Counting points of bounded height in monoid orbits

Wade Hindes

Received: 10 June 2021 / Accepted: 5 March 2022 / Published online: 13 April 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract
Given a set of endomorphisms on $\mathbb{P}^N$, we establish an upper bound on the number of points of bounded height in the associated monoid orbits. Moreover, we give a more refined estimate with an associated lower bound when the monoid is free. Finally, we show that most sets of rational functions in one variable satisfy these more refined bounds.

Mathematics Subject Classification Primary 37P15 · 37P05 · Secondary 11G50 · 11D45

1 Introduction
Let $H$ be the absolute multiplicative Weil height on $\mathbb{P}^N(\overline{\mathbb{Q}})$ and let $K$ be a number field. Then given a subset $X \subseteq \mathbb{P}^N(\overline{\mathbb{Q}})$ of interest in some context, the growth rate of the number of $K$-points in $X$ of bounded height,

$$X(K, B) := \# \{ Q \in X \cap \mathbb{P}^N(K) : H(Q) \leq B \},$$

is known to encode interesting invariants of $X$ and $K$. For instance, if $X = \mathbb{P}^N(K)$, then $X(K, B) \sim C_{K,N}B^{(N+1)[K:Q]}$ where $C_{K,N}$ depends on the regulator, class group, etc. of $K$. If $X$ is an abelian variety, then $X(K, B) \sim C_{K,X} \log(B)^{r/2}$ where $r$ is the rank of the Mordell–Weil group $X(K)$. If $X$ is a smooth curve of genus at least 2, then $X(K, B) \sim C_{K,X}$. More generally, if $X$ is a thin set, i.e., a proper Zariski closed subset or the image of some generically finite morphism of degree at least two, then Theorem 3 in [27, §13.1] implies that

$$X(K, B) \ll B^{(N+1)/2}[K:Q] \log(B).$$

(1)

This problem is also an active topic of research in arithmetic dynamics, where orbits play the role of $X$; see [2, 15, 18, 31, 32] for examples on Markoff varieties, K3 surfaces, and projective space. In this setting, the growth of heights in orbits provides a coarse tool for understanding the dynamical system itself. For example, suppose that $\phi$ is a dominant rational self-map of $\mathbb{P}^N$ with dynamical degree $\delta_\phi$, a classical invariant measuring the geometric complexity...
of the system. Then, for points $P \in \mathbb{P}^N(K)$ such that the orbit $\text{Orb}_\phi(P) = \{ \phi^n(P) \}_{n \geq 0}$ is Zariski dense, the Kawaguchi–Silverman Conjecture predicts that

$$\# \{ Q \in \text{Orb}_\phi(P) : H(Q) \leq B \} \sim \log(\delta_\phi)^{-1} \log \log(B);$$

(2)

see [18] for the relevant definitions and background and see [22, 23, 29] for recent progress on this problem. Likewise there is work on this topic for random dynamical systems [15]: if $S = \{ \phi_1, \ldots, \phi_s \}$ is a set of endomorphisms equipped with a probability measure $\nu$, then for almost every sequence $\gamma$ of elements of $S$, we have the analogous asymptotic to (2):

$$\# \{ Q \in \text{Orb}_\gamma(P) : H(Q) \leq B \} \sim \log(\delta_{S,\nu})^{-1} \log \log(B), \quad \delta_{S,\nu} = \prod_{\phi \in S} \text{deg}(\phi)^{\nu(\phi)}.$$  

(3)

In this paper, we study the problem of counting points of bounded height in monoid (or semigroup) orbits in $\mathbb{P}^N$, that is, counting all of the points of bounded height obtained by applying all possible compositions of maps within a fixed set $S$ to a given initial point $P$; compare to [2, 32]. Intuitively, one expects that if the maps in $S$ are related in some way (for instance, if they commute), then this should cut down the number of possible points in the associated orbits. However, for most $S$ we expect to see no relations (free monoids), and with this in mind, we have the following result; here and throughout, $M_S$ denotes the monoid generated under composition by a set $S$ of endomorphisms of $\mathbb{P}^N$ defined over $\overline{\mathbb{Q}}$.

**Theorem 1.1** Let $S = \{ \phi_1, \ldots, \phi_s \}$ be a set of endomorphisms on $\mathbb{P}^N(\overline{\mathbb{Q}})$ with distinct degrees all at least two. If $M_S$ is free, then for all $\epsilon > 0$ there exists an effectively computable positive constant $b = b(S, \epsilon)$ and a constant $B_S$ depending only on $S$ such that

$$(\log B)^b \ll \# \{ f \in M_S : H(f(P)) \leq B \} \ll (\log B)^{b+\epsilon}$$

holds for all $P \in \mathbb{P}^N(\overline{\mathbb{Q}})$ with $H(P) > B_S$. Moreover, the implicit constants and error terms depend on $P$ and are effectively computable if $B_S$ is.

**Remark 1** When $S = \{ \phi_1, \phi_2 \}$ generates a free monoid with $\deg(\phi_1) = 2$ and $\deg(\phi_2) = 3$, then we give explicit computations for the bounds in Theorem 1.1 in Example 1 below.

In particular, we can use the upper bound in Theorem 1.1 on the number of functions in the free case to give an upper bound on the number of points of bounded height in arbitrary dynamical orbits; compare to (1), to [2, Theorem 4.15], and to the asymptotic for abelian varieties above. In what follows, $\text{Orb}_S(P) = \{ f(P) : f \in M_S \}$ denotes the total orbit of $P$ under the monoid $M_S$.

**Corollary 1.2** Let $S = \{ \phi_1, \ldots, \phi_s \}$ be a set of endomorphisms on $\mathbb{P}^N(\overline{\mathbb{Q}})$ all of degree at least two (and distinct if $s \geq 2$). Then there exists an effectively computable positive constant $b$ and a constant $B_S$ depending only on $S$ such that

$$\# \{ Q \in \text{Orb}_S(P) : H(Q) \leq B \} \ll (\log B)^b$$

holds for all $P \in \mathbb{P}^N(\overline{\mathbb{Q}})$ with $H(P) > B_S$.

**Remark 2** Although we expect that $(\log B)^b$ is also a lower bound for some choice of $b$ and most $S$ (see Conjecture 1.3 and Theorem 1.6 below), we note that it is only an upper bound in general, even for $s \geq 2$. For instance, if $M_S$ is a free commutative monoid (e.g., if $S$ is a certain set of monic power maps), then the asymptotic height growth rate in orbits will be a constant times $(\log \log(B))^s$; see [15, §5] for details. This matches the case of a single map (also a commutative monoid); see also (2) and (3) above.
Motivated by the upper and lower bounds in Theorem 1.1, we conjecture the following exact asymptotic for the number of points (not functions) of bounded height in total orbits associated to free monoids:

**Conjecture 1.3** Let $S = \{\phi_1, \ldots, \phi_s\}$ be a set of endomorphisms on $\mathbb{P}^N(\overline{\mathbb{Q}})$ with distinct degrees all at least two. If $M_S$ is free, then there exist constants $a_P = a(S, P)$ and $b = b(S)$ such that

$$\lim_{B \to \infty} \frac{\#\{Q \in \text{Orb}_S(P) : H(Q) \leq B}\} {\log B}^b = a_P$$

holds for all sufficiently generic $P \in \mathbb{P}^N(\overline{\mathbb{Q}})$ (i.e., all $P \in \mathbb{P}^N(\overline{\mathbb{Q}})$ outside of the union of a proper Zariski closed subset and a set of points of bounded height).

**Remark 3** Hence, we expect most monoid orbits in $\mathbb{P}^N$ to exhibit similar height growth as: orbits on Markoff varieties [32], orbits on $K^3$ surfaces in $\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$ given by $(2, 2, 2)$-forms [2, Theorem 4.5], and Mordell–Weil groups of abelian varieties. However in these cases, the relevant monoids (or the underlying varieties themselves) form groups, and there is less need to distinguish between counting functions and points. For instance if there are inverses in $M_S$, distinct functions that agree at a point determine a non-trivial fixed point, and these fixed points can typically be controlled. On the other hand in the case of abelian varieties (where one considers the monoid generated by multiplication maps), distinct functions that agree at a point determine a torsion point. Thus this situation may be avoided by throwing away a set of bounded height.

To motivate our conjecture, we restrict our attention to morphisms of $\mathbb{P}^1$. To state our results in this setting, recall that $w \in \mathbb{P}^1(\mathbb{C})$ is called a *critical value* of $\phi \in \mathbb{C}(x)$ if $\phi^{-1}(w)$ contains fewer than $\deg(\phi)$ elements. Likewise, we call a critical value $w$ of $\phi$ *simple* if $\phi^{-1}(w)$ contains exactly $\deg(\phi) - 1$ points. In particular, we have the corresponding notions for sets:

**Definition 1.4** Let $S = \{\phi_1, \ldots, \phi_s\}$ be a set of rational maps on $\mathbb{P}^1$ and let $C_{\phi_i}$ denote the set of critical values of $\phi_i$. Then $S$ is called *critically separate* if $C_{\phi_i} \cap C_{\phi_j} = \emptyset$ for all $i \neq j$. Moreover, $S$ is called *critically simple* if every critical value of every $\phi \in S$ is simple.

As evidence for Conjecture 1.3 above, we establish the following weak version for generic sets of rational maps on $\mathbb{P}^1$; see Theorems 1.3 and 1.4 of [25] for a justification of generic. In particular, we are able to count points instead of just functions.

**Theorem 1.5** Let $S = \{\phi_1, \ldots, \phi_s\}$ be a set of rational maps on $\mathbb{P}^1(\overline{\mathbb{Q}})$ with distinct degrees all at least four. If $S$ is critically separate and critically simple, then $M_S$ is a free monoid and for all $\epsilon > 0$ there exists an effectively computable positive constant $b = b(S, \epsilon)$ and a constant $B_S$ depending only on $S$ such that

$$\left\langle \frac{\log B}{b} \right\rangle \ll \#\{Q \in \text{Orb}_S(P) : H(Q) \leq B\} \ll \left\langle \frac{\log B}{b+\epsilon} \right\rangle$$

holds for all $P \in \mathbb{P}^1(\overline{\mathbb{Q}})$ with $H(P) > B_S$.

Finally, since the theorem above does not directly apply to sets of polynomials, we give a different proof in this case which works quite generically. In what follows, for $m \geq 2$ the polynomials $Z_m = x^m$ are called *cyclic* polynomials and the polynomials $T_m$ satisfying $T_m(x + x^{-1}) = x^m + x^{-m}$ are called *Chebychev* polynomials (of the first kind).
**Theorem 1.6** Let $S = \{\phi_1, \ldots, \phi_s\}$ be a set of polynomials defined over $\overline{\mathbb{Q}}$ with distinct degrees. Moreover, assume the following conditions:

1. Each $\phi_i \in S$ is not of the form $R \circ E \circ L$ for some polynomial $R \in \overline{\mathbb{Q}}[x]$, some cyclic or Chebychev polynomial $E$, and some linear $L \in \overline{\mathbb{Q}}[x]$.
2. For all $i \neq j$, we have that $\phi_j \neq \phi_i \circ P$ for all non-linear $P \in \overline{\mathbb{Q}}[x]$.

