ORBIT COUNTING IN POLARIZED DYNAMICAL SYSTEMS

WADE HINDES

Abstract. We extend recent orbit counts for finitely generated semigroups acting on \( \mathbb{P}^N \) to certain infinitely generated, polarized semigroups acting on projective varieties. We then apply these results to semigroup orbits generated by some infinite sets of unicritical polynomials.

1. Introduction

Given a projective variety \( V \), a height function \( H : V \to \mathbb{R} \), and a subset \( X \subseteq V \) of some interest, it is often useful to understand the growth rate of the number of points in \( X \) of bounded height. For instance, if \( X \) is a set or rational or integral points, then this growth rate is known to encode many interesting arithmetic and geometric invariants of \( V \) (like its dimension, genus, or rank of an associated Mordell-Weil group); for examples, see [2, 8, 10, 11, 20]. Likewise, when \( X \) is a dynamical orbit generated by a collection of self maps of \( V \), then the growth rate on the number of points in \( X \) of bounded height frequently detects dynamical degrees [14, 17] as well as other invariants [1, 21]. In this paper, we take up this dynamical orbit counting problem, generalizing the main results from [13] in two ways: first we allow infinitely generated semigroups, and second we allow \( V \) to be any projective variety with a polarizable set of maps (not necessarily \( \mathbb{P}^N \)). As an application, we obtain precise estimates for the number of points of bounded height in many new orbits, including those generated by the (possibly infinite) set of polynomials of the form \( z^q + 1 \), where \( q \) is a Mersenne prime; see Corollary 1.4 below.

To describe our results, we fix some notation. Let \( K \) be a global field and let \( V \) be a projective variety defined over \( K \). Then, given a collection \( S \) of endomorphisms of \( V \), we define \( M_S \) to be the semigroup generated by \( S \) under composition together with the identity function (formally a monoid). In particular, to a point \( P \in V \) we associate an orbit,

\[
\text{Orb}_S(P) = \{f(P) : f \in M_S\},
\]

and study the growth rate of the number of points in \( \text{Orb}_S(P) \) of bounded height. There is very little known about this problem in general, even for sets \( S \) of polynomials in one-variable. So, as in the case of iterating a single morphism [6], we make this problem more tractable by assuming that our dynamical system is polarized: there is a divisor class \( \eta \in \text{Pic}(V) \otimes \mathbb{R} \) and a collection of real numbers \( d_\phi \) such that \( \phi^*(\eta) = d_\phi \eta \) for all \( \phi \in S \). For in this case, there is a minimal constant \( C(V, \eta, \phi) \) such that

\[
|h_\eta(\phi(P)) - d_\phi h_\eta(P)| \leq C(V, \eta, \phi)
\]

for all \( P \in V \); here \( h_\eta \) is a height function associated to \( \eta \). Then, in order to generalize the techniques in [13] for finitely generated semigroups acting on \( \mathbb{P}^N \) to infinitely generated polarized semigroups acting on arbitrary projective varieties, we define the following conditions:

**Definition 1.1.** A set of endomorphisms \( S \) on a projective variety \( V \) is called **height controlled** with respect to a divisor class \( \eta \in \text{Pic}(V) \otimes \mathbb{R} \) if:

1. for all \( \phi \in S \) there exists a real number \( d_\phi > 1 \) such that \( \phi^*(\eta) = d_\phi \eta \),
2. the corresponding height constants are bounded: \( \sup_{\phi \in S} C(V, \eta, \phi) \) is finite.

2010 Mathematics Subject Classification: Primary: 37P15, 37P05. Secondary: 11G50, 11D45.
Moreover, if in addition the set of real numbers $T := \{\log(d_\phi) : \phi \in S\}$ is uniformly discrete, i.e., there exists a constant $\delta_T > 0$ such that $|\log(d_\phi) - \log(d_\psi)| > \delta_T$ for all distinct $\phi, \psi$ in $S$, then we say that $S$ is height controlled and uniformly log-discrete with respect to $\eta$.

**Remark 1.** The first condition, which we call height controlled but has also been called bounded, has been used in several places to study both semigroup and random sequential orbits; see, for instance, [12, 14, 16, 19]. However the second condition is a new and technical one, allowing us to control the dominant poles of some associated meromorphic generating functions; see Section 2 for details.

**Example 1.** Let $S = \{\phi_1, \ldots, \phi_s\}$ be any finite set of endomorphisms of $\mathbb{P}^N$. If $\deg(\phi) \geq 2$ for all $\phi \in S$ and $\deg(\phi) \neq \deg(\psi)$ for all distinct $\phi, \psi \in S$, then $S$ is height controlled and uniformly log-discrete. More generally, let $S = \{\phi_1, \ldots, \phi_s\}$ be any finite set of endomorphisms on a projective variety $V$ and assume that there is a divisor class $\eta \in \text{Pic}(V) \otimes \mathbb{R}$ such that for each $\phi \in S$ there is a real number $d_\phi > 1$ satisfying $\phi^*(\eta) = d_\phi \eta$. If in addition $d_\phi \neq d_\psi$ for all distinct $\phi, \psi \in S$, then $S$ is height controlled and uniformly log-discrete with respect to $\eta$.

**Example 2.** For an infinite example, fix $c \in \mathbb{Q}$ and $a, b \in \mathbb{N}$. Then the set of unicritical polynomials with constant term $c$ and degree $d = a^n + b$ for some $n \geq 1$,

$$S_{a,b,c} := \{x^d + c : d = a^n + b \text{ for some } n \geq 1\},$$

is a height controlled and uniformly log-discrete set of endomorphisms of $\mathbb{P}^1$.

With this terminology in place we are ready to state our first result, an upper bound for the number of points of bounded height in semigroup orbits generated by height controlled and uniformly log-discrete sets. Moreover, if in addition the semigroup is free, then we can give an associated lower bound. In particular, since generic compositional semigroups tend to be free (e.g., see [13, §4]), it is reasonable to expect that the bounds we give below apply more generally, providing a template for future work. In what follows, $H_\eta = \exp \circ h_\eta$ denotes the multiplicative Weil height associated to $\eta$.

**Theorem 1.2.** Let $K$ be a global field, let $V$ be a projective variety over $K$, and let $S$ be a height controlled and uniformly log-discrete set of endomorphisms on $V$ with respect to a divisor class $\eta \in \text{Pic}(V) \otimes \mathbb{R}$. Then the following statements hold:

1. There is a positive constant $b = b(S, \eta)$ and a constant $B_{S,\eta}$ such that

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

for all $P \in V$ with $H_\eta(P) > B_{S,\eta}$.

2. If $M_S$ is free (non-abelian), then for all $\epsilon > 0$ there is a positive constant $b = b(S, \eta, \epsilon)$ and a constant $B_{S,\eta}$ such that

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

for all $P \in V$ with $H_\eta(P) > B_{S,\eta}$.

Here in both statements, the implicit constants depend on $P$, $S$, $\eta$ (and $\epsilon$ for statement (2)).

Of course, one would like upper and lower bounds for the number of points of bounded height in orbits, not just functions yielding a bounded height relation. However, to do this for a given basepoint $P$ we need to control the number of distinct functions $f, g \in M_S$ that can agree at $P$, i.e., $f(P) = g(P)$. This is quite challenging in higher dimensions, even for finitely generated semigroups. On the other hand, it is tractable in dimension one due to Siegel’s integral point theorem and the polynomial classification theorem of Bilu-Tichy [3]; essentially, it suffices to control the $s$-integral points on the finitely many curves $f(x) = g(y)$ where $f$ and $g$ are in the generating set; see [15]. However, this problem remains quite difficult (even for $\mathbb{P}^1$) when $M_S$ is infinitely generated, since there are infinitely many curves to consider. Nevertheless, there is a non-trivial case where it is possible, namely when we restrict ourselves to polynomials of the form $f(z) = z^d + 1$ with prime exponent and constant term one. In this case, we combine Theorem 1.2 above with the bounds in [3] on the $s$-integral solutions to the
Catalan equations \( x^n - y^m = 1 \) to prove the following result. Below, \( H : \mathbb{P}^1 \rightarrow \mathbb{R} \) denotes the standard absolute and multiplicative Weil height.

**Theorem 1.3.** Let \( K \) be a number field and let \( w_K \) be the number of roots of unity in \( K \). Given any uniformly log-discrete set \( q \) of rational prime numbers, define

\[
S_{K,q} = \{ z^q + 1 : q \in q \text{ and } \gcd(q, w_K) = 1 \}.
\]

Then for all \( \epsilon > 0 \) there is a positive constant \( b = b(K, q, \epsilon) \) such that

\[
(\log B)^b \ll \# \{ Q \in \text{Orb}_{S_{K,q}}(P) : H(Q) \leq B \} \ll (\log B)^b + \epsilon
\]

for all \( P \in \mathbb{P}^1(K) \) with \( H(P) > 4 \).

