{u} \circ \hat{b}$ for some $\hat{u} \in \mathbb{C}[X]$. A complete decomposition of $\hat{b}$ contains no indecomposable of degree $\deg(u_k)$ (since $\deg(\hat{b}) = LR$ is coprime to $\deg(u_k)$), so Corollary 2.12 implies $\hat{u}$ is indecomposable, which proves (4.2.2).

Next recall that $f = a \circ (\hat{g} \circ \tilde{u} \circ \hat{h} \circ d)$ where $\deg(\hat{g}) = \deg(c)$ and $\deg(\hat{h}) = \deg(b)$. Since also $f = a \circ (\hat{u} \circ \hat{b} \circ c \circ d)$, by Corollary 2.9 there is a linear $\ell \in \mathbb{C}[X]$ such that $\ell \circ \hat{g} \circ \tilde{u} \circ \hat{h} \circ d = \hat{u} \circ \hat{b} \circ c \circ d$. Putting $\tilde{c} = \ell \circ \hat{g}$ and $\tilde{b} = \hat{h}$, we obtain $\tilde{c} \circ \tilde{u} \circ \tilde{b} \circ d = \hat{u} \circ \hat{b} \circ c \circ d$. As above, Corollary 2.12 implies $\tilde{u}$ is indecomposable, which proves (4.2.3).

Assertions (4.2.4) and (4.2.5) follow by symmetry. ∎

Proposition 4.2 enables us to control the cumulative effect of a sequence of Ritt moves. We do this in three results: Proposition 4.3 addresses the case that some indecomposable is neither cyclic nor dihedral; Proposition 4.4 the case that some indecomposable is dihedral; and the easier Lemma 4.6 handles the case that every indecomposable is cyclic.

Proposition 4.3. *Let $\mathcal{U} = (u_1, \dots, u_r)$ and $\mathcal{V}$ be two complete decompositions of $f \in \mathbb{C}[X]$. Pick k with $1 \le k \le r$, and put $n = LR(\mathcal{U}, \mathcal{V}, k, k)$ and $m = RL(\mathcal{U}, \mathcal{V}, k, k)$. If u_k is neither cyclic nor dihedral, then there exist $a, d, \tilde{h} \in \mathbb{C}[X]$ and $\beta, \delta \in \mathbb{C}$ and $s \ge 0$ with $\gcd(s, mn) = 1$ such that*

$$u_1 \circ \cdots \circ u_{k-1} = a \circ X^n \circ (X + \delta)$$
$$u_k = (X - \delta) \circ X^s \tilde{h}(X^n)^m \circ (X + \beta)$$
$$u_{k+1} \circ \cdots \circ u_r = (X - \beta) \circ X^m \circ d.$$

In particular, $mn < \deg(u_k)$.

Proof. Let $a, b, c, d, \overline{c}, \overline{u}, \dot{c}, \tilde{u}, \dot{b}$ be as in Proposition 4.2, so n and m are coprime to each other and to $\deg(u_k)$. Since $u_k \circ c = \overline{c} \circ \overline{u}$ and $\deg(\overline{c}) = \deg(c) = m$ is coprime to $\deg(u_k)$, Theorem 2.17 implies (because u_k is not cyclic or dihedral) that $u_k = \ell_1 \circ X^s h(X)^m \circ \ell_2$ and $c = \ell_2^{\langle -1 \rangle} \circ X^m \circ \ell_3$ for some $h \in \mathbb{C}[X] \setminus X\mathbb{C}[X]$, some linear $\ell_j \in \mathbb{C}[X]$, and some $s \ge 0$ which is coprime to m. By replacing h and ℓ_3 by scalar multiples of themselves, we may assume $\ell_1 = X - \delta$ and $\ell_2 = X + \beta$ with $\beta, \delta \in \mathbb{C}$. If $n = 1$ then the result follows upon replacing a by $a \circ b \circ (X - \delta)$ and d by $\ell_3 \circ d$. A similar argument applies if $m = 1$, so assume $m, n > 1$. Since u_k is neither cyclic nor dihedral, Lemma 3.8 implies that $u_k \circ c = \overline{c} \circ \overline{u}$ is neither cyclic nor dihedral. Now $b \circ (\overline{c} \circ \overline{u}) = (\dot{c} \circ \tilde{u}) \circ \dot{b}$, and also $\deg(\overline{c} \circ \overline{u}) = \deg(\dot{c} \circ \tilde{u})$ is coprime to $\deg(b) = n$, so by Theorem 2.17 we have $b = \ell_4 \circ X^n \circ \ell_5$ and $\overline{c} \circ \overline{u} = \ell_5^{\langle -1 \rangle} \circ X^{\hat{s}} \hat{h}(X^n) \circ \ell_6$ for some $\hat{h} \in \mathbb{C}[X] \setminus X\mathbb{C}[X]$, some linear $\ell_j \in \mathbb{C}[X]$, and some $\hat{s} \ge 0$ which is coprime to n. As above, we may assume $\ell_5 = X + \gamma$ with $\gamma \in \mathbb{C}$. Thus

$$(X - \delta) \circ X^{ms} h(X^m)^m \circ \ell_3 = u_k \circ c = \overline{c} \circ \overline{u} = (X - \gamma) \circ X^{\hat{s}} \hat{h}(X^n) \circ \ell_6,$$

so by Lemma 3.20 we have $\gamma = \delta$ and $h(X^m)^m \in \mathbb{C}[X^n]$. Since $\gcd(m, n) = 1$, it follows that $h \in \mathbb{C}[X^n]$, which gives the result once we replace a by $a \circ \ell_4$ and d by $\ell_3 \circ d$. □

Proposition 4.4. *Let $\mathcal{U} = (u_1, \dots, u_r)$ and $\mathcal{V}$ be two complete decompositions of some $f \in \mathbb{C}[X]$. Pick k, i such that $1 \le k, i \le r$ and u_k is dihedral.*

(4.4.1) *If $i > k$ and $RL(\mathcal{U}, \mathcal{V}, i-1, k) > 2$, then $u_k \circ u_{k+1} \circ \cdots \circ u_i$ is dihedral.*

(4.4.2) *If $i < k$ and $LR(\mathcal{U}, \mathcal{V}, i+1, k) > 2$, then $u_i \circ u_{i+1} \circ \cdots \circ u_k$ is dihedral.*

Proof. Since the proofs of the two parts are similar, we just give the details for (4.4.1). So assume that $n := RL(\mathcal{U}, \mathcal{V}, i-1, k)$ satisfies $n > 2$. It suffices to prove the result in case i is chosen as large as possible so that this inequality holds (by Lemma 3.8). Thus we may assume $i \in \mathcal{RL}(\mathcal{U}, \mathcal{V}, i-1, k)$. Write $\hat{n} := RL(\mathcal{U}, \mathcal{V}, k, k)$, so $n \mid \hat{n}$. With notation as in Proposition 4.2, we have $u_{k+1} \circ \cdots \circ u_r = c \circ d$ where $\deg(c) = \hat{n}$, and also $u_k \circ c = \overline{c} \circ \overline{u}$ where $\deg(\overline{c}) = \deg(c)$ and $\gcd(\deg(c), \deg(u_k)) = 1$. Write $u_k = \ell_1 \circ T_m \circ \ell_2$ with $m > 2$ and ℓ_1, ℓ_2 linear. Then Lemma 3.16 implies $c = \ell_2^{\langle -1 \rangle} \circ \epsilon T_{\hat{n}} \circ \ell_3$ for some linear ℓ_3 and some $\epsilon \in \{1, -1\}$. Let $g = u_{k+1} \circ u_{k+2} \circ \cdots \circ u_{i-1}$

(and let $g = X$ if $i = k+1$). Then $c \circ d = g \circ h$ where $h = u_i \circ \cdots \circ u_r$. By Lemma 2.8, we have $c = a \circ c_0$ and $g = a \circ g_0$, and also $d = d_0 \circ b$ and $h = h_0 \circ b$ and $c_0 \circ d_0 = g_0 \circ h_0$, where $a, b, c_0, d_0, g_0, h_0 \in \mathbb{C}[X]$ satisfy $\deg(a) = \gcd(\deg(c), \deg(g))$ and $\deg(b) = \gcd(\deg(d), \deg(h))$. Lemma 3.8 implies that $a = \ell_2^{\langle -1 \rangle} \circ \epsilon T_s \circ \ell_4$ and $c_0 = \ell_4^{\langle -1 \rangle} \circ T_{\hat{n}/s} \circ \ell_3$ for some linear ℓ_4; by replacing c_0, g_0 and a by $\ell_4 \circ c_0$, $\ell_4 \circ g_0$ and $a \circ \ell_4^{\langle -1 \rangle}$, we may assume $\ell_4 = X$.

