

# LINEAR RELATIONS BETWEEN POLYNOMIAL ORBITS

---

DRAGOS GHIOCA, THOMAS J. TUCKER, and MICHAEL E. ZIEVE

## Abstract

We study the orbits of a polynomial  $f \in \mathbb{C}[X]$ , namely, the sets  $\{\alpha, f(\alpha), f(f(\alpha)), \dots\}$  with  $\alpha \in \mathbb{C}$ . We prove that if two nonlinear complex polynomials  $f, g$  have orbits with infinite intersection, then  $f$  and  $g$  have a common iterate. More generally, we describe the intersection of any line in  $\mathbb{C}^d$  with a  $d$ -tuple of orbits of nonlinear polynomials, and we formulate a question which generalizes both this result and the Mordell–Lang conjecture.

## 1. Introduction

One of the main topics in complex dynamics is the behavior of complex numbers  $x$  under repeated application of a polynomial  $f \in \mathbb{C}[X]$ . The basic object of study is the orbit  $\mathcal{O}_f(x) := \{x, f(x), f(f(x)), \dots\}$ . The theme of many results is that there are hidden interactions between different orbits of a polynomial  $f$ : for instance, the crude geometric shape of all orbits is determined by the orbits of critical points (see [8, Section 9]). However, little is known about the interaction between orbits of distinct polynomials. In this paper we determine when two such orbits have infinite intersection.

### THEOREM 1.1

*Pick  $x, y \in \mathbb{C}$  and nonlinear  $f, g \in \mathbb{C}[X]$ . If  $\mathcal{O}_f(x) \cap \mathcal{O}_g(y)$  is infinite, then  $f$  and  $g$  have a common iterate.*

Here the  $n$ th iterate  $f^{(n)}$  of  $f$  is defined as the  $n$ th power of  $f$  under the operation  $a(X) \circ b(X) := a(b(X))$ . We say  $f$  and  $g$  have a common iterate if  $f^{(n)} = g^{(m)}$  for some  $n, m > 0$ . Note that if  $f, g \in \mathbb{C}[X]$  have a common iterate, and  $\mathcal{O}_f(x)$  is infinite, then  $\mathcal{O}_f(x) \cap \mathcal{O}_g(y)$  is infinite whenever it is nonempty. In light of Ritt's

DUKE MATHEMATICAL JOURNAL

Vol. 161, No. 7, © 2012 DOI [10.1215/00127094-1598098](https://doi.org/10.1215/00127094-1598098)

Received 25 January 2011. Revision received 15 September 2011.

2010 *Mathematics Subject Classification*. Primary 37F10; Secondary 11C08, 14G99.

Ghioca's work partially supported by the Natural Sciences and Engineering Research Council of Canada.

Tucker's and Zieve's work partially supported by National Science Foundation grants DMS-080172 and DMS-0903420, respectively.

classification of polynomials with a common iterate (see [29]), the conclusion of our result can be refined as follows.

**COROLLARY 1.2**

*Pick  $x, y \in \mathbb{C}$  and nonlinear  $f, g \in \mathbb{C}[X]$ . If  $\mathcal{O}_f(x) \cap \mathcal{O}_g(y)$  is infinite, then  $f = \ell_1 \circ h^{(m)}$  and  $g = \ell_2 \circ h^{(n)}$  for some  $h \in \mathbb{C}[X]$ , some linear  $\ell_1, \ell_2 \in \mathbb{C}[X]$ , and some positive integers  $m$  and  $n$ .*

This corollary can be sharpened at the cost of a more complicated formulation (see Proposition 3.10). We remark that the nonlinearity hypothesis cannot be removed since, for instance,  $\mathcal{O}_{X+1}(0)$  contains  $\mathcal{O}_{X^2}(2)$ .

In our previous paper [18], we proved Theorem 1.1 in the special case  $\deg(f) = \deg(g)$ . In the present paper we prove Theorem 1.1 by combining the result from [18] with several new ingredients.

We can interpret Theorem 1.1 as describing when the Cartesian product  $\mathcal{O}_f(x) \times \mathcal{O}_g(y)$  has infinite intersection with the diagonal  $\Delta := \{(z, z) : z \in \mathbb{C}\}$ . The conclusion says that this occurs just when there exist positive integers  $n, m$  such that  $\Delta$  is preserved by the map  $(f^{(n)}, g^{(m)}) : \mathbb{C}^2 \rightarrow \mathbb{C}^2$  defined by  $(z_1, z_2) \mapsto (f^{(n)}(z_1), g^{(m)}(z_2))$ . Our next result generalizes this to products of more than two orbits.

**THEOREM 1.3**

*Let  $d$  be a positive integer, let  $x_1, \dots, x_d \in \mathbb{C}$ , let  $L$  be a line in  $\mathbb{C}^d$ , and let  $f_1, \dots, f_d \in \mathbb{C}[X]$  satisfy  $\deg(f_i) > 1$  for  $i = 1, \dots, d$ . If the Cartesian product  $\mathcal{O}_{f_1}(x_1) \times \dots \times \mathcal{O}_{f_d}(x_d)$  has infinite intersection with  $L$ , then there are nonnegative integers  $m_1, \dots, m_d$  such that  $\sum_{i=1}^d m_i > 0$  and*

$$(f_1^{(m_1)}, \dots, f_d^{(m_d)})(L) = L.$$

When Theorem 1.3 applies, we can describe the intersection of  $L$  with the product of orbits. Our description involves the following more general notion of orbits.

*Definition 1.4*

If  $\Omega$  is a set and  $T$  is a set of maps  $\Omega \rightarrow \Omega$ , then for  $\omega \in \Omega$  the orbit of  $\omega$  under  $T$  is  $\mathcal{O}_T(\omega) := \{t(\omega) : t \in T\}$ .

Recall that a *semigroup* is a set with an associative binary relation; in this paper, all semigroups are required to contain an identity element. Thus, for  $f \in \mathbb{C}[X]$  and  $\omega \in \mathbb{C}$ , the orbit  $\mathcal{O}_f(\omega)$  equals  $\mathcal{O}_S(\omega)$ , where  $S$  is the cyclic semigroup  $\langle f \rangle$  generated by the map  $f : \mathbb{C} \rightarrow \mathbb{C}$ ; in general, if  $S = \langle \Phi \rangle$ , then we write  $\mathcal{O}_\Phi(\alpha)$  in place of

$\mathcal{O}_S(\alpha)$ . Theorem 1.3 enables us to describe the intersection of a line and a product of orbits.

**COROLLARY 1.5**

Let  $\alpha \in \mathbb{C}^d$ , let  $f_1, \dots, f_d \in \mathbb{C}[X]$  satisfy  $\deg(f_i) > 1$  for  $i = 1, \dots, d$ , and let  $L$  be a line in  $\mathbb{C}^d$ . Let  $S$  be the semigroup generated by the maps  $\rho_i: \mathbb{C}^d \rightarrow \mathbb{C}^d$  with  $1 \leq i \leq d$ , where  $\rho_i$  acts as the identity on each coordinate of  $\mathbb{C}^d$  except the  $i$ th, on which it acts as  $f_i$ . Then the intersection of  $\mathcal{O}_S(\alpha)$  with  $L$  is  $\mathcal{O}_T(\alpha)$ , where  $T$  is the union of finitely many cosets of cyclic subsemigroups of  $S$ .

It is natural to seek analogues of Corollary 1.5 for other semigroups of endomorphisms of a variety. We propose the following question, in which  $\mathbb{N}_0$  denotes the set of nonnegative integers.

*Question 1.6*

Let  $X$  be a variety defined over  $\mathbb{C}$ , let  $V$  be a closed subvariety of  $X$ , let  $S$  be a finitely generated commutative subsemigroup of  $\text{End } X$ , and let  $\alpha \in X(\mathbb{C})$ . Do the following hold?

- (a) The intersection  $V \cap \mathcal{O}_S(\alpha)$  can be written as  $\mathcal{O}_T(\alpha)$ , where  $T$  is the union of finitely many cosets of subsemigroups of  $S$ .
- (b) For any choice of generators  $\Phi_1, \dots, \Phi_r$  of  $S$ , let  $Z$  be the set of tuples  $(n_1, \dots, n_r) \in \mathbb{N}_0^r$  for which  $\Phi_1^{n_1} \cdots \Phi_r^{n_r}(\alpha)$  lies in  $V$ ; then  $Z$  is the union of finitely many sets of the form  $z_i + (G_i \cap \mathbb{N}_0^r)$ , where each  $G_i$  is a subgroup of  $\mathbb{Z}^r$  and each  $z_i$  is in  $\mathbb{N}_0^r$ .

When  $S$  is cyclic, both parts of this question specialize to an earlier conjecture posed by the first two authors in [17] (which refines an earlier conjecture of Denis [12]). This previous conjecture for cyclic semigroups has been verified in several cases: [2], [3], [5], [6], [12], [17], [18], [24], [25], [33]; for instance, if  $X = \mathbb{G}_a \times \mathbb{G}_m$  and  $S$  is cyclic, then Question 1.6 is equivalent to the classical Skolem–Mahler–Lech theorem on zeros of linear recurrent sequences. However, it is unknown whether Question 1.6 always has an affirmative answer when  $S$  is cyclic. We formulate Question 1.6 to have a single question which simultaneously generalizes three different things: the earlier “cyclic semigroup” conjecture, Corollary 1.5 of the present paper, and the Mordell–Lang conjecture on intersections of subgroups and subvarieties of semiabelian varieties (which was proved by Faltings [14] and Vojta [35]). The Mordell–Lang conjecture immediately implies the special case of Question 1.6 in which  $X$  is semiabelian and  $S$  consists of translations (and conversely, the Mordell–Lang conjecture follows at once from this special case of Question 1.6). Inspired by

this reformulation of the Mordell–Lang conjecture, we showed in [19] that Question 1.6 often has an affirmative answer when  $X$  is semiabelian and  $S$  consists of algebraic group endomorphisms; for instance, this is true if each endomorphism in  $S$  has diagonalizable Jacobian at the identity of  $X$ . Note that Corollary 1.5 is equivalent to part (a) of Question 1.6 in case  $V$  is a line,  $X = \mathbb{A}^d$ , and  $S$  is generated by endomorphisms  $\rho_1, \dots, \rho_d$  as in Corollary 1.5. In this situation, part (a) implies part (b); in fact, we know no example where (a) holds but (b) fails (plainly (b) implies (a)). We do not know any other cases in which Question 1.6 has been resolved. We emphasize that the methods used to resolve Question 1.6 in these various settings are completely different from one another. It would thus be of great interest to prove a common generalization of all these results, thereby connecting the various results as well as the methods involved in their proofs. In particular, the fact that Question 1.6 has an affirmative answer in such disparate settings suggests that it may have an affirmative answer quite often, and perhaps it has an affirmative answer in a single setting which includes all the settings described above.

One crucial distinction between the results of the present paper and most of the results described above is that here we treat ramified endomorphisms. In fact, besides our previous paper [18], all other papers addressing Question 1.6 in the ramified case have used a  $p$ -adic analytic approach (and have treated only cyclic semigroups  $S$ ). However, the recent paper [5] gives a comprehensive treatment of this  $p$ -adic approach, which implies, in particular, that the approach seldom applies to semigroups containing ramified endomorphisms. By contrast, every nonidentity endomorphism occurring in Corollary 1.5 is ramified.

In Section 12 we present several examples in which part (a) of Question 1.6 does not hold; we do not know any general conjecture predicting when it should hold. We also explain how Question 1.6 relates to the existence of positive-dimensional subvarieties of  $V$  that are invariant under nontrivial subsemigroups of  $S$ .

In the case  $S = \langle \Phi \rangle$  is cyclic, Question 1.6 fits into Zhang’s far-reaching system of dynamical conjectures [37]. Zhang’s conjectures include dynamical analogues of the Manin–Mumford and Bogomolov conjectures for abelian varieties (now theorems of Raynaud [27], [28], Ullmo [34], and Zhang [36]), as well as a conjecture on the existence of a Zariski-dense orbit for a large class of endomorphisms  $\Phi$ . Let  $Y$  denote the union of the proper subvarieties of  $X$  which are preperiodic under  $\Phi$ . Then [37, Conjecture 4.1.6] asserts that  $X \neq Y$  if  $X$  is an irreducible projective variety and  $\Phi$  admits a polarization; the conclusion of Question 1.6 implies that  $\mathcal{O}_\Phi(\alpha) \cap V$  is finite whenever  $\alpha \in X(\mathbb{C}) \setminus Y(\mathbb{C})$  and  $V$  is a proper closed subvariety of  $X$ . Results in the direction of Zhang’s conjecture have been proved by Amerik, Bogomolov, and Rovinsky [1, Corollary 2.7]. We remark further that, if Question 1.6 has a positive answer for an irreducible quasi-projective variety  $X$  and a cyclic semigroup  $S = \langle \Phi \rangle$ ,

then any Zariski-dense orbit  $\mathcal{O}_S(\alpha)$  is *critically dense*, in the sense that each infinite subset of  $\mathcal{O}_S(\alpha)$  is Zariski-dense in  $X$ ; this concept arose in [21, Definition 3.6] and [11, Section 5] (see also [3, Corollary 5.5]).

In our previous paper [18], we proved Theorem 1.1 in the case  $\deg(f) = \deg(g)$ . The proof went as follows. First we used a specialization argument to show it suffices to prove the result when  $f, g, x, y$  are all defined over a number field  $K$ . Then in fact they are defined over some ring  $A$  of  $S$ -integers of  $K$ , where  $S$  is a finite set of primes; this implies that  $\mathcal{O}_f(x)$  and  $\mathcal{O}_g(y)$  lie in  $A$ . Thus, for each  $n$ , the equation  $f^{(n)}(X) = g^{(n)}(Y)$  has infinitely many solutions in  $A \times A$ , so by Siegel’s theorem the polynomial  $f^{(n)}(X) - g^{(n)}(Y)$  has an absolutely irreducible factor in  $K[X, Y]$  which has genus zero and has at most two points at infinity. A result of Bilu and Tichy [7] describes the polynomials  $F, G \in K[X]$  for which  $F(X) - G(Y)$  has such a factor. This gives constraints on the shape of  $f^{(n)}$  and  $g^{(n)}$ ; by combining the information obtained from different values of  $n$  and using elementary results about polynomial decomposition, we deduced that either  $f$  and  $g$  have a common iterate or there is a linear  $\ell \in K[X]$  such that  $(\ell \circ f \circ \ell^{(-1)}, \ell \circ g \circ \ell^{(-1)}) = (\alpha X^r, \beta X^r)$ . Finally, we proved the result directly for this last type of polynomials  $f, g$ .