Then $M_S$ is a free monoid and for all $\epsilon > 0$ there exists an effectively computable positive constant $b = b(S, \epsilon)$ and a constant $B_S$ depending only on $S$ such that

$$(\log B)^b \ll \# \{ Q \in \text{Orb}_S(P) : H(Q) \leq B \} \ll (\log B)^{b+\epsilon}$$

holds for all $P \in \mathbb{P}^1(\overline{\mathbb{Q}})$ with $H(P) > B_S$.

**Remark 4** The strength of our estimates in this paper rely greatly on the fact that the semigroups we consider are free. Moreover, it can be difficult to verify this freeness in practice. Nevertheless, there is a long history of proving the freeness of semigroups and groups [7] as well as recent progress for the type of semigroups that we consider here [3, 16, 25].

We briefly outline the proofs of our main results. Assuming $M_S$ is free, the first step (in all cases and in all dimensions) is to count the number of functions yielding a bounded height relation, $f \in M_S$ with $H(f(P)) \leq B$. To do this, we replace such functions (via Tate’s telescoping Lemma 2.2) with functions of bounded log-degree. We then approximate the number of functions of bounded log-degree by the number of some associated integer compositions; see Sect. 2 for details. In particular, by analyzing the poles of known generating functions for these integer compositions, we prove Theorem 1.1.

The next step is to show that $M_S$ is free, at least generically. Here, we restrict our attention to dimension one. For rational functions, this follows from the genus calculations in [25] and Picard’s theorem (precluding the existence of non-constant meromorphic maps from the complex plane to curves of genus at least 2). For polynomials, conditions (1) and (2) of Theorem 1.6 and the integral point classification theorems in [4] and [33] together imply that $M_S$ is free. Lastly, to pass from counting functions to counting points, we prove that the number of functions evaluating $P$ to a given point $Q$,

$$\# \{ f \in M_S : f(P) = Q \},$$

is uniformly bounded; see Lemma 4.8 below. Here we use Faltings’ Theorem and the aforementioned integral point theorems. In particular, we deduce the desired estimate on the number of points of bounded height in orbits from Theorem 1.1 and the bound on (4).

### 2 Auxiliary results

Before we count functions yielding a bounded height relation (as in Theorem 1.1), we first recall some basic facts about heights and generating functions. To help motivate these facts, we remind the reader of our overall strategy to count functions. The main idea, consistent with earlier work on orbits attached to sequences in [12, 13, 18], is that the logarithmic height of a point $f(P) \in \text{Orb}_S(P)$ is roughly determined by the size of $\text{deg}(f)$, as long as the initial point $P$ is sufficiently generic; see Lemma 2.2 below. With this in mind, to count functions $f \in M_S$ with $\log H(f(P)) \leq B$, we should in some sense simply be counting the number of $f$’s of bounded degree. In particular, when $M_S$ is a free monoid and the degrees of the maps in $S$ are distinct, we can relate the number of $f \in M_S$ with bounded degree to the
number of restricted integer compositions of bounded size, once we approximate $\log \text{deg}(\phi)$ for all $\phi \in S$ by rational numbers. Finally, we use generating functions (and the location of their poles via Lemmas 2.6 and 2.5 below) to estimate the number of restricted integer compositions of bounded size. These facts together imply Theorem 1.1. With this sketch in place, we move on and review some basic facts about heights.

**Remark 5** Since multiplicative heights tend to grow exponentially when evaluating functions, it is convenient to use the logarithmic height $h = \log \circ H$ (instead of $H$) to state certain height estimates in dynamics. However, since height-counting on varieties is usually done with multiplicative heights, we convert back to $H$ at the end of the proof of Theorem 1.1, to be consistent with similar results in the literature.

Suppose that $\phi : \mathbb{P}^N(\overline{\mathbb{Q}}) \to \mathbb{P}^N(\overline{\mathbb{Q}})$ is a morphism defined over $\overline{\mathbb{Q}}$ of degree $d_\phi$. Then it is well known that

$$h(\phi(P)) = d_\phi h(P) + O_\phi(1) \quad \text{for all } P \in \mathbb{P}^N(\overline{\mathbb{Q}}); \quad (5)$$

see, for instance, [30, Theorem 3.11]. With this in mind, we let

$$C(\phi) := \sup_{P \in \mathbb{P}^N(\overline{\mathbb{Q}})} \left| h(\phi(P)) - d_\phi h(P) \right| \quad (6)$$

be the smallest constant needed for the bound in (5). Then, in order to control height growth rates when composing arbitrary elements of a set of endomorphisms, we define the following fundamental notion; compare to [12, 13, 17].

**Definition 2.1** A set $S$ of endomorphisms of $\mathbb{P}^N(\overline{\mathbb{Q}})$ is called *height controlled* if the following properties hold:

1. $d_S := \inf \{d_\phi : \phi \in S\}$ is at least 2.
2. $C_S := \sup \{C(\phi) : \phi \in S\}$ is finite.

**Remark 6** We note first that any finite set of morphisms of degree at least 2 is height controlled. To construct infinite collections, let $T$ be any non-constant set of maps on $\mathbb{P}^1$ and let $S_T = \{\phi \circ x^d : \phi \in T, d \geq 2\}$. Then $S_T$ is height controlled and infinite; a similar construction works for $\mathbb{P}^N$ in any dimension.

**Remark 7** Although the results in this paper are for finite $S$, we include the notion of height controlled sets to motivate future work. For instance, many of the tools used below: canonical heights, generating functions, etc. work perfectly well for infinite sets. However, the generating functions that appear in this case are not rational, which adds some subtlety.

As in the case of iterating a single function, it is Tate’s telescoping Lemma (generalized below) that allows us to transfer information back and forth between heights and degrees; for a proof, see [12, Lemma 2.1].

**Lemma 2.2** Let $S$ be a height controlled set of endomorphisms of $\mathbb{P}^N(\overline{\mathbb{Q}})$, and let $d_S$ and $C_S$ be the corresponding height controlling constants. Then for all $f \in M_S$,

$$\left| \frac{h(f(Q))}{\text{deg}(f)} - h(Q) \right| \leq \frac{C_S}{d_S - 1} \quad \text{for all } Q \in \mathbb{P}^N(\overline{\mathbb{Q}}).$$

\textcopyright Springer
Now that we have a tool to pass from functions yielding a bounded height relation to functions of bounded degree (via Lemma 2.2), we next relate counting functions of bounded degree to counting restricted integer compositions; this is essentially achieved by the fact that \( \log \deg(F \circ G) = \log \deg(F) + \log \deg(G) \) for all endomorphisms \( F \) and \( G \). However, to make this idea precise, we briefly discuss integer compositions, a classical object of study in combinatorics. For more details, see [8, §I.3.1].

Let \( T \subseteq \mathbb{N}_{>0} \) be a collection of positive integers (not necessarily finite). Then a restricted composition of an integer \( n \) with summands in \( T \) (or a \( T \)-restricted composition of \( n \)) is an ordered collection of elements in \( T \) whose sum is \( n \). For instance, \( 5 = 2 + 3 \) and \( 5 = 3 + 2 \) are two different restricted compositions of \( 5 \) for the set \( T = \{2, 3\} \). Given \( n \), let \( f_n^T \) be the number of distinct ways of writing \( n \) as a composition with summands (parts) in \( T \). Then to give an asymptotic for \( f_n^T \), one can try and understand the ordinary generating function \( f^T(z) = \sum_n f_n^T z^n \). In particular, if in addition \( f^T(z) \) is a rational or meromorphic function, then the radius of convergence of the generating function, determined by the poles of \( f^T(z) \), can be used to deduce an asymptotic for \( f_n^T \). Luckily, the generating functions for restricted compositions are particularly simple rational functions; see Proposition I.1 in [8].

**Proposition 2.3** The ordinary generating function of the number of compositions having summands restricted to a set \( T \subseteq \mathbb{N}_{>0} \) is given by

\[
f^T(z) = \frac{1}{1 - \sum_{n \in T} z^n}.
\]

As mentioned above, once we have an expression for \( f^T(z) \) as a rational function, we can use the poles of \( f^T(z) \) to estimate the \( f_n^T \). Specifically, we have the following Theorem, a simple consequence of partial fractions and Newton expansion. In what follows, if \( \mathcal{F}(z) = \sum_n a_n z^n \) is a power series expansion about \( z = 0 \) for a meromorphic function \( \mathcal{F} \), then we use the notation \( [z^n] \mathcal{F}(z) = a_n \) to extract coefficients.

**Theorem 2.4** (Expansion of rational functions) If \( \mathcal{F}(z) \) is a rational function that is analytic at zero and has poles at points \( \alpha_1, \alpha_2, \ldots, \alpha_m \), then its coefficients (as a power series about 0) are a sum of exponential-polynomials: there exist \( m \) polynomials \( \{\Pi_j(x)\}_{j=1}^m \) such that for \( n \) larger than some fixed \( n_0 \),

\[
[z^n] \mathcal{F}(z) = \sum_{j=1}^m \Pi_j(n) \alpha_j^{-n}.
\]

Furthermore, the degree of \( \Pi_j \) is equal to the order of the pole of \( \mathcal{F} \) at \( \alpha_j \) minus one.

In particular, after combining Proposition 2.3 and Theorem 2.4, we see that to obtain an asymptotic formula for the number of integer compositions whose parts are restricted to the set \( \{n_1, \ldots, n_s\} \), we must control the roots of smallest modulus of \( g(z) \).

With this in mind, we have the following elementary proposition.

**Lemma 2.5** Let \( n_1, n_2, \ldots, n_s \) be positive integers satisfying \( \gcd(n_1, n_2, \ldots, n_s) = 1 \). Then the polynomial \( g(z) = 1 - (z^{n_1} + z^{n_2} + \cdots + z^{n_s}) \) has a unique complex root \( \alpha \) of smallest modulus. Moreover, \( \alpha \) is the unique positive real root of \( g \), and \( \alpha \) has multiplicity one.

**Proof** We first show that any positive real root \( \alpha \) of \( g \) is a root of smallest modulus for \( g \) (clearly \( g \) has a positive root by the Intermediate Value Theorem). This is a simple consequence of
Rouché’s Theorem: let \( r < \alpha \), let \( p(z) = -1 - z^{n_1} - \cdots - z^{n_s} \), and let \( q(z) = 2 \). Then for all \(|z| = r\), we have that
\[
|p(z)| = |1 - 1 - z^{n_1} - \cdots - z^{n_s}| \leq 1 + |z|^{n_1} + \cdots + |z|^{n_s} \\
= 1 + r^{n_1} + \cdots + r^{n_s} \\
< 1 + \alpha^{n_1} + \cdots + \alpha^{n_s} = 2 - (1 - \alpha^{n_1} - \cdots - \alpha^{n_s}) = |q(z)|
\]
by the triangle inequality and since \( \alpha \) is a root of \( g \). In particular, \( p \) and \( q \) are holomorphic functions on the disc \( D_r \) of radius \( r \) such that \(|p(z)| < |q(z)|\) on the boundary \( D_r \). Hence, Rouché’s Theorem implies that \( q \) and \( q + p = g \) have the same number of roots inside \( D_r \). Therefore, \( g \) has no complex roots in \( D_r \), and \( \alpha \) is a root of smallest modulus for \( g \). On the other hand, it is clear that \( g \) restricted to the positive real numbers is strictly decreasing. Hence, \( g \) has only one positive real root. Likewise, it is easy to see that \( g'(\alpha) < 0 \) (since \( \alpha \) is positive). Hence, \( \alpha \) must be a root of multiplicity one for \( g \).