**Remark 2.** As a first step in the proof of Theorem 1.3, we show that \( M_{S_{K,q}} \) is a free semigroup. In fact, we prove a stronger statement: if \( k \) is any field of characteristic zero and

\[
S = \{ z^d + c : d \geq 2, c \in k^* \},
\]

then \( M_S \) is free; see Theorem 3.1 in Section 3.

For a concrete and more classically looking arithmetic application of Theorem 1.3, consider the Mersenne primes (i.e., those of the form \( q = 2^n - 1 \)). Then, after applying some additional approximation techniques discussed in Remark 7 below, we prove an explicit version of Theorem 1.3 in this case; for simplicity, we state this result for \( K = \mathbb{Q} \).

**Corollary 1.4.** Let \( r \in \mathbb{Q} \) with \( H(r) > 4 \) and let \( m \) be the set of Mersenne primes. Then

\[
(\log B)^{0.60839} \ll \# \{ t = ((r^q + 1)^{q_2} \cdots + 1)^{q_n} + 1 : q_1, \ldots, q_n \in m, H(t) \leq B \} \ll (\log B)^{0.60872}.
\]

Here the implicit constants depend on the initial point \( r \in \mathbb{Q} \).

At present, there are only 51 known Mersenne primes, and it is unknown whether or not there are infinitely many such primes - there are, however, probabilistic heuristics suggesting that there are infinitely many \([7, \S 1.3.1]\). With this in mind, it is (to the author at least) amusing that we can give a reasonably good estimate for the number of points of bounded height in the orbits in Corollary 1.4 while remaining agnostic on the infinitude of Mersenne primes, an age-old question.

**Remark 3.** Certainly there exist many provably infinite, uniformly log-discrete sets of primes (where Theorem 1.3 applies). To construct explicit examples, fix any prime \( q \) and any real number \( \delta > 0 \). Now let \( q_0 = q \) and define

\[
q_{n+1} = \min \{ q \in \mathbb{N} : q \text{ is prime and } \log(q) > \log(q_n) + \delta \},
\]

recursively. Then it is easy to check that \( q(q_0, \delta) = \bigcup_{n=0}^{\infty} \{ q_n \} \) is an infinite and uniformly log-discrete set of primes. However, we prefer to display the Mersenne primes as our chief example (even though we cannot be sure if they are infinite or not), due to their prominent place in the history of arithmetic.

2. Counting functions in semigroups

Similar to the case of semigroups acting on projective space \([13, \text{Section 2}]\), the overall strategy used to prove Theorem 1.2 is the following. First, aided by a suitable generalization of Tate’s Telescoping Lemma, we relate the functions \( f \in M_S \) yielding a bounded height relation \( h(f(P)) \leq B \) to the functions having bounded log degree (or more precisely bounded log \( \eta \)-degree, where \( \eta \) is the common eigendivisor class in Definition 1.1 above). Then, we approximate the number of functions of bounded log degree by the number of some related, restricted integer compositions (a standard combinatorial object); to do this we must choose suitable rational approximations for each \( \log(d_\phi) \) and \( \phi \in S \). Next we count restricted integer compositions by analyzing the dominant poles of some associated generating functions. Finally, we show that our estimates for the number of functions of bounded log degree (coming from integer compositions) become better and better as we refine our rational approximations for the \( \log(d_\phi) \).
Remark 4. However, there is some added subtlety that arises in this more general setting (i.e., infinitely generated polarized semigroups), largely coming from the fact that the aforementioned combinatorial generating functions are now meromorphic (as opposed to simply rational, as in the case when $S$ is finite $[13, \S2]$).

With this sketch in place, we begin with a few auxiliary results, starting with a generalization of Tate’s Telescoping Lemma. In what follows, given a function $f = \phi_1 \circ \cdots \circ \phi_m$ in $MS$ for some $\phi_i \in S$, then we write $deg_\eta(f)$ for the quantity $\prod_{i=1}^m d_{\phi_i}$, satisfying $f^*(\eta) = deg_\eta(f)\eta$; here $\eta \in \text{Pic}(V) \otimes \mathbb{R}$ is the common eigendivisor class given in Definition 1.1 for height controlled sets. Then we have the following tool for relating heights and degrees.

**Lemma 2.1.** Let $S$ be a height controlled set of endomorphisms of $V$, and let $d_S := \inf_{\phi \in S} \{d_\phi\}$ and $C_S := \sup_{\phi \in S} C(V, \eta, \phi)$ be the corresponding height controlling constants. Then

$$\left| \frac{h_\eta(f(Q))}{\deg_\eta(f)} - h_\eta(Q) \right| \leq \frac{C_S}{d_S - 1}$$

for all $f \in M_S$ and all $Q \in V(K)$.

See $[12, \text{Lemma 2.1}]$ for a proof of this fact. Having related heights and degrees, we may reduce counting $\# \{f \in MS : h_\eta(f(P)) \leq B\}$ to counting functions of bounded log degree. Moreover if $M_S$ is free and $d_\phi \neq d_\psi$ all $\phi \neq \psi$ in $S$, then this is equivalent to counting finite sequences $(\log(d_{\phi_1}), \ldots, \log(d_{\phi_m}))$ such that $\sum_{i=1}^m \log(d_{\phi_i}) \leq B$. This type of problem is typical in combinatorics, except that the summands $\log(d_{\phi_i})$ are not positive integers. Nevertheless, by approximating the $\log(d_{\phi_i})$ by suitable rational numbers, we can estimate the number of sequences in the $\log(d_{\phi_i})$ with bounded sum by the number of some associated sequences of positive integers with bounded sum (for which there are combinatorial tools).

To discuss the relevant combinatorial tools, we fix some notation. Let $T$ be a set of positive real numbers, and let $\text{Seq}(T)$ be the set of finite sequences of elements of $T$. Then for any sequence $\omega = (t_1, \ldots, t_m)$ with $t_i \in T$, we define

$$|\omega|_T = \sum_{i=1}^m t_i, \quad \omega = (t_1, \ldots, t_m) \in \text{Seq}(T).$$

Of particular historical interest is the case when $T \subseteq \mathbb{N}$, especially as it relates to the problem of counting restricted integer compositions: for a fixed $n \in \mathbb{N}$, count the number of $\omega \in \text{Seq}(T)$ satisfying $|\omega|_T = n$. Luckily this is a standard problem in analytic combinatorics $[9, \S1.3.1]$, and can be attacked (asymptotically) via generating functions and an analysis of their associated poles and residues.

**Theorem 2.2.** Let $T$ be a set of positive integers with at least two elements. For $n > 0$ let

$$a_{T,n} := \# \{\omega \in \text{Seq}(T) : |\omega|_T = n\}.$$  

Then the ordinary generating function for the $a_{T,n}$ is

$$F_T(z) := \sum_n a_{T,n} z^n = \frac{1}{1 - (\sum_{t \in T} z^t)}.$$  

Moreover, if $\gcd(T) = 1$, then

$$a_{T,n} = \frac{1}{\alpha G_T^\prime(\alpha)} \alpha^{-n} \left(1 + O(A^n)\right),$$

where $G_T := \sum_{t \in T} z^t$, $0 < \alpha < 1$ is the unique solution to $G_T(\alpha) = 1$, and $0 < A < 1$.

**Proof.** See $[9, \text{Theorem V.1}]$ and $[9, \text{Example V.2}]$ on page 297 for a justification of why Theorem V.1 applies in this case. \hfill $\square$

**Remark 5.** Here $\gcd(T) = 1$ means that not all elements in $T \subseteq \mathbb{N}$ are a multiple of some common divisor $d > 1$. In particular if $\gcd(T) = 1$, if $\beta = \alpha^{-1}$, and if $0 < \epsilon < 1$, then

$$\frac{(1 - \epsilon)\beta}{(\beta - 1)G_T^\prime(\frac{1}{\beta})} \beta^R - O, (1) \leq \# \{\omega \in \text{Seq}(T) : |\omega|_T \leq R\} \leq \frac{(1 + \epsilon)\beta^3}{(\beta - 1)G_T^\prime(\frac{1}{\beta})} \beta^R + O, (1)$$
for all $R$ sufficiently large; here we sum the expression for $a_{T,n}$ in Theorem 2.2 over $n \leq [R]$ (for the upper bound) and over $n \leq [R]$ (for the lower bound).

Next, we prove a technical lemma that allows us to choose suitable rational approximations for each $\log(\alpha_\phi)$ and $\phi \in S$ (in order to use Theorem 2.2 above); compare to [13, Lemma 2.6].

**Lemma 2.3.** Let $T$ be a positive set of real numbers with at least 2 elements and let $\delta > 0$. Then for all $t \in T$ there are integers $n_{t,\delta}, m_{t,\delta}, u_\delta \in \mathbb{N}$ depending on $\delta$ such that:

1. $t - \delta < \frac{n_{t,\delta}}{u_\delta} < t < \frac{m_{t,\delta}}{u_\delta} < t + \delta$.
2. $\gcd(T_1, \delta) = 1 = \gcd(T_2, \delta)$ where $T_1, \delta = \{n_{t,\delta} : t \in T\}$ and $T_2, \delta = \{m_{t,\delta} : t \in T\}$.