We use two approaches to prove versions of Theorem 1.1 in the case  $\deg(f) \neq \deg(g)$ , both of which rely on the fact that the result is known when  $\deg(f) = \deg(g)$ . Our first approach utilizes canonical heights to reduce the problem to the case  $\deg(f) = \deg(g)$  treated in [18]; this approach does not work when  $f, g, x, y$  are defined over a number field, but works in essentially every other situation (cf. Theorem 8.1). As remarked by a referee, this reduction to the case  $\deg(f) = \deg(g)$  can be proved by similar methods for some other classes of rational maps, and can likely be proved for some classes of maps between higher-dimensional varieties.

Our second approach uses delicate results about polynomial decomposition to obtain the full Theorem 1.1. In this proof we do not use the full strength of the result from [18]; instead, we just use the main polynomial decomposition result from that paper. In particular, our proof of Theorem 1.1 does not depend on the complicated specialization argument used in [18]. We now describe the second approach in more detail.

Our proof of Theorem 1.1 uses a similar strategy to that in [18], but here the polynomial decomposition work is much more difficult. The main reason for this is that, when analyzing functional equations involving  $f^{(n)}$  and  $g^{(n)}$  in the case  $\deg(f) = \deg(g)$ , we could use the fact that if  $A, B, C, D \in \mathbb{C}[X] \setminus \mathbb{C}$  satisfy  $A \circ B = C \circ D$  and  $\deg(A) = \deg(C)$ , then  $C = A \circ \ell$  and  $D = \ell^{(-1)} \circ B$  for some linear  $\ell \in \mathbb{C}[X]$ . When  $f$  and  $g$  have distinct degrees, one must use a different approach. Our proof relies on the full strength of the new description given in [38] for the collection of all decompositions of a polynomial; in addition, we use several new types of polynomial

decomposition arguments in the present paper. As above, for every  $m, n$  we find that  $f^{(n)}(X) - g^{(m)}(Y)$  has a genus-zero factor with at most two points at infinity. We show that this implies that either  $f$  and  $g$  have a common iterate or there is a linear  $\ell \in \mathbb{C}[X]$  such that  $(\ell \circ f \circ \ell^{(-1)}, \ell \circ g \circ \ell^{(-1)})$  is either  $(\alpha X^r, \beta X^s)$  or  $(\pm T_r, \pm T_s)$ , where  $T_r$  is the degree- $r$  Chebyshev polynomial of the first kind. We then use a consequence of Siegel's theorem to handle these last possibilities.

The contents of this paper are as follows. In the next section we state the results of Siegel and Bilu and Tichy, and we deduce some consequences. In Section 3 we present the results about polynomial decomposition used in this paper. In the following two sections we prove that if  $f, g \in \mathbb{C}[X]$  with  $\deg(f), \deg(g) > 1$  are such that, for every  $n, m > 0$ ,  $f^{(n)}(X) - g^{(m)}(Y)$  has a genus-zero factor with at most two points at infinity, then either  $f$  and  $g$  have a common iterate or some linear  $\ell \in \mathbb{C}[X]$  makes  $(\ell \circ f \circ \ell^{(-1)}, \ell \circ g \circ \ell^{(-1)})$  have the form  $(\alpha X^r, \beta X^s)$  or  $(\pm T_r, \pm T_s)$ . Then in Section 6 we conclude the proof of Theorem 1.1, and in Section 7 we prove Theorem 1.3 and Corollary 1.5. In the next several sections we give an alternate proof of Theorem 1.1 in case  $x, y, f, g$  cannot be defined over a number field; this proof uses canonical heights to reduce the problem to the case  $\deg(f) = \deg(g)$  treated in our previous paper, and does not rely on any difficult polynomial decomposition arguments. In the final section we discuss related problems.

### Notation

Throughout this paper,  $f^{(n)}$  denotes the  $n$ th iterate of the polynomial  $f$ , with the convention  $f^{(0)} = X$ . When  $f$  has degree 1, we denote the functional inverse of  $f$  by  $f^{(-1)}$ ; this is again a linear polynomial. By  $T_n$  we mean the (normalized) degree- $n$  Chebyshev polynomial of the first kind, which is defined by the equation  $T_n(X + X^{-1}) = X^n + X^{-n}$ ; the classical Chebyshev polynomial  $C_n$  defined by  $C_n(\cos \theta) = \cos n\theta$  satisfies  $2C_n(X/2) = T_n(X)$ . We write  $\mathbb{N}$  for the set of positive integers,  $\mathbb{N}_0$  for the set of nonnegative integers, and  $\mathbb{Z}_{>1}$  for the set of integers greater than 1. We write  $\overline{K}$  for an algebraic closure of the field  $K$ . We say that  $\Phi(X, Y) \in K[X, Y]$  is absolutely irreducible if it is irreducible in  $\overline{K}[X, Y]$ . In this case we let  $C$  be the completion of the normalization of the curve  $\Phi(X, Y) = 0$  and define the genus of  $\Phi(X, Y)$  to be the (geometric) genus of  $C$ . Likewise we define the points at infinity on  $\Phi(X, Y)$  to be the points in  $C(\overline{K})$  which correspond to places of  $\overline{K}(C)$  extending the infinite place of  $\overline{K}(X)$ . In this paper, all subvarieties are closed.

## 2. Integral points on curves

The seminal result on curves with infinitely many integral points is the theorem proved in 1929 by Siegel [32]; we use the following generalization due to Lang [23, Theorems 8.2.4 and 8.5.1].

**THEOREM 2.1**

Let  $K$  be a finitely generated field of characteristic zero, and let  $R$  be a finitely generated subring of  $K$ . Let  $C$  be a smooth, projective, geometrically irreducible curve over  $K$ , and let  $\phi$  be a nonconstant function in  $K(C)$ . Suppose there are infinitely many points  $P \in C(K)$  which are not poles of  $\phi$  and which satisfy  $\phi(P) \in R$ . Then  $C$  has genus zero and  $\phi$  has at most two distinct poles.

We use this result in two ways. One is in the form of the following consequence due to Lang [22].

**COROLLARY 2.2**

Let  $a, b \in \mathbb{C}^*$ , and let  $\Gamma$  be a finitely generated subgroup of  $\mathbb{C}^* \times \mathbb{C}^*$ . Then the equation  $ax + by = 1$  has at most finitely many solutions  $(x, y) \in \Gamma$ .

This result is proved by applying Theorem 2.1 to the genus-1 curves  $a\alpha X^3 + b\beta Y^3 = 1$ , where  $(\alpha, \beta)$  runs through a finite subset of  $\Gamma$  which surjects onto  $\Gamma/\Gamma^3$ .

To describe the other way we apply Theorem 2.1, we introduce the following terminology.

*Definition 2.3*

A Siegel polynomial over a field  $K$  is an absolutely irreducible polynomial  $\Phi(X, Y) \in K[X, Y]$  for which the curve  $\Phi(X, Y) = 0$  has genus zero and has at most two points at infinity. A Siegel factor of a polynomial  $\Psi(X, Y) \in K[X, Y]$  is a factor of  $\Psi$  which is a Siegel polynomial over  $K$ .

*Remark*

What we call Siegel polynomials were called exceptional polynomials in [7]; we use new terminology because the term exceptional polynomial has been used with a different meaning in several papers (see, e.g., [20]).

*Remark*

Clearly a Siegel polynomial over  $K$  maintains the Siegel property over  $\overline{K}$ . Further, an irreducible  $\Phi \in \overline{K}[X, Y]$  is a Siegel polynomial if and only if  $\Phi(\phi, \psi) = 0$  for some Laurent polynomials  $\phi, \psi \in \overline{K}(Z)$  which are not both constant (recall that the Laurent polynomials in  $\overline{K}(Z)$  are the elements of the form  $F/Z^n$  with  $F \in \overline{K}[Z]$  and  $n \in \mathbb{N}_0$ ). We do not know a reference for this fact, so we sketch the proof. If  $\Phi$  is a Siegel polynomial, then the function field of the curve  $\Phi(X, Y) = 0$  (over  $\overline{K}$ ) has the form  $\overline{K}(Z)$ , so  $X = \phi(Z)$  and  $Y = \psi(Z)$  for some  $\phi, \psi \in \overline{K}(Z)$ ; then  $\Phi(\phi, \psi) = 0$  and  $\phi, \psi$  are not both constant. Since  $\Phi(X, Y) = 0$  has at most two points at infinity

ity, at most two points of  $\overline{K} \cup \{\infty\}$  are poles of either  $\phi$  or  $\psi$ . By making a suitable linear fractional change to  $Z$ , we may assume that  $\phi$  and  $\psi$  have no poles except possibly 0 and  $\infty$ , which implies  $\phi$  and  $\psi$  are Laurent polynomials. Conversely, suppose  $\Phi(\phi, \psi) = 0$  for some Laurent polynomials  $\phi, \psi \in \overline{K}(Z)$  which are not both constant. Then the function field of  $\Phi(X, Y) = 0$  is a subfield  $F$  of  $\overline{K}(Z)$ , and each infinite place of  $F$  lies under either  $Z = 0$  or  $Z = \infty$ ; so indeed  $F$  has genus zero with at most two points at infinity.

#### COROLLARY 2.4

*Let  $R$  be a finitely generated integral domain of characteristic zero, let  $K$  be the field of fractions of  $R$ , and pick  $\Phi(X, Y) \in K[X, Y]$ . Suppose there are infinitely many pairs  $(x, y) \in R \times R$  for which  $\Phi(x, y) = 0$ . Then  $\Phi(X, Y)$  has a Siegel factor over  $K$ .*

Although this result is well known to arithmetic geometers, we sketch a proof for the reader's convenience.

#### *Proof*

The hypotheses imply that  $\Phi(X, Y)$  has an irreducible factor  $\Psi(X, Y)$  in  $\overline{K}[X, Y]$  which has infinitely many roots in  $R \times R$ . By replacing  $\Psi$  by a scalar multiple, we may assume that some coefficient of  $\Psi$  equals 1. Since any  $\sigma \in \text{Gal}(\overline{K}/K)$  fixes  $\Phi$ , the polynomial  $\Psi^\sigma$  is an absolutely irreducible factor of  $\Phi$ . Moreover, every root of  $\Psi$  in  $R \times R$  is also a root of  $\Psi^\sigma$ ; since there are infinitely many such roots, it follows (e.g., by Bezout's theorem) that  $\Psi^\sigma$  is a scalar multiple of  $\Psi$ . But since  $\Psi$  has a coefficient equal to 1, the corresponding coefficient of  $\Psi^\sigma$  is also 1, so  $\Psi^\sigma = \Psi$ . Thus  $\Psi$  is fixed by  $\text{Gal}(\overline{K}/K)$ , so  $\Psi \in K[X, Y]$ , whence  $\Psi$  is the desired Siegel factor.  $\square$

In light of Siegel's theorem, there has been intensive study of polynomials  $\Phi(X, Y)$  having a Siegel factor. As noted above, a nonzero polynomial  $\Phi \in \overline{K}[X, Y]$  has a Siegel factor if and only if  $\Phi(\phi, \psi) = 0$  for some Laurent polynomials  $\phi, \psi \in \overline{K}(X)$  which are not both constant. Especially strong results have been obtained in the case  $\Phi(X, Y) = F(X) - G(Y)$  with  $F, G \in K[X]$ ; in this case the problem amounts to solving the functional equation  $F \circ \phi = G \circ \psi$  in polynomials  $F, G \in K[X]$  and Laurent polynomials  $\phi, \psi \in \overline{K}[X]$ . Using Ritt's classical results on such functional equations, together with subsequent results of Fried [15] and Schinzel [31, Theorem 8], Bilu and Tichy [7, Theorem 9.3] proved the following definitive result in this case.

#### THEOREM 2.5

*Let  $K$  be a field of characteristic zero, and pick  $F, G \in K[X]$  for which  $F(X) - G(Y)$*

has a Siegel factor in  $K[X, Y]$ . Then  $F = E \circ F_1 \circ \mu$  and  $G = E \circ G_1 \circ \nu$ , where  $E, \mu, \nu \in K[X]$  with  $\deg(\mu) = \deg(\nu) = 1$ , and either  $(F_1, G_1)$  or  $(G_1, F_1)$  is one of the following pairs (in which  $m, n \in \mathbb{N}$ ,  $a, b \in K^*$ , and  $p \in K[X] \setminus \{0\}$ ):

- (2.5.1)  $(X^m, aX^r p(X)^m)$  with  $r$  a nonnegative integer coprime to  $m$ ;
- (2.5.2)  $(X^2, (aX^2 + b)p(X)^2)$ ;
- (2.5.3)  $(D_m(X, a^n), D_n(X, a^m))$  with  $\gcd(m, n) = 1$ ;
- (2.5.4)  $(a^{-m/2}D_m(X, a), -b^{-n/2}D_n(X, b))$  with  $\gcd(m, n) = 2$ ;
- (2.5.5)  $((aX^2 - 1)^3, 3X^4 - 4X^3)$ ;
- (2.5.6)  $(D_m(X, a^{n/d}), -D_n(X \cos(\pi/d), a^{m/d}))$  where  $d = \gcd(m, n) \geq 3$  and  $\cos(2\pi/d) \in K$ .

Here  $D_n(X, Y)$  is the unique polynomial in  $\mathbb{Z}[X, Y]$  such that  $D_n(U + V, UV) = U^n + V^n$ . Note that, for  $\alpha \in K$ , the polynomial  $D_n(X, \alpha) \in K[X]$  is monic of degree  $n$ . The defining equation implies that  $D_n(X, 0) = X^n$  and  $\alpha^n D_n(X, 1) = D_n(\alpha X, \alpha^2)$  for  $\alpha \in \mathbb{C}^*$ . Since  $T_n(Y + Y^{-1}) = Y^n + Y^{-n}$ , we have

$$D_n(\alpha X, \alpha^2) = \alpha^n T_n(X) \quad \text{for any } n \in \mathbb{N} \text{ and } \alpha \in \mathbb{C}^*. \tag{2.6}$$

For our application to orbits of complex polynomials, we only need the case  $K = \mathbb{C}$  of Theorem 2.5. We now state a simpler version of the result in this case.