We next show that \( \alpha \) is the unique complex root of \( g \) of smallest modulus. This portion of the proof of Lemma 2.5 follows from results and arguments in [8, IV.6], namely the “Daffodil Lemma” [8, IV.1] and the proof of [8, Proposition IV.3] on the commensurability of dominant directions for rational generating functions arising from regular languages. To see this, suppose that \( \zeta = \alpha e^{i\theta} \) is another root of smallest modulus of \( g \). Let \( f(z) = z^{n_1} + \cdots + z^{n_s} \), so that \( \zeta \) satisfies \(|f(\zeta)| = |1| = 1 = f(\alpha) = f(|\zeta|)\). In particular, [8, Lemma IV.1] implies that \( \theta = 2\pi r/p \) for some integers \( 0 \leq r < p \) with \( \gcd(r, p) = 1 \) (when \( r \neq 0 \)). Moreover, \( f \) admits \( p \) as a span; see [8, Definition IV.5]. In particular (since \( f \) admits \( p \) as a span), \( f(z) = z^a h(z^p) \) for some polynomial \( h \) and some non-negative integer \( a \). Note also that \( \gcd(a, p) = 1 \), since \( \gcd(n_1, \ldots, n_s) = 1 \) by assumption. On the other hand,
\[
1 = f(\zeta) = \zeta^a h(\zeta^p) = (\alpha e^{i2\pi r/p})^a h((\alpha e^{i2\pi r/p})^p) = e^{i2\pi ar/p} \alpha^a h(\alpha^p) \\
= e^{i2\pi ar/p} f(\alpha) = e^{i2\pi ar/p}.
\]
Hence, \( ar/p \in \mathbb{Z} \). But this is impossible unless \( r = 0 \), since \( \gcd(ar, p) = 1 \) otherwise. In particular, \( \zeta = \alpha \) and \( \alpha \) is the unique complex root of \( g \) of smallest modulus as claimed. \( \square \)

Lastly, we include a technical result that allows us to approximate the number of bounded compositions whose parts are restricted to the set of non-integers \( \{\log \deg(\phi_1), \ldots, \log \deg(\phi_s)\} \), a task that is equivalent to counting the number of functions in \( M_S \) of bounded degree, by integer compositions whose parts satisfy the gcd condition needed to apply Lemma 2.5.

**Lemma 2.6** Let \( c_1 < c_2 < \cdots < c_s \) be distinct positive real numbers. Then for all \( \delta > 0 \) there exist positive integers \( n_1, \ldots, n_s, m_1, \ldots, m_s \) and \( u \) such that the following conditions hold:

1. \( c_i - \delta \leq \frac{n_i}{u} < c_i < \frac{m_i}{u} \leq c_i + \delta \).
2. \( \gcd(n_1, \ldots, n_s) = 1 = \gcd(m_1, \ldots, m_s) \).

**Remark 8** In particular, we may assume that \( n_1 < \cdots < n_s < m_1 < \cdots < m_s \) by choosing \( \delta \) sufficiently small.

**Proof** Let \( \delta > 0 \). Then the prime number theorem implies that \( \pi(x+\delta x) - \pi(x) \to \infty \) as \( x \to \infty \). In particular, for large \( u \) one can find a prime \( n_i \) in the interval \( [uc_i - \delta u, uc_i] \) and a prime \( m_i \) in the interval \( [c_i u, c_i u + \delta u] \). It is straightforward to check that the \( n_1, \ldots, n_s, m_1, \ldots, m_s \) satisfy the desired conditions when \( \delta \) is sufficiently small. \( \square \)
3 Counting functions

With the necessary background in place, we are ready to prove the bounds on the number of functions \( f \in M_S \) yielding a bounded height relation from the Sect. 1.

**Proof of Theorem 1.1** Let \( S = \{ \phi_1, \ldots, \phi_s \} \) be a finite set of endomorphisms on \( \mathbb{P}^N \) all of degree at least 2, and suppose that the monoid \( M_S \) generated by \( S \) under composition is free. We begin by defining some lengths on \( M_S \), which we then relate to integer compositions. Given any vector \( v = (v_1, \ldots, v_s) \in \mathbb{R}^s \) of positive weights, we define \( l_{S,v}(\phi_i) = v_i \) for \( \phi_i \in S \) and extend \( l_{S,v} \) to all \( f \in M_S \) by:

\[
l_{S,v}(f) = \sum_{j=1}^{n} l_{S,v}(\theta_j), \quad \text{where} \ f = \theta_1 \circ \theta_2 \circ \cdots \circ \theta_n \text{ for some } \theta_j \in S. \tag{7}
\]

**Remark 9** Note that since \( S \) is a free basis of \( M_S \), there is a unique way to write \( f \) as a composition of elements of \( S \). In particular, \( l_{S,v} \) is well-defined. Alternatively, in the non-free case one can define \( l_{S,v}(f) \) by taking an inf over the possible expressions in (7).

On the other hand, since \( M_S \) is a set of functions, there is a natural choice of weighting given by \( c = (c_1, \ldots, c_s) \) where \( c_i = \log \deg(\phi_i) \); moreover, we assume \( c_1 < c_2 < \cdots < c_s \). In particular, it follows from the fact that \( \deg(F \circ G) = \deg(F) \cdot \deg(G) \) for morphisms that

\[
l_{S,c}(f) = \log \deg(f) \quad \text{for all } f \in M_S, \tag{8}
\]

independent of the generating set. However, non-integer weights (like logs of integers) appear sparingly in the literature, and so we approximate the growth rate of \( l_{S,c} \) (which relates to the growth rate of heights in orbits via Tate’s telescoping argument) using integer weights.

To wit, choose positive integers \( n_1, \ldots, n_s, m_1, \ldots, m_s \) and \( u \) depending on \( \delta \) as in Lemma 2.6 and Remark 8. Then it follows by construction that

\[ u^{-1}l_{S,n}(f) \leq l_{S,c}(f) \leq u^{-1}l_{S,m}(f) \]

for all \( f \in M_S \). Hence,

\[ \{ f \in M_S : l_{S,m}(f) \leq uB \} \subseteq \{ f \in M_S : l_{S,c}(f) \leq B \} \subseteq \{ f \in M_S : l_{S,n}(f) \leq uB \} \]

holds for all positive \( B \); here \( n = (n_1, \ldots, n_s) \) and \( m = (m_1, \ldots, m_s) \). Now given a positive integer \( n \) we define

\[ L_n := \# \{ f \in M_S : l_{S,m}(f) = n \} \quad \text{and} \quad U_n := \# \{ f \in M_S : l_{S,n}(f) = n \}. \tag{10} \]

In particular, since \( n \) and \( m \) are integer weight vectors, it follows from (8), (9) and (10) that

\[
\sum_{n=0}^{uB} L_n \leq \# \{ f \in M_S : \log \deg(f) \leq B \} \leq \sum_{n=0}^{uB} U_n. \tag{11}
\]

Here \( [uB] \) denotes the nearest integer to \( uB \). On the other hand, since \( S \) generates \( M_S \) as a free monoid, we can identify \( M_S \) with the set of finite sequences of elements of \( S \) (or any set). Specifically, \( L_n \) (respectively \( U_n \)) represents the number of ways of writing \( n \) as the sum of a sequence of elements in \( \{ m_1, \ldots, m_s \} \) (respectively in \( \{ n_1, \ldots, n_s \} \)). Such sequences have been extensively studied in combinatorics [8, §I.3.1] and are called restricted integer compositions. Specifically, generating functions for these compositions are known; see Proposition 2.3 above. In particular,

\[
L_n = \left[ z^n \right] \frac{1}{1 - (z^{m_1} + \cdots + z^{m_s})} \quad \text{and} \quad U_n = \left[ z^n \right] \frac{1}{1 - (z^{n_1} + \cdots + z^{n_s})}. \tag{12}
\]
As a reminder, \([z^n]\mathcal{F}(z)\) denotes the operation of extracting the coefficient of \(z^n\) in the formal power series \(\mathcal{F}(z) = \sum f_n z^n\); see [8, p. 19]. On the other hand, since \(\gcd(n_1, \ldots, n_s) = 1\) and \(\gcd(m_1, \ldots, m_s) = 1\) by construction, Lemma 2.5 implies that both of the rational functions in (12) have unique poles of smallest modulus (and these poles are positive real numbers of multiplicity one). Let \(\alpha_1, \ldots, \alpha_r\) be the roots of \(g_n(z) = 1 - (z^{n_1} + \cdots + z^{n_s})\) arranged in increasing order of modulus and let \(\beta_1, \ldots, \beta_r\) be the roots of \(g_m(z) = 1 - (z^{m_1} + \cdots + z^{m_s})\) arranged in increasing order of modulus. Then Theorem 2.4 and (12) together imply that

\[
L_n = \kappa_1 \beta_1^{-n} + p_2(n) \beta_2^{-n} + \cdots + p_r(n) \beta_r^{-n} \quad \text{and} \quad U_n = \tau_1 \alpha_1^{-n} + q_2(n) \alpha_2^{-n} + \cdots + q_r(n) \alpha_r^{-n}
\]

for some constants \(\kappa_1\) and \(\tau_1\) and some polynomials \(p_i, q_j \in \mathbb{C}[z]\). Explicitly,

\[
\kappa_1 = \frac{-1}{\beta_1 \ g_m'(\beta_1)} \quad \text{and} \quad \tau_1 = \frac{-1}{\alpha_1 \ g_n'(\alpha_1)}.
\]

Here we use the residue method for extracting partial fraction coefficients and Newton’s expansion; see the proof of [8, Theorem IV.9]. Moreover, the expressions in (13) and (14) hold simultaneously for all \(n > n_0\) for some constant \(n_0 \in \mathbb{N}\). In particular, by summing (13) and using the triangle inequality (for both sums and differences) we see that

\[
\kappa_2 \beta_1^{-m} - \kappa_3 m^{r_3} |\beta_2|^{-m} - \kappa_4 \leq \sum_{n=0}^{m} L_n \quad \text{and} \quad \sum_{n=0}^{m} U_n \leq \tau_2 \alpha_1^{-m} + \tau_3 m^{r_3} |\alpha_2|^{-m} + \tau_4
\]

holds for all \(m\) sufficiently large. Again, in the interest of being as explicit as possible (at least for the main terms), we have that

\[
\kappa_2 = \frac{\kappa_1 (\frac{1}{\beta_1})}{(\frac{1}{\beta_1}) - 1} = \frac{-1}{\beta_1 (1 - \beta_1) \ g_m'(\beta_1)} \quad \text{and} \quad \tau_2 = \frac{\tau_1 (\frac{1}{\alpha_1})}{(\frac{1}{\alpha_1}) - 1} = \frac{-1}{\alpha_1 (1 - \alpha_1) \ g_n'(\alpha_1)},
\]

obtained by summing the corresponding geometric series. Moreover, \(r_3\) (respectively \(r_4\)) is the maximum of the multiplicities of the roots of \(g_m\) (respectively \(g_n\)) minus one. Hence, after taking \(m = \lfloor Bu \rfloor\), combining (11) and (15), and absorbing \(u\) into the relevant constants, we see that

\[
\kappa_5 \ C_1^B - \kappa_6 \ B^{r_3} \ C_2^B - \kappa_4 \leq \# \{ f \in M_S : \log \deg(f) \leq B \} \leq \tau_5 \ C_3^B + \tau_6 \ B^{r_3} \ C_4^B + \tau_4
\]

holds for all \(B\) sufficiently large; here we use also that \(Bu - 1 \leq \lfloor Bu \rfloor \leq Bu + 1\), so that (some) of the relevant constants are given explicitly by

\[
C_1 = \frac{1}{\beta_1^u}, \quad \kappa_5 = \kappa_2 \beta_1 = \frac{1}{(\beta_1 - 1) \ g_m'(\beta_1)}, \quad C_2 = \frac{1}{|\beta_2|^u},
\]

\[
C_3 = \frac{1}{\alpha_1^u}, \quad \tau_5 = \frac{\tau_2}{\alpha_1} = \frac{1}{\alpha_1^2 (\alpha_1 - 1) \ g_n'(\alpha_1)}, \quad C_4 = \frac{1}{|\alpha_2|^u}.
\]