By Lemma 4.1, for $k+1 \le j \le i-1$, if $\gcd(\deg(u_j), n) > 1$ then $\sigma(j) < \sigma(k)$; since $\hat{n}/n$ is the product of $\deg(u_j)$ over all j for which $k+1 \le j \le i-1$ and $\sigma(j) < \sigma(k)$, it follows that $\deg(g)/(\hat{n}/n)$ is coprime to n. Plainly $\hat{n}/n$ divides $\gcd(\deg(g), \hat{n}) = s$. Now $s' := s/(\hat{n}/n)$ divides $\deg(g)/(\hat{n}/n)$, and so is coprime to n, and $s'(\hat{n}/n) = s$ divides $\hat{n}$ so $s' \mid n$, whence $s' = 1$ and $\hat{n} = ns$.

By definition, $\gcd(\deg(c_0), \deg(g_0)) = \gcd(\deg(d_0), \deg(h_0)) = 1$ and $c_0 \circ d_0 = g_0 \circ h_0$. Since $c_0 = T_{\hat{n}/s} \circ \ell_3 = T_n \circ \ell_3$, Lemma 3.16 implies that $g_0 \circ h_0 = \hat{\epsilon} T_{\hat{m}} \circ \ell_5$ for some linear $\ell_5 \in \mathbb{C}[X]$ and some $\hat{\epsilon} \in \{1, -1\}$. Thus $u_k \circ g \circ h_0 = u_k \circ a \circ (g_0 \circ h_0) = \ell_1 \circ T_n \circ \epsilon T_s \circ \hat{\epsilon} T_{\hat{m}} \circ \ell_5$ is dihedral. Now n is divisible by $\deg(u_i)$ (since $i \in \mathcal{RL}$), and $\deg(h_0) = \deg(c_0) = n$, so by applying Corollary 2.9 to the decompositions $h_0 \circ b = h = u_i \circ (u_{i+1} \circ \cdots \circ u_r)$ we find that $h_0 = u_i \circ h_1$ for some $h_1 \in \mathbb{C}[X]$. Finally, since $u_k \circ g \circ h_0$ is dihedral, Lemma 3.8 implies that $u_k \circ g \circ u_i$ is dihedral, as desired. $\square$

Remark 4.5. Proposition 4.4 would not be true if we only required that $RL(\mathcal{U}, \mathcal{V}, i-1, k) \ge 2$, since for instance $T_3 \circ (-2 + X(X+1)^2) \circ X^2 = T_2 \circ T_3 \circ X(X^2+1)$ is not dihedral.

The proof of Proposition 4.4 can be adapted to apply when u_k is cyclic, although it leads to a result with a rather complicated formulation. Instead of doing this, we give a result which applies in the one situation not covered by the previous two results, namely when every u_k is cyclic. By Lemma 3.9, the composition $u \circ v$ of cyclic polynomials is cyclic if and only if the finite branch point of v equals the finite ramification point of u. Conversely, we now show that if each u_k in a complete decomposition is cyclic, then we can group together blocks of consecutive u_k's whose composition is cyclic, and any two u_k's whose relative positions are interchanged via a sequence of Ritt moves must lie in the same block.

Lemma 4.6. *Let $\mathcal{U} = (u_1, \ldots, u_r)$ be a complete decomposition of $f \in \mathbb{C}[X]$ in which each u_i is cyclic. Pick k with $1 \le k < r$, and suppose the finite ramification point of u_k differs from the finite branch point of u_{k+1}. Then for any complete decomposition $\mathcal{V}$ of f, and any i with $1 \le i \le r$, we have $\sigma_{\mathcal{U},\mathcal{V}}(j) \le k$ if and only if $j \le k$.*

Proof. We first show that, in any Ritt move $u_i \circ u_{i+1} = c \circ d$, the composition $u_i \circ u_{i+1}$ is cyclic. The Ritt move cannot be of type (2.17.1), since in that case u_i and u_{i+1} would be equivalent to Chebychev polynomials, and thus would have degree 2 (by Lemma 3.14), contradicting the fact that their degrees are coprime. Thus the Ritt move is of type (2.17.2), so Lemma 3.15 implies that $u_i \circ u_{i+1}$ is cyclic. It follows that c and d are cyclic, and moreover (by Lemma 3.9) the finite branch point of u_{i+1} equals the finite ramification point of u_i (so $i \neq k$). Furthermore, the finite branch point of $u_i \circ u_{i+1}$ equals that of both u_i and c, and the finite ramification point of $u_i \circ u_{i+1}$ equals that of both u_{i+1} and d.

Let $\mathcal{W} = (w_1, \dots, w_r)$ be a Ritt neighbor of $\mathcal{U}$, so $w_j = u_j$ for $j \notin \{i, i+1\}$ and $u_i \circ u_{i+1} = w_i \circ w_{i+1}$. Suppose first that $u_i = w_i \circ \ell$ and $u_{i+1} = \ell^{\langle -1 \rangle} \circ w_{i+1}$ for some linear ℓ. Then w_i and w_{i+1} are cyclic, u_i and w_i have the same finite branch point, u_{i+1} and w_{i+1} have the same finite ramification point, and if $i = k$ then the finite branch point of w_{k+1} differs from the finite ramification point of w_k. In the previous paragraph we showed that these properties also hold if $u_i \circ u_{i+1} = w_i \circ w_{i+1}$ is a Ritt move, in which case we must have $i \neq k$. By Corollary 2.11, it follows that these properties hold in every case. Thus, in every case, the finite branch point of w_{k+1} differs from the finite ramification point of w_k, both w_k and w_{k+1} are cyclic, and $\sigma_{\mathcal{U},\mathcal{W}}(j) \leq k$ if and only if $j \leq k$. By induction, the same properties hold if $\mathcal{U}$ and $\mathcal{W}$ are contained in a finite sequence of complete decompositions of f in which any two decompositions are Ritt neighbors. Thus, the result follows from Theorem 2.1. $\square$

We can now give our new description of the collection of all complete decompositions of a polynomial. We begin with a decomposition $\mathcal{U} = (u_1, \dots, u_r)$ of f in which each u_i is either indecomposable or cyclic or dihedral. We then move cyclic factors as far to the right as possible, by the following procedure. For each $i = 1, 2, \dots,$ do the following: if $u_i \circ u_{i+1}$ is cyclic or dihedral then replace u_i and u_{i+1} by $u_i \circ u_{i+1}$ and repeat step i. Otherwise, if $u_i = g \circ X^m \circ \ell$ with $g \in \mathbb{C}[X]$ and ℓ linear and $m > 1$ maximal, and $u_{i+1} = \ell^{\langle -1 \rangle} \circ X^s h(X^n) \circ \hat{\ell}$ with $\hat{\ell}$ linear, h nonconstant, $s \geq 0$, and n is maximal, then put $k := \gcd(n, m)$, and assume $k > 1$. If h is a monomial then replace u_i and u_{i+1} by $g \circ X^{m/k}$ and $h(X^n)^k \circ \hat{\ell}$; otherwise replace u_i and u_{i+1} by $g \circ X^{m/k}$ and $X^s h(X^{n/k})^k$ and $X^k \circ \hat{\ell}$ unless $g \circ X^{m/k}$ is linear, in which case replace the new u_i and u_{i+1} by their composition and repeat step i. Having moved all cyclic factors to the right, now move some of them to the left as follows. For each $i = |\mathcal{U}|, \dots, 2$ do the following. If $u_{i-1} \circ u_i$ is cyclic then replace u_{i-1} and u_i by $u_{i-1} \circ u_i$. Otherwise make no change except perhaps in case $u_i = \ell \circ X^m \circ g$ with $g \in \mathbb{C}[X]$ and ℓ linear and $m > 1$ maximal, and $u_{i-1} = \hat{\ell} \circ X^s h(X)^n \circ \ell^{\langle -1 \rangle}$ with $s > 0$, $h \in \mathbb{C}[X] \setminus \mathbb{C}$, and n maximal. In this case, either make no change or choose a divisor $k > 1$ of $\gcd(m, n)$. If $u_{i-1} \circ X^k$ is dihedral, then replace u_{i-1} and u_i by $u_{i-1} \circ X^k$ and $X^{m/k} \circ g$, unless the latter polynomial is linear in which case replace u_{i-1} and u_i by their composition. Otherwise replace u_{i-1} and u_i by $\hat{\ell} \circ X^k$, $X^s h(X^k)^{n/k}$, and $X^{m/k} \circ g$, and repeat step i unless $X^{m/k} \circ g$ is linear, in which case replace the new u_i and u_{i+1} by their composition. Finally, expand $\mathcal{U}$ into a complete decomposition by replacing each cyclic or dihedral u_i by one of the following types of complete decompositions: if $u_i = \ell_1 \circ X^n \circ \ell_2$ then choose any permutation $(p_1, \dots, p_s)$ of the prime factors (counted with multiplicities) of n, and replace u_i by $\ell_1 \circ X^{p_1}$, X^{p_2}, $\dots$, $X^{p_s} \circ \ell_2$; and similarly if u_i is dihedral.