**COROLLARY 2.7**

For nonconstant  $F, G \in \mathbb{C}[X]$ , if  $F(X) - G(Y)$  has a Siegel factor in  $\mathbb{C}[X, Y]$ , then  $F = E \circ F_1 \circ \mu$  and  $G = E \circ G_1 \circ \nu$ , where  $E, \mu, \nu \in \mathbb{C}[X]$  with  $\deg(\mu) = \deg(\nu) = 1$ , and either  $(F_1, G_1)$  or  $(G_1, F_1)$  is one of the following pairs (in which  $m, n \in \mathbb{N}$  and  $p \in \mathbb{C}[X] \setminus \{0\}$ ):

- (2.7.1)  $(X^m, X^r p(X)^m)$  where  $r \in \mathbb{N}_0$  is coprime to  $m$ ;
- (2.7.2)  $(X^2, (X^2 + 1)p(X)^2)$ ;
- (2.7.3)  $(T_m, T_n)$  with  $\gcd(m, n) = 1$ ;
- (2.7.4)  $(T_m, -T_n)$  with  $\gcd(m, n) > 1$ ;
- (2.7.5)  $((X^2 - 1)^3, 3X^4 - 4X^3)$ .

*Proof*

Let  $E, F_1, G_1, \mu, \nu$  satisfy the conclusion of Theorem 2.5. In light of (2.6), if a pair  $(f, g)$  has the form of one of (2.5.1)–(2.5.6), then there are linear  $\ell_i \in \mathbb{C}[X]$  for which  $(\ell_1 \circ f \circ \ell_2, \ell_1 \circ g \circ \ell_3)$  has the form of one of (2.7.1)–(2.7.5). This implies that  $(F, G)$  has the desired form, since we can replace  $E$  by  $E \circ \ell_1$  and replace  $(\mu, \nu)$  by either  $(\ell_2 \circ \mu, \ell_3 \circ \nu)$  or  $(\ell_3 \circ \mu, \ell_2 \circ \nu)$ . □

*Remark*

The converse of Corollary 2.7 is also true; since it is not needed for the present paper,

we only sketch the proof. It suffices to show that, for each pair  $(f, g)$  satisfying one of (2.7.1)–(2.7.5), we have  $f \circ \phi = g \circ \psi$  for some Laurent polynomials  $\phi, \psi \in \mathbb{C}(X)$  which are not both constant. For this, observe that

$$\begin{aligned} X^m \circ X^r p(X^m) &= X^r p(X)^m \circ X^m; \\ X^2 \circ (X + (4X)^{-1})p(X - (4X)^{-1}) &= (X^2 + 1)p(X)^2 \circ (X - (4X)^{-1}); \\ T_m \circ T_n &= T_n \circ T_m; \\ T_m \circ (X^n + X^{-n}) &= -T_n \circ ((\zeta X)^m + (\zeta X)^{-m}), \quad \text{where } \zeta^{mn} = -1; \end{aligned}$$

and

$$\begin{aligned} (X^2 - 1)^3 \circ \frac{X^2 + 2X + X^{-1} - (2X)^{-2}}{\sqrt{3}} \\ = (3X^4 - 4X^3) \circ \frac{(X + 1 - (2X)^{-1})^3 + 4}{3}. \end{aligned}$$

*Remark*

Our statement of Theorem 2.5 differs slightly from [7, Theorem 9.3], since there is a mistake in the definition of specific pairs in [7]. (The terms  $a^{m/d}$  and  $a^{n/d}$  should be interchanged.) The proof of [7, Theorem 9.3] contains some minor errors related to this point, but they are easy to correct. Also, although the sentence in [7] following the definition of specific pairs is false for odd  $n$  (because implication (9) is false for odd  $n$ ), neither this nor (9) is used in the paper [7].

**3. Polynomial decomposition**

Our proof relies on several results about decompositions of polynomials. Especially, we make crucial use of the following result proved in [38, Theorem 1.4].

**THEOREM 3.1**

*Pick  $f \in \mathbb{C}[X]$  with  $\deg(f) = n > 1$ , and suppose there is no linear  $\ell \in \mathbb{C}[X]$  such that  $\ell \circ f \circ \ell^{(-1)}$  is either  $X^n$  or  $T_n$  or  $-T_n$ . Let  $r, s \in \mathbb{C}[X]$  and  $d \in \mathbb{N}$  satisfy  $r \circ s = f^{(d)}$ . Then we can write*

$$r = f^{(i)} \circ \tilde{r}$$

and

$$s = \tilde{s} \circ f^{(j)},$$

where  $\tilde{r}, \tilde{s} \in \mathbb{C}[X]$  and  $i, j, k \in \mathbb{N}_0$  satisfy  $\tilde{r} \circ \tilde{s} = f^{(k)}$  and  $k \leq \log_2(n + 2)$ .

The proof of this result relies on the full strength of the new description given in [38] for the collection of all decompositions of a polynomial; this in turn depends on the classical results of Ritt [30] among other things. By contrast, all the other polynomial decomposition results we need can be proved fairly quickly from first principles.

The next result follows from results of Engstrom [13]. (For a proof using methods akin to Ritt’s [30], see [38, Corollary 2.9].)

LEMMA 3.2

*Pick  $a, b, \tilde{a}, \tilde{b} \in \mathbb{C}[X] \setminus \mathbb{C}$  with  $a \circ b = \tilde{a} \circ \tilde{b}$ . If  $\deg(\tilde{a}) \mid \deg(a)$ , then  $a = \tilde{a} \circ t$  for some  $t \in \mathbb{C}[X]$ . If  $\deg(\tilde{b}) \mid \deg(b)$ , then  $b = t \circ \tilde{b}$  for some  $t \in \mathbb{C}[X]$ .*

We often use the above two results in conjunction with one another.

COROLLARY 3.3

*Pick  $f \in \mathbb{C}[X]$  with  $\deg(f) = n > 1$ , and assume there is no linear  $\ell \in \mathbb{C}[X]$  such that  $\ell \circ f \circ \ell^{(-1)}$  is either  $X^n$  or  $T_n$  or  $-T_n$ . Then there is a finite subset  $S$  of  $\mathbb{C}[X]$  such that, if  $r, s \in \mathbb{C}[X]$  and  $d \in \mathbb{N}$  satisfy  $r \circ s = f^{(d)}$ , then*

- *either  $r = f \circ t$  (with  $t \in \mathbb{C}[X]$ ) or  $r \circ \ell \in S$  (with  $\ell \in \mathbb{C}[X]$  linear);*
- *either  $s = t \circ f$  (with  $t \in \mathbb{C}[X]$ ) or  $\ell \circ s \in S$  (with  $\ell \in \mathbb{C}[X]$  linear).*

As an immediate consequence of the functional equation defining  $T_n$ , we see that  $T_n$  is either an even or odd polynomial.

LEMMA 3.4

*For any  $n \in \mathbb{N}$ , we have  $T_n(-X) = (-1)^n T_n(X)$ .*

Note that  $X^d \circ X^e = X^{de}$  and  $T_d \circ T_e = T_{de}$ . By Lemma 3.2, these are essentially the only decompositions of  $X^n$  and  $T_n$ .

LEMMA 3.5

*If  $n \in \mathbb{N}$  and  $f, g \in \mathbb{C}[X]$  satisfy  $f \circ g = X^n$ , then  $f = X^d \circ \ell$  and  $g = \ell^{(-1)} \circ X^{n/d}$  for some linear  $\ell \in \mathbb{C}[X]$  and some positive divisor  $d$  of  $n$ . If  $n \in \mathbb{N}$  and  $f, g \in \mathbb{C}[X]$  satisfy  $f \circ g = T_n$ , then  $f = T_d \circ \ell$  and  $g = \ell^{(-1)} \circ T^{n/d}$  for some linear  $\ell \in \mathbb{C}[X]$  and some positive divisor  $d$  of  $n$ .*

The following simple result describes the linear relations between polynomials of the form  $X^n$  or  $T_n$  (see [38, Lemmas 3.13 and 3.14]).

LEMMA 3.6

Pick  $n \in \mathbb{N}$  and linear  $a, b \in \mathbb{C}[X]$ .

(3.6.1) If  $n > 1$  and  $a \circ X^n \circ b = X^n$ , then  $b = \beta X$  and  $a = X/\beta^n$  for some  $\beta \in \mathbb{C}^*$ .

(3.6.2) If  $n > 2$ , then  $a \circ X^n \circ b \neq T_n$ .

(3.6.3) If  $n > 2$  and  $a \circ T_n \circ b = T_n$ , then  $b = \epsilon X$  and  $a = \epsilon^n X$  for some  $\epsilon \in \{1, -1\}$ .

The previous two results have the following consequence (see [38, Corollary 3.10]).

LEMMA 3.7

Pick  $r, s \in \mathbb{Z}$  and linear  $\ell, \ell_1, \ell_2 \in \mathbb{C}[X]$ . If  $r, s > 1$  and  $X^r \circ \ell \circ X^s = \ell_1 \circ X^{rs} \circ \ell_2$ , then  $\ell = \alpha X$  for some  $\alpha \in \mathbb{C}^*$ . If  $r, s > 2$  and  $T_r \circ \ell \circ T_s = \ell_1 \circ T_{rs} \circ \ell_2$ , then  $\ell = \epsilon X$  for some  $\epsilon \in \{1, -1\}$ .

We also need to know the possible decompositions of polynomials of the form  $X^i h(X)^n$  (see [38, Lemma 3.11]).

LEMMA 3.8

If  $a \circ b = X^i h(X)^n$  with  $h \in \mathbb{C}[X] \setminus \{0\}$  and coprime  $i, n \in \mathbb{N}$ , then  $a = X^j \hat{h}(X)^n \circ \ell$  and  $b = \ell^{(-1)} \circ X^k \tilde{h}(X)^n$  for some  $j, k \in \mathbb{N}$  and some  $\hat{h}, \tilde{h}, \ell \in \mathbb{C}[X]$  with  $\ell$  linear.

The following result presents situations where the shape of a polynomial is determined by the shape of one of its iterates.

LEMMA 3.9

Pick  $f, \ell, \hat{\ell} \in \mathbb{C}[X]$  with  $r := \deg(f) > 1$  and  $\ell, \hat{\ell}$  linear, and pick  $n \in \mathbb{Z}_{>1}$ .

(3.9.1) If  $f^{(n)} = \ell \circ X^{rn} \circ \hat{\ell}$ , then  $f = \ell \circ \alpha X^r \circ \ell^{(-1)}$  for some  $\alpha \in \mathbb{C}^*$ .

(3.9.2) If  $f^{(n)} = \ell \circ T_{r^n} \circ \hat{\ell}$  and  $\{r, n\} \neq \{2\}$ , then  $f = \ell \circ T_r \circ \epsilon \ell^{(-1)}$  for some  $\epsilon \in \{1, -1\}$ .

*Proof*

If  $f^{(n)} = \ell \circ X^{rn} \circ \hat{\ell}$ , then  $f = \ell \circ X^r \circ \bar{\ell}$  for some linear  $\bar{\ell}$  (by Lemma 3.2). Likewise  $f^{(2)} = \ell \circ X^{r^2} \circ \bar{\ell}$ , so Lemma 3.7 implies that  $\bar{\ell} \circ \ell = \beta X$  for some  $\beta \in \mathbb{C}^*$ . Hence  $f = \ell \circ X^r \circ \beta \ell^{(-1)}$ .

Henceforth suppose  $f^{(n)} = \ell \circ T_{r^n} \circ \hat{\ell}$  and  $n > 1$ . As above,  $f = \ell \circ T_r \circ \bar{\ell}$  and  $f^{(2)} = \ell \circ T_{r^2} \circ \bar{\ell}$ , so if  $r > 2$  then Lemma 3.7 implies  $\bar{\ell} \circ \ell = \epsilon X$  for some  $\epsilon \in \{1, -1\}$ , whence  $f = \ell \circ T_r \circ \epsilon \ell^{(-1)}$ .

Now assume that  $r = 2$  and  $n > 2$ . Then  $f = \ell \circ T_2 \circ \bar{\ell}$  and  $f^{(3)} = \ell \circ T_8 \circ \bar{\ell}$ . Writing  $\ell^{(-1)} \circ f^{(3)} = (T_2 \circ \bar{\ell} \circ \ell) \circ (T_2 \circ \bar{\ell} \circ \ell) \circ (T_2 \circ \bar{\ell}) = T_2 \circ T_2 \circ (T_2 \circ \bar{\ell})$ ,

Lemma 3.2 implies there are linear  $\mu, \lambda \in \mathbb{C}[X]$  such that  $T_2 \circ \bar{\ell} = \lambda^{(-1)} \circ T_2 \circ \tilde{\ell}$  and  $T_2 \circ \bar{\ell} \circ \ell = \mu^{(-1)} \circ T_2 \circ \lambda$  and  $T_2 \circ \bar{\ell} \circ \ell = T_2 \circ \mu$ . Since  $T_2 = (X - 2) \circ X^2$ , by Lemma 3.6 the equality  $T_2 \circ \mu = \mu^{(-1)} \circ T_2 \circ \lambda$  implies that  $\mu \circ \lambda^{(-1)} = \beta X$  and  $\mu = -2 + (X + 2)/\beta^2$  for some  $\beta \in \mathbb{C}^*$ . Likewise, from  $\lambda \circ T_2 \circ \bar{\ell} \circ \tilde{\ell}^{(-1)} = T_2$  we get  $\lambda = -2 + (X + 2)/\alpha^2$  for some  $\alpha \in \mathbb{C}^*$ ; but also  $\lambda = \beta^{-1}\mu$ , so since  $\lambda$  and  $\mu$  fix  $-2$ , it follows that  $\beta = 1$ . Thus  $\mu = X$ , so we have  $T_2 \circ \bar{\ell} \circ \ell = T_2$  and thus  $\bar{\ell} \circ \ell = \epsilon X$  with  $\epsilon \in \{1, -1\}$ , and the result follows.  $\square$

*Remark*

The hypothesis  $\{r, n\} \neq \{2\}$  is needed in (3.9.2): for any linear  $\ell(x) \in \mathbb{C}[X]$  and any  $\alpha \in \mathbb{C}^* \setminus \{1, -1\}$ , the polynomial  $f = \ell \circ T_2 \circ (-2 + \alpha^2(X + 2)) \circ \ell^{(-1)}$  satisfies  $f^{(2)} = \ell \circ T_4 \circ (-2\alpha + \alpha^3(X + 2)) \circ \ell^{(-1)}$ , but  $f \neq \ell \circ T_2 \circ \pm \ell^{(-1)}$ .