We note in particular that \(C_1 > C_2\) and \(C_3 > C_4\), since \(\beta_1 < |\beta_2|\) and \(\alpha_1 < |\alpha_2|\) by construction. Now suppose that \(P \in \mathbb{P}^N(\mathbb{Q})\) is such that \(h(P) > b_S := C_S/(d_S - 1)\), where
\( C_S \) and \( d_S \) are the constants from Definition 2.1 above. Then, Tate’s telescoping Lemma 2.2 implies that

\[
\deg(f)(h(P) - b_S) \leq h(f(P)) \leq \deg(f)(h(P) + b_S).
\]

Therefore, for all \( B \) we have the subset relations:

\[
\begin{align*}
\left\{ f \in M_S : \log \deg(f) &\leq \log \left( \frac{B}{h(P) + b_S} \right) \right\} \subseteq \left\{ f \in M_S : h(f(P)) \leq B \right\} \\
\subseteq \left\{ f \in M_S : \log \deg(f) &\leq \log \left( \frac{B}{h(P) - b_S} \right) \right\}.
\end{align*}
\]

(19)

In particular, if we replace \( B \) with \( \log(B/(h(P) + B_S)) \) on the left side of (17), replace \( B \) with \( \log(B/(h(P) - B_S)) \) on the right side of (17), and apply the change of base formulas for logarithms, then we deduce from (17) and (19) that

\[
\kappa_5 \left( \frac{B}{h(P) + b_S} \right)^{\log(C_1)} - \kappa_6 \log \left( \frac{B}{h(P) + b_S} \right)^{r_3} \left( \frac{B}{h(P) + b_S} \right)^{\log(C_2)} - \kappa_4
\]

\[
\leq \# \left\{ f \in M_S : h(f(P)) \leq B \right\}
\]

(20)

\[
\leq \tau_5 \left( \frac{B}{h(P) - b_S} \right)^{\log(C_3)} + \tau_6 \log \left( \frac{B}{h(P) - b_S} \right)^{r_4} \left( \frac{B}{h(P) - b_S} \right)^{\log(C_4)} + \tau_4.
\]

holds for all \( B \) sufficiently large and all initial points \( P \) such that \( h(P) > b_S \). Moreover, since most height counting problems on varieties are stated in terms of multiplicative heights, we replace \( B \) with \( \log B \) in (20) to obtain

\[
\left( \frac{\kappa_5}{(h(P) + b_S)^{\log(C_1)}} \right)^{\log(B)} - \left( \frac{\kappa_6}{(h(P) + b_S)^{\log(C_2)}} \right) \log \left( \frac{\log B}{h(P) + b_S} \right)^{r_3} \left( \frac{\log B}{h(P) + b_S} \right)^{\log(C_2)} - \kappa_4
\]

\[
\leq \# \left\{ f \in M_S : H(f(P)) \leq B \right\}
\]

(21)

\[
\leq \left( \frac{\tau_5}{(h(P) - b_S)^{\log(C_3)}} \right)^{\log(B)} + \left( \frac{\tau_6}{(h(P) - b_S)^{\log(C_4)}} \right) \log \left( \frac{\log B}{h(P) - b_S} \right)^{r_4} \left( \frac{\log B}{h(P) - b_S} \right)^{\log(C_4)} + \tau_4.
\]

Hence, after renaming the constants above, we see that there exist positive constants \( a_1(S, P, \delta), a_2(S, P, \delta), b_1(S, \delta), b_2(S, \delta) \) and \( B_S := e^{b_S} \) such that

\[
a_1(\log B)^{b_1} + o(\log B)^{b_1}) \leq \# \left\{ f \in M_S : H(f(P)) \leq B \right\} \leq a_2(\log B)^{b_2} + o(\log B)^{b_2}
\]

(22)

holds for all \( P \in \mathbb{P}^N(\overline{\mathbb{Q}}) \) with \( H(P) \geq B_S \). Moreover, \( b_1 \) and \( b_2 \) depend only on the set \( S \) and \( \delta \), and \( a_1 \) and \( a_2 \) (and the lower order terms) depend on \( S \), \( \delta \) and \( P \). Specifically, (14), (16) and (18) together imply

\[
a_1 = \frac{1}{(\beta_1 - 1) g_m'(\beta_1) \log \left( B_S H(P) \right)^{\log(\beta_1^{-u})}}, \quad b_1 = \log(\beta_1^{-u}),
\]

\[
a_2 = \frac{1}{\alpha_1^2 (\alpha_1 - 1) g_m'(\alpha_1) \log \left( \frac{H(P)}{B_S} \right)^{\log(\alpha_1^{-u})}}, \quad b_2 = \log(\alpha_1^{-u}).
\]

(23)
Moreover, since roots of polynomials can be approximated to any accuracy effectively, \( b_1 \) and \( b_2 \) can be computed effectively (also integers as in Lemma 2.6 can be produced effectively for all \( \delta \)). Therefore, to complete the proof of Theorem 1.1, we need only show that the difference \( b_2 - b_1 > 0 \) can be made arbitrarily small (by letting \( \delta \) go to zero); see (30) below. Then we set \( b = b_1 \) and \( b_2 = b_1 + \epsilon \) to deduce the claim in Theorem 1.1.

To do this, we use the Mean Value Theorem applied to the functions \( f(x) = -g_m(x) \) and \( h(x) = u \log(x) \) on the intervals \([\alpha_1, \beta_1]\). With this in mind, we begin with a few estimates, all of which follow easily from part (1) of Lemma 2.6:

\[
\frac{2\delta}{c_1} < \frac{2\delta u}{n_1} < \frac{2\delta}{c_1 - \delta}, \quad 1 < \frac{m_1}{n_1} < \frac{c_1 + \delta}{c_1 - \delta}, \quad \frac{u}{m_1} < \frac{1}{c_1}.
\]  

(24)

To simplify the expressions that follow, let \( \alpha = \alpha_1 \) and \( \beta = \beta_1 \). Then we set \( \delta = 0 \) can be computed effectively (also integers as in Lemma 2.6 can be produced effectively)

\[
\frac{1}{s} \leq \alpha.
\]  

(25)

In particular, (24) and (25) together imply the following lower bound on the derivative:

\[
f'(\alpha) = m_s \alpha^{m_s - 1} + \cdots + m_1 \alpha^{m_1 - 1} \geq m_1 \alpha^{m_1 - 1} \geq m_1 \left( \frac{1}{s} \right)^{\frac{m_1}{n_1}} \geq m_1 \left( \frac{1}{s} \right)^{\frac{c_1 + \delta}{c_1 - \delta}}.
\]  

(26)

Similarly, (24) and (25) together imply that:

\[
f(\alpha) = \alpha^{m_s - \frac{m_s}{u}} \cdot \alpha^{n_s} + \cdots + \alpha^{m_1 - \frac{m_1}{u}} \cdot \alpha^{n_1} - 1
\geq \alpha^{2\delta u} \cdot \alpha^{n_s} + \cdots + \alpha^{2\delta u} \cdot \alpha^{n_1} - 1
= \alpha^{2\delta u} (\alpha^{n_s} + \cdots + \alpha^{n_1}) - 1
= \alpha^{2\delta u} - 1 \geq \left( \frac{1}{s} \right)^{\frac{2\delta u}{n_1}} - 1 \geq \left( \frac{1}{s} \right)^{\frac{2\delta}{c_1 - \delta}} - 1.
\]

(27)

Here, we use also that \( 0 \leq \frac{m_i}{u} - \frac{n_i}{u} \leq 2\delta \) by construction; see Lemma 2.6 part (1). In particular, we deduce the following key upper bound:

\[
f(\alpha) \leq 1 - \left( \frac{1}{s} \right)^{\frac{2\delta}{c_1 - \delta}}.
\]  

(28)

We are now ready to apply the Mean Value Theorem to \( f(x) \) on \([\alpha, \beta] \). Specifically,

\[
m_1 \left( \frac{1}{s} \right)^{\frac{c_1 + \delta}{c_1 - \delta}} \leq f'(\alpha) = \min_{\alpha \leq x \leq \beta} f'(x) \leq \frac{f(\beta) - f(\alpha)}{\beta - \alpha} = \frac{-f(\alpha)}{\beta - \alpha} \leq 1 - \left( \frac{1}{s} \right)^{\frac{2\delta}{c_1 - \delta}}
\]

follows from (26), (27), and the Mean Value Theorem. Therefore, we have the estimate:

\[
0 \leq \beta - \alpha \leq \frac{1 - \left( \frac{1}{s} \right)^{\frac{2\delta}{c_1 - \delta}}}{m_1 \left( \frac{1}{s} \right)^{\frac{c_1 + \delta}{c_1 - \delta}}}
\]

(28)

Likewise, the Mean Value Theorem for \( h(x) = u \log(x) \) on \([\alpha, \beta] \), (25), and the fact that \( n_1 > 0 \) together yield

\[
0 \leq \frac{h(\beta) - h(\alpha)}{\beta - \alpha} \leq \max_{\alpha \leq x \leq \beta} h'(x) = h'(\alpha) = u \alpha^{-1} \leq su.
\]  

(29)
Hence, after combining (23), (24), (28) and (29), we deduce that
\[
0 \leq b_2 - b_1 = h(\beta) - h(\alpha) \leq s u \cdot 
\]
\[
1 - \left( \frac{1}{2} \right)^{\frac{2S}{C_1^2 - 3}} = \frac{m_1}{\left( \frac{1}{2} \right)^{\frac{2S}{C_1^2 - 3}}} \cdot \frac{1 - \left( \frac{1}{2} \right)^{\frac{2S}{C_1^2 - 3}}}{m_1} \leq \frac{s}{c_1} \cdot \frac{1 - \left( \frac{1}{2} \right)^{\frac{2S}{C_1^2 - 3}}}{\left( \frac{1}{2} \right)^{\frac{2S}{C_1^2 - 3}}} \quad (30)
\]
However, the upper bound in (30) goes to zero as \( \delta \) goes to zero. Therefore, the exponents \( b_1 \) and \( b_2 \) in (22) can be made arbitrarily close.

**Remark 10** If \( S \) has only two maps (\( s = 2 \)), then the trinomials \( g_n(z) = 1 - z^{n1} - z^{n2} \) and \( g_m(z) = 1 - z^{m1} - z^{m2} \) must have non-zero discriminant (in fact, here we need only that \( n_1 \neq n_2 \) and \( m_1 \neq m_2 \), making no assumptions on gcd’s); this fact follows easily from the discriminant formula in [11, Theorem 4]. In particular, \( r_3 \) and \( r_4 \) from (17) and (21) must be zero. Hence, we obtain simpler bounds for the number of functions of bounded degree (hence, also for the number of points of bounded height in orbits). For instance,
\[
\kappa_5 C_1^B - \kappa_6 C_2^B - \kappa_4 \leq \# \{ f \in M_S : \log \deg(f) \leq B \} \leq \tau_5 C_3^B + \tau_6 C_4^B + \tau_4
\]
holds for all \( B \) sufficiently large.