The results of this section and the previous section show that this procedure yields a representative of every equivalence class of complete decompositions of f. The results of Section 3.4 control the different ways of writing the various polynomials u_i in the forms required in the procedure.

We now prove a refinement of Theorem 1.4. Here we write $\mathcal{Z}$ for the set of polynomials of degree at least 2 which are equivalent to either $X^s h(X^n)$ or $X^s h(X)^n$ for some $h \in \mathbb{C}[X]$ and some coprime positive integers s, n with $n > 1$. Note that $\mathcal{Z}$ contains X^m for every $m > 1$, and $\mathcal{Z}$ contains T_m for every odd $m > 1$ (and also for $m = 2$). Thus, an indecomposable is in $\mathcal{Z}$ if and only if it occurs in a Ritt move. Recall that $f^{\langle k \rangle}$ denotes the k^{th} iterate of f.

Theorem 4.7. *Pick $f \in \mathbb{C}[X]$ with $n = \deg(f) > 1$. Let $a, b \in \mathbb{C}[X] \setminus \mathbb{C}$ and $k > 1$ satisfy $r \circ s = f^{\langle k \rangle}$, and assume there is no $g \in \mathbb{C}[X]$ for which either $a = f \circ g$ or $b = g \circ f$. Let $\mathcal{U} = (u_1, \dots, u_r)$ be a complete decomposition of f. Then, for each i with $1 \le i \le r$, we have:*

(4.7.1) *If $u_i \notin \mathcal{Z}$ then $m \le 2$.*

(4.7.2) *If $k > 1$ and u_i is neither cyclic nor dihedral, then either $n \ge 6\deg(u_i) \ge 6(2^{k-2} + 1)$ or $n \ge 2\deg(u_i) \ge 2^k + 2$.*

(4.7.3) *If $k > 3$ and u_i is dihedral, then $f = \ell \circ \epsilon T_n \circ \ell^{\langle -1 \rangle}$ for some linear $\ell \in \mathbb{C}[X]$ and some $\epsilon \in \{1, -1\}$.*

(4.7.4) *If $k > 2$ and u_j is cyclic for every $1 \le j \le r$, then $f = \ell \circ X^n \circ \ell^{\langle -1 \rangle}$ for some linear $\ell \in \mathbb{C}[X]$.*

Proof. Since $f^{\langle k \rangle} = a \circ b$, by Corollary 2.9 the nonexistence of g implies $\deg(f) \nmid \deg(a)$ and $\deg(f) \nmid \deg(b)$. Extend $\mathcal{U}$ to a complete decomposition $\mathcal{U}^k = (u_1, \dots, u_{kr})$ of $f^{\langle k \rangle}$, by putting $u_i = u_{i-r}$ for $r + 1 \le i \le kr$. Let $\mathcal{V} = (v_1, \dots, v_{kr})$ be a complete decomposition of $f^{\langle k \rangle}$ such that $a = v_1 \circ \dots \circ v_e$ for some e. The decompositions $\mathcal{U}^k$ and $\mathcal{V}$ will be implicit in what follows: for instance, we will write $\sigma(i)$, $\mathcal{LR}(i, j)$, and $RL(i, j)$ in place of $\sigma_{\mathcal{U}^k, \mathcal{V}}(i)$, $\mathcal{LR}(\mathcal{U}^k, \mathcal{V}, i, j)$, and $RL(\mathcal{U}^k, \mathcal{V}, i, j)$.

Since $\deg(f) \nmid \deg(a)$, there is an I with $1 \le I \le t$ such that $\sigma(I) > e$. It follows from Lemma 4.1 that $\sigma(I + jr) > e$ for $0 \le j < k$. Since $\deg(f) \nmid \deg(b)$, there exists J with $1 \le J \le r$ such that $\sigma(J + (k-1)r) \le e$, so $\sigma(J + jr) < e$ for $0 < j < k - 1$. In particular, $\sigma(J + (k-1)r) < \sigma(I)$, so Lemma 4.1 implies $\deg(u_I)$ and $\deg(u_J)$ are coprime. Thus, for $1 \le i \le r$, we have $\deg(f) \ge 2\deg(u_i)$; if u_i is neither cyclic nor dihedral then $\deg(u_i) \ge 4$, so (4.7.2) holds for $k = 2$.

Suppose henceforth that $k > 2$. For $1 \le i \le kr$, we write $\mathcal{LR}(i)$ and $LR(i)$ in place of $\mathcal{LR}(i, i)$ and $LR(i, i)$, and we define $\mathcal{RL}(i)$ and $RL(i)$ similarly. Pick $1 \le i \le r$. If $\sigma(i) > e$ then $J + jr \in \mathcal{RL}(i)$ for $0 < j < k$, so

$\deg(u_J)^{k-1} \mid RL(i)$; in particular, $\deg(u_J)^{k-1} \mid RL(I)$. If $\sigma(i+(k-1)r) \le e$ then $\deg(u_I)^{k-1} \mid LR(i+(k-1)r)$; thus, $\deg(u_I)^{k-1} \mid LR(J+(k-1)r)$. If $\sigma(i) \le e < \sigma(i+(k-1)r)$ then there is a unique m with $0 \le m < k-1$ such that $\sigma(i+mr) \le e < \sigma(i+(m+1)r)$. Thus $I+jr \in \mathcal{LR}(i+mr)$ for $0 \le j < m$, so $\deg(u_I)^m \mid LR(i+mr)$; similarly $\deg(u_J)^{k-m-2} \mid RL(i+(m+1)r)$.

Pick i with $1 \le i \le r$. Proposition 4.2 implies that $u_i \circ c = \overline{c} \circ \overline{u}$ for some $c, \overline{c}, \overline{u} \in \mathbb{C}[X]$ such that $\deg(c) = \deg(\overline{c})$ is coprime to $\deg(u_i)$, where $\deg(c)$ is the largest element of $\{RL(i+mr) : 0 \le m < k\}$. Similarly, $b \circ u_i = \hat{u} \circ \hat{b}$ for some $\overline{b}, \hat{b}, \hat{u} \in \mathbb{C}[X]$ such that $\deg(\overline{b}) = \deg(\hat{b})$ is coprime to $\deg(u_i)$, where $\deg(\overline{b})$ is the largest element of $\{LR(i+mr) : 0 \le m < k\}$. We showed above that $\overline{b}$ and c are not both linear; thus Theorem 2.17 implies $u_i \in \mathcal{Z}$, which proves (4.7.1). In fact, either $\deg(\overline{b})\deg(c) \ge \min(\deg(u_I), \deg(u_J))^{k-1} \ge 2^{k-1}$ or there is some m with $0 \le m < k-1$ such that $\deg(\overline{b})\deg(c) \ge \deg(u_I)^m \deg(u_J)^{k-m-2} \ge 2^{k-2}$. If $\deg(\overline{b})\deg(c) < 2^{k-1}$ then $i \notin \{I, J\}$, so since $\deg(u_I)$ and $\deg(u_J)$ are coprime we obtain $\deg(f) \ge \deg(u_i)\deg(u_I)\deg(u_J) \ge 6\deg(u_i)$.