Moreover if $T$ is uniformly discrete, then we can assume $n_{t,\delta} \neq n_{t',\delta}$ and $m_{t,\delta} \neq m_{t',\delta}$ for all distinct $t, t' \in T$ by choosing $\delta$ sufficiently small.

**Remark 6.** Recall that a set of real numbers is uniformly discrete if there exists a positive constant $\delta_T$ such that $|t - t'| > \delta_T$ for all distinct $t, t' \in T$.

**Proof.** Clearly integers $n_{t,\delta}, m_{t,\delta}, u_\delta \in \mathbb{N}$ satisfying condition (1) exists for all $\delta > 0$. Therefore, to find integers satisfying both (1) and (2), we choose integers satisfying (1) and deform them to ensure that both conditions hold. Specifically, fix an integer $r > 0$ and two arbitrary elements $t_1, t_2 \in T$. Then define $v = (n_{t_2,\delta} \cdot m_{t_2,\delta})^r$ and $n_{t_1,\delta} = n_{t_2,\delta} \cdot v + 1, \quad n_{t_2,\delta} = n_{t_2,\delta} \cdot v, \quad m_{t_1,\delta} = m_{t_2,\delta} \cdot v + 1, \quad m_{t_2,\delta} = m_{t_2,\delta} \cdot v, \quad u_\delta = u_\delta \cdot v$

for all $t \neq t_1$. Note that $m_{t_2,\delta} > m_{t_2,\delta}$ by condition (1) so that $v \geq 2$. Now, we check that

$$\frac{n_{t_1,\delta}}{u_\delta} = \frac{n_{t_2,\delta} + 1}{u_\delta v}, \quad \frac{m_{t_1,\delta}}{u_\delta} = \frac{m_{t_2,\delta} + 1}{u_\delta v}, \quad \frac{n_{t_1,\delta}}{u_\delta} = \frac{n_{t_2,\delta}}{u_\delta} = \frac{m_{t_1,\delta}}{u_\delta} = \frac{m_{t_2,\delta}}{u_\delta}$$

for all $t \neq t_1$. In particular, the new integers $n_{t_1,\delta}, n_{t_2,\delta}, u_\delta \in \mathbb{N}$ also satisfy condition (1) for all $r$ sufficiently large. On the other hand, we will show that $\gcd(n_{t_1,\delta}, m_{t_2,\delta}) = 1$ and $\gcd(m_{t_1,\delta}, m_{t_2,\delta}) = 1$. To see this, suppose $p$ is a prime such that $p | n_{t_1,\delta}$ and $p | m_{t_2,\delta}$. Then since $p | n_{t_2,\delta}$, either $p | n_{t_2,\delta}$ or $p | v$. But if $p | n_{t_2,\delta}$, then $p | v$ also by definition of $v$. Hence, in either case $p | v$. Therefore, both $p | n_{t_1,\delta}$ and $p | m_{t_2,\delta}$. However, it then follows that $p | (n_{t_1,\delta} - n_{t_1,\delta} \cdot v) = 1$, a contradiction. In particular, we deduce that $\gcd(n_{t_1,\delta}, m_{t_1,\delta}) = 1$ as claimed. The same argument shows that $\gcd(m_{t_1,\delta}, m_{t_2,\delta}) = 1$. In particular, the new integers $n_{t_1,\delta}, n_{t_2,\delta}, u_\delta \in \mathbb{N}$ satisfy both conditions (1) and (2) for all $r$ sufficiently large. Moreover, it is straightforward to check that $n_{t_1,\delta} \neq n_{t_2,\delta}$ and $m_{t_1,\delta} \neq m_{t_2,\delta}$ for all $0 < \delta < \delta_T$ by condition (1) alone; here $\delta_T$ is the constant from Remark 6, guaranteed to exist since $T$ is uniformly discrete.

With these preliminaries in place, we are now ready to prove our first main result, a count on the number of functions $f \in M_S$ yielding a bounded height relation $H_\eta(f(P)) \leq B$ when $M_S$ is free. This is statement (2) of Theorem 1.2 from the introduction.

**Theorem 2.4.** Let $K$ be a global field, let $V$ be a projective variety over $K$, and let $S$ be a height controlled and uniformly log-discrete set of endomorphisms on $V$ with respect to a divisor class $\eta \in \text{Pic}(V) \otimes \mathbb{R}$. If $M_S$ is free (non-abelian), then for all $\epsilon > 0$ there is a positive constant $b = b(S, \eta, \epsilon)$ and a constant $B_{S, \eta}$ such that

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

for all $P \in V$ with $H_\eta(P) > B_{S, \eta}$. Here the implicit constants depend on $P, S, \eta$ and $\epsilon$.

**Proof.** Assume that $S$ is height controlled and uniformly log-discrete (with respect to a divisor class $\eta$) and that $M_S$ is free. In particular, if

$$T = \{ \log \deg_\eta(\phi) : \phi \in S \},$$

then we have a bijection of sets $\chi : M_S \to \text{Seq}(T)$ given by

$$\chi(\theta_1 \circ \cdots \circ \theta_m) = (\log \deg_\eta(\theta_1), \ldots, \log \deg_\eta(\theta_m)), \quad \theta_i \in S.$$
Moreover, since $\deg_{\eta}(F \circ G) = \deg_{\eta}(F) \deg_{\eta}(G)$ for all $F, G \in M_S$ by definition of $\deg_{\eta}(\cdot)$, we see from the identification in (2) that $\log \deg_{\eta}(f) = |x(f)|_{T}$ for all $f \in M_S$; here $|\cdot|_{T}$ is defined as in (1). Therefore,

\[
(3) \quad \# \{ f \in M_S : \log \deg_{\eta}(f) \leq B \} = \# \{ \omega \in \text{Seq}(T) : |\omega|_{T} \leq B \}
\]

holds for all $B$. Now for all $\delta > 0$ and $t \in T$, we may choose $n_{t, \delta}, m_{t, \delta}, u_{t} \in \mathbb{N}$ as in Lemma 2.3. Moreover, we may assume that $\delta > 0$ is sufficiently small to ensure that $n_{t, \delta} \neq n_{t', \delta}$ and $m_{t, \delta} \neq m_{t', \delta}$ for all distinct $t, t' \in T$. In particular, the map $\chi_1 : \text{Seq}(T) \to \text{Seq}(T_{1, \delta})$ given by

\[
\chi_1((t_1, \ldots, t_r)) = (n_{t_1, \delta}, \ldots, n_{t_r, \delta})
\]

is injective (in-fact, bijective by definition). Moreover, part (1) of Lemma 2.3 implies that

\[
\frac{|\chi_1(\omega)|_{T_{1, \delta}}}{u_{\delta}} \leq |\omega|_{T} \quad \text{for all } \omega \in \text{Seq}(T).
\]

Therefore, we deduce that

\[
(4) \quad \# \{ \omega \in \text{Seq}(T) : |\omega|_{T} \leq B \} \leq \# \{ \omega_1 \in \text{Seq}(T_{1, \delta}) : |\omega_1|_{T_{1, \delta}} \leq u_{\delta} B \}.
\]

Similarly, the map $\chi_2 : \text{Seq}(T_{2, \delta}) \to \text{Seq}(T)$ given by

\[
\chi_2((m_{t_1, \delta}, \ldots, m_{t_r, \delta})) = (t_1, \ldots, t_r)
\]

is well-defined and injective, and hence

\[
(5) \quad \# \{ \omega_2 \in \text{Seq}(T_{2, \delta}) : |\omega_2|_{T_{2, \delta}} \leq u_{\delta} B \} \leq \# \{ \omega \in \text{Seq}(T) : |\omega|_{T} \leq B \}
\]

follows from part (1) of Lemma 2.3. Therefore, we deduce from (3), (4) and (5) that

\[
(6) \quad \# \{ \omega_2 \in \text{Seq}(T_{2, \delta}) : |\omega_2|_{T_{2, \delta}} \leq u_{\delta} B \} \leq \# \{ f \in M_S : \log \deg_{\eta}(f) \leq B \}
\]

\[
\leq \# \{ \omega_1 \in \text{Seq}(T_{1, \delta}) : |\omega_1|_{T_{1, \delta}} \leq u_{\delta} B \}.
\]