**Example 1** In particular, if \( S = \{ \phi_1, \phi_2 \} \) with \( \deg(\phi_1) = 2 \) and \( \deg(\phi_2) = 3 \), then we use the crude approximations
\[
\frac{79}{115} < \log(2) < \frac{80}{115} \quad \text{and} \quad \frac{126}{115} < \log(3) < \frac{127}{115}
\]
as inputs to Lemma 2.6 to obtain some explicit bounds for Theorem 1.1. Specifically,
\[
\left( \frac{1.46457}{\log(B_S H(P))^{0.78437}} \right) \log(B)^{0.78437} + o(\log(B)^{0.78437})
\]
\[
\leq \# \{ f \in M_S : H(f(P)) \leq B \}
\]
\[
\leq \left( \frac{1.48541}{\log(H(P))^{0.79232}} \right) \log(B)^{0.79232} + o(\log(B)^{0.79232})
\]
holds for all \( P \in \mathbb{P}^N(\mathbb{Q}) \) of sufficiently large height; here we use (18), (21) and Magma [6] to approximate roots of polynomials.

Lastly, we can use the bounds in Theorem 1.1 on the number of functions in free monoids satisfying a bounded height relation to give an upper bound on the number of points of bounded height in arbitrary monoid orbits.

**Proof of Corollary 1.2.** Let \( S = \{ \phi_1, \ldots, \phi_s \} \) be a set of endomorphisms all of degree at least 2. If \( s = 1 \) (i.e., \( S = \{ \phi \} \) contains just one map), then one may use the canonical height [30, §3.4] associated to \( \phi \) to reach the desired bound. Namely, the fact that \( |\hat{h}_\phi - h| \leq c_\phi \) and that \( \hat{h}(\phi^n(P)) = d_\phi^n \hat{h}_\phi(P) \) together imply that
\[
\left\{ n : n \leq \log_{d_\phi} \left( \frac{\log(B) - c_\phi}{\hat{h}_\phi(P)} \right) \right\} \subseteq \left\{ Q \in \text{Orb}(P) : H(Q) \leq B \right\}
\]
\[
\subseteq \left\{ n : n \leq \log_{d_\phi} \left( \frac{\log(B) + c_\phi}{\hat{h}_\phi(P)} \right) \right\}
\]
for all non-preperiodic \( P \). On the other hand, if \( P \) is preperiodic, then \( \text{Orb}_\phi(P) \) is finite. In particular, the number of points with (multiplicative) height at most \( B \) is certainly bounded above by a constant times \( \log \log(B) \ll \log(B) \) as claimed; hence, \( b = 1 \) in this case.

Now assume that \( s \geq 2 \), and let \( F_S \) be the free monoid generated by \( S \) under concatenation. Then, given a word \( w = \theta_1 \ldots \theta_n \in F_S \), we can define an action of \( w \) on \( \mathbb{P}^N(\overline{\mathbb{Q}}) \) via \( w \cdot P = \theta_1 \circ \cdots \circ \theta_n(P) \). Likewise, we define the degree of \( w \) to be \( \deg(\theta_1 \circ \cdots \circ \theta_n) \). In particular, (by counting words of bounded degree) it is straightforward to see that we can replace \( M_S \) with \( F_S \) in the proof of Theorem 1.1 and deduce that

\[
a_1 \log(B)^{b_1} + o(\log(B)^{b_1}) \leq \#\{w \in F_S : H(w \cdot P) \leq B\} \leq a_2 \log(B)^{b_2} + o(\log(B)^{b_2})
\]

for some constants \( a_1(P), a_2(P), b_1 \) and \( b_2 \) (whenever \( H(P) > B_S \), as before); here we can choose any \( \delta \) small enough to separate the logs of distinct degrees:

\[
\left| \log \deg(\phi) - \log \deg(\psi) \right| > \delta
\]

for all distinct \( \phi, \psi \in S \); see Remark 8. In particular, since every point \( Q \in \text{Orb}_S(P) \) is of the form \( Q = w \cdot P \) for some \( w \in F_S \), we have that

\[
\#\{Q \in \text{Orb}_S(P) : H(Q) \leq B\} \leq \#\{w \in F_S : H(w \cdot P) \leq B\} \leq a_2 \log(B)^{b_2} + o(\log(B)^{b_2}).
\]

Therefore, the number of points in \( \text{Orb}_S(P) \) with height at most \( B \) is \( \ll \log(B)^{b_2} \).

**Remark 11** It is likely that the statement and proof of Theorem 1.1 hold for height controlled sets of simultaneously polarizable maps on any projective variety. The main arithmetic ingredient, Tate’s telescoping Lemma 2.2, works perfectly well with this level of generality; see [12, Lemma 2.1]. Moreover, the other components of the proof (generating functions and diophantine approximation of degrees) don’t depend on \( \mathbb{P}^N \).

### 4 Counting points in orbits in dimension one

In this section, we prove Theorems 1.5 and 1.6 on monoid orbits over \( \mathbb{P}^1 \). To do this, we first show that the relevant sets of maps generate free monoids under composition. For critically separate and simple sets of rational maps, this follows directly from the main results of [25].

**Proposition 4.1** Let \( S = \{\phi_1, \ldots, \phi_s\} \) be a set of rational maps on \( \mathbb{P}^1(\mathbb{C}) \) all of degree at least four. If \( S \) is critically separate and critically simple, then \( M_S \) is a free.

**Proof** Suppose that \( f_1 = \theta_1 \circ \cdots \circ \theta_n = \tau_1 \circ \cdots \circ \tau_m = g_1 \) for some \( \theta_i, \tau_j \in S \). Without loss of generality, we may assume that \( n \geq m \). Clearly if \( n = m = 1 \), then \( \theta_1 = \tau_1 \) and there is nothing to prove. Therefore, we may assume that \( n \geq m > 1 \). Write \( f_2 = \theta_2 \circ \cdots \circ \theta_n \) and \( f_3 = \tau_2 \circ \cdots \circ \tau_m \) so that \( \theta_1(f_2) = \tau_1(f_3) \). However, since \( f_2 \) and \( f_3 \) are non-constant and \( S \) is critically separate, [25, Theorem 1.1] implies that \( \theta_1 = \tau_1 \). Likewise since \( S \) is critically simple and \( \deg(\theta_1) \geq 4 \), we see that \( f_2 = g_2 \) by [25, Theorem 1.3]. Repeating the argument above now for \( f_2 \) and \( g_2 \) (instead of \( f_1 \) and \( f_2 \)), we see that \( \theta_2 = \tau_2 \) and \( \theta_3 \circ \cdots \circ \theta_n = \tau_3 \circ \cdots \circ \tau_m \). We can clearly keep going to deduce that \( \theta_i = \tau_i \) for all \( 1 \leq i \leq m \). Finally, by equating degrees given by the original relation \( f_1 = g_1 \), we see that \( \deg(\theta_{m+1}) \cdots \deg(\theta_n) = 1 \), a contradiction unless \( n = m \). This completes the proof that \( M_S \) is free. \( \square \)
To prove that $M_S$ is free for the sets of polynomials in Theorem 1.6, it suffices to show that the curves
\[ C_k : \frac{\phi_k(x) - \phi_k(y)}{x - y} = 0 \quad \text{and} \quad C_{i,j} : \phi_i(x) = \phi_j(y) \quad \text{for } i \neq j \quad (31) \]
have only finitely many integral points over number fields. To do this, we need the integral point classification theorems in [4] and [33]. However, to put these results in context, we first recall the definition of Siegel factors and Siegel’s integral point theorem.

**Definition 4.2** 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 $K$.

The following result explains the relevance of Siegel factors in this context and is one of the most important results in arithmetic geometry; see Theorems 8.2.4 and 8.5.1 in [20].

**Theorem 4.3** (Siegel) Let $R$ be a finitely generated integral domain of characteristic zero, let $K$ be the field of fractions of $R$, and let $\Phi(x, y) \in K[x, y]$. Then there are only finitely many pairs $(x, y) \in R \times R$ for which $\Phi(x, y) = 0$ unless $\Phi(x, y)$ has a Siegel factor over $K$.

**Remark 12** Clearly if $K$ is a number field (viewed inside the complex numbers) and $\Phi(x, y)$ has no Siegel factors over $\mathbb{C}$, then $\Phi(x, y)$ has no Siegel factors over $K$. Therefore, to prove that the equation $\Phi(x, y) = 0$ has only finitely many solutions $(x, y)$ in some ring of $\mathcal{O}$-integers $R \subset K$, it suffices to show that $\Phi(x, y)$ has no Siegel factors over $\mathbb{C}$.

To use Siegel’s integral point theorem to prove that $M_S$ is free for the polynomials in Theorem 1.6, we need the following theorem of Bilu and Tichy [4, Theorem 10.1], which classifies the polynomials $\Phi(x, y) = F(x) - G(y)$ having a Siegel factor.

**Theorem 4.4** For non-constant $F, G \in \mathbb{C}[x]$, if $F(x) - G(y)$ has a Siegel Factor in $\mathbb{C}[x, y]$ then $F = R \circ E_1 \circ L_1$ and $G = R \circ E_2 \circ L_2$, where $R, L_1, L_2 \in \mathbb{C}[x]$ with $\deg(L_1) = \deg(L_2) = 1$ and either $(E_1, E_2)$ or $(E_2, E_1)$ is one of the following pairs (here $m, n \geq 1$ and $p \in \mathbb{C}[x] \setminus \{0\}$):

(a) $(x^m, x^r p(x)^m)$, where $r \in \mathbb{N}$ is coprime to $m$;
(b) $(x^2, (x^2 + 1)p(x)^2)$;
(c) $(T_m, T_n)$ with $\gcd(m, n) = 1$;
(d) $(T_m, -T_n)$ with $\gcd(m, n) > 1$;
(e) $((x^2 - 1)^3, 3x^4 - 4x^3)$.

**Remark 13** Technically, the statement above is a simplified version of [4, Theorem 10.1] taken from [10, Corollary 2.7]. For a more detailed description of the classification of pairs $(F, G)$ such that $F(x) - G(y)$ has a Siegel factor (with the relevant fields of definition taken into account), see [4].

With the necessary background in place, we are ready to prove that the generic sets of polynomials from the introduction generate free monoids under composition.

**Proposition 4.5** Let $S = \{\phi_1, \ldots, \phi_s\}$ be a set of polynomials defined over a number field $K$ with distinct degrees, and let $S$ be a finite set of places of $K$ (containing the archimedean ones) over which the coefficients of each $\phi_i$ are defined. Moreover, assume the following conditions:
(1) Each $\phi_i \in S$ is not of the form $R \circ E \circ L$ for some polynomial $R \in \mathbb{C}[x]$, some cyclic or Chebychev polynomial $E$, and some linear $L \in \mathbb{C}[x]$.

(2) For all $i \neq j$, we have that $\phi_j \neq \phi_i \circ P$ for all non-linear $P \in \mathbb{C}[x]$.

Then, the curves in (31) have at most finitely many $S$-integral points for all $i \neq j$ and all $k$. Furthermore, $M_S$ is a free monoid.