If u_i is neither cyclic nor dihedral then Theorem 2.17 implies u_i is equivalent to both $X^s h(X)^{\deg(c)}$ and $X^{\hat{s}} \hat{h}(X^{\deg(\overline{b})})$, where $s, \hat{s} \ge 0$ and $h, \hat{h} \in \mathbb{C}[X] \setminus X\mathbb{C}[X]$ satisfy $\gcd(s, \deg(c)) = 1 = \gcd(\hat{s}, \deg(\overline{b}))$. By Lemma 3.22 it follows that u_i is equivalent to $X^{\tilde{s}} \tilde{h}(X^{\deg(\overline{b})})^{\deg(c)}$ for some $\tilde{h} \in \mathbb{C}[X] \setminus X\mathbb{C}[X]$ and some $\tilde{s} > 0$ which is coprime to $\deg(\overline{b})\deg(c)$. Thus $\deg(u_i) \ge 1 + \deg(\overline{b})\deg(c)$, which implies (4.7.2).

Now suppose u_i is dihedral and $k \ge 4$. If there is some j with $1 \le j \le r$ for which u_j is dihedral and $RL(2r, j) > 2$, then Proposition 4.4 implies $u_j \circ \cdots \circ u_{2r+1}$ is dihedral. Putting $h := u_j \circ \cdots \circ u_r$, we have $h \circ f \circ u_1 = \ell_1 \circ T_{mns} \circ \ell_2$ for some linear $\ell_1, \ell_2 \in \mathbb{C}[X]$, where $m = \deg(h)$ and $s = \deg(u_1)$. Note that $s > 1$ and $m > 2$ (since u_j dihedral). By Lemma 3.8, we have

$$\begin{aligned} h &= \ell_1 \circ T_m \circ \ell_3, \\ f &= \ell_3^{\langle -1 \rangle} \circ T_n \circ \ell_4, \qquad \text{and} \\ u_1 &= \ell_4^{\langle -1 \rangle} \circ T_s \circ \ell_2 \qquad \text{for some linear } \ell_3, \ell_4 \in \mathbb{C}[X]. \end{aligned}$$

Putting $g := u_1 \circ \cdots \circ u_{j-1}$, we have $g \circ h = f$, so $T_m \circ \ell_3 = \ell_5 \circ T_m \circ \ell_4$ for some linear ℓ_5. Now Lemma 3.13 implies $\ell_4 \circ \ell_3^{\langle -1 \rangle} = \epsilon X$ with $\epsilon \in \{1, -1\}$, so $f = \ell_4^{\langle -1 \rangle} \circ \epsilon T_n \circ \ell_4$, as desired.

So assume there is no j as above. Since $\deg(u_I)$ and $\deg(u_J)$ are coprime, they cannot both be even; by symmetry, we may assume $\deg(u_I)$ is odd. Since $J+2r, J+3r \in \mathcal{RL}(2r, I)$, our asumption on nonexistence of j implies u_I is not dihedral. This assumption also implies $\sigma(i) < e$, since otherwise $J+2r, J+3r \in \mathcal{RL}(2r, i)$. If $I < i$ then, since $\sigma(i) < e < \sigma(I)$, (4.4.2) would imply $u_I \circ \cdots \circ u_i$ is dihedral, which by Lemma 3.8 would imply u_I dihedral, contradiction. Thus $I > i$, and similarly $\sigma(I) < \sigma(i+r)$. Since $\sigma(J+2r) < \sigma(J+3r) \le e < \sigma(I) < \sigma(i+r)$, we have $J+2r, J+3r \in$

$\mathcal{RL}(2r, i+r)$, so Proposition 4.4 implies $u_{i+r} \circ \cdots \circ u_{2r+1}$ is dihedral; since $i + r < I + r < 2r + 1$, it follows that u_{I+r} is dihedral, contradiction.

Finally, suppose every u_j is cyclic. Since $\sigma(I) > \sigma(J + 2r)$, by Lemma 4.6 the finite ramification point of u_j equals the finite branch point of u_{j+1} for $I \le j < J + 2r$, and hence also for $1 \le j < kr$. Thus $f^{\langle k \rangle}$ is cyclic (by Lemma 3.9), so f is cyclic, whence $f = \ell_1 \circ X^n \circ \ell_2$ with ℓ_1, ℓ_2 linear. Then the finite ramification point of u_r equals that of f, namely $\ell_2^{\langle -1 \rangle}(0)$; likewise the finite branch point of $u_1 = u_{r+1}$ equals that of f, namely $\ell_1(0)$. Since these points coincide, $\ell_2 \circ \ell_1$ fixes 0, and so has the form αX with $\alpha \in \mathbb{C}^*$, whence $f = \ell_2^{\langle -1 \rangle} \circ \alpha X^n \circ \ell_2$. By replacing ℓ_2 by $\alpha^{1/(1-n)} \ell_2$, we may assume $\alpha = 1$, proving (4.7.4). $\square$

We now give examples showing that the conclusion of Theorem 4.7 cannot be improved.

Example 4.8. The exceptions in (4.7.3) and (4.7.4) cannot be avoided. First, by Corollary 2.9, if n is a prime power then the hypotheses of Theorem 4.7 cannot hold. Now assume $n > 1$ is not a prime power. For any linear $\ell \in \mathbb{C}[X]$, if we put $f := \ell \circ X^n \circ \ell^{\langle -1 \rangle}$ then $f^{\langle k \rangle} = \ell \circ X^{n^k} \circ \ell^{\langle -1 \rangle}$, so if $e > 1$ is a prime power dividing n such that $\gcd(e, n/e) = 1$, then $a := \ell \circ X^{(n/e)^k}$ and $b := X^{e^k} \circ \ell^{\langle -1 \rangle}$ satisfy $f^{\langle k \rangle} = a \circ b$ and $\deg(f) \nmid \deg(a), \deg(b)$. Note that $\deg(a), \deg(b) \to \infty$ as $k \to \infty$.

Similar remarks apply to $f := \ell \circ \epsilon T_n \circ \ell^{\langle -1 \rangle}$ for any linear $\ell \in \mathbb{C}[X]$ and any $\epsilon \in \{1, -1\}$.

Example 4.9. The bounds in (4.7.2) are best possible. For instance, pick an integer $m > 1$ and let $f_i := X(1 + X^{2^i})^{2^{m-i}}$ for $0 \le i \le m$. Then $X^2 \circ f_i = f_{i-1} \circ X^2$, so $f := f_m \circ X^2$ and $k := m + 1$ satisfy $f^{\langle k \rangle} = a \circ b$ for $a := f_m \circ f_{m-1} \circ \cdots \circ f_0$ and $b := X^{2^{m+1}}$. Here $\deg(f) = 2^k + 2$ does not divide $\deg(a)$ or $\deg(b)$. By Lemma 3.2, f_m is neither cyclic nor dihedral, since it has more than two finite branch points (because the 2^m roots of the derivative $f_m'(X)$ have distinct images under f_m). It follows from (4.7.2) that f_m is indecomposable; alternately, indecomposability of f_m is equivalent to primitivity of $\mathrm{Mon}(f_m)$, which holds because $\mathrm{Mon}(f_m) = S_{1+2^m}$ (as follows from Hilbert's theorem on monodromy groups of Morse polynomials, cf. [21, §III] or [37, §4.4]). Likewise, $\hat{f} := X^3 \circ f_m \circ X^2$ and $\hat{k} := m + 2$ satisfy $\hat{f}^{\langle \hat{k} \rangle} = \hat{a} \circ \hat{b}$ where $\hat{a} := X^3 \circ f_m \circ X^3 \circ f_{m-1} \circ X^3 \circ \cdots \circ f_0 \circ X^3$ and $\hat{b} := X^2 \circ f_0 \circ X^{2^{m+1}}$; here $\deg(\hat{f}) = 6(2^{\hat{k}-2} + 1)$ does not divide $\deg(\hat{a})$ or $\deg(\hat{b})$.