However, the sets $T_{i, \delta} \subseteq \mathbb{N}$ and so we can count the right and left sides of (6) using generating functions. To make this precise let $G_{i, \delta}(z) = G_{T_{i, \delta}}(z) = \sum_{t \in T_{i, \delta}} z^{t}$, let $\beta_{i, \delta}$ satisfy $G_{T_{i, \delta}}(\beta_{i, \delta}^{-1}) = 1$, and let $0 < \epsilon' < 1$. Then Theorem 2.2 and Remark 5, together with (6) imply

\[
(7) \quad \frac{(1-\epsilon')\beta_{2, \delta}}{(\beta_{2, \delta}-1) G_{T_{2, \delta}}(\frac{\beta_{2, \delta}}{\beta_{2, \delta}-1})} \beta_{2, \delta}^{u_{\delta} B} - O_{\epsilon'}(1) \leq \# \{ f \in M_S : \log \deg_{\eta}(f) \leq B \}
\]

\[
\leq \frac{(1+\epsilon')\beta_{1, \delta}}{(\beta_{1, \delta}-1) G_{T_{1, \delta}}(\frac{\beta_{1, \delta}}{\beta_{1, \delta}-1})} \beta_{1, \delta}^{u_{\delta} B} + O_{\epsilon'}(1)
\]

holds for all $B$ sufficiently large (depending on $\delta$ and $\epsilon'$). On the other hand, a count on the number of functions of bounded log degree yields a count on the number of functions determining a bounded height relation (by Tate’s Telescoping Lemma). Specifically, suppose that $P \in V(K)$ is such that $h_{\eta}(P) > b_{S, \eta} := c_{S, \eta}/(d_{S, \eta} - 1)$, where $c_{S, \eta}$ and $d_{S, \eta}$ are the constants from Definition 1.1 above. Then, Tate’s Telescoping Lemma 2.1 implies that

\[
\deg_{\eta}(f)(h_{\eta}(P) - b_{S, \eta}) \leq h_{\eta}(f(P)) \leq \deg_{\eta}(f)(h_{\eta}(P) + b_{S, \eta}).
\]

Therefore, for all $B$ we have the subset relations:

\[
(8) \quad \left\{ f \in M_S : \log \deg_{\eta}(f) \leq \log \left( \frac{B}{h_{\eta}(P) + b_{S, \eta}} \right) \right\} \subseteq \left\{ f \in M_S : h_{\eta}(f(P)) \leq B \right\}
\]

\[
\subseteq \left\{ f \in M_S : \log \deg_{\eta}(f) \leq \log \left( \frac{B}{h_{\eta}(P) - b_{S, \eta}} \right) \right\}.
\]

In particular, if we replace $B$ with $\log(B/(h_{\eta}(P) + B_{S, \eta}))$ on the left side of (7), replace $B$ with $\log(B/(h_{\eta}(P) - B_{S, \eta}))$ on the right side of (7), and apply the change of base formulas for logarithms, then we deduce from (7) and (8) that
\[
\frac{(1-\epsilon')\beta_{2,\delta}}{(\beta_{2,\delta} - 1) G_{T_2,\delta}(\frac{1}{\beta_{2,\delta}})}\left(\frac{B}{h_\eta(P) + b_{S,\eta}}\right)^{u_4 \log(\beta_{2,\delta})} = O_{\epsilon',\delta}(1)
\]

\[
\leq \# \{ f \in M_S : h_\eta(f(P)) \leq B \}
\]

\[
\leq \frac{(1+\epsilon')\beta_{1,\delta}^3}{(\beta_{1,\delta} - 1) G_{T_1,\delta}(\frac{1}{\beta_{1,\delta}})}\left(\frac{B}{h_\eta(P) - b_{S,\eta}}\right)^{u_4 \log(\beta_{1,\delta})} + O_{\epsilon',\delta}(1)
\]

Moreover, since point counting on varieties is normally phrased in terms of multiplicative heights, we replace \( B \) with \( \log B \) in the bounds directly above to deduce that

\[(9) \quad C_{(2,P,\epsilon',\delta)} \log(B)^{b_{2,\delta}} - O_{\epsilon',\delta}(1) \leq \# \{ f \in M_S : H_\eta(f(P)) \leq B \} \leq C_{(1,P,\epsilon',\delta)} \log(B)^{b_{1,\delta}} + O_{\epsilon',\delta}(1).
\]

Here the explicit constants \( C_{(i,P,\epsilon',\delta)} \) and the inexplicit \( O_{\epsilon',\delta}(1) \) both depend on \( S \) and \( \eta \) as well, though we leave this dependence off in the notation. Specifically,

\[(10) \quad C_{(2,P,\epsilon',\delta)} = \frac{(1-\epsilon')\beta_{2,\delta}}{(\beta_{2,\delta} - 1) G_{T_2,\delta}(\frac{1}{\beta_{2,\delta}})(h_\eta(P) + b_{S,\eta})^{u_4 \log(\beta_{2,\delta})}}, \quad b_{2,\delta} = u_4 \log(\beta_{2,\delta}),
\]

\[C_{(1,P,\epsilon',\delta)} = \frac{(1+\epsilon')\beta_{1,\delta}^3}{(\beta_{1,\delta} - 1) G_{T_1,\delta}(\frac{1}{\beta_{1,\delta}})(h_\eta(P) - b_{S,\eta})^{u_4 \log(\beta_{1,\delta})}}, \quad b_{1,\delta} = u_4 \log(\beta_{1,\delta}).
\]

In particular by fixing any \( 0 < \epsilon' < 1 \), we obtain the desired shape of the bound on heights in orbits in Theorem 2.4 from (9). It therefore remains to show that the positive difference \( b_{1,\delta} - b_{2,\delta} \) can be made arbitrarily small (i.e., \( < \epsilon \)) by letting \( \delta \) tend to zero. To see this we need the following convenient lower bound on \( \alpha_{1,\delta} \); compare to [13, Equation (26)].

**Lemma 2.5.** Let \( T \) and \( T_{1,\delta} \) be as above and let \( t_1 := \min(T) \). Then there exist positive constants \( b_T \) and \( c_T \) (depending only on \( T \)) such that \( c_T \frac{t}{t_1} \leq \alpha_{1,\delta} \) for all \( 0 \leq \delta \leq \delta_T \).

**Proof.** We begin with a few observations about \( T \). Since \( T \) is uniformly discrete, we may choose a constant \( \delta_T > 0 \) such that \( |t - t'| > \delta_T \) for all distinct \( t, t' \in T \). Hence for all \( 0 < \delta < \delta_T \), it follows from Lemma 2.3 condition (1) that \( n_{t,\delta} \neq n_{t',\delta} \) for all \( t \neq t' \). Therefore, the association \( t \mapsto n_{t,\delta} \) is a bijection of sets \( T \mapsto T_{1,\delta} \). In particular, \( T \) is countable. Moreover, \( T \) must have a unique minimum element. To see this, note that for any fixed \( t \in T \) there can be at most finitely many elements of \( T \) less than \( t \). Explicitly, we have that \( \# \{ t' \in T : t' < t \} \leq \frac{1}{\delta_T} \) by definition of \( \delta_T \) and the fact that every element of \( T \) is positive. Putting these facts together, we may enumerate \( T = \{ t_1, t_2, \ldots \} \) and assume that \( t_1 < t_2 < \ldots \), i.e., the \( t_i \)'s are arranged in increasing order. Now assume that \( \delta < \delta_T \). Then Lemma 2.3 condition (1) implies that

\[
\frac{t - \delta_T}{t_1} < \frac{t - \delta}{t_1} < \frac{n_{t,\delta}}{n_{t_1,\delta}}
\]

for all \( t \in T \). Moreover writing \( t_m - \delta_T \) as a telescoping sum, we see that

\[
t_m - \delta_T = (t_m - t_{m-1}) + (t_{m-1} - t_{m-2}) + \cdots + (t_2 - t_1) + (t_1 - \delta_T) > (m-1)\delta_T + (t_1 - \delta_T).
\]

In particular, if we define the constant \( e_T := \delta_T/t_1 \) (independent of \( \delta \)), then the bounds above together imply that

\[(m-2)e_T + 1 < \frac{t_m - \delta_T}{t_1} < \frac{n_{t_m,\delta}}{n_{t_1,\delta}}.
\]
for all indices $m > 1$ (this bound remains true for $m = 1$, but we do not use it in that case).
With this bound in mind, consider the auxiliary function
\[ g_T(z) = z + \sum_{m \geq 2} z e^{(m-2)\epsilon + 1} = z + z \sum_{i=0}^{\infty} (z e^\epsilon)^i. \]

In particular, it is clear that $g_T$ is convergent on $[0, 1]$; in fact, $g_T = z + z/(1 - e^\epsilon)$ on this interval. With these preliminaries in place we reach the key fact:

(11) \[ 1 = \sum_{n_1, \delta \in T, \epsilon} (\alpha_{i, \delta})^{n_1, \delta} = \sum_{t \in T} (\alpha_{i, \delta})^{n_1, \delta} = (\alpha_{i, \delta})^{n_1, \delta} + \sum_{t \in T} ((\alpha_{i, \delta})^{n_1, \delta})^{n_2, \delta} \leq g_T(\alpha_{i, \delta}^{n_1, \delta}), \]

since $(\alpha_{i, \delta})^{n_1, \delta} \in (0, 1)$ and smaller exponents yield larger values on this interval. On the other hand, $g_T$ is continuous, $g_T(0) = 0$ and $1 \leq g_T(\alpha_{i, \delta}^{n_1, \delta})$ by (11). Therefore, we may choose a constant $c_T \in (0, 1)$ such that $g_T(c_T) = 1/2$ by the Intermediate Value Theorem. Most importantly since $g_T$ is an increasing function we must have that $c_T \leq (\alpha_{i, \delta})^{n_1, \delta}$, independent of $\delta$ as claimed.