Proof The statement that each $C_k$ has at most finitely many $S$-integral points follows directly from condition (1), Theorem 4.3, and [33, Theorem 1.1]; Zannier has recently shown that such curves have at least 3 points at infinity over $\mathbb{C}$ and thus cannot have a Siegel factor over any number field. As for the curves $C_{i,j}$ for $i \neq j$, we use Theorem 4.3, Remark 12, and Theorem 4.4. Specifically, the pairs (b), (d), and (e) in Theorem 4.4 are ruled out by condition (1) only, simply by examining first coordinates; for case (e), we use also that $(x^2 - 1)^3 = R \circ E$, where $R(x) = (x - 1)^3$ and $E(x) = x^2$ is cyclic. Likewise, case (c) is ruled out by condition (1) unless $m = 1 = n$. However in this case, $\phi_i = R \circ L_1$ and $\phi_j = R \circ L_2$ for some $R \in \mathbb{C}[x]$ and some linear $L_1, L_2 \in \mathbb{C}[x]$ (the middle polynomials are both $T_1(x) = x$). But then $\deg(\phi_i) = \deg(R) = \deg(\phi_j)$, contradicting our assumption that the maps in $S$ have distinct degrees. Finally, it remains to rule out case (a). If case (a) holds, then again condition (1) forces $m = 1$. Hence up to reordering, $\phi_i = R \circ L_1$ and $\phi_j = R \circ E_2 \circ L_2$. But then $\phi_1 = R \circ L_1 \circ L_1^{-1} \circ E_2 \circ L_2 = \phi_i \circ P$, where $P = L_1^{-1} \circ E_2 \circ L_2$. Moreover, $\deg(P) > 1$ since $\deg(\phi_j) \neq \deg(\phi_i)$ by assumption. Therefore, $\phi_j = \phi_i \circ P$ for some non-linear $P \in \mathbb{C}[x]$, contradicting condition (2) of Proposition 4.5.

From here to show that $M_S$ is free, we use a similar argument to Proposition 4.1. Suppose that $f_1 = \theta_1 \circ \cdots \circ \theta_m = \tau_1 \circ \cdots \circ \tau_m = g_1$ for some $\theta_i, \tau_j \in S$. Without loss of generality, we may assume that $n \geq m$. Clearly if $n = m = 1$, then $\theta_1 = \tau_1$ and there is nothing to prove. Therefore, we may assume that $n \geq m > 1$. Write $f_2 = \theta_2 \circ \cdots \circ \theta_n$ and $g_2 = \tau_2 \circ \cdots \circ \tau_m$ so that $\theta_1(f_2) = \tau_1(g_2)$. Now if $\theta_1 \neq \tau_1$ and $O_S$ denotes the set of $S$-integral elements of $K$, then $(x, y) = (f_2(z), g_2(z))$ is an $S$-integral point on $C_{i,j}$ (where $\theta_1 = \phi_i$ and $\tau_1 = \phi_j$) for all $z \in O_S$. However, this contradicts the first conclusion of Proposition 4.5, and we deduce that $\theta_1 = \tau_1$. On the other hand if $f_2 \neq g_2$, then $f_2(z) \neq g_2(z)$ for infinitely many $z \in O_S$. Moreover, for such $z$ we have that $(x, y) = (f_2(z), g_2(z))$ is an $S$-integral point on $C_{i,j}$, also a contradiction. In particular, it must be the case that $f_2 = g_2$. Now replacing the pair $(f_1, g_1)$ with $(f_2, g_2)$ and repeating the argument above (until the $\tau$’s are exhausted) we deduce that $\theta_i = \tau_i$ for all $1 \leq i \leq m$. Finally, by equating degrees given by the original relation $f_1 = g_1$, we see that $\deg(\theta_{m+1}) \cdots \deg(\theta_n) = 1$, a contraction unless $n = m$. This completes the proof that $M_S$ is free. 

To complete the main remaining step in the proof of Theorems 1.5 and 1.6, (i.e., to pass from counting functions to counting points), we need to show that $f(P) = g(P)$ occurs rarely for $f, g \in M_S$ and $P$ of large enough height. This is largely achieved by ensuring that the rational or integral points on the curves in (31) are finite. Of course, we already know this for polynomials. However for critically separate and critically simple sets of rational maps, the finiteness we seek follows from the genus calculations in [25] and Faltings’ theorem:

Proposition 4.6 Let $S = \{\phi_1, \ldots, \phi_3\}$ be a set of rational maps on $\mathbb{P}^1(\overline{\mathbb{Q}})$ all of degree at least 4. If $S$ is critically separate and critically simple, then the curves in (31) have at most finitely many rational points over any number field.

Proof Since $S$ is critically separate, [25, Proposition 3.1] implies that each $C_{j,k}$ is an irreducible curve for all $j \neq k$. Likewise, it is shown on [25, p 208] that the genus of $C_{j,k}$ is
given by \((\deg(\phi_j) - 1)(\deg(\phi_k) - 1) \geq 9\). Hence, the \(C_{j,k}\) have at most finitely many rational points over any number field by Faltings’ theorem. Likewise, [25, Corollary 3.6] implies that each \(C_i\) is an irreducible curve. Moreover, it is shown on [25, p 210] that the genus of \(C_i\) is \((\deg(\phi_i) - 2)^2 \geq 4\). Hence, the \(C_i\) also have at most finitely many rational points over any number field by Faltings’ theorem.

In particular, we are now ready to prove our orbit counts for \(\mathbb{P}^1\) from the Sect. 1.

**Proof of Theorems 1.5 and 1.6.** Suppose that \(S\) is a critically separate and critically simple set of rational functions or that \(S\) is a set of polynomials satisfying the conditions of Theorem 1.6. Then in particular, \(M_S\) is free by Proposition 4.1 in the rational function case and Proposition 4.5 in the polynomial case. Hence, Theorem 1.1 implies that the number of functions \(f \in M_S\) satisfying \(H(f(P)) \leq B\), has the desired growth rate (in either case), whenever \(P\) has large enough height.

To pass from functions to points, we need to control when \(f(P) = g(P)\) is possible for \(f, g \in M_S\). With this in mind, let \(R_P \subset K\) be a ring of \(S\)-integers in some number field \(K\) (not the same \(S\) as the set of functions) containing \(P\) and the coefficients of the maps in \(S\). Then define the quantities

\[
\kappa_P := \max \left\{ h(x) : (x, y) \in C_i(K) \text{ or } (x, y) \in C_{j,k}(K) \text{ for some } y \in K \text{ and some } i, j, k \right\}
\]

in the rational function case and

\[
\kappa_P := \max \left\{ h(x) : (x, y) \in C_i(R_P) \text{ or } (x, y) \in C_{j,k}(R_P) \text{ for some } y \in R_P \text{ and some } i, j, k \right\}
\]

in the polynomial case. Then \(\kappa_P\) is finite by Propositions 4.6 and 4.5 in either case. Now given \(f = \theta_1 \circ \theta_2 \circ \cdots \circ \theta_n \in M_S\), define the length of \(f\) to be \(\ell(f) = n\); note that this quantity is well-defined since \(M_S\) is free. Moreover letting \(v = (1, \ldots, 1)\), we see that \(\ell = \ell_{S,v}\) in our earlier notation. Next, recall the constant \(b_S\) given by \(b_S = C_S/(d_S - 1)\), where \(C_S\) and \(d_S\) are the height constants in Definition 2.1 above. Then, Tate’s telescoping Lemma 2.2 implies that if \(h(\rho(P)) \leq \kappa_P\) for some \(P\) with \(h(P) > 2b_S\) and some \(\rho \in M_S\), then

\[
2^{\ell(\rho)}b_S \leq \deg(\rho)(h(P) - b_S) \leq h(\rho(P)) \leq \kappa_P.
\]

Hence, the length of such \(\rho\) is bounded; specifically, \(\ell(\rho) \leq \max \{1, \log_2(\kappa_P/b_S)\}\) := \(r_P\), from which we deduce the following fact.

**Lemma 4.7** Suppose that \(S\) satisfies the conditions of Theorems 1.5 or 1.6 and let \(\rho \in M_S\). If \(\ell(\rho) > r_P\), \(h(P) > 2b_S\), and \(\theta(\rho(P)) = \tau(P')\) for some \(P' \in R_P\) and some \(\theta, \tau \in S\), then \(\theta = \tau\) and \(\rho(P) = P'\).

In particular, this allows us to control the number of functions in \(M_S\) that can agree at \(P\).

**Lemma 4.8** Suppose that \(S\) satisfies the conditions of Theorems 1.5 or 1.6 and \(h(P) > 2b_S\). Then there is a constant \(t_{P,S}\) depending only on \(P\) and \(S\) such that

\[
\# \left\{ f \in M_S : f(P) = Q \right\} \leq t_{P,S}
\]

holds for all but finitely many \(Q \in \text{Orb}_S(P)\).

**Proof** Fix a number field \(K\) containing the coefficients of the maps in \(S\) and over which \(P\) is defined. Let \(d_S = \max\{\deg(\phi) : \phi \in S\}\) and suppose that \(Q \in \text{Orb}_S(P)\) satisfies

\[
h(Q) > d_S^{r_P+1}(h(P) + b_S),
\]
true of all but finitely many $Q$ by Northcott’s Theorem; each $Q \in \text{Orb}_S(P) \subseteq \mathbb{P}_1(K)$ by construction of $K$. Then, it follows from Tate’s telescoping Lemma 2.2 that $\ell(f) > r_P + 1$ for all $f \in M_S$ with $f(P) = Q$; otherwise,

$$h(Q) = h(f(P)) \leq \deg(f)(h(P) + b_S) \leq \deg(\ell(f))(h(P) + b_S) \leq \deg r_P + 1(h(P) + b_S),$$

a contradiction. In particular, each function taking the value of $Q$ at $P$ has length strictly larger than $r_P + 1$. Now, let $f_Q \in M_S$ be a function of smallest length taking the value of $Q$ at $P$. Then $\ell(f_Q) > r_P + 1$ and we may write $f_Q = \tau_1 \circ \cdots \circ \tau_m \circ \rho_Q$ for some $\tau_i \in S$, some $m \geq 1$, and some $\rho_Q \in M_S$ of length $r_P + 1$. Likewise, for any other $f \in M_S$ with $f(P) = Q$, we may write $f = \theta_1 \circ \cdots \circ \theta_m \circ q_f \circ \rho_f$ for some $\theta_i \in S$, some $q_f \in M_S$, and some $\rho_f \in M_S$ of length $r_P + 1$; here we use the minimality of the length of $f_Q$. Then $f(P) = f_Q(P)$ implies:

$$\theta_1 \circ \cdots \circ \theta_m \circ q_f \circ \rho_f(P) = \tau_1 \circ \cdots \circ \tau_m \circ \rho_Q(P). \quad (33)$$

Now for all $1 \leq i \leq m$, let $\rho_i = \theta_{i+1} \circ \cdots \circ \theta_m \circ q_f \circ \rho_f$ and $P'_i = \tau_{i+1} \circ \cdots \circ \tau_m \circ \rho_Q(P)$. In particular, (33) becomes

$$\theta_1 \circ \rho_1(P) = \tau_1(P').$$

On the other hand, $P'_i \in R_P$ by definition of $R_P$ and $\ell(\rho_i) \geq \ell(\rho_f) = r_P + 1 > r_P$ for all $i$. Hence, Lemma 4.7 applied to $\rho = \rho_1$, $P' = P'_1$, $\theta = \theta_1$, and $\tau = \tau_1$ implies that $\theta_1 \circ \rho_1 = \tau_1$ and $\rho_1(P) = P'_1$. Therefore,

$$\theta_2 \circ \cdots \circ \theta_m \circ q_f \circ \rho_f(P) = \tau_2 \circ \cdots \circ \tau_m \circ \rho_Q(P).$$