Example 4.10. The bound on k in (4.7.3) cannot be improved in general. Pick coprime odd $e, s > 1$, and put $f := X^2 \circ (X + 2) \circ T_e \circ (X - 2) \circ X^s$. Then $f^{\langle 3 \rangle} = a \circ b$ where $a := X^2 \circ (X + 2) \circ T_e \circ (X - 2) \circ X^4$ and $b := X^s \circ T_e \circ X^s \circ (X + 2) \circ T_e \circ (X - 2) \circ X^s$. Note that $\deg(f) \nmid \deg(a), \deg(b)$, and also f is neither cyclic nor dihedral (by Lemmas 3.8 and 3.14).

Example 4.11. The bounds on k in (4.7.1) and (4.7.4) cannot be improved. For instance, pick any $g \in \mathbb{C}[X] \setminus \mathbb{C}$ and put $f := X^2 \circ g \circ X^3$; then $f^{\langle 2 \rangle} = a \circ b$ where $a := X^2 \circ g \circ X^2$ and $b := X^3 \circ g \circ X^3$, and plainly $\deg(f) \nmid \deg(a), \deg(b)$. If $g = X + 1$ then f is the composition of cyclic indecomposables, but f is not cyclic (by Lemma 3.9). The hypotheses of (4.7.1) are satisfied whenever $f \notin \mathcal{Z}$, which holds for a Zariski-dense sublocus of the locus of polynomials g of any prescribed degree greater than 3; explicitly, $g := X^4 + X^2 + X$ is not in $\mathcal{Z}$.

We now deduce Theorem 1.4 from Theorem 4.7. We need the following simple result.

Lemma 4.12. *Pick $f \in \mathbb{C}[X]$ of degree $n > 1$. Pick $a, b \in \mathbb{C}[X]$ and $e > 0$ such that $a \circ b = f^{\langle e \rangle}$. Then there exist $\hat{a}, \hat{b} \in \mathbb{C}[X]$ and $i, j, k \geq 0$ such that*

$$a = f^{\langle i \rangle} \circ \hat{a} \quad \text{and} \quad b = \hat{b} \circ f^{\langle j \rangle} \quad \text{and} \quad \hat{a} \circ \hat{b} = f^{\langle k \rangle},$$

and also $\hat{a} \neq f \circ h$ and $\hat{b} \neq h \circ b$ for every $h \in \mathbb{C}[X]$.

Proof. Let $i, j \geq 0$ be maximal such that $\deg(f)^i \mid \deg(a)$ and $\deg(g)^j \mid \deg(b)$. Then Corollary 2.9 implies $a = g^{\langle i \rangle} \circ \hat{a}$ and $b = \hat{b} \circ g^{\langle j \rangle}$ for some $\hat{a}, \hat{b} \in \mathbb{C}[X]$. Thus $f^{\langle e \rangle} = a \circ b = f^{\langle i \rangle} \circ \hat{a} \circ \hat{b} \circ f^{\langle j \rangle}$, so $f^{\langle i \rangle} \circ (\hat{a} \circ \hat{b}) = f^{\langle e-j \rangle} = f^{\langle i \rangle} \circ f^{\langle e-j-i \rangle}$. If $i = 0$ then $a \circ b = f^{\langle e-j \rangle}$, so the required properties hold since $\deg(f)$ does not divide $\deg(a)$ or $\deg(b)$. Henceforth assume $i > 0$. By (2.9.3), there is a linear $\ell \in \mathbb{C}[X]$ for which $\ell \circ (a \circ b) = f^{\langle e-j-i \rangle}$ and $f^{\langle i \rangle} \circ \ell = f^{\langle i \rangle}$. Upon replacing a by $\ell \circ a$, we obtain the desired conclusion. □

Proof of Theorem 1.4. Lemma 4.12 gives everything but the bound on k. So suppose $a \circ b = f^{\langle k \rangle}$ with $k \geq 0$, where there is no $g \in \mathbb{C}[X]$ for which either $a = f \circ g$ or $b = g \circ f$. By Corollary 2.9, neither $\deg(a)$ nor $\deg(b)$ is divisible by $\deg(f)$. If f is indecomposable, Corollary 2.12 implies that each of a and b is either linear or the composition of indecomposables having the same degree as f; thus a and b must be linear, so $k = 0$, whence $k < \log_2(n)$. Now let $\mathcal{U} = (u_1, \ldots, u_r)$ be a complete decomposition of f, and assume $r > 1$. Then $n = \deg(f)$ satisfies $n \geq 2\deg(u_i) \geq 4$ for every i. If some u_i is neither cyclic nor dihedral, then (4.7.2) implies $k < \log_2(n)$. If every u_i is cyclic then (4.7.4) implies $k \leq 2 \leq \log_2(n)$. Finally, if some u_i is dihedral then (4.7.3) implies $k \leq 3$, so $k \leq \log_2(n)$ whenever $n \geq 8$. Since n is composite, the only possible exceptions are $n = 6$ (for which $k = \log_2(n+2)$) and $n = 4$. But if $n = 4$ then the degrees of a and b are powers of 2 which are not divisible by 4, so $\deg(a), \deg(b) \leq 2$ and thus $k \leq 1 < \log_2(n)$. □

Remark 4.13. The above proof shows that if $n \neq 6$ then the bound on k can be improved to $k \leq \log_2(n)$. This improvement is not possible for $n = 6$, since $f = T_3 \circ 2T_2$ satisfies $f^{\langle 3 \rangle} = (T_3 \circ 2T_3 \circ (4T_3 + 6)) \circ (T_4 \circ 2T_2)$ (and f is neither cyclic nor dihedral).

5. Related topics

We now briefly discuss some related topics. First, any polynomial (or rational function) over any field has only finitely many equivalence classes of decompositions. However, in most situations we know much less about these decompositions than we do in the case of polynomials over $\mathbb{C}$.

5.1. **Decomposition of rational functions.** Ritt [33, 34] studied decompositions of rational functions over $\mathbb{C}$. He recalled [34, p. 222] that the groups A_4, S_4, and A_5 act as groups of automorphisms of $\mathbb{C}(x)$, with fixed field $\mathbb{C}(f)$ where the equivalence classes of decompositions of f are in bijection with the (increasing) chains of subgroups of the relevant group. Since these groups contain distinct-length maximal chains of subgroups, the rational function analogue of Theorem 2.1 is not true. Further examples of distinct-length complete decompositions can be produced from group actions on the $j = 0$ and $j = 1728$ elliptic curves. There are only a few known theorems limiting the possibilities, the best being Ritt's classification of pairs of commuting rational functions [33]. For the current state of knowledge, see [27].

5.2. **Decomposition of polynomials over other fields.** All results and proofs in this paper work over arbitrary algebraically closed fields of characteristic zero. All but three of our results remain valid over an arbitrary algebraically closed field whose characteristic does not divide the degree of the relevant polynomials; the exceptions are Lemmas 3.1, 3.2, and 3.9. This generalization only presents difficulties for Theorem 2.17, where it was done by Zannier (cf. [39] or [36, §1.4]). There are versions of all results in this paper (with the above three exceptions) over any field K whose characteristic does not divide the degree of the polynomial $f \in K[X]$ under consideration; this is because in this situation every decomposition of f over the algebraic closure $\overline{K}$ is equivalent to a decomposition over K [26, §2]. (For instance, see [36, p. 25] for a version of Theorem 2.17 in this situation.) However, new phenomena occur when the characteristic divides $\deg(f)$:

- An indecomposable polynomial over K can decompose over $\overline{K}$; however, Guralnick and Saxl [19] proved this can only happen for polynomials of degree either a power of the characteristic, or 21 or 55. All examples of degree 21 or 55 were determined in [20]. Several families of examples of degree a power of the characteristic were given in [4], in addition to some partial classification results.
- Two complete decompositions of f can have distinct lengths [9, p. 98]; see [4] for further examples, and [3] for classes of indecomposables which cannot occur in any such examples.
- There are decomposable odd polynomials which are not the composition of two nonlinear odd polynomials [4].

Several of the results from Section 2.2 remain valid for decompositions into monic polynomials over any ring in which the degrees of the polynomials are units. We will expand on this point elsewhere.