Although it is not used in this paper, for the reader’s convenience we recall Ritt’s [29, p. 356] description of polynomials with a common iterate.

PROPOSITION 3.10 (Ritt)

Let  $f_1, f_2 \in \mathbb{C}[X]$  with  $d_i := \deg(f_i) > 1$  for each  $i \in \{1, 2\}$ . For  $m_1, m_2 \in \mathbb{N}$ , we have  $f_1^{(m_1)} = f_2^{(m_2)}$  if and only if  $f_1(X) = -\beta + \epsilon_1 g^{(n_1)}(X + \beta)$  and  $f_2(X) = -\beta + \epsilon_2 g^{(n_2)}(X + \beta)$  for some  $n_1, n_2 \in \mathbb{N}$  with  $n_1 m_1 = n_2 m_2$ , some  $g \in X^r \mathbb{C}[X^s]$  (with  $r, s \in \mathbb{N}_0$ ), and some  $\epsilon_1, \epsilon_2, \beta \in \mathbb{C}$  with  $\epsilon_i^s = 1$  and  $\epsilon_i^{(d_i^{m_i} - 1)/(d_i - 1)} = 1$  for each  $i \in \{1, 2\}$ .

**4. Commensurable polynomials**

In this section we analyze  $f, g \in \mathbb{C}[X]$  which are *commensurable*, in the sense that for every  $m \in \mathbb{N}$  there exist  $n \in \mathbb{N}$  and  $h_1, h_2 \in \mathbb{C}[X]$  such that  $f^{(n)} = g^{(m)} \circ h_1$  and  $g^{(n)} = f^{(m)} \circ h_2$ . Plainly two polynomials with a common iterate are commensurable; we give an explicit description of all other pairs of commensurable polynomials. In fact, we need only assume half of the commensurability hypothesis.

PROPOSITION 4.1

Pick  $f, g \in \mathbb{C}[X]$  for which  $r := \deg(f)$  and  $s := \deg(g)$  satisfy  $r, s > 1$ . Suppose that, for every  $m \in \mathbb{N}$ , there exists  $n \in \mathbb{N}$  and  $h \in \mathbb{C}[X]$  such that  $g^{(n)} = f^{(m)} \circ h$ . Then either  $f$  and  $g$  have a common iterate, or there is a linear  $\ell \in \mathbb{C}[X]$  such that  $(\ell \circ f \circ \ell^{(-1)}, \ell \circ g \circ \ell^{(-1)})$  is either  $(\alpha X^r, X^s)$  (with  $\alpha \in \mathbb{C}^*$ ) or  $(T_r \circ \hat{\epsilon} X, T_s \circ \epsilon X)$  (with  $\hat{\epsilon}, \epsilon \in \{1, -1\}$ ).

*Remark*

The converse of Proposition 4.1 holds if and only if every prime factor of  $r$  is also a factor of  $s$ .

Our proof of Proposition 4.1 consists of a reduction to the case  $r = s$ . The case  $r = s$  of Proposition 4.1 was analyzed in our previous paper [18], as one of the main ingredients in our proof of Theorem 1.1 in the case  $\deg(f) = \deg(g)$ . The following result is [18, Proposition 3.3].

PROPOSITION 4.2

Let  $F, G \in \mathbb{C}[X]$  satisfy  $\deg(F) = \deg(G) = r > 1$ . Suppose that, for every  $m \in \mathbb{N}$ , there is a linear  $\ell_m \in \mathbb{C}[X]$  such that  $G^{(m)} = F^{(m)} \circ \ell_m$ . Then either  $F$  and  $G$  have a common iterate, or there is a linear  $\ell \in \mathbb{C}[X]$  for which  $\ell \circ F \circ \ell^{(-1)} = \alpha X^r$  and  $\ell \circ G \circ \ell^{(-1)} = \beta X^r$  with  $\alpha, \beta \in \mathbb{C}^*$ .

By Lemma 3.2, this implies the case  $r = s$  of Proposition 4.1. Note that Chebyshev polynomials are given special mention in the conclusion of Proposition 4.1, but not in the conclusion of Proposition 4.2; this is because  $T_r(X)$  and  $T_r(-X)$  have the same second iterate.

*Proof of Proposition 4.1*

First assume that  $\ell \circ g \circ \ell^{(-1)} = X^s$  for some linear  $\ell \in \mathbb{C}[X]$ . Then  $g^{(n)} = f^{(2)} \circ h$  becomes  $\ell^{(-1)} \circ X^{s^n} \circ \ell = f^{(2)} \circ h$ , so Lemma 3.5 implies  $f^{(2)} = \ell^{(-1)} \circ X^{r^2} \circ \hat{\ell}$  for some linear  $\hat{\ell} \in \mathbb{C}[X]$ . Now Lemma 3.9 implies  $f = \ell^{(-1)} \circ \alpha X^r \circ \ell$  for some  $\alpha \in \mathbb{C}^*$ , so the result holds in this case.

Next assume that  $\ell \circ g \circ \ell^{(-1)} = T_s \circ \epsilon X$  for some linear  $\ell \in \mathbb{C}[X]$  and some  $\epsilon \in \{1, -1\}$ . Then we can use the fact that  $T_s(-X) = (-1)^s T_s(X)$  to rewrite  $g^{(n)} = f^{(3)} \circ h$  as  $\ell^{(-1)} \circ T_{s^n} \circ \epsilon^n \ell = f^{(3)} \circ h$ . As above, Lemma 3.5 implies that  $f^{(3)} = \ell^{(-1)} \circ T_{r^3} \circ \hat{\ell}$  for some linear  $\hat{\ell} \in \mathbb{C}[X]$ . Then Lemma 3.9 implies  $f = \ell^{(-1)} \circ T_r \circ \hat{\epsilon} \ell$  with  $\hat{\epsilon} \in \{1, -1\}$ , so the result holds in this case.

Henceforth assume that there is no linear  $\ell \in \mathbb{C}[X]$  for which  $\ell \circ g \circ \ell^{(-1)}$  is either  $X^s$  or  $T_s$  or  $T_s(-X)$ . For  $m \in \mathbb{N}$ , let  $n \in \mathbb{N}$  be minimal for which  $g^{(n)} = f^{(m)} \circ h$  with  $h \in \mathbb{C}[X]$ , and let  $h_m \in \mathbb{C}[X]$  satisfy  $g^{(n)} = f^{(m)} \circ h_m$ . Minimality of  $n$  implies there is no  $t \in \mathbb{C}[X]$  with  $h_m = t \circ g$ , so by Corollary 3.3 there is a bound on  $\deg(h_m)$  depending only on  $g$ . In particular, this implies that there are distinct  $m, M \in \mathbb{N}$  for which  $\deg(h_m) = \deg(h_M)$ . Assuming that  $m < M$  and equating degrees in the identities  $g^{(n)} = f^{(m)} \circ h_m$  and  $g^{(N)} = f^{(M)} \circ h_M$ , it follows that  $\deg(g)^{N-n} = \deg(f)^{M-m}$ .

Let  $S = c(M - m)$  with  $c \in \mathbb{N}$ , and write  $g^{(R)} = f^{(S)} \circ h_S$ . Since  $h_S \neq t \circ g$  for every  $t \in \mathbb{C}[X]$ , Lemma 3.2 implies  $\deg(g) \nmid \deg(h_S)$ , so we must have  $R = c(N - n)$  and  $\deg(h_S) = 1$ . Thus,  $F := f^{(M-m)}$  and  $G := g^{(N-n)}$  satisfy the hypotheses of Proposition 4.2, so either  $F$  and  $G$  have a common iterate (so  $f$  and  $g$  do as well) or there is a linear  $\ell \in \mathbb{C}[X]$  for which  $\ell \circ G \circ \ell^{(-1)} = \beta X^{\deg(G)}$  (with  $\beta \in \mathbb{C}^*$ ). In the

latter case, Lemma 3.9 implies that there is a linear  $\hat{\ell} \in \mathbb{C}[X]$  such that  $\hat{\ell} \circ g \circ \hat{\ell}^{(-1)} = X^s$ , contradicting our assumption on the form of  $g$ . □

**5. Noncommensurable polynomials**

In this section we classify the noncommensurable pairs of polynomials  $(f, g)$  for which each polynomial  $f^{(n)}(X) - g^{(n)}(Y)$  has a Siegel factor (in the sense of Definition 2.3).

**PROPOSITION 5.1**

*Pick  $f, g \in \mathbb{C}[X]$  for which  $r := \deg(f)$  and  $s := \deg(g)$  satisfy  $r, s > 1$ . Assume that there exists  $m \in \mathbb{N}$  with these properties:*

(5.1.1)  $g^{(n)} \neq f^{(m)} \circ h$  for every  $h \in \mathbb{C}[X]$  and  $n \in \mathbb{N}$ ; and

(5.1.2) *there are infinitely many  $j \in \mathbb{N}$  for which  $f^{(mj)}(X) - g^{(mj)}(Y)$  has a Siegel factor in  $\mathbb{C}[X, Y]$ .*

*Then there is a linear  $\ell \in \mathbb{C}[X]$  for which  $(\ell \circ f \circ \ell^{(-1)}, \ell \circ g \circ \ell^{(-1)})$  is either  $(X^r, \alpha X^s)$  (with  $\alpha \in \mathbb{C}^*$ ) or  $(\epsilon_1 T_r, \epsilon_2 T_s)$  (with  $\epsilon_1, \epsilon_2 \in \{1, -1\}$ ).*

*Remark*

The converse of Proposition 5.1 holds if and only if some prime factor of  $r$  is not a factor of  $s$ .

*Remark*

The pair  $(\epsilon_1 T_r, \epsilon_2 T_s)$  in the conclusion of Proposition 5.1 differs slightly from the pair  $(T_r \circ \hat{\epsilon} X, T_s \circ \epsilon X)$  in the conclusion of Proposition 4.1. The latter pairs are special cases of the former pairs, but if  $r$  and  $s$  are even, then  $(T_r, -T_s)$  cannot be written in the latter form (even after conjugation by a linear).

*Proof of Proposition 5.1*

Let  $\mathcal{J}$  be the (infinite) set of  $j \in \mathbb{N}$  for which  $f^{(mj)}(X) - g^{(mj)}(Y)$  has a Siegel factor in  $\mathbb{C}[X, Y]$ . For  $j \in \mathcal{J}$ , Corollary 2.7 implies that there are  $A_j, B_j, C_j \in \mathbb{C}[X]$  and linear  $\mu_j, \nu_j \in \mathbb{C}[X]$  such that  $f^{(mj)} = A_j \circ B_j \circ \mu_j$  and  $g^{(mj)} = A_j \circ C_j \circ \nu_j$ , where either  $(B_j, C_j)$  or  $(C_j, B_j)$  has the form of one of (2.7.1)–(2.7.5).

We split the proof into two cases, depending on whether the degrees of the polynomials  $A_j$  are bounded.

*Case 1:  $\{\deg(A_j) : j \in \mathcal{J}\}$  is infinite*

In this case there is an infinite subset  $\mathcal{J}_0$  of  $\mathcal{J}$  such that  $j \mapsto \deg(A_j)$  is a strictly increasing function on  $\mathcal{J}_0$ . Replacing  $\mathcal{J}$  by  $\mathcal{J}_0$ , it follows that  $\deg(A_j)$  exceeds any prescribed bound whenever  $j \in \mathcal{J}$  is sufficiently large. By (5.1.1), for  $j \in \mathcal{J}$  we can-

not have  $A_j = f^{(m)} \circ h$  with  $h \in \mathbb{C}[X]$ . Applying Corollary 3.3 to the decomposition  $(f^{(m)})^{(j)} = A_j \circ (B_j \circ \mu_j)$ , and recalling that  $\deg(A_j) \rightarrow \infty$ , it follows that for sufficiently large  $j$  we have either

$$f^{(mj)} = \ell_j \circ X^{r^{mj}} \circ \ell_j^{(-1)} \quad \text{or} \quad f^{(mj)} = \ell_j \circ T_{r^{mj}} \circ \epsilon_j \ell_j^{(-1)},$$

where  $\ell_j \in \mathbb{C}[X]$  is linear and  $\epsilon_j \in \{1, -1\}$ . Thus, by Lemma 3.9, either

$$f = \ell^{(-1)} \circ X^r \circ \ell \tag{5.2}$$

or

$$f = \ell^{(-1)} \circ T_r \circ \epsilon \ell \tag{5.3}$$

for some linear  $\ell \in \mathbb{C}[X]$  and some  $\epsilon \in \{1, -1\}$ . It remains to determine the shape of  $g$ . To this end, note that in the cases (5.2) and (5.3), respectively, we have

$$f^{(n)} = \ell^{(-1)} \circ X^{r^n} \circ \ell$$

and

$$f^{(n)} = \ell^{(-1)} \circ T_{r^n} \circ \epsilon^n \ell,$$

where in the latter case we have used Lemma 3.4. Since  $f^{(mj)} = A_j \circ (B_j \circ \mu_j)$ , Lemma 3.5 implies that for every  $j \in \mathcal{J}$  there is a linear  $\hat{\ell}_j \in \mathbb{C}[X]$  such that

$$A_j = \ell^{(-1)} \circ X^{\deg(A_j)} \circ \hat{\ell}_j \quad \text{if (5.2) holds,} \tag{5.4}$$

and

$$A_j = \ell^{(-1)} \circ T_{\deg(A_j)} \circ \hat{\ell}_j \quad \text{if (5.3) holds.} \tag{5.5}$$

If  $A_j = g^{(3)} \circ h$  for some  $j \in \mathcal{J}$  and  $h \in \mathbb{C}[X]$ , then by Lemma 3.5 there is a linear  $\tilde{\ell} \in \mathbb{C}[X]$  such that

$$g^{(3)} = \ell^{(-1)} \circ X^{s^3} \circ \tilde{\ell} \quad \text{if (5.4) holds,}$$

and

$$g^{(3)} = \ell^{(-1)} \circ T_{s^3} \circ \tilde{\ell} \quad \text{if (5.5) holds.}$$

By Lemma 3.9, there are  $\alpha \in \mathbb{C}^*$  and  $\hat{\epsilon} \in \{1, -1\}$  such that

$$g = \ell^{(-1)} \circ \alpha X^s \circ \ell \quad \text{if (5.2) holds,}$$

and

$$g = \ell^{(-1)} \circ T_s \circ \hat{\epsilon} \ell \quad \text{if (5.3) holds.}$$

This completes the proof in case  $A_j = g^{(3)} \circ h$ .