Repeating the same argument, this time with $\rho = \rho_2$, $P' = P'_2$, etc., we see that Lemma 4.7 implies that $\theta_2 \circ \rho_2 = \tau_2$ and $\rho_2(P) = P'_2$. We can clearly continue this argument ($m$-times) and obtain that

$$q_f \circ \rho_f(P) = \rho_Q(P) \quad \text{and} \quad \theta_i \circ \rho_i = \tau_i \quad \text{for all} \ 1 \leq i \leq m. \quad (34)$$

On the other hand, Tate’s Telescoping Lemma 2.2 and the fact that $h(P) > 2b_S$ imply the lower bound

$$2^{r_P + 1}b_S \leq \deg(\rho_f)(h(P) - b_S) \leq h(\rho_f(P)). \quad (35)$$

Likewise, we have the upper bound

$$h(\rho_Q(P)) \leq \deg(\rho_Q)(h(P) + b_S) \leq \deg r_P + 1(h(P) + b_S). \quad (36)$$

Hence, after combining (34), (35) and (36) with Lemma 2.2 applied to the map $q_f$, we see that

$$\deg(q_f)(2^{r_P + 1} - 1)b_S \leq \deg(q_f)(h(\rho_f(P)) - b_S) \leq h(q_f \circ \rho_f(P)) = h(\rho_Q(P)) \leq \deg r_P + 1(h(P) + b_S).$$

In particular, dividing both sides of the inequality above by $(2^{r_P + 1} - 1)b_S$, we deduce that

$$2^{\ell(q_f)} \leq \deg(q_f) \leq \frac{\deg r_P + 1(h(P) + b_S)}{(2^{r_P + 1} - 1)b_S}. \quad (37)$$

Hence the length of $q_f$ is bounded. But $S$ is a finite set of maps, so the number of possible $q_f$’s is finite. Likewise, the length of $\rho_f$ is $r_P + 1$ is bounded, and so there are only finitely many possible $\rho_f$’s. In summation, we have shown that if $f \in M_S$ is any function with
\[ f(P) = Q, \text{ then } f = \tau_1 \circ \cdots \circ \tau_m \circ q_f \circ \rho_f \text{ such that: the } \tau_i \text{ are fixed, and the number of possible } q_f \text{'s and } \rho_f \text{'s are bounded independently of } Q. \text{ Specifically, we have that} \]

\[ \# \{ f \in M_S : f(P) = Q \} \leq s^{r_p + 1} \left( \frac{d_P^{r_p + 1}(h(P) + b_S)}{(2^p + 1)(h(P) + b_S)} \right)^{r_p - 1} \]

holds for all \( Q \in \text{Orb}_S(P) \) with \( h(Q) > d_P^{r_p + 1}(h(P) + b_S) \), which proves the claim.

We now finish the proof of Theorems 1.5 and 1.6. Note that Lemma 4.8 implies that:

\[ \left( \log B \right)^b \ll \# \{ Q \in \text{Orb}_S(P) : H(Q) \leq B \} \ll \left( \log B \right)^b + \epsilon \]

as desired.

**Remark 14** In higher dimensions, it is possible that one can attack Conjecture 1.3 in a similar manner to that above, provided that one can give a reasonable condition ensuring that the set of rational/integral points on the variety \( V_{f, g} = \{ (P, Q) \in \mathbb{P}^N \times \mathbb{P}^N : f(P) = g(Q) \} \) is not Zariski dense (for all distinct \( f, g \in M_S \) of some fixed length) in \( V_{f, g} \). To do this, it may be necessary to assume the Bombieri–Lang Conjecture.

### 5 More on free monoids

Since the results and conjecture within this paper are chiefly concerned with free monoids, it would be useful to have an explicit way to check that a given set of morphisms of \( \mathbb{P}^N \) generates a free monoid (especially when \( n \geq 2 \)). One possibility, by analogy with Proposition 4.1 above in dimension one, is to try and use generalizations of Picard’s Theorem in higher dimensions; see for instance [21]. On the other hand, for sets of polynomials there is a much more straightforward way of ensuring that the corresponding monoid is free, without reference to critical values or to compositional decompositions as in Theorem 1.6. Namely, it suffices to check that the set of degrees and the set of leading coefficients both form multiplicatively independent sets. This is perhaps known to the experts. However without a reference, we include a proof for completeness and to motivate possible generalizations to affine morphisms in higher dimensions. Our argument is inspired by the proof of [16, Lemma 3.2].

**Theorem 5.1** Let \( S = \{ \phi_1, \ldots, \phi_s \} \) be a set of polynomials defined over a field \( K \) of characteristic zero, and let \( a_i x^{d_i} \) denote the leading term of \( \phi_i \). If \( \{ d_1, \ldots, d_s \} \) is a multiplicatively independent set in \( \mathbb{Z} \) and \( \{ a_1, \ldots, a_s \} \) is a multiplicatively independent set in \( K^* \), then \( M_S \) is a free monoid.

**Proof.** As the statement of the theorem suggests, it suffices to study the monoid generated by the leading terms in \( S \). To make this statement precise, we note the following lemma:
Lemma 5.2 Let $S = \{\phi_1, \ldots, \phi_s\}$ be a set of polynomials defined over a field $K$, let $a_i x^{d_i}$ denote the leading term of $\phi_i$, and let $S' = \{a_1 x^{d_1}, \ldots, a_s x^{d_s}\}$. If $M_{S'}$ is a free monoid, then $M_S$ is a free monoid.

Proof This statement is a simple consequence of the fact that $\text{lt}(f \circ g) = \text{lt}(f) \circ \text{lt}(g)$ for all $f, g \in K[x]$; here $\text{lt}(\cdot)$ denotes the leading term of a polynomial. To see this, suppose that $M_{S'}$ is a free monoid and that there is some relation

$$\theta_1 \circ \theta_2 \circ \cdots \circ \theta_n = \tau_1 \circ \tau_2 \circ \cdots \circ \tau_m$$

(38)

for some $\theta_i, \tau_j \in S$. Then, in particular, we have an equality of leading terms,

$$\text{lt}(\theta_1) \circ \text{lt}(\theta_2) \circ \cdots \circ \text{lt}(\theta_n) = \text{lt}(\tau_1) \circ \text{lt}(\tau_2) \circ \cdots \circ \text{lt}(\tau_m).$$

But this is a relation in $M_{S'}$, which is free on the letters in $S'$. Therefore, $n = m$ and $\text{lt}(\theta_i) = \text{lt}(\tau_i)$. However, again since $M_{S'}$ is free, $\text{lt}(\theta_i) = \text{lt}(\tau_i)$ implies that $\theta_i = \tau_i$. Hence the relation in (38) is a trivial one. \hfill \Box

Now back to the proof of Theorem 5.1. In particular, in light of Lemma 5.2, we may assume that $S = \{\phi_1, \ldots, \phi_s\}$ is a set of monomials with $\phi_i = a_i x^{d_i}$, that $\{d_1, \ldots, d_s\}$ is a multiplicatively independent set in $\mathbb{Z}$, and that $\{a_1, \ldots, a_s\}$ is a multiplicatively independent set in $K^*$. Now, given $F = \theta_1 \circ \cdots \circ \theta_n \in M_S$ and $\phi \in S$, we define $e_\phi(F) = \# \{j \mid \theta_j = \phi\}$ to be the number of $\phi$’s appearing in the string defining $F$ (strictly speaking this is an abuse of notation; $e_\phi$ is a function on words). In particular, if there is a relation $F = G$ for some $F, G \in M_S$, then we see that

$$d_1^{e_{\phi_1}(F)} \cdots d_s^{e_{\phi_s}(F)} = \deg(F) = \deg(G) = d_1^{e_{\phi_1}(G)} \cdots d_s^{e_{\phi_s}(G)}.$$  

(39)

However, the $d_i$’s are multiplicatively independent by assumption, so that $e_{\phi_i}(F) = e_{\phi_i}(G)$ for all $i$. In particular, the strings defining $F$ and $G$ have the same length (i.e., the total number of letters from $S$) equal to $n = \sum e_{\phi_i}(F)$. Hence,

$$F = \theta_1 \circ \cdots \circ \theta_n = \tau_1 \circ \cdots \circ \tau_n = G \quad \text{for some } \theta_i, \tau_i \in S.$$  

(40)

Moreover, $e_{\phi_i}(F) = e_{\phi_i}(G)$ for all $i$. From here, we will show that $\theta_i = \tau_i$ by induction on the length $n$. The $n = 1$ case is clear. For $n > 1$, if (40) holds then

$$F' \circ \theta = F = G = G' \circ \tau$$  

(41)

for some $\theta, \tau \in S$ and some monomials $F'$ and $G'$ given by strings of length $n - 1$ of elements of $S$. We proceed in cases.

Case(1): Suppose that $\theta = \tau$, and write $\theta = ax^d, F' = a_{F'} x^{\deg(F')} = a_{G'} x^{\deg(F')}$. Here we use that $\deg(F) = \deg(G)$ and $\theta = \tau$, so that $\deg(F') = \deg(G')$. Therefore, (41) becomes

$$a_{F'} x^{\deg(F')} = a_{G'} x^{\deg(F')}.$$  

and we deduce that $a_{F'} = a_{G'}$. However, then $F' = a_{F'} x^{\deg(F')} = a_{G'} x^{\deg(F')} = G'$ and $F', G' \in M_S$ are polynomials obtained by composing strings of elements of $S$ of length $n - 1$. In particular, we may deduce that $\theta_i = \tau_i$ for all $i < n$ by induction. On the other hand, $\theta_n = \tau = \tau_n$ by construction. Therefore, $\theta_i = \tau_i$ for all $i \leq n$ as claimed.

Case(2): Suppose that $\theta \neq \tau$. We fix some notation. Given a string $\theta_1 \ldots \theta_m$ of elements of $S$, write

$$f = \theta_1 \circ \cdots \circ \theta_m = a f x^{\deg(f)} = (a_1^{n_1} \cdots a_s^{n_s}) x^{\deg(f)}.$$  

Springer
Then define the \(a_i\)-degree of \(f\) (or more accurately, the \(a_i\)-degree of the corresponding string) to be \(\deg_a(f) = n_i\). Note that this construction is well-defined since the leading coefficient \(a_f\) is in the (multiplicative) semigroup generated by the \(a_i\’s\) and the \(a_i\’s\) are multiplicatively independent by assumption. Now write \(\theta = ax^d\). Then we will show that \(\deg_a(F) \neq \deg_a(G)\), a contradiction, using (41), the fact that \(\theta \neq \tau\), and the following elementary observations about \(a\)-degrees:

**Lemma 5.3** Let \(S\) be as in Theorem 5.1 and let \(\theta = ax^d \in S\). Then the following statements hold:

1. If \(f_1, f_2, g \in M_S\), \(\deg_a(f_1) \leq \deg_a(f_2)\), and \(\deg(f_1) \leq \deg(f_2)\), then \(\deg_a(f_1 \circ g) \leq \deg_a(f_2 \circ g)\).