From this point, we can complete the proof of Theorem 1.2 by following the proof of [13, Theorem 1.1]. To do this, we use the Mean Value Theorem applied to the functions $G_2, \delta(\alpha) = \sum_{t \in T} z^{n_1, \delta}$ and $h_\delta(z) = u_\delta \log z$ on the intervals $[\alpha_{i, \delta}, \alpha_{2, \delta}]$. Here and for the remainder of the proof, we assume that $0 < \delta < \min \{ \delta_T, t_1 \}$; recall that $t_1 := \min(T)$. In particular, the sets $T$, $T_1, \delta$ and $T_2, \delta$ are all in bijection, and so we can use any of them for indexing sets for the countable, absolutely convergent sums below. We begin with a few estimates, all of which follow easily from part (1) of Lemma 2.3.

(12) \[ \frac{2\delta}{t_1} < \frac{2\delta u_\delta}{m_{t_1, \delta}} < \frac{2\delta}{t_1 - \delta}, \quad 1 < \frac{m_{t_1, \delta}}{m_{t_2, \delta}} < \frac{t_1 + \delta}{t_1 - \delta}, \quad \frac{u_\delta}{m_{t_1, \delta}} < \frac{1}{t_1}. \]

Likewise, we have the lower bound on $\alpha_{i, \delta}$ from Lemma 2.5:

(13) \[ c_T^{-n_1, \delta} \leq \alpha_{i, \delta}. \]

In particular, (12) and (13) together imply the following lower bound on the derivative:

(14) \[ G_2, \delta'(\alpha_{i, \delta}) = \sum_{t \in T} m_{t, \delta}(\alpha_{i, \delta})^{n_1, \delta - 1} \geq m_{t_1, \delta}(\alpha_{i, \delta})^{n_1, \delta - 1} \]
\[ \geq m_{t_1, \delta}(\alpha_{i, \delta})^{n_1, \delta} \geq m_{t_1, \delta} c_T^{m_{t_1, \delta} - n_1, \delta} \geq m_{t_1, \delta} e^\epsilon c_T^{e^\epsilon - 1}. \]

Similarly, (12) and (13) together imply that:

\[ G_2, \delta'(\alpha_{s, \delta}) = \sum_{t \in T} \alpha_{i, \delta}^{m_{t, \delta}} = \sum_{t \in T} \alpha_{i, \delta}^{m_{t_1, \delta}} \cdot \alpha_{i, \delta}^{m_{t_2, \delta}} \]
\[ \geq \sum_{t \in T} \alpha_{i, \delta}^{2\delta u_\delta} \cdot \alpha_{i, \delta}^{m_{t_1, \delta}} \]
\[ = \alpha_{i, \delta}^{2\delta u_\delta} \cdot \alpha_{i, \delta}^{m_{t_1, \delta}} \]
\[ = \alpha_{i, \delta}^{2\delta u_\delta} \geq c_T^{2\delta u_\delta} \geq c_T^{2\delta u_\delta - 1}. \]

Here, we use also that $0 < m_{t_1, \delta} - m_{t_2, \delta} < 2\delta$ by construction; see Lemma 2.3 part (1). In particular, we deduce the following key upper bound:

(15) \[ G_2, \delta(\alpha_{2, \delta}) - G_2, \delta(\alpha_{1, \delta}) = 1 - G_2, \delta(\alpha_{1, \delta}) \leq 1 - c_T^{2\delta u_\delta}. \]

We are now ready to apply the Mean Value Theorem to $G_2, \delta(\alpha)$ on $[\alpha_{1, \delta}, \alpha_{2, \delta}]$. Specifically,

\[ \frac{m_{t_1, \delta} c_T^{n_1, \delta - 1}}{\alpha_{i, \delta}^{n_1, \delta}} \leq G_2'(\alpha_{i, \delta}) = \min_{\alpha_{i, \delta} \leq \alpha \leq \alpha_{2, \delta}} G_2'(\alpha) \leq \frac{G_2(\alpha_{2, \delta}) - G_2(\alpha_{1, \delta})}{\alpha_{2, \delta} - \alpha_{1, \delta}} \leq 1 - c_T^{2\delta u_\delta}. \]
follows from (14), (15), and the Mean Value Theorem. Therefore, we have the estimate:

\begin{equation}
0 \leq \alpha_{2,\delta} - \alpha_{1,\delta} \leq \frac{1 - c\delta}{m_{t_1,\delta} c T} \cdot \frac{1}{t_1^{1+\gamma}} .
\end{equation}

Likewise, the Mean Value Theorem for $h_\delta(x) = u_\delta \log(x)$ on $[\alpha_{1,\delta}, \alpha_{2,\delta}]$, (13), and the fact that $n_{1,\delta} > 0$ together yield

\begin{equation}
0 \leq \frac{h_\delta(\alpha_{2,\delta}) - h_\delta(\alpha_{1,\delta})}{\alpha_{2,\delta} - \alpha_{1,\delta}} \leq \max_{\alpha_{1,\delta} \leq z \leq \alpha_{2,\delta}} h_\delta'(x) = h_\delta'(\alpha_{1,\delta}) = u_\delta(\alpha_{1,\delta})^{-1} \leq \frac{u_\delta}{c T} .
\end{equation}

Hence, after combining (10),(12), (16) and (17), we deduce that

\begin{equation}
0 \leq b_{1,\delta} - b_{2,\delta} = h_\delta(\alpha_{2,\delta}) - h_\delta(\alpha_{1,\delta}) \leq \frac{u_\delta}{c T} \cdot \frac{1 - (c T)^{rt}}{m_{1,\delta} c T} \cdot \frac{1}{t_1^{1+\gamma}} \leq \frac{1}{c T} \cdot \frac{1 - (c T)^{rt}}{(c T)^{rt}} \cdot \frac{1}{t_1^{1+\gamma}} .
\end{equation}

However, the upper bound in (18) goes to zero as $\delta$ goes to zero (since $c T > 0$). Therefore, the exponents $b_{1,\delta}$ and $b_{2,\delta}$ in (9) can be made arbitrarily close as claimed.

**Remark 7.** To approximate $b_{1,\delta}$ and $b_{2,\delta}$ in practice, we first approximate the solutions to $G_\delta(z) = 1$ by similar solutions for associated cutoff functions (which are rational). To wit, let $N > 0$ be any positive integer and define the rational functions

\begin{equation*}
G_{1,\delta,N}(z) := \sum_{n \in N \in \mathbb{N}, \alpha_{1,\delta}} z^n + \frac{z^N}{1 - z} \quad \text{and} \quad G_{2,\delta,N}(z) := \sum_{m \in N \in \mathbb{N}, \alpha_{2,\delta}} z^m
\end{equation*}

associated to $G_{1,\delta}$ and $G_{2,\delta}$ respectively. Then since $n_t < m_t$ for all $t \in T$ by construction, we see that

\begin{equation}
G_{2,\delta,N}(z) \leq G_{2,\delta}(z) < G_{1,\delta}(z) \leq G_{1,\delta,N}(z) \quad \text{for all } 0 < z < 1 .
\end{equation}

Now let $\alpha_{i,\delta} \in (0,1)$ satisfy $G_{i,\delta}(\alpha_{i,\delta}) = 1$ and let $\alpha_{i,\delta,N} \in (0,1)$ satisfy $G_{i,\delta,N}(\alpha_{i,\delta,N}) = 1$. In particular, $\beta_{i,\delta} = (\alpha_{i,\delta})^{-1}$ and $b_{1,\delta} = u_\delta \log(\beta_{1,\delta})$ by definition of $\beta_{1,\delta}$ and (10) above. Likewise, we define $\beta_{i,\delta,N} = (\alpha_{i,\delta,N})^{-1}$ and $b_{i,\delta,N} = u_\delta \log(\beta_{i,\delta,N})$. Then (19) implies that

\begin{equation}
\alpha_{1,\delta,N} \leq \alpha_{1,\delta} < \alpha_{2,\delta} \leq \alpha_{2,\delta,N}
\end{equation}

\begin{equation}
\beta_{2,\delta,N} \leq \beta_{2,\delta} < \beta_{1,\delta} \leq \beta_{1,\delta,N}
\end{equation}

\begin{equation}
b_{2,\delta,N} \leq b_{2,\delta} < b_{1,\delta} \leq b_{1,\delta,N} .
\end{equation}

Here we use also that both the $G_{i,\delta}$ and the $G_{i,\delta,N}$ are strictly increasing on $(0,1)$. Moreover, $\alpha_{1,\delta,N}$ and $\alpha_{2,\delta,N}$ are algebraic, and we can (in practice) estimate these quantities using known root finding algorithms (and Magma [4]) to any degree of accuracy. In this way, we can give explicit bounds for the growth rate of $\# \{ f \in M_S : H_\eta(f(P)) \leq B \}$ in practice. We use this in section 3 to give explicit bounds for the number of points of bounded height in semigroup orbits generated by the polynomials $z^q + 1$, where $q$ is in various subsets of the prime numbers (e.g., the Mersenne primes); see, for instance, Corollary 1.4 from the introduction.