Now suppose that  $A_j \neq g^{(3)} \circ h$  for every  $j \in \mathcal{J}$  and  $h \in \mathbb{C}[X]$ . Since  $(g^{(3)})^{(mj)} = g^{(3mj)} = A_j \circ (C_j \circ v_j \circ g^{(2mj)})$ , and moreover  $\deg(A_j) \rightarrow \infty$  as  $j \rightarrow \infty$ , Corollary 3.3 implies that either

$$g^{(3)} = \tilde{\ell} \circ X^{s^3} \circ \tilde{\ell}^{(-1)}$$

or

$$g^{(3)} = \tilde{\ell} \circ T_{s^3} \circ \tilde{\epsilon} \tilde{\ell}^{(-1)},$$

where  $\tilde{\ell} \in \mathbb{C}[X]$  is linear and  $\tilde{\epsilon} \in \{1, -1\}$ . By Lemma 3.9, either

$$g = \tilde{\ell} \circ \beta X^s \circ \tilde{\ell}^{(-1)} \tag{5.6}$$

or

$$g = \tilde{\ell} \circ T_s \circ \hat{\epsilon} \tilde{\ell}^{(-1)}, \tag{5.7}$$

where  $\beta \in \mathbb{C}^*$  and  $\hat{\epsilon} \in \{1, -1\}$ . Thus, for  $n \in \mathbb{N}$ , we have

$$g^{(n)} = \tilde{\ell} \circ \beta^{1+s+\dots+s^{n-1}} X^{s^n} \circ \tilde{\ell}^{(-1)} \quad \text{if (5.6) holds,}$$

and

$$g^{(n)} = \tilde{\ell} \circ T_{s^n} \circ \hat{\epsilon}^n \tilde{\ell}^{(-1)} \quad \text{if (5.7) holds.}$$

Applying Lemma 3.5 to the decomposition  $g^{(mj)} = A_j \circ (C_j \circ v_j)$ , we see that there is a linear  $\tilde{\ell}_j \in \mathbb{C}[X]$  such that

$$A_j = \tilde{\ell} \circ X^{\deg(A_j)} \circ \tilde{\ell}_j \quad \text{if (5.6) holds,} \tag{5.8}$$

and

$$A_j = \tilde{\ell} \circ T_{\deg(A_j)} \circ \tilde{\ell}_j \quad \text{if (5.7) holds.} \tag{5.9}$$

Choose  $j \in \mathcal{J}$  for which  $\deg(A_j) > 2$ .

If (5.3) holds then so does (5.5), so Lemma 3.6 implies (5.8) does not hold, whence (5.9) and (5.7) hold; Lemma 3.6 implies further that  $\tilde{\ell} = \ell^{(-1)} \circ \delta X$  for some  $\delta \in \{1, -1\}$ . But then

$$\begin{aligned}
 g &= \ell^{(-1)} \circ \delta T_s \circ \hat{\epsilon} \delta \ell \\
 &= \ell^{(-1)} \circ \delta^{1+s} \hat{\epsilon}^s T_s \circ \ell,
 \end{aligned}$$

which completes the proof in this case.

Finally, if (5.2) holds, then so does (5.4), so Lemma 3.6 implies (5.9) does not hold, whence (5.8) and (5.6) hold; moreover,  $\tilde{\ell} = \ell^{(-1)} \circ \gamma X$  for some  $\gamma \in \mathbb{C}^*$ . But then

$$\begin{aligned}
 g &= \ell^{(-1)} \circ \gamma \beta X^s \circ \gamma^{-1} \ell \\
 &= \ell^{(-1)} \circ \gamma^{1-s} \beta X^s \circ \ell,
 \end{aligned}$$

which completes the proof in Case 1.

*Case 2:  $\{\deg(A_j) : j \in \mathcal{J}\}$  is finite*

Suppose first that  $e := \gcd(\deg(f), \deg(g))$  satisfies  $e > 1$ . In this case,  $\gcd(\deg(f^{(mj)}), \deg(g^{(mj)})) = e^{mj} \rightarrow \infty$  as  $j \rightarrow \infty$ , and since  $\deg(A_j)$  is bounded it follows that  $\gcd(\deg(B_j), \deg(C_j)) \rightarrow \infty$ . For any nonconstant  $F, G \in \mathbb{C}[X]$  such that  $(F, G)$  has any of the forms (2.7.1)–(2.7.5) other than (2.7.4), we observe that  $\gcd(\deg(F), \deg(G)) \leq 2$ ; thus, for every sufficiently large  $j \in \mathcal{J}$ , either  $(B_j, C_j)$  or  $(C_j, B_j)$  has the form (2.7.4). For any such  $j$ , after perhaps replacing  $(A_j, B_j, C_j)$  by  $(A_j(-X), -B_j, -C_j)$ , we find that  $B_j = T_{\deg(B_j)}$  and  $C_j = -T_{\deg(C_j)}$ . Since  $f^{(mj)} = A_j \circ T_{\deg(B_j)} \circ \mu_j$  and  $\deg(A_j)$  is bounded, for sufficiently large  $j \in \mathcal{J}$  we must have  $r^3 \mid \deg(B_j)$ ; applying Lemma 3.2 to the decomposition  $f^{(mj-3)} \circ f^{(3)} = (A_j \circ T_{\deg(B_j)/r^3}) \circ (T_{r^3} \circ \mu_j)$  gives  $f^{(3)} = \ell_j \circ T_{r^3} \circ \mu_j$  with  $\ell_j \in \mathbb{C}[X]$  linear. Lemma 3.9 implies  $f = \ell_j \circ T_r \circ \epsilon \ell_j^{(-1)}$  with  $\epsilon \in \{1, -1\}$ ; then  $\ell_j \circ T_{r^3} \circ \mu_j = f^{(3)} = \ell_j \circ T_{r^3} \circ \epsilon \ell_j^{(-1)}$ , so Lemma 3.6 implies  $\mu_j = \delta \epsilon \ell_j^{(-1)}$  for some  $\delta \in \{1, -1\}$  with  $\delta^r = 1$ . But then  $A_j \circ T_{\deg(B_j)} \circ \mu_j = f^{(mj)} = \mu_j^{(-1)} \circ \delta \epsilon T_{r^{mj}} \circ \delta \epsilon^{mj+1} \mu_j$ , so Lemma 3.5 implies there is a linear  $\tilde{\ell} \in \mathbb{C}[X]$  such that  $A_j \circ \tilde{\ell} = \mu_j^{(-1)} \circ \delta \epsilon T_{\deg(A_j)}$  and  $\tilde{\ell}^{(-1)} \circ T_{\deg(B_j)} \circ \mu_j = T_{\deg(B_j)} \circ \delta \epsilon^{mj+1} \mu_j$ . Then  $\tilde{\ell} \in \{X, -X\}$ , so  $\mu_j \circ A_j = \tilde{\epsilon} T_{\deg(A_j)}$  with  $\tilde{\epsilon} \in \{1, -1\}$ . The same argument shows that  $v_j \circ A_j = \hat{\epsilon} T_{\deg(A_j)}$  for some  $\hat{\epsilon} \in \{1, -1\}$ , so  $\hat{\epsilon} v_j = \tilde{\epsilon} \mu_j$ . From above,  $f = \mu_j^{(-1)} \circ \epsilon_0 T_r \circ \mu_j$  with  $\epsilon_0 \in \{1, -1\}$ , and similarly  $g = v_j^{(-1)} \circ \epsilon_1 T_s \circ v_j$  with  $\epsilon_1 \in \{1, -1\}$ , so  $g = \mu_j^{(-1)} \circ \epsilon_2 T_s \circ \mu_j$  with  $\epsilon_2 \in \{1, -1\}$ , and the result follows.

Henceforth suppose that  $\gcd(\deg(f), \deg(g)) = 1$ . In this case, for  $j \in \mathcal{J}$  we have  $\deg(A_j) = 1$  and  $\gcd(\deg(B_j), \deg(C_j)) = 1$ ; by examining (2.7.1)–(2.7.5), we see that one of  $(B_j, C_j)$  and  $(C_j, B_j)$  must have the form of either (2.7.1) or (2.7.3).

Suppose that there is some  $j \in \mathcal{J}$  with  $j > 2/m$  such that either  $(B_j, C_j)$  or  $(C_j, B_j)$  has the form (2.7.3). For any such  $j$  we have

$$(B_j, C_j) = (T_{\deg(B_j)}, T_{\deg(C_j)});$$

since  $A_j$  is linear, this implies that

$$f^{(mj)} = A_j \circ T_{r^{mj}} \circ \mu_j$$

and

$$g^{(mj)} = A_j \circ T_{s^{mj}} \circ \nu_j.$$

By Lemma 3.9, we have

$$f = A_j \circ T_r \circ \epsilon_j \circ A_j^{(-1)}$$

and

$$g = A_j \circ T_s \circ \bar{\epsilon}_j \circ A_j^{(-1)}$$

for some  $\epsilon_j, \bar{\epsilon}_j \in \{1, -1\}$ , so the result holds.

Now suppose that, for every  $j \in \mathcal{J}$  with  $j > 2/m$ , either  $(B_j, C_j)$  or  $(C_j, B_j)$  has the form (2.7.1). For any such  $j$ , we have

$$\{B_j, C_j\} = \{X^n, X^i p(X)^n\},$$

where  $p \in \mathbb{C}[X]$  and  $i \in \mathbb{N}_0$  satisfy  $\gcd(i, n) = 1$ . Since  $n$  is the degree of either  $f^{(mj)}$  or  $g^{(mj)}$ , we have  $n \in \{r^{mj}, s^{mj}\}$ , so  $n > 1$  and thus  $i > 0$ . Lemmas 3.5 and 3.8 imply that

$$\{f^{(2)}, g^{(2)}\} = \{A_j \circ X^{\tilde{n}} \circ \mu, A_j \circ X^{\tilde{i}} \tilde{p}(X)^n \circ \nu\}, \tag{5.10}$$

where  $\tilde{i}, \tilde{n} \in \mathbb{N}$  and  $\mu, \nu, \tilde{p} \in \mathbb{C}[X]$  with  $\mu, \nu$  linear. We may assume that  $j$  satisfies

$$\min(r, s)^{mj} > \max(r, s)^2.$$

Since  $n \in \{r^{mj}, s^{mj}\}$ , it follows that  $n > \max(r, s)^2$ , so we must have  $\tilde{p} \in \mathbb{C}^*$ . Applying Lemma 3.9 to (5.10), we conclude that

$$(f, g) = (A_j \circ \hat{\alpha} X^r \circ A_j^{(-1)}, A_j \circ \hat{\beta} X^s \circ A_j^{(-1)})$$

for some  $\hat{\alpha}, \hat{\beta} \in \mathbb{C}^*$ . Finally, after replacing  $A_j$  by  $A_j \circ \gamma X$  for suitable  $\gamma \in \mathbb{C}^*$ , we may assume  $\hat{\alpha} = 1$ , which completes the proof. □

### 6. Proof of Theorem 1.1

In this section we conclude the proof of Theorem 1.1. Our strategy is to combine the results of the previous two sections with Siegel’s theorem to reduce to the case that the pair  $(f, g)$  has one of the two forms

$$(X^r, \beta X^s), \quad \text{with } \beta \in \mathbb{C}^* \text{ and } r, s \in \mathbb{Z}_{>1}; \tag{6.1}$$

$$(\epsilon_1 T_r, \epsilon_2 T_s), \quad \text{with } \epsilon_1, \epsilon_2 \in \{1, -1\} \text{ and } r, s \in \mathbb{Z}_{>1}. \tag{6.2}$$

We then use Corollary 2.2 (which is a consequence of Siegel’s theorem) to handle these two possibilities.

PROPOSITION 6.3

*Pick  $f, g \in \mathbb{C}[X]$  for which  $r := \deg(f)$  and  $s := \deg(g)$  satisfy  $r, s > 1$ . Assume that, for every  $n \in \mathbb{N}$ , the polynomial  $f^{(n)}(X) - g^{(n)}(Y)$  has a Siegel factor in  $\mathbb{C}[X, Y]$ . Then either  $f$  and  $g$  have a common iterate or there is a linear  $\ell \in \mathbb{C}[X]$  such that  $(\ell \circ f \circ \ell^{(-1)}, \ell \circ g \circ \ell^{(-1)})$  has one of the forms (6.1) or (6.2).*

*Proof*

This follows from Propositions 4.1 and 5.1. □

COROLLARY 6.4

*Pick  $x, y \in \mathbb{C}$  and nonlinear  $f, g \in \mathbb{C}[X]$ . If  $\mathcal{O}_f(x) \cap \mathcal{O}_g(y)$  is infinite, then either  $f$  and  $g$  have a common iterate or there is a linear  $\ell \in \mathbb{C}[X]$  such that  $(\ell \circ f \circ \ell^{(-1)}, \ell \circ g \circ \ell^{(-1)})$  has one of the forms (6.1) or (6.2).*

*Proof*

Let  $R$  be the ring generated by  $x, y$  and the coefficients of  $f$  and  $g$ , and let  $K$  be the field of fractions of  $R$ . Note that both  $R$  and  $K$  are finitely generated. Since  $\mathcal{O}_f(x) \cap \mathcal{O}_g(y)$  is infinite, for each  $n \in \mathbb{N}$  the equation  $f^{(n)}(X) = g^{(n)}(Y)$  has infinitely many solutions in  $\mathcal{O}_f(x) \times \mathcal{O}_g(y) \subseteq R \times R$ . By Siegel’s theorem (Corollary 2.4), for each  $n \in \mathbb{N}$  the polynomial  $f^{(n)}(X) - g^{(n)}(Y)$  has a Siegel factor in  $K[X, Y]$ . Now the conclusion follows from the previous result. (Note that  $f$  and  $g$  are nonconstant since  $\mathcal{O}_f(x)$  and  $\mathcal{O}_g(y)$  are infinite.) □

*Proof of Theorem 1.1*

By Corollary 6.4, it suffices to prove Theorem 1.1 in case there is a linear  $\ell \in \mathbb{C}[X]$  for which  $(\tilde{f}, \tilde{g}) := (\ell \circ f \circ \ell^{(-1)}, \ell \circ g \circ \ell^{(-1)})$  has one of the forms (6.1) or (6.2). But then

$$\mathcal{O}_{\tilde{f}}(\ell(x)) \cap \mathcal{O}_{\tilde{g}}(\ell(y)) = \ell(\mathcal{O}_f(x)) \cap \ell(\mathcal{O}_g(y)) = \ell(\mathcal{O}_f(x) \cap \mathcal{O}_g(y))$$

is infinite, so Proposition 6.5 implies that  $\tilde{f}^{(i)} = \tilde{g}^{(j)}$  for some  $i, j \in \mathbb{N}$ , whence  $f^{(i)} = g^{(j)}$ . □

PROPOSITION 6.5

Pick  $f, g \in \mathbb{C}[X]$  such that  $(f, g)$  has one of the forms (6.1) or (6.2). If there are  $x, y \in \mathbb{C}$  for which  $\mathcal{O}_f(x) \cap \mathcal{O}_g(y)$  is infinite, then  $f$  and  $g$  have a common iterate.