2. Let \(f \in M_S\), and suppose that \(e_{\theta}(f) = e \geq 1\). Then \(\deg_a(f) \leq \frac{d^e-1}{d-1} \cdot \frac{\deg(f)}{d^e}\).

We grant Lemma 5.3 for now and return to the proof later. To see that \(\deg_a(F) \neq \deg_a(G)\) in Case 2, let \(e = e_{\theta}(F) = e_{\theta}(G)\) be the number of \(\theta\’s\) appearing in the strings defining \(F\) and \(G\). Then, writing \(F = F' \circ \theta\) as in (41), we see that

\[
\deg_a(F) = \deg_a(F') + \deg(F') \deg_a(\theta) = \deg_a(F') + \deg(F') \geq \deg(F').
\]  

(42)

On the other hand, Lemma 5.3 part (2) applied to \(f = G'\) implies that

\[
\deg_a(G) = \deg_a(G') + \deg(G') \deg_a(\tau) = \deg_a(G') \leq \frac{d^e-1}{d-1} \cdot \frac{\deg(G')}{{\frac{d^e}{d^e}}}. 
\]  

(43)

Here we use that \(G = G' \circ \tau\) and that \(\deg_a(\tau) = 0\), since \(\theta \neq \tau\) and the leading coefficients of the elements in \(S\) are multiplicatively independent. Therefore, if \(\deg_a(F) = \deg_a(G)\), then (41), (42), (43) together imply that

\[
\deg(\tau) \deg(F) = \deg(\tau) d \deg(F') \\
\leq \deg(\tau) d \deg_a(F) \\
= \deg(\tau) d \deg_a(G) \\
\leq \frac{d^e-1}{d-1} \cdot \frac{\deg(\tau) \deg(G')}{d^e} \\
\leq \frac{d^e-1}{d-1} \cdot \frac{\deg(G)}{d^e} \\
< \frac{d^e-1}{d-1} \cdot \frac{d}{d^e} \\
= \frac{d}{d-1} \deg(G).
\]  

(44)

However, \(F = G\) so that \(\deg(F) = \deg(G)\). In particular, (44) implies that

\[
2 \leq \frac{d}{d-1} < \frac{d}{d-1} \leq 2,
\]

a contradiction. Therefore, \(\deg_a(F) \neq \deg_a(G)\) and Case 2 is incompatible with (41). Therefore, any relation in \(M_S\) must be of the form in Case 1. However, since we have settled Theorem 5.1 in this case by induction, \(M_S\) is a free monoid as claimed. \(\square\)

We now include a proof of Lemma 5.3 regarding \(a\)-degrees.
(Lemma 5.3) The first statement is a simple consequence of the definition of $a$-degrees. Suppose that $f_1, f_2, g \in M_S$, that $\deg_a(f_1) \leq \deg_a(f_2)$, and that $\deg(f_1) \leq \deg(f_2)$. Then

$$\deg_a(f_1 \circ g) = \deg_a(f_1) + \deg_a(g) \cdot \deg(f_1) \leq \deg_a(f_2) + \deg_a(g) \cdot \deg(f_2) = \deg_a(f_2 \circ g)$$

as claimed. For the second statement, let $f \in M_S$ and suppose that $e_0(f) = e \geq 1$. Then, we may write

$$f = g_{t+1} \circ \theta^t \circ g_t \circ \cdots \circ g_2 \circ \theta^1 \circ g_1$$

for some $g_i \in M_S$ with $\deg_a(g_i) = 0$, some $t \geq 1$, and some $r_i \geq 0$ with $\sum_{i=1}^t r_i = e$. We will show by induction on $t$ that

$$\deg_a(f) \leq \deg_a(g_{e+1} \circ g_e \cdots \circ g_1 \circ \theta^e),$$

from which statement (2) of the Lemma easily follows. If $t = 1$, then

$$\deg_a(g_2 \circ \theta^t \circ g_1) = \deg(g_2) \deg_a(\theta^t) \leq \deg(g_2) \deg(g_1) \deg_a(\theta^t) \leq \deg_a(g_2 \circ g_1 \circ \theta^t).$$

Here we use that $\deg_a(g_i) = 0$. On the other hand, assume that $t > 1$ and that (46) is true for polynomials of the form in (45) with $t - 1$ appearances of substrings of the form $\theta^r$. Then given $f$ as in (45), let $f_1 = g_{t+1} \circ \theta^t \circ g_t \circ \theta^{r_{t-1}} \circ \cdots \circ g_2$, let $f_2 = g_{t+1} \circ \cdots \circ g_2 \circ \theta^{r_{t-1}}$ where $r = \sum_{i=2}^t r_i$, and let $g = \theta^{r_1} \circ g_1$. Then $f = f_1 \circ g$ and $\deg(f_1) = \deg(f_2)$. Hence, part 1 of Lemma 5.3 and the induction hypothesis together imply that

$$\deg_a(f) = \deg_a(f_1 \circ g) \leq \deg(f_2 \circ g) = \deg_a((g_{t+1} \circ \cdots \circ g_2) \circ \theta^e \circ g_1).$$

(47)

On the other hand letting $g' = g_{t+1} \cdots \circ g_2$, we see that the $t = 1$ case above applied to $g' \circ \theta^e \circ g_1$ in place of $f$ implies that

$$\deg_a((g_{t+1} \circ \cdots \circ g_2) \circ \theta^e \circ g_1) \leq \deg_a((g_{t+1} \circ \cdots \circ g_1) \circ \theta^e).$$

(48)

Therefore after combining (47) and (48), we establish (46) as claimed. Finally, the bound in part 2 of Lemma 5.3 follows easily from (46), the fact that

$$\deg_a(\theta^e) = (d^{e-1} + \cdots + d + 1) = (d^e - 1)/(d - 1),$$

and that $\deg(g_{t+1} \circ \cdots \circ g_1) = \deg(f)/d^e$. This completes the proof.

Remark 15 Although most sets generate free monoids, it would be interesting to study the height growth rates in monoid orbits which are not free (or free commutative). For polynomial semigroups there is a Tits alternative [3], so either $M_S$ contains a free subsemigroup or it has linear growth. As a test case, one might consider the following example from [16, Remark 1.5]: let $\omega$ be a primitive cube root of unity and let $F(x) = x^2$ and $G(x) = \omega x^2$. Then the monoid generated by $S = \{F, G\}$ has three independent relations: $F^2 = G^2$, $F^2 \circ G = G \circ F^2$, and $G \circ F \circ G = F \circ G \circ F$.

Acknowledgements We thank Yuri Bilu, Andrew Bridy, Alexander Evetts, Joseph Silverman, and Umberto Zannier for discussions related to this paper. We also thank the authors of [16]; Lemma 3.2 in their paper inspired the proof of Theorem 5.1. Finally, we thank the anonymous referee for their insightful comments.

References

1. Avanzi, R., Zannier, U.: The equation $f(X) = f(Y)$ in rational functions $X = X(t), Y = Y(t)$. Compos. Math. 139(3), 263–295 (2003)
2. Baragar, A.: Rational points on $K^3$ surfaces in $\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$. Math. Ann. 305(3), 541–558 (1996)
3. Bell, J.P., Huang, K., Peng, W., Tucker, T.J.: A Tits alternative for rational functions. arXiv:2103.09994 (preprint)
4. Bilu, Y., Tichy, R.: The diophantine equation $f(x) = g(y)$. Acta Arith. 95(3), 261–288 (2000)
5. Bombieri, E., Gubler, W.: Heights in Diophantine Geometry, New Math. Monographs, vol. 4. Cambridge University Press, Cambridge (2006)
6. Bosma, W., Cannon, J., Playoust, C.: The Magma algebra system I: the user language. J. Symb. Comput. 24(3–4), 235–265 (1997)
7. Cantat, S.: Sur les groupes de transformations birationnelles des surfaces. Ann. Math. (2) 174(1), 299–340 (2011)
8. Flajolet, P., Sedgewick, R.: Analytic Combinatorics. Cambridge University Press, Cambridge (2009)
9. Fried, M.: On a conjecture of Schur. Mich. Math. J. 17, 41–50 (1970)
10. Ghioca, D., Tucker, T.J.; Zieve, M.E.: Linear relations between polynomial orbits. Duke Math. J. 161(7), 1379–1410 (2012)
11. Greenfield, G., Drucker, D.: On the discriminant of a trinomial. Linear Algebra Appl. 62, 105–112 (1984)
12. Healey, V., Hindes, W.: Stochastic canonical heights. J. Number Theory 201, 228–256 (2019)
13. Hindes, W.: Dynamical and arithmetic degrees for random iterations of maps on projective space. Math. Proc. Camb. Philos. Soc., pp. 1–17 (2021). https://doi.org/10.1017/S0305004120000250
14. Hindes, W.: Finite orbit points for sets of quadratic polynomials. Int. J. Number Theory 15(8), 1693–1719 (2019)
15. Hindes, W.: Dynamical height growth: left, right, and total orbits. Pac. J. Math. (to appear)
16. Jiang, Z., Zieve, M.: Functional equations in polynomials, REU project
17. Kawaguchi, S.: Canonical heights for random iterations in certain varieties. Int. Math. Res. Not., Article ID rnm023 (2007)
18. Kawaguchi, S., Silverman, J.H.: On the dynamical and arithmetic degrees of rational self-maps of algebraic varieties. J. Reine Angew. Math. (Crelles J.) 713, 21–48 (2016)
19. Lang, S.: Diophantine Geometry. Springer, Berlin (1982)
20. Lang, S.: Fundamentals of Diophantine Geometry. Springer Science & Business Media, New York (2013)
21. Levin, A.: Generalizations of Siegel’s and Picard’s theorems. Ann. Math. 170, 609–655 (2009)
22. Matsuzawa, Y.: On upper bounds of arithmetic degrees. Am. J. Math. 142(6), 1797–1820 (2020)
23. Mello, J.: The dynamical and arithmetical degrees for eigensystems of rational self-maps. Bull. Braz. Math. Soc. 51(2), 569–596 (2020)
24. Mueller, P.: Permutation groups with a cyclic two-orbits subgroup and monodromy groups of Laurent polynomials. Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) 12(2), 369–398 (2013)
25. Pakovich, F.: Algebraic curves $P(x) - Q(y) = 0$ and functional equations. Complex Var. Ellipt. Equ. 56(1–4), 199–213 (2011)
26. Schinzel, A.: Polynomials with Special Regard to Reducibility. Cambridge University Press, Cambridge (2000)
27. Serre, J.-P.: Lectures on the Mordell–Weil theorem. Aspects of Mathematics, 3rd edn. Friedr. Vieweg & Sohn, Braunschweig (1997). Translated from the French and edited by Martin Brown from notes by Michel Waldschmidt. Foreword by Brown and Serre
28. Serre, J.-P.: Lectures on the Mordell–Weil Theorem, 2nd edn. Vieweg (1990)
29. Silverman, J.H.: Arithmetic and dynamical degrees on abelian varieties. J. Thör. Nombres Bordeaux 29(1), 151–167 (2017)
30. Silverman, J.H.: The Arithmetic of Dynamical Systems, vol. 241. Springer GTM (2007)
31. Silverman, J.H.: Rational points on $K^3$ surfaces: a new canonical height. Invent. Math. 105(1), 347–373 (1991)
32. Zagier, D.: On the number of Markoff numbers below a given bound. Math. Comput. 39(160), 709–723 (1982)
33. Zannier, U.: Integral points on curves $(f(x) - f(y))/(x - y)$ (forthcoming)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.