Lastly, we can use the bounds in Theorem 2.4 on the number of functions in free semigroups satisfying a bounded height relation to give an upper bound on the number of points of bounded height in arbitrary semigroup orbits. This is part (1) of Theorem 1.2 from the introduction.

**Corollary 2.6.** Let $K$ be a global field, let $V$ be a projective variety over $K$, and let $S$ be a height controlled and uniformly log-discrete set of endomorphisms on $V$ with respect to a divisor class $\eta \in \text{Pic}(V) \otimes \mathbb{R}$. Then there is a positive constant $b = b(S, \eta)$ and a constant $B_{S,\eta}$ such that

\begin{equation*}
\# \{ Q \in \text{Orb}_S(P) : H_\eta(Q) \leq B \} \ll (\log B)^b
\end{equation*}
for all $P \in V$ with $H_q(P) > B_{S,q}$.

**Proof.** Suppose $S$ is a height controlled and uniformly log-discrete set of endomorphisms on $V$. We may assume that $S$ has at least two elements: otherwise the desired bound, in fact with a power of $\log(B)$ replaced with $\log(\log(B))$, follows easily from the celebrated results of Call and Silverman [6]. Now let $F_S$ be the free semigroup 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 $V$ via $w \cdot P = \theta_1 \circ \cdots \circ \theta_n(P)$. Likewise, we define the $n$-degree of $w$ to be $\deg_n(\theta_1 \circ \cdots \circ \theta_n)$. 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 2.4, choose any $0 < \delta < \delta_T$, and deduce that

$$\log(B)^{b_{2,\delta}} \ll \#\{w \in F_S : H_q(w \cdot P) \leq B\} \ll \log(B)^{b_{1,\delta}}$$

for all $P$ with $H_q(P) > B_S$ as before; see (9). 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

$$\#\{Q \in \text{Orb}_S(P) : H_q(Q) \leq B\} \leq \#\{w \in F_S : H_q(w \cdot P) \leq B\} \ll \log(B)^{b_{1,\delta}}.$$

Therefore, the number of points in $\text{Orb}_S(P)$ with height at most $B$ is $\ll \log(B)^{b_{1,\delta}}$.□

3. Free semigroups generated by unicritical polynomials

In order to apply both the upper and lower bounds from Theorem 1.2 to some explicit, infinitely generated examples, we need three properties to hold: the full semigroup is free, the underlying generating set is height controlled, and the set of degrees of the generating set is uniformly log-discrete. This last property is the easiest to ensure; if we can find an infinitely generated example that is free and height controlled, then we can restrict to a subset of the generating set on which all three properties hold. Likewise freeness, though it can be nontrivial to prove, should hold generically; see, for instance, [13, Proposition 4.1]. Therefore, the property of being height controlled is the most restrictive of the three (for infinite generating sets). Nevertheless, we can produce examples of infinite height controlled sets using unicritical polynomials as in [12, Example 2.7]. With this in mind, we first show that these polynomials generate free semigroups under composition. This fact is perhaps well known to the experts (at least in the finitely generated case), but without a suitable reference we include a proof here:

**Theorem 3.1.** Let $k$ be a field of characteristic zero and let $S = \{z^d + c : d \geq 2, c \in k^*\}$. Then the semigroup generated by $S$ under composition is free.

As a first step, we note that there are no non-constant solutions to a Cassels-Catalan equation over the polynomial ring $k[z]$, a simple consequence of Mason’s ABC theorem [18].

**Lemma 3.2.** Let $k$ be a field of characteristic zero, let $a, b, c \in k^*$, and let $m, n \geq 2$. Then there are no non-constant solutions $x, y \in k[z]$ to the equation $ax^n + by^m = c$.

**Proof.** Suppose that $x, y \in k[z]$ are non-constant solutions to the equation $ax^n + by^m = c$. Note that since $a, b, c \in k^*$, the triple $(A, B, C) = (ax^n, by^m, c)$ is pairwise coprime in $k[z]$. In particular, Mason’s ABC Theorem over $k[z]$ implies that

$$\max\{n \deg(x), m \deg(y)\} = \max\{\deg(A), \deg(B), \deg(C)\}$$

(21)

$$\leq \deg(\text{rad}(ABC)) - 1 \leq \deg(x) + \deg(y) - 1;$$

here rad$(f)$ denotes the number of distinct roots of $f \in k[z]$. On the other hand $n, m \geq 2$, hence

$$2 \deg(x) \leq \deg(x) + \deg(y) - 1 \quad \text{and} \quad 2 \deg(y) \leq \deg(x) + \deg(y) - 1.$$ 

In particular, these inequalities together imply that

$$\deg(x) \leq \deg(y) - 1 \leq \deg(x) - 2,$$

a contradiction. Therefore, there are no non-constant solutions $x, y \in k[z]$ to the equation $ax^n - by^m = c$ as claimed. □
(Proof of Theorem 3.1). Now suppose that we have a relation
\[(22) \quad \theta_1 \circ \cdots \circ \theta_r = \tau_1 \circ \cdots \circ \tau_k \]
for some \( \theta_i, \tau_j \in S \). We will show that \( r = k \) and \( \theta_i = \tau_i \) for all \( 1 \leq i \leq r \). To do this, write \( \theta_1 = z^{d_1} + c_1 \) and \( \tau_1 = z^{d_2} + c_2 \). Clearly if \( n = m = 1 \), then \( \theta_1 = \tau_1 \) and there is nothing to prove. Therefore, we may assume without loss of generality that \( r \geq k \) and \( r \geq 2 \). Let \( f_1 = \theta_2 \circ \cdots \circ \theta_r \), \( g_1 = \tau_2 \circ \cdots \circ \tau_k \) if \( k \geq 2 \), and let \( g_1 \) be the identity if \( k = 1 \) (all non-constant polynomials). Then (22) implies that
\[(23) \quad f_1^{d_1} - g_1^{d_2} = c_2 - c_1.\]
In particular, \( c_1 = c_2 \), since otherwise \((x, y) = (f_1(z), g_1(z))\) is a non-constant solution to the Cassels-Catalan equation given by \((a, b, c, n, m) = (1, -1, c_1 - c_2, d_1, d_2)\), which contradicts Lemma 3.2. Therefore, (22) becomes \( f_1^{d_1} = g_1^{d_2} \). Now suppose \( d = \gcd(d_1, d_2) \) so that \( d_1 = e_1d \) and \( d_2 = e_2d \) for some coprime positive integers \( e_1 \) and \( e_2 \). Then
\[(f_1^{e_1})^d = f_1^{d_1} = g_1^{d_2} = (g_1^{e_2})^d\]
and thus \((f_1^{e_1})/(g_1^{e_2})\) is a \( d \)-th root of unity (in particular, a constant function). Hence, we may write \( f_1^{e_1} = \zeta_d g_1^{e_2} \) for some \( d \)-th root of unity \( \zeta_d \). On the other hand \( f_1, g_1 \in M_S \), and so \( f_1 \) and \( g_1 \) are both monic polynomials. Likewise, \( f_1^{e_1} \) and \( g_1^{e_2} \) are monic. In particular, the relation \( f_1^{e_1} = \zeta_d g_1^{e_2} \) forces \( \zeta_d = 1 \), and we deduce that \( f_1^{e_1} = g_1^{e_2} \) for some coprime exponents \( e_1 \) and \( e_2 \). Therefore after writing \( 1 = u e_1 + v e_2 \) for some \( u, v \in \mathbb{Z} \), we see that
\[f_1 = f_1^{u e_1 + v e_2} = (f_1^{u e_1})^{f_1^{v e_2}} = (g_1^{v e_2})^{f_1^{u e_1}} = (f_1^{u e_1} g_1^{v e_2})^{f_1}.\]
Likewise, it is easy to check that \( g_1 = (f_1^{v e_2} g_1^{u e_1})^{e_1} \). In particular, if we let \( u = f_1^{v e_2} g_1^{u e_1} \in k(z) \), then we have shown that
\[(24) \quad f_1 = u^{e_1} \quad \text{and} \quad g_1 = u^{e_2}.\]
Note also that \( u \) must in fact lie in the polynomial ring, since it is integral over \( k[z] \) and \( k[z] \) is integrally closed. Now then suppose that either \( e_1 \) or \( e_2 \) is strictly bigger than one (we’ll show that this is impossible). Suppose first that \( e_2 > 1 \). Then writing \( f_1 = f_2^{d_3} + c_3 \) for some \( f_2 \in M_S \) (possibly the identity function), we see that (24) implies that
\[(25) \quad f_2^{d_3} + c_3 = u^{e_2}.\]
However, this contradicts Lemma 3.2 since in that case \((x, y) = (f_2(z), u(z))\) is a non-constant solution to the Cassels-Catalan equation given by \((a, b, c, n, m) = (1, -1, -c_3, d_3, e_2)\); here we use crucially that \( c_3 \neq 0 \) by definition of \( S \). Hence, \( e_2 \) must be equal to one. Repeating the same argument when \( k \geq 2 \) with the relation \( g_1 = u^{e_1} \) from (24), we see that \( e_1 \) must equal one also. On the other hand if \( k = 1 \), then \( g_1 = z \) and (24) immediately implies that \( e_1 = 1 \) (since \( z \) is irreducible). In particular \( e_1 = 1 = e_2 \) in either case, and it follows by definition of \( e_1 \) and \( e_2 \) that \( d_1 = d = d_2 \). Hence, coupled with the fact that we have already shown that \( c_1 = c_2 \), we deduce that \( \theta_1 = \tau_1 \). Therefore, (23) implies that \( f_1^{d_1} = g_1^{d_2} \) so that \( f_1 = \zeta_d g_1 \) for some root of unity \( \zeta_d \). However, again \( f_1 \) and \( g_1 \) are both in \( M_S \) and therefore monic.
In particular, \( \zeta_d = 1 \) and \( f_1 = g_1 \). Note that this precludes the possibility that \( k = 1 \), since otherwise \( d_3 = \deg(f_1) = \deg(g_1) = 1 \) by simply equating degrees.
To recap, we have shown that \( \theta_1 \circ \cdots \circ \theta_r = \tau_1 \circ \cdots \circ \tau_k \) for some \( \theta_i, \tau_j \in S \) and some \( r \geq 2 \), then \( k \geq 2 \), \( \theta_1 = \tau_1 \), and \( \theta_2 \circ \cdots \circ \theta_r = f_1 = g_1 = \tau_2 \circ \cdots \circ \tau_k \). Repeating this argument until the \( \tau \)'s are eliminated (possible since \( r \geq k \)), we see that \( \theta_i = \tau_i \) for all \( 1 \leq i \leq k \). On the other hand if \( r > k \), then after equating the degrees in the original expression \( \theta_1 \circ \cdots \circ \theta_r = \tau_1 \circ \cdots \circ \tau_k \), we see that \( \deg(\theta_{k+1}) \cdots \deg(\theta_r) = 1 \). But, this is impossible since each \( \theta \) in \( S \) has degree at least 2. Hence, \( r = k \) and the expression in (22) is the trivial one. Therefore, \( M_S \) is free as claimed.