5.3. **Monodromy groups of indecomposable polynomials.** In light of Theorem 1.3, it is of interest to determine the possible monodromy groups of indecomposable polynomials. This was done in [11, 29], according to which the possible groups are cyclic, dihedral, alternating, symmetric, and finitely many other groups of small degree. The analogous problem in positive characteristic is much more difficult: a reduced list of group-theoretic possibilities is given in [18], and there are families of indecomposable polynomials whose monodromy groups are quite different from the groups occurring in characteristic zero (see [1, 20] and the references therein). The latter families have remarkable properties: for instance, they include infinite families of pairs (f, g) of non-equivalent indecomposables such that $f(X) - g(Y)$ is reducible; and also they include several families of polynomials $f \in \mathbb{F}_q[X]$ for which the map $\alpha \mapsto f(\alpha)$ induces a bijection on $\mathbb{F}_{q^k}$ for infinitely many k.

5.4. **Algorithms.** Zippel [42] discovered a deterministic polynomial-time algorithm for finding a complete decomposition of a rational function f over an arbitrary field K. In case f is a polynomial of degree not divisible by the characteristic of K, the algorithm in [15] (following [24] and [26]) obtains such a decomposition in essentially linear time. By combining this algorithm with Ritt's results, one can compute representatives of all equivalence classes of complete decompositions of f by means of $\mathcal{O}(\deg(f)^3)$ arithmetic operations. Our results yield a faster algorithm, with optimal complexity. We will present the details elsewhere.

Appendix: Ritt's second theorem

We now prove Theorem 2.17.

In Section 2 we showed that many problems about polynomial decomposition reduce to questions about subgroups of the inertia group at infinity. However, there is no such reduction for the present question: besides the ramification at infinity, we need to keep track of the ramification at finite points as well. The problem amounts to the determination of all genus-zero curves of the form $a(X) = c(Y)$ with a, c polynomials of coprime degrees. We solve it by comparing contributions to the Riemann–Hurwitz formula for the covers $\mathbb{P}^1 \to \mathbb{P}^1$ corresponding to each of a, b, c, d, where $a \circ b = c \circ d$.

Proof of Theorem 2.17. Pick $a, b, c, d \in \mathbb{C}[X] \setminus \mathbb{C}$ such that $a \circ b = c \circ d$ and $\gcd(\deg(a), \deg(c)) = \gcd(\deg(b), \deg(d)) = 1$. Write $m := \deg(c)$ and $n := \deg(a)$, so $\gcd(m, n) = 1$ and also $m = \deg(b)$ and $n = \deg(d)$. The result is clear if $\min(m, n) = 1$, so assume $m, n > 1$. Let x be transcendental over $\mathbb{C}$, and put $t = a(b(x))$.

Let $P_1, \dots, P_k$ be the finite branch points of $\mathbb{P}^1_x \to \mathbb{P}^1_t$. For any i with $1 \le i \le k$, let $Q^i_1, \dots, Q^i_{q(i)}$ be the points of $\mathbb{P}^1_{b(x)}$ lying over P_i, and let α^i_j be the ramification index of Q^i_j/P_i. Likewise, let $R^i_1, \dots, R^i_{r(i)}$ be the points of $\mathbb{P}^1_{d(x)}$ lying over P_i, and let β^i_J be the ramification index of R^i_J/P_i. Then $n = \sum_{j=1}^{q(i)} \alpha^i_j$ and $m = \sum_{J=1}^{r(i)} \beta^i_J$. By Lemma 3.1, each point S of $\mathbb{P}^1_x$ lying over both Q^i_j and R^i_J has ramification index $\operatorname{lcm}(\alpha^i_j, \beta^i_J)$ in $\mathbb{P}^1_x \to \mathbb{P}^1_t$, and hence has ramification index $\operatorname{lcm}(\alpha^i_j, \beta^i_J)/\beta^i_J$ in $\mathbb{P}^1_x \to \mathbb{P}^1_{d(x)}$. Moreover, the number of such points S is $\gcd(\alpha^i_j, \beta^i_J)$. Thus, for each i, some α^i_j or β^i_J is greater than 1. Let A_i and B_i be the multisets $\{\alpha^i_1, \dots, \alpha^i_{q(i)}\}$ and $\{\beta^i_1, \dots, \beta^i_{r(i)}\}$, respectively. By applying the Riemann–Hurwitz formula to the covers $\mathbb{P}^1_{b(x)} \to \mathbb{P}^1_t$ and $\mathbb{P}^1_x \to \mathbb{P}^1_{d(x)}$, we obtain

$$n - 1 = \sum_{i=1}^{k} \sum_{j=1}^{q(i)} (\alpha^i_j - 1) = \sum_{i=1}^{k} (n - |A_i|) \quad \text{and}$$

$$\begin{aligned} n - 1 &= \sum_{i=1}^{k} \sum_{j=1}^{q(i)} \sum_{J=1}^{r(i)} \gcd(\alpha^i_j, \beta^i_J) \cdot \left(\frac{\operatorname{lcm}(\alpha^i_j, \beta^i_J)}{\beta^i_J} - 1 \right) \\ &= \sum_{i=1}^{k} \sum_{\alpha \in A_i} \sum_{\beta \in B_i} (\alpha - \gcd(\alpha, \beta)). \end{aligned}$$

Combined with the analogous expressions for $m - 1$, these equations imply that A_i and B_j satisfy the hypotheses of Lemma A below, in which we will determine the possibilities for A_i and B_j. We now determine the corresponding polynomials. If (C1) holds then some i has these properties: P_i has a unique preimage in $\mathbb{P}^1_{b(x)}$, there is a unique $\hat{J}$ for which $s := \beta^i_{\hat{J}}$ is coprime to n, and further $n \mid \beta^i_J$ for all $J \neq \hat{J}$. By replacing a and c by $\ell_1 \circ a$ and $\ell_1 \circ c$ with ℓ_1 linear, we may assume $P_i = 0$. By replacing a and b by $a \circ \ell_2$ and $\ell_2^{\langle -1 \rangle} \circ b$, we may assume $b = 0$ is the unique root of a (and also that $b = 1$ lies above $t = 1$), and likewise we may assume $d = 0$ is the unique root of c having multiplicity s. Then $a = X^n$ and $c = X^s H(X)^n$ for some $H \in \mathbb{C}[X]$. For any I, j, and J, each point of $\mathbb{P}^1_x$ lying over Q^I_j and R^I_J has ramification index $\alpha^I_j / \gcd(\alpha^I_j, \beta^I_J)$ in $\mathbb{P}^1_x \to \mathbb{P}^1_{d(x)}$. Since $\alpha^i_1 = n$ divides β^i_J whenever $Q^i_J \neq 0$, and also $\alpha^I_j = 1$ if $I \neq i$ (because $n - 1 = \sum_I (n - |A_I|)$ and $n = \sum_j \alpha^I_j$), it follows that $d = 0$ is the unique finite branch point of d. Upon composing b and d on the right with a linear, we may assume $d = X^n$. Then $a \circ b = c \circ d$ becomes $b^n = X^{sn} H(X^n)^n$, whence $b = \zeta X^s H(X^n)$ with $\zeta^n = 1$. Replacing H by ζH puts the quadruple (a, b, c, d) in the form (2.17.2). By symmetry, (c, d, a, b) has this form (after composing with linears) if (C2) holds. So assume the A_i and B_j satisfy (C3). Then $f = a \circ b$ has just two finite branch points, and every finite ramification point has ramification index at most 2, so Lemma 3.2 implies that f is dihedral. Now the result follows from Lemma 3.8. $\square$

Lemma A. *Pick coprime $m, n > 1$, and let $A_1, \dots, A_k$ and $B_1, \dots, B_k$ be multisets of positive integers such that, for each i, either A_i or B_i (or both) contains an integer greater than 1. Suppose further that*

$$\text{(H1)} \qquad \sum_{\alpha \in A_i} \alpha = n \quad \textit{and} \quad \sum_{\beta \in B_j} \beta = m \quad \textit{for each } 1 \le i \le k;$$

$$\text{(H2)} \qquad \sum_{i=1}^{k} (n - |A_i|) = n - 1 = \sum_{i=1}^{k} \sum_{\alpha \in A_i} \sum_{\beta \in B_i} (\alpha - \gcd(\alpha, \beta)); \quad \textit{and}$$

$$\text{(H3)} \qquad \sum_{i=1}^{k} (m - |B_i|) = m - 1 = \sum_{i=1}^{k} \sum_{\alpha \in A_i} \sum_{\beta \in B_i} (\beta - \gcd(\alpha, \beta)).$$

Then one of these holds:

(C1) *For some i we have $A_i = \{n\}$, one element of B_i is coprime to n, and all other elements of B_i are divisible by n; or*
(C2) *For some i we have $B_i = \{m\}$, one element of A_i is coprime to m, and all other elements of A_i are divisible by m; or*
(C3) *$k = 2$ and the largest element of $A_1 \cup A_2 \cup B_1 \cup B_2$ is 2.*

Proof. If $|A_1| = 1$ then (H1) implies $A_1 = \{n\}$; thus, by (H2), at most one element $\hat{\beta}$ of B_1 is not divisible by n. Since n is coprime to $m = \sum_{\beta \in B_1} \beta$, it follows that n is coprime to $\hat{\beta}$, so (C1) holds. Similarly, if $|B_1| = 1$ then (C2) holds. Henceforth we assume $|A_i|, |B_i| > 1$ for each i; by (H2), we have $|A_i| < n$ for at least two values i (so $k > 1$), and also $|B_j| < m$ for at least two values j. We may assume $|A_1|, |A_2| < n$.