Proof

Assuming  $\mathcal{O}_f(x) \cap \mathcal{O}_g(y)$  is infinite, let  $M$  be the set of pairs  $(m, n) \in \mathbb{N} \times \mathbb{N}$  for which  $f^{(m)}(x) = g^{(n)}(y)$ . Note that any two elements of  $M$  have distinct first coordinates, since if  $M$  contains  $(m, n_1)$  and  $(m, n_2)$  with  $n_1 \neq n_2$ , then  $g^{(n_1)}(y) = g^{(n_2)}(y)$  so  $\mathcal{O}_g(y)$  would be finite. Likewise, any two elements of  $M$  have distinct second coordinates, so there are elements  $(m, n) \in M$  in which  $\min(m, n)$  is arbitrarily large.

Suppose  $(f, g)$  has the form (6.1). Since  $f^{(m)}(x) = x^{r^m}$  and  $\mathcal{O}_f(x)$  is infinite,  $x$  is neither zero nor a root of unity. We compute

$$g^{(n)}(y) = \beta \frac{s^n - 1}{s - 1} y^{s^n};$$

putting  $y_1 := \beta_1 y$  where  $\beta_1 \in \mathbb{C}^*$  satisfies  $\beta_1^{s-1} = \beta$ , it follows that  $g^{(n)}(y) = y_1^{s^n} / \beta_1$ , so infinitude of  $\mathcal{O}_g(y)$  implies that  $y_1$  is neither zero nor a root of unity.

A pair  $(m, n) \in \mathbb{N} \times \mathbb{N}$  lies in  $M$  if and only if

$$x^{r^m} = \beta \frac{s^n - 1}{s - 1} y^{s^n}, \tag{6.6}$$

or equivalently

$$\beta_1 x^{r^m} = y_1^{s^n}. \tag{6.7}$$

Since (6.7) holds for two pairs  $(m, n) \in M$  which differ in both coordinates, we have  $x^a = y_1^b$  for some nonzero integers  $a, b$ . By choosing  $a$  to have minimal absolute value, it follows that the set  $S := \{(u, v) \in \mathbb{Z}^2 : \beta_1 x^u = y_1^v\}$  has the form  $\{(c + ak, d + bk) : k \in \mathbb{Z}\}$  for some  $c, d \in \mathbb{Z}$ . For  $(m, n) \in M$  we have  $(r^m, s^n) \in S$ , so  $(r^m - c)/a = (s^n - d)/b$ . Since  $M$  is infinite, Corollary 2.2 implies that  $c/a = d/b$ . In particular, every  $(m, n) \in M$  satisfies  $br^m = as^n$ . Pick two pairs  $(m, n)$  and  $(m + m_0, n + n_0)$  in  $M$  with  $m_0, n_0 \in \mathbb{N}$ . Then  $r^{m_0} = s^{n_0}$ , and  $S$  contains both  $(r^m, s^n)$  and  $(r^{m+m_0}, s^{n+n_0})$ , so

$$y_1^{s^n} x^{-r^m} = \beta_1 = y_1^{s^{n+n_0}} x^{-r^{m+m_0}},$$

and thus

$$(y_1^{s^n})^{s^{n_0}-1} = (x^{r^m})^{r^{m_0}-1}.$$

Since  $r^{m_0} = s^{n_0}$ , it follows that  $\beta_1^{s^{n_0}-1} = 1$ , so  $f^{(m_0)} = g^{(n_0)}$ .

Now suppose  $(f, g)$  has the form (6.2). Then (by Lemma 3.4) for any  $m, n \in \mathbb{N}$  there exist  $\epsilon_3, \epsilon_4 \in \{1, -1\}$  such that  $(f^{(m)}, g^{(n)}) = (\epsilon_3 T_{r^m}, \epsilon_4 T_{s^n})$ . Since  $\mathcal{O}_f(x) \cap$

$\mathcal{O}_g(y)$  is infinite, we can choose  $\delta \in \{1, -1\}$  such that  $T_{r^m}(x) = \delta T_{s^n}(y)$  for infinitely many  $(m, n) \in \mathbb{N} \times \mathbb{N}$ . Pick  $x_0, y_0 \in \mathbb{C}^*$  such that  $x_0 + x_0^{-1} = x$  and  $y_0 + y_0^{-1} = y$ . Then there are infinitely many pairs  $(m, n) \in \mathbb{N} \times \mathbb{N}$  for which

$$x_0^{r^m} + x_0^{-r^m} = \delta(y_0^{s^n} + y_0^{-s^n}),$$

so we can choose  $\epsilon \in \{1, -1\}$  such that

$$x_0^{r^m} = \delta y_0^{\epsilon s^n} \tag{6.8}$$

for infinitely many  $(m, n) \in \mathbb{N} \times \mathbb{N}$ . Moreover, since  $\mathcal{O}_f(x)$  and  $\mathcal{O}_g(y)$  are infinite, neither  $x_0$  nor  $y_0$  is a root of unity, so distinct pairs  $(m, n) \in \mathbb{N} \times \mathbb{N}$  which satisfy (6.8) must differ in both coordinates. Now (6.8) is a reformulation of (6.7), so we conclude as above that  $r^{m_0} = s^{n_0}$  for some  $m_0, n_0 \in \mathbb{N}$  such that  $\delta^{s^{n_0}-1} = 1$ . If  $s$  is odd, it follows that  $f^{(2m_0)} = g^{(2n_0)}$ . If  $s$  is even, then we cannot have  $\delta = -1$ ; since  $f^{(m)} = \epsilon_1 T_{r^m}$  and  $g^{(n)} = \epsilon_2 T_{s^n}$ , it follows that  $\epsilon_1 = \epsilon_2$ , so  $f^{(m_0)} = g^{(n_0)}$ .  $\square$

*Remark*

If  $(f, g)$  has the form (6.1) or (6.2), then  $f^n(X) - g^m(Y)$  has a Siegel factor in  $\mathbb{C}[X, Y]$  for every  $n, m \in \mathbb{N}$ . (In fact,  $f^n(X) - g^m(Y)$  is the product of irreducible Siegel polynomials.) So the results of the previous two sections give no information. To illustrate Theorem 1.1 for such  $(f, g)$ , consider  $(f, g) = (X^2, X^3)$ . In this case, for any  $n, m \in \mathbb{N}$ , the equation  $f^n(X) = g^m(Y)$  has infinitely many solutions in  $\mathbb{Z} \times \mathbb{Z}$ . However, for any  $x_0, y_0 \in \mathbb{C}$ , each such equation has only finitely many solutions in  $\mathcal{O}_f(x_0) \times \mathcal{O}_g(y_0)$ . In particular, each such equation has only finitely many solutions in  $\mathcal{O}_f(2) \times \mathcal{O}_g(2) = \{(2^{2^a}, 2^{3^b}) : a, b \in \mathbb{N}_0\}$ , but has infinitely many solutions in  $2^{\mathbb{N}_0} \times 2^{\mathbb{N}_0}$ . The underlying principle is that orbits are rather thin subsets of  $\mathbb{C}$ .

**7. A multivariate generalization**

In this section we show that Theorem 1.1 implies Theorem 1.3 and Corollary 1.5.

*Proof of Theorem 1.3*

We use induction on  $d$ . If  $d = 1$ , then  $L(\mathbb{C}) = \mathbb{C}$ , so  $f(L) = L$ . Now assume that the result holds for lines in  $\mathbb{C}^{d-1}$ .

If all points of  $L$  take the same value  $z_d$  on the last coordinate, then  $L = L_0 \times \{z_d\}$  for some line  $L_0 \subset \mathbb{C}^{d-1}$ . By the inductive hypothesis, there exist nonnegative integers  $m_1, \dots, m_{d-1}$  (not all zero) such that  $L_0$  is invariant under  $(f_1^{(m_1)}, \dots, f_{d-1}^{(m_{d-1})})$ . Then  $L$  is invariant under  $(f_1^{(m_1)}, \dots, f_{d-1}^{(m_{d-1})}, f_d^{(0)})$ , as desired.

Henceforth assume that  $L$  projects surjectively onto each coordinate. Then any point of  $L$  is uniquely determined by its value at any prescribed coordinate. Since  $L$  contains infinitely many points on  $\mathcal{O}_{f_1}(x_1) \times \dots \times \mathcal{O}_{f_d}(x_d)$ , it follows that  $\mathcal{O}_{f_i}(x_i)$

is infinite for each  $i$ . For each  $i = 2, \dots, d$ , let  $\pi_i : \mathbb{C}^d \rightarrow \mathbb{C}^2$  be the projection onto the first and  $i$ th coordinates of  $\mathbb{C}^d$ . Then  $L_i := \pi_i(L)$  is a line in  $\mathbb{C}^2$  having infinite intersection with  $\mathcal{O}_{f_1}(x_1) \times \mathcal{O}_{f_i}(x_i)$ . Since  $L$  projects surjectively onto each coordinate,  $L_i$  is given by the equation  $X_i = \sigma_i(X_1)$  for some degree-one  $\sigma_i \in \mathbb{C}[X]$ . For any  $k, \ell \in \mathbb{N}$  such that

$$(f_1^{(k)}(x_1), f_i^{(\ell)}(x_i)) \in L_i,$$

we have  $(\sigma_i \circ f_1 \circ \sigma_i^{(-1)})^{(k)}(\sigma_i(x_1)) = f_i^{(\ell)}(x_i)$ . Thus, by Theorem 1.1 there exist  $m_i, n_i \in \mathbb{N}$  such that

$$(\sigma_i \circ f_1 \circ \sigma_i^{(-1)})^{(m_i)} = f_i^{(n_i)}.$$

Let  $M_1$  be the least common multiple of all the  $m_i$ , and for each  $i \geq 2$  define  $M_i := (n_i M_1)/m_i$ . Then

$$(\sigma_i \circ f_1 \circ \sigma_i^{(-1)})^{(M_1)} = f_i^{(M_i)},$$

so for any  $y_1 \in \mathbb{C}$  we have

$$f_i^{(M_i)}(\sigma_i(y_1)) = \sigma_i \circ f_1^{(M_1)}(y_1).$$

Since  $L$  is defined by the  $(d - 1)$  equations  $X_i = \sigma_i(X_1)$ , it follows that  $L$  is invariant under  $(f_1^{(M_1)}, \dots, f_d^{(M_d)})$ . □

*Proof of Corollary 1.5*

Arguing inductively as in the above proof, we may assume that the projection of  $L$  onto each coordinate of  $\mathbb{C}^d$  is surjective. Thus each point of  $L$  is uniquely determined by its value on any prescribed coordinate. By Theorem 1.3,  $L$  is preserved by  $\rho_1^{m_1} \cdots \rho_d^{m_d}$  for some nonnegative integers  $m_1, \dots, m_d$  which are not all zero. Without loss of generality, assume that  $m_1 > 0$ . For each  $k$  with  $1 \leq k \leq m_1$ , let  $U_k$  be the set of tuples  $(n_1, \dots, n_d) \in (\mathbb{N}_0)^d$  such that  $n_1 \equiv k \pmod{m_1}$  and  $\rho_1^{n_1} \cdots \rho_d^{n_d}(\alpha)$  lies on  $L$ . If  $U_k$  is nonempty, pick  $(n_1, \dots, n_d) \in U_k$  for which  $n_1$  is minimal; then  $U_k$  contains  $V_k := \{(n_1 + jm_1, \dots, n_d + jm_d) : j \in \mathbb{N}_0\}$ , and the set  $Z_k$  of values  $\rho_1^{u_1} \cdots \rho_d^{u_d}(\alpha)$  for  $(u_1, \dots, u_d) \in U_k$  is the same as the corresponding set for  $(u_1, \dots, u_d) \in V_k$ . Thus  $Z_k$  is the orbit of  $\alpha$  under  $\langle \rho_1^{m_1} \cdots \rho_d^{m_d} \rangle \rho_1^{n_1} \cdots \rho_d^{n_d}$ , which is a coset of a cyclic subsemigroup. □

**8. Function field case, second proof**

We now turn our attention to the following result.

**THEOREM 8.1**

*Let  $K$  be a field of characteristic 0, let  $f, g \in K[X]$  be polynomials of degree greater*

than one, and let  $x_0, y_0 \in K$ . Assume that there is no linear  $\mu \in \overline{K}[X]$  for which  $\mu^{(-1)}(x_0), \mu^{(-1)}(y_0) \in \overline{\mathbb{Q}}$  and both  $\mu^{(-1)} \circ f \circ \mu$  and  $\mu^{(-1)} \circ g \circ \mu$  are in  $\overline{\mathbb{Q}}[X]$ . If  $\mathcal{O}_f(x_0) \cap \mathcal{O}_g(y_0)$  is infinite, then  $f$  and  $g$  have a common iterate.

Theorem 8.1 may be viewed as the “function field” part of our Theorem 1.1. We give an alternate proof of Theorem 8.1 by using the theory of heights. In the next two sections we review canonical heights associated to nonlinear polynomials. Then in Section 11 we prove Theorem 8.1 by reducing it to the case  $\deg(f) = \deg(g)$  handled in our previous paper [18, Theorem 1.1]. Here we avoid the intricate arguments about polynomial decomposition used in the first part of the present paper; instead our proof relies on a result of Lang, already used in the proof of Proposition 6.5, which is itself a consequence of Siegel’s theorem.

### 9. Canonical heights associated to polynomials

In this section we recall some standard terminology about heights. First, a *global field* is either a number field or a function field of transcendence degree 1 over another field. Any global field  $E$  comes equipped with a standard set  $M_E$  of absolute values  $|\cdot|_v$  which satisfy a product formula

$$\prod_{v \in M_E} |x|_v^{N_v} = 1 \quad \text{for every } x \in E^*,$$

where  $N : M_E \rightarrow \mathbb{N}$  and  $N_v := N(v)$  (cf. [23] for details).