Remark 8. Note that if we include the constant terms \( c = 0 \) in \( S \), then \( M_S \) will not be free (monic power maps commute). Likewise, if we allow non-monic unital critical polynomials, then \( M_S \) may fail to be free. For instance, if \( \phi_1 = z^2 + 1 \) and \( \phi_2 = -\phi_1 = -z^2 - 1 \), then \( \phi_1 \circ \phi_1 = \phi_1 \circ \phi_2 \) is a non-trivial compositional relation. In this way, Theorem 3.1 is in some sense the strongest possible statement.
In particular, we can put Theorem 3.1 together with [15, Lemma 12] to produce infinitely generated semigroups for which the upper and lower bounds in Theorem 1.2 hold. Recall that \( \mathfrak{d} \subset \mathbb{N} \) is called uniformly log-discrete if there is a constant \( \delta \) such that \( |\log d - \log d'| > \delta \) for all distinct \( d, d' \in \mathfrak{d} \).

**Corollary 3.3.** Let \( \mathfrak{d} \subset \mathbb{N} \) be uniformly log-discrete, and let \( c \in \overline{\mathbb{Q}}^+ \). Then
\[
S_{\mathfrak{d}, c} := \{ z^d + c : d \in \mathfrak{d}, d \geq 2 \}
\]
is height controlled and uniformly log-discrete. Moreover, the semigroup generated by \( S_{\mathfrak{d}, c} \) under composition is free. In particular, statement (2) of Theorem 1.2 holds for \( S_{\mathfrak{d}, c} \).

**Proof.** It follows from [15, Lemma 12] that
\[
(25) \quad |h(\phi(P)) - \deg(\phi)h(P)| \leq h(c) + \log 2
\]
for all \( P \in \mathbb{P}^1(\overline{\mathbb{Q}}) \) and all \( \phi \in S_{\mathfrak{d}, c} \). Hence, \( S_{\mathfrak{d}, c} \) is height controlled. Likewise, \( S_{\mathfrak{d}, c} \) is uniformly log-discrete by our assumption on \( \mathfrak{d} \). Finally, the semigroup generated by \( S_{\mathfrak{d}, c} \) under composition is free since \( S_{\mathfrak{d}, c} \subset S \) and \( M_S \) is free by Theorem 3.1. \( \square \)

**Remark 9.** In particular, \( S_{a, b, c} := \{ z^d + c : d = a^n + b \text{ for some } n \geq 1 \} \), from Example 2 above satisfy Corollary 3.3 and Theorem 1.2 part (2).

**Remark 10.** Technically, if we include finitely many \( c \)'s (i.e., constant terms), then \( S_{\mathfrak{d}, c} \) is still height controlled. However, allowing more than one \( c \) precludes \( S_{\mathfrak{d}, c} \) from being uniformly log-discrete.

4. **Orbits generated by unicritical polynomials: the Catalan case**

Now that we know that polynomials of the form \( z^d + c \) generate free semigroups, we would like to use Theorem 1.2 to count points (not just functions) of bounded height in orbits. To do this for any given basepoint \( P \), we need to control the number of distinct functions \( f \) and \( g \) in the relevant semigroup that can agree at \( P \), i.e., \( f(P) = g(P) \). Perhaps unsurprisingly, this is possible for finitely generated semigroups over \( \mathbb{P}^1 \) by invoking Siegel’s integral point theorem and the polynomial classification theorem of Bilu-Tichy [3]; essentially, you need only control the integral points on the finitely many curves \( f(x) = g(y) \) where \( f \) and \( g \) are in the generating set; see [13] for details. However, the problem is harder when one allows infinitely generated semigroups (which is in some sense the point of this paper), since there are infinitely many curves to control. On the other hand, there is a case where this is possible, namely when we restrict ourselves to polynomials of the form \( z^d + 1 \) (with prime exponent and constant term one); to do this, we use known results for integral solutions to the Catalan equations.

With this in mind, we set some notation used throughout this section. Fix a number field \( K \) and let \( \omega_K \) be the number of roots of unity in \( K \). Then we let \( S_K \) be the set of polynomials
\[
(26) \quad S_K := \{ z^q + 1 : q \text{ is prime and } \gcd(q, \omega_K) = 1 \}
\]
with prime degree (coprime to \( \omega_K \)) and constant term one. Likewise, since we need our generating sets to be uniformly log-discrete to apply Theorem 1.2, we consider subsets of \( S_K \) also. Namely, given a uniformly log-discrete set of primes \( q \) (meaning there is a constant \( \delta > 0 \) such that \( |\log(q) - \log(q')| > \delta \) for all distinct \( q, q' \in q \), we define
\[
(27) \quad S_{K, q} := \{ z^q + 1 : q \in q \text{ and } \gcd(q, \omega_K) = 1 \}.
\]
Then to pass from counting functions in \( M_{S_K} \) and \( M_{S_{K, q}} \) to counting points in orbits, we need the following result, which bounds the solutions to the Catalan equation over suitable rings of integers.

**Theorem 4.1 (Brindza [5]).** Let \( K \) be a number field, let \( \mathfrak{a} \) be a finite set of places of \( K \) containing the archimedean ones, and let \( \mathfrak{a}_{K, \mathfrak{a}} \) be the ring of \( \mathfrak{a} \)-integers of \( K \). Then there exists an effectively computable constant \( \kappa_\mathfrak{a} > 2 \) such that if
\[
(28) \quad x^n - y^m = 1
\]
for some \( n, m \in \mathbb{N} \) with \( nm > 4 \) and some \( x, y \in \mathfrak{a}_{K, \mathfrak{a}} \) which are not roots of unity, then
\[
\max\{ h(x), h(y), n, m \} \leq \kappa_\mathfrak{a}.
\]
Now, we may use Brindza’s bound on \( s \)-integral solutions to the Catalan equation to control the set of functions with common value at \( P \) in the semigroups generated by \((26)\) and \((27)\). In particular, in this step we do not need to assume that the set of degrees form a uniformly log-discrete set; our argument works for the full semigroup generated by \( z^2 + 1 \) and \( q \) prime, as long as \( \gcd(q, w_K) = 1 \). In what follows, given \( f = \theta_1 \circ \cdots \circ \theta_n \in M_{S_K} \) for some \( \theta_i \in S_K \), we define the length of \( f \) to be \( \ell(f) = n \). Note that this is well-defined since \( M_{S_K} \) is free.