Now suppose that, for each i with $1 \le i \le k$, we have

$$\text{(P1)} \qquad n - |A_i| = \sum_{\alpha \in A_i} \sum_{\beta \in B_i} (\alpha - \gcd(\alpha, \beta)).$$

We first show that $|B_i| \le (m+1)/2$ for $i \in \{1, 2\}$. If $1 \notin B_i$ then (H1) implies the stronger inequality $|B_i| \le m/2$, so assume $1 \in B_i$. Since $|A_i| < n$, by (P1) and (H1) we see that B_i contains precisely one copy of 1, and every other element of B_i is divisible by every element of A_i. But some element of A_i is at least 2 (since $|A_i| < n$), so all but one element of B_i is at least 2, whence $|B_i| \le (m+1)/2$ with equality just when $B_i = \{1, 2, 2, \dots, 2\}$ and every element of A_i is at most 2.

Now $m - 1 = \sum_{i=1}^{k} (m - |B_i|) \ge (m - |B_1|) + (m - |B_2|) \ge m - 1$, so for $i \in \{1, 2\}$ we have $|B_i| = (m+1)/2$, whence $B_i = \{1, 2, 2, \dots, 2\}$ and every element of A_i is at most 2. Moreover, if $k > 2$ then $|B_3| = m$, so $B_3 = \{1, 1, \dots, 1\}$; thus (P1) says that $n - |A_3| = \sum_{\alpha \in A_3} m(\alpha - 1) = m(n - |A_3|)$, so $|A_3| = n$ and $A_3 = \{1, 1, \dots, 1\}$, contradiction. This gives (C3), and concludes the proof if (P1) holds for every i.

For each i with $1 \le i \le k$, and each $\hat\alpha \in A_i$ and $\hat\beta \in B_i$, define

$$z(i,\hat\alpha) := 1 - \hat\alpha + \sum_{\beta\in B_i} (\hat\alpha - \gcd(\hat\alpha,\beta))$$

$$y(i,\hat\beta) := 1 - \hat\beta + \sum_{\alpha\in A_i} (\hat\beta - \gcd(\alpha,\hat\beta))$$

$$Z(i) := \sum_{\alpha\in A_i} z(i,\alpha) = |A_i| - n + \sum_{\alpha\in A_i}\sum_{\beta\in B_i} (\alpha - \gcd(\alpha,\beta))$$

$$Y(i) := \sum_{\beta\in B_i} y(i,\beta) = |B_i| - m + \sum_{\alpha\in A_i}\sum_{\beta\in B_i} (\beta - \gcd(\alpha,\beta)).$$

Thus $\sum_{i=1}^k Z(i) = 0$, and we have already proved the result if every $Z(i) = 0$, so we may assume $Z(1) < 0$ (because $Z(i) = 0$ if $|A_i| = n$). Likewise we may assume $Y(I) < 0$ for some I. We will deduce a contradiction. We compute

$$\begin{aligned}
\sum_{i=1}^k \Big(mn - \sum_{\alpha\in A_i}\sum_{\beta\in B_i} \gcd(\alpha,\beta)\Big) &= \sum_{i=1}^k \sum_{\alpha\in A_i}\sum_{\beta\in B_i} (\alpha\beta - \gcd(\alpha,\beta)) \\
&= m - 1 + \sum_{i=1}^k \sum_{\alpha\in A_i}\sum_{\beta\in B_i} (\alpha\beta - \beta) \\
&= m - 1 + \sum_{i=1}^k m(n - |A_i|) \qquad \text{(P2)} \\
&= m - 1 + m(n-1) = mn - 1.
\end{aligned}$$

(In the setting of Theorem 2.17, this is Riemann–Hurwitz for $\mathbb{P}^1_x \to \mathbb{P}^1_t$). If $Z(i) < 0$ then $1 \notin B_i$, so $|B_i| \le m/2$ and thus $\sum_{\alpha\in A_i}\sum_{\beta\in B_i} \gcd(\alpha,\beta) \le |B_i| \sum_{\alpha\in A_i} \alpha = |B_i| n \le nm/2$; similarly, the same conclusion holds if $Y(i) < 0$. But (P2) implies there is at most one i satisfying this conclusion, so $I = 1$ and $Y(i), Z(i) \ge 0$ for $i > 1$. Since $\sum_{i=1}^k Z(i) = 0$, we have $Z(i) \le -Z(1)$ for $i > 1$, and likewise $Y(i) \le -Y(1)$. Since $Z(1) < 0$, also $z(1,\alpha) < 0$ for some $\alpha \in A_1$. Thus α is not coprime to any element of B_1, and α divides all but at most one element of B_1, so there exists $D > 1$ dividing both α and every element of B_1. Then $\sum_{\beta\in B_1} \beta = m$ is divisible by D, so D is coprime to n and thus some $\alpha' \in A_1$ is not divisible by D. For $\beta \in B_1$ we have $\gcd(\alpha',\beta) \le \beta/2$, so $\sum_{\beta\in B_1}(\beta - \gcd(\alpha',\beta)) \ge m/2$. Also $|B_1| \le m/D \le m/2$, so $m - |B_1| = m/2 + \delta$ with $\delta \ge 0$, and similarly $n - |A_1| = n/2 + \gamma$ with $\gamma \ge 0$. Thus $Y(1) \ge |B_1| - m + \sum_{\beta\in B_1}(\beta - \gcd(\alpha',\beta)) \ge -\delta$, so $Y(i) \le \delta$ for any $i > 1$. For any $i > 1$ we have $n - |A_i| \le n - 1 - (n/2 + \gamma)$ (by (H2)), so $|A_i| \ge n/2 + \gamma + 1$, whence the number of 1's in A_i is at least $\sum_{\alpha\in A_i}(2-\alpha) = 2|A_i| - n \ge 2(\gamma+1)$. Thus $Y(i) \ge |B_i| - m + 2(\gamma+1)(m - |B_i|)$, so $\delta \ge (2\gamma+1)(m - |B_i|)$. Since $|B_i| < m$ for some $i > 1$, we obtain $\delta \ge 2\gamma+1$. Similarly, $\gamma \ge 2\delta + 1 \ge 4\gamma + 3$, which is impossible since $\gamma \ge 0$. $\square$

Remark. The proof of Theorem 2.17 becomes simpler if we assume in addition that a and b are indecomposable, or more generally that neither a nor c has the form $\ell \circ X^e \circ f$ with ℓ linear, $e > 1$, and f not a power of a linear polynomial. The latter condition is equivalent to requiring that, for each i, if $|A_i| > 1$ then the elements of A_i have no common factor exceeding 1; and similarly for B_i. If $|A_i| = 1$ then the beginning of the proof of the Lemma shows (C1) holds. So assume $|A_i| > 1$ for every i, and similarly $|B_i| > 1$. Since the elements of A_i have gcd 1, for $\beta \in B_i$ we have $y(i, \beta) \geq 0$, with equality just when β is coprime to an element of A_i and divides all other elements of A_i. Thus $Y(i) \geq 0$; since $\sum_{i=1}^{k} Y(i) = 0$, it follows that $Y(i) = 0$ for every i, so $y(i, \beta) = 0$ for every i and β, whence the above equality condition holds. In particular, if we pick i such that $|B_i| < m$, then B_i contains an element $\beta > 1$, so $|A_i| \leq (n+1)/2$. Since $n - 1 = \sum_i (n - |A_i|)$, there are at most two values i for which $|B_i| < m$, so there are exactly two and each satisfies $|A_i| = (n+1)/2$, whence $A_i = \{1, 2, \dots, 2\}$ and further the largest element of B_i is 2. For any other i, (H2) implies $|A_i| = n$, so A_i and B_i consist solely of 1's; this contradicts our hypothesis, so $k = 2$ and thus (C3) holds. This proves Lemma A, and the theorem follows as above.