If  $E$  is a global field, the logarithmic Weil height of  $x \in \overline{E}$  (with respect to  $E$ ) is defined as (see [23, p. 52])

$$h_E(x) = \frac{1}{[E(x) : E]} \cdot \sum_{v \in M_E} \sum_{\substack{w|v \\ w \in M_{E(x)}}} \log \max\{|x|_w^{N_w}, 1\}.$$

#### Definition 9.1

Let  $E$  be a global field, let  $\phi \in E[X]$  with  $\deg(\phi) > 1$ , and let  $z \in \overline{E}$ . The canonical height  $\widehat{h}_{\phi,E}(z)$  of  $z$  with respect to  $\phi$  (and  $E$ ) is

$$\widehat{h}_{\phi,E}(z) := \lim_{k \rightarrow \infty} \frac{h_E(\phi^{(k)}(z))}{\deg(\phi)^k}.$$

Call and Silverman [10, Theorem 1.1] proved the existence of the above limit, using boundedness of  $|h_E(\phi(x)) - (\deg \phi)h_E(x)|$  and a telescoping sum argument due to Tate. We usually write  $h(x)$  and  $\widehat{h}_\phi(x)$  rather than  $h_E(x)$  and  $\widehat{h}_{\phi,E}(x)$ ; this should not cause confusion. We use the following properties of the canonical height.

PROPOSITION 9.2

Let  $E$  be a global field, let  $\phi \in E[X]$  be a polynomial of degree greater than 1, and let  $z \in \overline{E}$ . Then

- (a) for each  $k \in \mathbb{N}$ , we have  $\widehat{h}_\phi(\phi^{(k)}(z)) = \deg(\phi)^k \cdot \widehat{h}_\phi(z)$ ;
- (b)  $|h(z) - \widehat{h}_\phi(z)|$  is bounded by a function which does not depend on  $z$ ;
- (c) if  $E$  is a number field, then  $z$  is preperiodic if and only if  $\widehat{h}_\phi(z) = 0$ .

*Proof*

Part (a) is clear; for (b) see [10, Theorem 1.1]; and for (c) see [10, Corollary 1.1.1]. □

Part (c) of Proposition 9.2 is not true if  $E$  is a function field with constant field  $E_0$ , since  $\widehat{h}_\phi(z) = 0$  whenever  $z \in E_0$  and  $\phi \in E_0[X]$ . But these are essentially the only counterexamples in the function field case (cf. Lemma 10.7).

**10. Canonical heights in function fields**

The setup for this section is as follows:  $E$  is a field, and  $K$  is a function field of transcendence degree 1 over  $E$ .

First we note that for each place  $v \in M_K$  of the function field  $K$ , we may assume that every  $z \in K$  satisfies  $\log |z|_v \in \mathbb{Q}$ . (We use  $c := e^{-1}$  in the definition of absolute values on function fields from [23, p. 62].)

Let  $\phi \in K[X]$  be a polynomial of degree greater than 1. For each  $v \in M_K$ , we let

$$\widehat{h}_{\phi,v}(z) := \lim_{n \rightarrow \infty} \frac{\log \max\{|\phi^{(n)}(z)|_v^{N_v}, 1\}}{\deg(\phi)^n} \tag{10.1}$$

be the canonical local height of  $z \in K$  at  $v$ . Clearly, for all but finitely many  $v \in M_K$ , both  $z$  and all coefficients of  $\phi$  are  $v$ -adic integers. Hence, for such  $v \in M_K$ , we have  $\widehat{h}_{\phi,v}(z) = 0$ . It turns out that

$$\widehat{h}_\phi(z) = \sum_{v \in M_K} \widehat{h}_{\phi,v}(z). \tag{10.2}$$

For a proof of the existence of the limit in (10.1), and of the equality in (10.2), see [9].

The following result is crucial for Section 11.

LEMMA 10.3

For each  $z \in K$ , and for each  $\phi \in K[X]$  with  $\deg(\phi) > 1$ , we have  $\widehat{h}_\phi(z) \in \mathbb{Q}$ .

*Proof*

Since (10.2) is a finite sum, it suffices to prove that  $\widehat{h}_{\phi,v}(z) \in \mathbb{Q}$  for each  $v \in M_K$ . For some fixed  $v \in M_K$ , write  $\phi(x) = \sum_{i=0}^d \delta_i x^i$  with  $\delta_i \in K$  and  $\delta_d \neq 0$ , and let

$$M_v := \max \left\{ |\delta_d|_v^{-\frac{1}{d-1}}, \max_{0 \leq i < d} \left| \frac{\delta_i}{\delta_d} \right|_v^{\frac{1}{d-i}} \right\}.$$

The following claim follows from [16, Lemma 4.4]; we include a short proof for the sake of completeness.

CLAIM 10.4

If  $|x|_v > M_v$ , then  $\widehat{h}_{\phi,v}(x) = \log |x|_v + \frac{\log |\delta_d|_v}{d-1}$ .

*Proof of Claim 10.4*

From the definition of  $M_v$  and the fact that  $|x|_v > M_v$ , we obtain that

$$|\delta_d x^d|_v > \max_{0 \leq i < d} |\delta_i x^i|_v.$$

Thus  $|\phi(x)|_v = |\delta_d x^d|_v$ ; but since  $|x|_v > M_v \geq |\delta_d|_v^{-1/(d-1)}$ , it follows that  $|\phi(x)|_v > |x|_v$ . An easy induction now shows that

$$|\phi^{(n)}(x)|_v = |\delta_d|_v^{\frac{d^n-1}{d-1}} \cdot |x|_v^{d^n} \quad \text{for all } n \in \mathbb{N}.$$

Hence,

$$\widehat{h}_{\phi,v}(x) = \lim_{n \rightarrow \infty} \frac{\log |\phi^{(n)}(x)|_v}{d^n} = \log |x|_v + \frac{\log |\delta_d|_v}{d-1},$$

as desired. □

Finally, if  $\widehat{h}_{\phi,v}(z) > 0$ , then there exists  $n \in \mathbb{N}$  such that  $|\phi^{(n)}(z)|_v > M_v$ . Thus,

$$\widehat{h}_{\phi,v}(z) = \frac{\widehat{h}_{\phi,v}(\phi^{(n)}(z))}{d^n} = \frac{\log |\phi^{(n)}(z)|_v + \frac{\log |\delta_d|_v}{d-1}}{d^n} \tag{10.5}$$

is a rational number (by our convention on  $|\cdot|_v$ ), which concludes the proof of Lemma 10.3. □

The following result about canonical heights of non-preperiodic points for non-isotrivial polynomials is used later.

*Definition 10.6*

We say a polynomial  $\phi \in K[X]$  is isotrivial over  $E$  if there exists a linear  $\ell \in \overline{K}[X]$  such that  $\ell \circ \phi \circ \ell^{(-1)} \in \overline{E}[X]$ .

Benedetto proved that a non-isotrivial polynomial has nonzero canonical height at its non-preperiodic points [4, Theorem B].

LEMMA 10.7

Let  $\phi \in K[X]$  with  $\deg(\phi) \geq 2$ , and let  $z \in \overline{K}$ . If  $\phi$  is non-isotrivial over  $E$ , then  $\widehat{h}_\phi(z) = 0$  if and only if  $z$  is preperiodic for  $\phi$ .

We state one more preliminary result, which is proved in [18, Lemma 6.8].

LEMMA 10.8

Let  $\phi \in K[X]$  be isotrivial over  $E$ , and let  $\ell$  be as in Definition 10.6. If  $z \in \overline{K}$  satisfies  $\widehat{h}_\phi(z) = 0$ , then  $\ell(z) \in \overline{E}$ .

Definition 10.9

With the notation as in Lemma 10.8, we call the pair  $(\phi, z)$  isotrivial. Furthermore, if  $F \subset K$  is any subfield, and there exists a linear polynomial  $\ell \in \overline{K}[X]$  such that  $\ell \circ \phi \circ \ell^{(-1)} \in \overline{F}[X]$  and  $\ell(z) \in \overline{F}$ , then we call the pair  $(\phi, z)$  isotrivial over  $F$ .

11. Proof of Theorem 8.1

In this section we assume the hypotheses of Theorem 8.1. We first prove two easy claims.

CLAIM 11.1

Let  $E$  be any subfield of  $K$ , and assume that  $(f, x_0)$  and  $(g, y_0)$  are isotrivial over  $E$ . If  $\mathcal{O}_f(x_0) \cap \mathcal{O}_g(y_0)$  is infinite, then there exists a linear  $\mu \in \overline{K}[X]$  such that  $\mu \circ f \circ \mu^{(-1)}, \mu \circ g \circ \mu^{(-1)} \in \overline{E}[X]$  and  $\mu(x_0), \mu(y_0) \in \overline{E}$ .

Proof of Claim 11.1

We know that there exist linear  $\mu_1, \mu_2 \in \overline{K}[X]$  such that  $f_1 := \mu_1 \circ f \circ \mu_1^{(-1)} \in \overline{E}[X]$  and  $g_1 := \mu_2 \circ g \circ \mu_2^{(-1)} \in \overline{E}[X]$ , and  $x_1 := \mu_1(x_0) \in \overline{E}$  and  $y_1 := \mu_2(y_0) \in \overline{E}$ . Thus  $\mathcal{O}_{f_1}(x_1) = \mu_1(\mathcal{O}_f(x_0))$  and  $\mathcal{O}_{g_1}(y_1) = \mu_2(\mathcal{O}_g(y_0))$ . Since  $\mathcal{O}_f(x_0) \cap \mathcal{O}_g(y_0)$  is infinite, there are infinitely many pairs  $(z_1, z_2) \in \overline{E} \times \overline{E}$  such that  $\mu_1^{(-1)}(z_1) = \mu_2^{(-1)}(z_2)$ . Thus  $\mu := \mu_2 \circ \mu_1^{(-1)} \in \overline{E}[X]$ . Hence,

$$\mu_1 \circ g \circ \mu_1^{(-1)} = \mu^{(-1)}(\mu_2 \circ g \circ \mu_2^{(-1)})\mu \in \overline{E}[X],$$

and

$$\mu_1(y_0) = (\mu_1 \circ \mu_2^{(-1)})(y_1) = \mu^{(-1)}(y_1) \in \overline{E},$$

as desired. □

CLAIM 11.2

If  $\mathcal{O}_f(x_0) \cap \mathcal{O}_g(y_0)$  is infinite, then there exist subfields  $E \subset F \subset K$  such that  $F$  is a function field of transcendence degree 1 over  $E$ , and there exists a linear polynomial  $\mu \in \overline{K}[X]$  such that  $\mu \circ f \circ \mu^{(-1)}, \mu \circ g \circ \mu^{(-1)} \in \overline{F}[X]$ , and  $\mu(x_0), \mu(y_0) \in \overline{F}$ , and either  $(f, x_0)$  or  $(g, y_0)$  is not isotrivial over  $E$ .

*Proof of Claim 11.2*

Let  $K_0$  be a finitely generated subfield of  $K$  such that  $f, g \in K_0[X]$  and  $x_0, y_0 \in K_0$ . Then there exists a finite tower of field subextensions:

$$K_s \subset K_{s-1} \subset \dots \subset K_1 \subset K_0$$

such that  $K_s$  is a number field, and for each  $i = 0, \dots, s - 1$ , the extension  $K_i/K_{i+1}$  is finitely generated of transcendence degree 1. Using Claim 11.1 and the hypotheses of Theorem 8.1, we conclude that there exists  $i = 0, \dots, s - 1$ , and there exists a linear  $\mu \in \overline{K_0}[X]$  such that  $\mu \circ f \circ \mu^{(-1)}, \mu \circ g \circ \mu^{(-1)} \in \overline{K_i}[X]$ , and  $\mu(x_0), \mu(y_0) \in \overline{K_i}$ , and either  $(f, x_0)$  or  $(g, y_0)$  is not isotrivial over  $K_{i+1}$ . □

*Proof of Theorem 8.1*

Let  $E, F$ , and  $\mu$  be as in the conclusion of Claim 11.2. At the expense of replacing  $f$  and  $g$  with their respective conjugates by  $\mu$ , and at the expense of replacing  $F$  by a finite extension, we may assume that  $f, g \in F[X]$  and  $x_0, y_0 \in F$ , and  $(f, x_0)$  is not isotrivial over  $E$ .

Let  $d_1 := \deg(f)$  and  $d_2 := \deg(g)$ . We construct the canonical heights  $\widehat{h}_f$  and  $\widehat{h}_g$  associated to the polynomials  $f$  and  $g$ , with respect to the set of absolute values associated to the function field  $F/E$ . Because  $(f, x_0)$  is non-isotrivial and because  $x_0$  is not preperiodic for  $f$  (note that  $\mathcal{O}_f(x_0) \cap \mathcal{O}_g(y_0)$  is infinite), Lemma 10.7 yields  $H_1 := \widehat{h}_f(x_0) > 0$ . Moreover, if  $H_2 := \widehat{h}_g(y_0)$ , then using Lemma 10.3, we have that  $H_1, H_2 \in \mathbb{Q}$ . Because there exist infinitely many pairs  $(m, n) \in \mathbb{N} \times \mathbb{N}$  such that  $f^{(m)}(x_0) = g^{(n)}(y_0)$ , Proposition 9.2(a)–(b) yields that

$$|d_1^m \cdot H_1 - d_2^n \cdot H_2| \text{ is bounded} \tag{11.3}$$

for infinitely many pairs  $(m, n) \in \mathbb{N} \times \mathbb{N}$ . Because  $H_1, H_2 \in \mathbb{Q}$ , we conclude that there exist *finitely* many rational numbers  $\gamma_1, \dots, \gamma_s$  such that

$$\gamma_i = d_1^m \cdot H_1 - d_2^n \cdot H_2$$

for each pair  $(m, n)$  as in (11.3). (We are using the fact that there are finitely many rational numbers of bounded denominator and bounded absolute value.) Therefore, there exists a rational number  $\gamma := \gamma_i$  (for some  $i = 1, \dots, s$ ) such that

$$d_1^m H_1 - d_2^n H_2 = \gamma \tag{11.4}$$

for infinitely many pairs  $(m, n) \in \mathbb{N} \times \mathbb{N}$ . Hence, the line  $L \subset \mathbb{A}^2$  given by the equation  $H_1 \cdot X - H_2 \cdot Y = \gamma$  has infinitely many points in common with the rank-2 subgroup  $\Gamma := \{(d_1^{k_1}, d_2^{k_2}) : k_1, k_2 \in \mathbb{Z}\}$  of  $\mathbb{G}_m^2$ . Using Corollary 2.2, we obtain that  $\gamma = 0$ . Because there are infinitely many pairs  $(m, n)$  satisfying (11.4) and because  $H_1 \neq 0$ , we conclude that there exist positive integers  $m_0$  and  $n_0$  such that  $d_1^{m_0} = d_2^{n_0}$ ; thus,  $\deg(f^{(m_0)}) = \deg(g^{(n_0)})$ . Because  $\mathcal{O}_f(x_0) \cap \mathcal{O}_g(y_0)$  is infinite, we can find  $k_0, \ell_0 \in \mathbb{N}$  such that  $\mathcal{O}_{f^{(m_0)}}(f^{(k_0)}(x_0)) \cap \mathcal{O}_{g^{(n_0)}}(g^{(\ell_0)}(y_0))$  is infinite. Because  $\deg(f^{(m_0)}) = \deg(g^{(n_0)})$ , we can apply [18, Theorem 1.1] and conclude the proof of Theorem 8.1.  $\square$

*Remark*

Theorem 8.1 holds essentially by the same argument as given above, if  $x_0$  is not in the  $v$ -adic filled Julia set of  $f$ , where  $v$  is any place of a function field  $K$  over a field  $E$ .

*Remark*

One can show that if  $f$  is a linear polynomial and  $g$  is any non-isotrivial polynomial of degree larger than one, then  $\mathcal{O}_f(x_0) \cap \mathcal{O}_g(y_0)$  is finite. This assertion fails if  $g$  is isotrivial, as shown by the infinite intersection  $\mathcal{O}_{X+1}(0) \cap \mathcal{O}_{X^2}(2)$ .

**12. The dynamical Mordell–Lang problem**

In this section we discuss topics related to Question 1.6. We begin with examples in which this question has a negative answer. We then show how this question can be reformulated in terms of the existence of invariant subvarieties.

*12.1. Examples*

There are several situations where Question 1.6 has a negative answer. For instance, let  $S$  be the subsemigroup of  $\text{End } \mathbb{A}^2$  generated by the endomorphisms  $\Phi(x, y) = (2x, y)$  and  $\Psi(x, y) = (x, y^2)$ . If  $\Delta$  is the diagonal subvariety of  $\mathbb{A}^2$ , then  $\Delta(\mathbb{C}) \cap \mathcal{O}_S((1, 2)) = \{\Phi^{2^n} \Psi^n((1, 2)) : n \in \mathbb{N}_0\}$ , which yields a negative answer to Question 1.6. A similar example occurs for  $X = E \times E$  with  $E$  any commutative algebraic group, where  $\Phi(P, Q) = (P + P_0, Q)$  and  $\Psi(P, Q) = (P, 2Q)$  with  $P_0 \in E(\mathbb{C})$  a nontorsion point: letting  $\Delta$  be the diagonal in  $E \times E$ , and  $S$  the semigroup generated by  $\Phi$  and  $\Psi$ , we have  $\Delta(\mathbb{C}) \cap \mathcal{O}_S((0, P_0)) = \{\Phi^{2^n} \Psi^n((0, P_0)) : n \in \mathbb{N}_0\}$  (where zero is the identity element of the group  $E(\mathbb{C})$ ). One can produce similar examples in which  $S$  contains infinite-order elements which restrict to automorphisms on some positive-dimensional subvariety of  $X$ . However, as noted in Section 1, there is an important situation where  $S$  consists of automorphisms but Question 1.6 has an affirmative answer, namely, when  $S$  consists of translations on a semiabelian variety  $X$ .

## 12.2. Invariant subvarieties

The following result connects Question 1.6 with the topic of invariant subvarieties.

### LEMMA 12.1

If Question 1.6 has an affirmative answer for some  $X$ ,  $V$ ,  $S$ , and  $\alpha$ , then the Zariski closure of  $V(\mathbb{C}) \cap \mathcal{O}_S(\alpha)$  consists of finitely many points and finitely many positive-dimensional subvarieties  $V_0 \subset V$ , where for each  $V_0$  there is an infinite subsemigroup  $S_0$  of  $S$  such that  $S_0(V_0) \subset V_0$ .

### Proof

The intersection  $V(\mathbb{C}) \cap \mathcal{O}_S(\alpha)$  is the union of finitely many sets of the form  $T_0 := \mathcal{O}_{S_0, \Phi}(\alpha)$ , with  $\Phi \in S$  and  $S_0$  a subsemigroup of  $S$ . For any such  $T_0$ , let  $V_0$  be the Zariski closure of  $T_0$ , so  $V_0 \subset V$  and  $S_0(V_0) \subset V_0$  (since  $S_0(T_0) \subset T_0$ ). The result follows.  $\square$

Note that, if  $\Phi$  is a closed morphism of  $X$  (as occurs, for instance, when  $\Phi$  is polarizable), and  $\Phi(V_0) \subset V_0$  for some subvariety  $V_0$  of  $X$ , then  $V_0 \supset \Phi(V_0) \supset \Phi^2(V_0) \supset \dots$  is a descending chain of closed subvarieties of  $X$ , so  $\Phi^{N+1}(V_0) = \Phi^N(V_0)$  for some  $N \in \mathbb{N}_0$ . Thus, in particular,  $V_0$  is  $\Phi$ -preperiodic.

Conversely, for any  $X$ ,  $V$ ,  $S$ , and  $\alpha$  which satisfy the conclusion of Lemma 12.1, if each  $S_0$  has finite index in  $S$  (as happens, for instance, when  $S$  is cyclic), then Question 1.6(a) has an affirmative answer for these choices of  $X$ ,  $V$ ,  $S$ , and  $\alpha$ . We do not know whether this implication holds when  $S_0$  has infinite index in  $S$ .

Finally, we remark that Medvedev and Scanlon [26] have classified the subvarieties  $V$  of  $\mathbb{A}^d$  such that  $(f_1, \dots, f_d)(V) \subset V$  for some  $f_i \in \mathbb{C}[X]$  of degree at least 2.

*Acknowledgments.* The authors thank the referees for their valuable comments.

## References

- [1] E. AMERIK, F. BOGOMOLOV, and M. ROVINSKY, *Remarks on endomorphisms and rational points*, *Compos. Math.* **147** (2011), 1819–1842. (1382)
- [2] J. P. BELL, *A generalised Skolem–Mahler–Lech theorem for affine varieties*, *J. London Math. Soc.* (2) **73** (2006), 367–379; *Correction*, *J. London Math. Soc.* (2) **78** (2008), 267–272. MR 2225492; MR 2427064 (1381)
- [3] J. P. BELL, D. GHIOCA, and T. J. TUCKER, *The dynamical Mordell–Lang problem étale maps*, *Amer. J. Math.* **132** (2010), 1655–1675. MR 2766180 (1381, 1383)
- [4] R. BENEDETTO, *Heights and preperiodic points of polynomials over function fields*, *Int. Math. Res. Not.* **62** (2005), 3855–3866. MR 2202175 (1405)
- [5] R. L. BENEDETTO, D. GHIOCA, B. HUTZ, P. KURLBERG, T. SCANLON, and T. J. TUCKER, *Periods of rational maps modulo primes*, to appear in *Math. Ann.*, preprint, arXiv:1107.2816v1 [math.AG] (1381, 1382)

- [6] R. L. BENEDETTO, D. GHIOCA, P. KURLBERG, and T. J. TUCKER, *A case of the dynamical Mordell–Lang conjecture*, *Math. Ann.* **352** (2012), 1–26. [MR 2885573](#) (1381)
- [7] Y. F. BILU and R. F. TICHY, *The Diophantine equation  $f(x) = g(y)$* , *Acta Arith.* **95** (2000), 261–288. [MR 1793164](#) (1383, 1385, 1386, 1388)
- [8] P. BLANCHARD, *Complex analytic dynamics of the Riemann sphere*, *Bull. Amer. Math. Soc.* **11** (1984), 85–141. [MR 0741725](#) (1379)
- [9] G. S. CALL and S. GOLDSTINE, *Canonical heights on projective space*, *J. Number Theory* **63** (1997), 211–243. [MR 1443758](#) (1403)
- [10] G. S. CALL and J. H. SILVERMAN, *Canonical heights on varieties with morphisms*, *Compositio Math.* **89** (1993), 163–205. [MR 1255693](#) (1402, 1403)
- [11] S. D. CUTKOSKY and V. SRINIVAS, *On a problem of Zariski on dimensions of linear systems*, *Ann. of Math. (2)* **137** (1993), 531–559. [MR 1217347](#) (1383)
- [12] L. DENIS, *Géométrie et suites récurrentes*, *Bull. Soc. Math. France* **122** (1994), 13–27. [MR 1259107](#) (1381)
- [13] H. T. ENGSTROM, *Polynomial substitutions*, *Amer. J. Math.* **63** (1941), 249–255. [MR 0003599](#) (1389)
- [14] G. FALTINGS, “The general case of S. Lang’s theorem” in *Barsotti Symposium in Algebraic Geometry (Abano Terme, 1991)*, *Perspect. Math.* **15**, Academic Press, San Diego, 1994, 175–182. [MR 1307396](#) (1381)
- [15] M. FRIED, *On a theorem of Ritt and related Diophantine problems*, *J. Reine Angew. Math.* **264** (1973), 40–55. [MR 0337915](#) (1386)
- [16] D. GHIOCA and T. J. TUCKER, *Equidistribution and integral points for Drinfeld modules*, *Trans. Amer. Math. Soc.* **360** (2008), 4863–4887. [MR 2403707](#) (1404)
- [17] ———, *Periodic points, linearizing maps, and the dynamical Mordell–Lang problem*, *J. Number Theory* **129** (2009), 1392–1403. [MR 2521481](#) (1381)
- [18] D. GHIOCA, T. J. TUCKER, and M. E. ZIEVE, *Intersections of polynomial orbits, and a dynamical Mordell–Lang theorem*, *Invent. Math.* **171** (2008), 463–483. [MR 2367026](#) (1380, 1381, 1382, 1383, 1392, 1402, 1405, 1407)
- [19] ———, *The Mordell–Lang question for endomorphisms of semiabelian varieties*, *J. Théor. Nombres Bordeaux* **23** (2011), 645–666. (1382)
- [20] R. M. GURALNICK, J. E. ROSENBERG, and M. E. ZIEVE, *A new family of exceptional polynomials in characteristic two*, *Ann. of Math. (2)* **172** (2010), 1367–1396. [MR 2680493](#) (1385)
- [21] D. S. KEELER, D. ROGALSKI, and T. J. STAFFORD, *Naïve noncommutative blowing up*, *Duke Math. J.* **126** (2005), 491–546. [MR 2120116](#) (1383)
- [22] S. LANG, *Integral points on curves*, *Inst. Hautes Études Sci. Publ. Math.* **6** (1960), 27–43. [MR 0130219](#) (1385)
- [23] ———, *Fundamentals of Diophantine Geometry*, Springer, New York, 1983. [MR 0715605](#) (1384, 1402, 1403)
- [24] C. LECH, *A note on recurring series*, *Ark. Mat.* **2** (1953), 417–421. [MR 0056634](#) (1381)
- [25] K. MAHLER, *Eine arithmetische Eigenschaft der Taylor-Koeffizienten rationaler Funktionen*, *Proc. Kon. Ned. Akad. Wetensch.* **38** (1935), 50–60. (1381)

- [26] A. MEDVEDEV and T. SCANLON, *Polynomial dynamics*, preprint, arXiv:0901.2352v2 [math.DS] (1408)
- [27] M. RAYNAUD, *Courbes sur une variété abélienne et points de torsion*, Invent. Math. **71** (1983), 207–233. MR 0688265 (1382)
- [28] ———, “Sous-variétés d’une variété abélienne et points de torsion” in *Arithmetic and Geometry, I*, Progr. Math. **35**, Birkhäuser, Boston, 1983, 327–352. MR 0717600 (1382)
- [29] J. F. RITT, *On the iteration of rational functions*, Trans. Amer. Math. Soc. **21** (1920), 348–356. MR 1501149 (1380, 1391)
- [30] ———, *Prime and composite polynomials*, Trans. Amer. Math. Soc. **23** (1922), 51–66. MR 1501189 (1389)
- [31] A. SCHINZEL, *Selected Topics on Polynomials*, University of Michigan Press, Ann Arbor, Mich., 1982. MR 0649775 (1386)
- [32] C. L. SIEGEL, *Über einige Anwendungen Diophantischer Approximationen*, Abh. Preuss. Akad. Wiss. Phys. Math. Kl. (1929), 41–69. (Reprinted as *Gesammelte Abhandlungen I*, Springer, Berlin, 1966, 209–266.) (1384)
- [33] T. SKOLEM, *Ein Verfahren zur Behandlung gewisser exponentialer Gleichungen und diophantischer Gleichungen*, Comptes rendus du 8e congrès des mathématiciens scandinaves, (1935), 163–188. (1381)
- [34] E. ULLMO, *Positivité et discrétion des points algébriques des courbes*, Ann. of Math. (2) **147** (1998), 167–179. MR 1609514 (1382)
- [35] P. VOJTA, *Integral points on subvarieties of semiabelian varieties. I*, Invent. Math. **126** (1996), 133–181. MR 1408559 (1381)
- [36] S. ZHANG, *Equidistribution of small points on abelian varieties*, Ann. of Math. (2) **147** (1998), 159–165. MR 1609518 (1382)
- [37] ———, “Distributions in algebraic dynamics” in *Surveys in Differential Geometry*, Vol. X, International Press, Boston, 2006, 381–430. MR 2408228 (1382)
- [38] M. E. ZIEVE and P. MÜLLER, *On Ritt’s polynomial decomposition theorems*, preprint, arXiv:0807.3578v1 [math.AG] (1383, 1388, 1389, 1390)

### Ghioca

Department of Mathematics, University of British Columbia, Vancouver, British Columbia V6T 1Z2, Canada; [dghioca@math.ubc.ca](mailto:dghioca@math.ubc.ca)

### Tucker

Department of Mathematics, University of Rochester, Rochester, New York 14627, USA; [ttucker@math.rochester.edu](mailto:ttucker@math.rochester.edu)

### Zieve

Department of Mathematics, University of Michigan, Ann Arbor, Michigan 48109, USA; [zieve@umich.edu](mailto:zieve@umich.edu)