Remark. Our proof of Theorem 2.17 is a simplified and rearranged version of Ritt's proof. Ritt's proof looks rather different, since he worked in terms of the monodromy group of the Riemann surface for $f(x) - z$, and gave a cumbersome description of elements of this group via their action on branches. This is logically equivalent to what we did above, but it was viewed by some as being unduly difficult. Consequently, several authors rewrote Ritt's proof in other languages, usually under the simplifying assumption that a, b, c, d are indecomposable. In this special case, Ritt's proof has been rewritten in terms of polynomial arithmetic ([26], [25, §2 of Ch. 4] and [8]), valuation theory [9], and group theory [29]. Ritt's proof of the full Theorem 2.17 has been translated into the language of polynomial arithmetic [35, §5], as well as into a language closer to ours [7, Thm. 6.1]. There is also a valuation-theoretic version of Ritt's proof [38], including a different proof of Lemma A. Finally, as in the previous remark, it is easier to prove Theorem 2.17 when neither $a - \alpha$ nor $b - \alpha$ is a nontrivial power of a nonlinear polynomial for any $\alpha \in \mathbb{C}$; one can deduce the full result from this by a different kind of argument [39] (see also [36, §1.4] or [31, §9]). A flawed attempt at such an approach is [13, Thm. 2]. Our proof is arranged quite differently from previous ones, and we hope this makes it more understandable.

References

[1] S. S. Abhyankar, *Symplectic groups and permutation polynomials, part II*, Finite Fields Appl. **8** (2002), 233–255.

[2] G. af Hällström, *Über halbvertauschbare Polynome*, Acta Acad. Abo. **21** (1957), no. 2, 20 pp.

[3] R. M. Beals, J. L. Wetherell and M. E. Zieve, *Polynomials with a common composite*, Israel J. Math., to appear, arXiv:0707.1552.

[4] R. M. Beals and M. E. Zieve, *Polynomial decomposition in characteristic* p, preprint, 2007.

[5] A. F. Beardon and T. W. Ng, *On Ritt's factorization of polynomials*, J. London Math. Soc. **62** (2000), 127–138.

[6] Y. F. Bilu, *Quadratic factors of* $f(x) - g(y)$, Acta Arith. **90** (1999), 341–355.

[7] Y. F. Bilu and R. F. Tichy, *The Diophantine equation* $f(x) = g(y)$, Acta Arith. **95** (2000), 261–288.

[8] F. Binder, *Characterization of polynomial prime bidecompositions: a simplified proof*, in: Contributions to General Algebra, 9, 61–72, Hölder-Pichler-Tempsky, Vienna, 1995.

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

[10] H. T. Engstrom, *Polynomial substitutions*, Amer. J. Math. **63** (1941), 249–255.

[11] W. Feit, *On symmetric balanced incomplete block designs with doubly transitive automorphism groups*, J. Combin. Theory Ser. A **14** (1973), 221–247.

[12] M. D. Fried, *On a conjecture of Schur*, Michigan Math. J. **17** (1970), 41–55.

[13] ———, *On a theorem of Ritt and related Diophantine problems*, J. Reine Angew. Math. **264** (1973), 40–55.

[14] M. D. Fried and R. E. MacRae, *On the invariance of chains of fields*, Illinois J. Math. **13** (1969), 165–171.

[15] J. von zur Gathen, *Functional decomposition of polynomials: the tame case*, J. Symb. Comp. **9** (1990), 281–299.

[16] D. Ghioca, T. J. Tucker and M. E. Zieve, *Intersections of polynomial orbits, and a dynamical Mordell-Lang conjecture*, Invent. Math. **171** (2008), 463–483, arXiv:0705.1954v2.

[17] ———, *Linear relations between polynomial orbits*, submitted for publication, arXiv:0807.3576.

[18] R. M. Guralnick and J. Saxl, *Monodromy groups of polynomials*, in: Groups of Lie Type and their Geometries, 125–150, Cambridge Univ. Press, Cambridge, 1995.

[19] ———, *Exceptional polynomials over arbitrary fields*, in: Algebra, Arithmetic and Geometry with Applications, 457–472, Springer, Berlin, 2004.

[20] R. M. Guralnick and M. E. Zieve, *Polynomials with* PSL(2) *monodromy*, submitted for publication, arXiv:0707.1835.

[21] D. Hilbert, *Über die Irreduzibilität ganzer rationaler Funktionen mit ganzzahligen Koeffizienten*, J. Reine Angew. Math. **110** (1892), 104–129. (Ges. Abh. II, 264–286)

[22] A. Horwitz, *Even compositions of entire functions and related matters*, J. Austral. Math. Soc. Ser. A **63** (1997), 225–237.

[23] A. L. Horwitz and L. A. Rubel, *When is the composition of two power series even?*, J. Austral. Math. Soc. (Series A) **56** (1994), 415–420.

[24] D. Kozen and S. Landau, *Polynomial decomposition algorithms*, J. Symb. Comp. **7** (1989), 445–456.

[25] H. Lausch and W. Nöbauer, Algebra of Polynomials, North-Holland, Amsterdam, 1973.

[26] H. Levi, *Composite polynomials with coefficients in an arbitrary field of characteristic zero*, Amer. J. Math. **64** (1942), 389–400.

[27] R. Lyons and M. E. Zieve, *The rational function analogues of Ritt's polynomial decomposition theorems*, in preparation.

[28] A. Medvedev and T. Scanlon, *Polynomial dynamics*, arXiv:0901.2352.

[29] P. Müller, *Primitive monodromy groups of polynomials*, in: Recent Developments in the Inverse Galois Problem, 385–401, Amer. Math. Soc., Providence, RI, 1995.

[30] F. Pakovich, *On polynomials sharing preimages of compact sets, and related questions*, Geom. Funct. Anal., **18** (2008), 163–183, arXiv:math/0603452.

[31] F. Pakovich, *Prime and composite Laurent polynomials*, arXiv:0710.3860v4.

[32] J. F. Ritt, *Prime and composite polynomials*, Trans. Amer. Math. Soc. **23** (1922), 51–66.
[33] ———, *Permutable rational functions*, Trans. Amer. Math. Soc. **25** (1923), 399–448.
[34] ———, *Equivalent rational substitutions*, Trans. Amer. Math. Soc. **26** (1924), 221–229.
[35] A. Schinzel, Selected Topics on Polynomials, University of Michigan Press, 1982.
[36] ———, Polynomials with Special Regard to Reducibility, Cambridge University Press, 2000.
[37] J.-P. Serre, Topics in Galois Theory, Jones and Bartlett, Boston, 1992.
[38] P. Tortrat, *Sur la composition des polynômes*, *Colloq. Math.*, **55** (1988), 329–353.
[39] U. Zannier, *Ritt's second theorem in arbitrary characteristic*, J. Reine Angew. Math. **445** (1993), 175–203.
[40] ———, *On a functional equation relating a Laurent series $f(x)$ to $f(x^m)$*, Aequat. Math. **55** (1998), 15–43.
[41] M. E. Zieve, *Decompositions of Laurent polynomials*, submitted for publication, arXiv:0710.1902.
[42] R. Zippel, *Rational function decomposition*, in: Proceedings of ISSAC 91, 1–6, ACM Press, New York, 1991.

Department of Mathematics, Rutgers University, Piscataway, NJ 08854, USA
E-mail address: `zieve@math.rutgers.edu`
URL: `www.math.rutgers.edu/~zieve/`

Institut für Mathematik, Universität Würzburg, Am Hubland, D-97074, Würzburg, Germany
E-mail address: `Peter.Mueller@mathematik.uni-wuerzburg.de`
URL: `www.mathematik.uni-wuerzburg.de/~mueller`