**Lemma 4.2.** Let \( S_K \) be as in \((26)\) and suppose that \( P \in P^1(K) \) satisfies \( h(P) > \log 4 \). Then there exists a constant \( d_{K,P} \) (depending only on \( K \) and \( P \)) such that if
\[
\theta \circ f(P) = \tau \circ g(P)
\]
for some \( \theta, \tau \in S_K \) and some \( f, g \in M_{S_K} \), then either \( \theta = \tau \) and \( f(P) = g(P) \) or \( \max \{ \deg(\theta \circ f), \deg(\tau \circ g) \} \leq d_{K,P} \).

**Proof.** Assume that \((29)\) holds and choose a set of places \( s \) so that \( P \) is \( s \)-integral. Now define \( d_{K,P} \) depending on the constant \( \kappa_s \) from Lemma 4.2 as follows:
\[
d_{K,P} := \max \left\{ \frac{\kappa_s^2}{\log(2)}, \frac{\kappa_s^2}{\log 2} \right\};
\]
here \( h_K^\text{min} \) is the minimum height of the elements in \( K \) which are not roots of unity. To proceed, we argue in cases. In what follows, we write \( \theta(z) = z^{d_1} + 1 \) and \( \tau(z) = z^{d_2} + 1 \) for some primes \( d_1 \) and \( d_2 \) coprime to the number of roots of unity in \( K \).

**Case (1):** Suppose that \( \ell(f) = \ell(g) = 0 \), i.e., \( f \) and \( g \) are both the identity function. Then \((29)\) implies that \( P^{d_1} = P^{d_2} \), and therefore \( P^{d_1 - d_2} = 1 \). Hence if \( d_1 \neq d_2 \), then \( P \) is a root of unity. But this implies that \( h(P) = 0 \), contradicting our assumption that \( h(P) > \log 4 \). In particular, \( \theta = \tau \) and \( f = g \) in this case.

**Case (2):** Suppose that either \( \ell(f) = 0 \) or \( \ell(g) = 0 \), but not both. Without loss assume that \( \ell(f) \geq 1 \) and \( \ell(g) = 0 \), so we may write \( f = \omega \circ f_1 \) for some \( \omega(z) = z^{d_2} + 1 \) in \( S_K \) and some \( f_1 \in M_{S.K} \). Then \((29)\) implies that
\[
f(f_1)^{d_1} = P^{d_2}.
\]
In particular if \( \tau = \theta \), then \( d_1 = d_2 \) and \( f(P)/P \) is a \( d_1 \)'st root of unity. But \( d_1 \) is coprime to \( w_K \) (the number of roots of unity in \( K \)) and \( f(P)/P \in K \), so that \( f(P) = P \). On the other hand, the lower bound in Tate’s Telescoping Lemma 2.1 and \((25)\) imply that
\[
3(h(P) - \log 2) \leq \deg(f)(h(P) - \log 2) \leq h(f(P)) = h(P).
\]
However, this implies that \( 2h(P) \leq 3 \log 2 \), again violating our assumption that \( h(P) > \log 4 \). In particular, it must be the case that \( \tau \neq \theta \), and thus \( d_1 \neq d_2 \) are distinct odd primes. Therefore, there are integers \( a_1 \) and \( a_2 \) such that \( a_1 d_1 + a_2 d_2 = 1 \), from which it follows from \((30)\) that
\[
f(P) = t^{d_2} \quad \text{and} \quad P = t^{d_1}, \quad \text{for} \quad t = f(P)^{a_2} \cdot P^{a_1}.
\]
Note that \( t \) is \( s \)-integral since \( P \) is \( s \)-integral and the ring of \( s \)-integers is integrally closed. On the other hand, \( f = \omega \circ f_1 = f_1^{d_3} + 1 \) so that the left side of \((31)\) implies that \( (x, y, n, m) = (t, f_1(P), d_2, d_3) \) is an \( s \)-integral solution to the Catalan equation \((28)\). Moreover, \( f_1(P) \) cannot be a root of unity, since \( \log 2 \leq \deg(f_1)(h(P) - \log 2) = h(f_1(P)) \) by Tate’s Telescoping Lemma and our assumption on the height of \( P \). Likewise, \( t \) cannot be a root of unity since then \( P \) (which equals \( t^{d_1} \)) must be. Therefore, Theorem 4.1 implies that
\[
\max \{ h(t), h(f_1(P)), d_2, d_3 \} \leq \kappa_s.
\]
However, again applying Tate’s telescoping Lemma to \( f_1 \), we deduce from \((32)\) that
\[
\deg(f_1) \log 2 \leq \deg(f_1)(h(P) - \log 2) = h(f_1(P)) \leq \kappa_s.
\]
In particular, \( \deg(f_1) \leq \kappa_s/\log(2) \). Moreover, \( d_3 \) is also bounded by \( \kappa_s \) so that
\[
\deg(f) = \deg(f_1) \cdot d_1 \leq \kappa_s^2/\log(2).
\]
Likewise, \( P = t^{d_1} \) implies that
\[
d_1 h_K^\text{min} \leq d_1 h(t) = h(t^{d_1}) = h(P);
\]
here we use that $t \in K$ is not a root of unity. In particular, $d_1 \leq h(P)/h_K^{\min}$. Combining this fact with the bounds on (32) and (33), we see that

$$\max\{\deg(\theta \circ f), \deg(\tau \circ g)\} = \max\{d_1 \cdot \deg(f), d_2 \} \leq \max\left\{\frac{\kappa_s^2 h(P)}{\log(2) h_K^{\min}}, \kappa_s\right\} \leq d_{K,P}.$$ 

Here we use that $h(P) \geq h_K^{\min}$ and $\kappa_s > 1$. This completes the proof of the claim in Case (2).

**Case (3):** Suppose that $\ell(f), \ell(g) \geq 1$, so that we may write $f = \omega_1 \circ f_1$ and $g = \omega_2 \circ g_1$, where $\omega_1(z) = z^{s_1} + 1$, $\omega_2(z) = z^{s_2} + 1$, and $f_1, g_1 \in M_{S_K}$. If $\tau = \theta$ then $d_1 = d_2$ and $f(P)/g(P)$ is a $d_1$’st root of unity. But $d_1$ coprime to $w_K$ and $f(P)/g(P) \in K$, so that $f(P) = g(P)$. This is the first possible conclusion of Lemma 4.2.

Therefore, we may assume that $\tau \neq \theta$ and $d_1 \neq d_2$ are distinct odd primes. In particular, there are integers $a_1$ and $a_2$ such that $a_1 d_1 + a_2 d_2 = 1$. On the other hand (29) implies that $f(P)^{d_1} = g(P)^{d_2}$ from which it follows that

$$f_1(P)^{\epsilon_1} + 1 = f(P) = t^{d_2} \quad \text{and} \quad g_1(P)^{\epsilon_2} + 1 = g(P) = t^{d_1},$$

where $t = f(P)^{s_2} \cdot g(P)^{s_1}$. Note that $f_1(P), g_1(P)$ and $t$ are all $s$-integral since $P$ is $s$-integral, the coefficients of all relevant polynomials are integral, and the ring of $s$-integrals is integrally closed (needed for $t$ only). Moreover $f_1(P), g_1(P), f(P), g(P)$ cannot be roots of unity since, as argued in Case 2, Tate’s telescoping Lemma and the fact that $h(P) > \log 4$ together imply that each of these points has height at least $\log 2$. There is nothing special about these functions; no points in $\text{Orb}_{S_K}(P)$ can be roots of unity. Likewise $t$ cannot be a root of unity since otherwise $f(P)$, which equals $t^{d_2}$, must be. Therefore, (34) implies that $(x, y, n, m) = (t, f_1(P), d_2, e_1)$ and $(x, y, n, m) = (t, g_1(P), d_1, e_2)$ are $s$-integral solutions to the Catalan equation. In particular, the bound in Theorem 4.1 implies that

$$\max\{h(t), h(f_1(P)), h(g_1(P)), d_1, d_2, e_1, e_2\} \leq \kappa_s.$$ 

On the other hand, again applying Tate’s telescoping Lemma to $f_1$, we deduce from (35) that

$$\deg(f_1) \log 2 \leq \deg(f_1)h(P) - h(f_1) \leq \kappa_s - \log 2.$$ 

In particular, $\deg(f_1) \leq \kappa_s/\log(2)$. Likewise, we may deduce that $\deg(g_1) \leq \kappa_s/\log(2)$. These facts coupled with (35) together imply that

$$\deg(\theta \circ f) = \deg(\theta \circ \omega_1 \circ f_1) = d_1 \cdot e_1 \cdot \deg(f_1) \leq \frac{\kappa_s^2}{\log 2}$$

and

$$\deg(\tau \circ g) = \deg(\tau \circ \omega_2 \circ g_1) = d_2 \cdot e_2 \cdot \deg(g_1) \leq \frac{\kappa_s^2}{\log 2}.$$ 

In particular, $\max\{\deg(\theta \circ f), \deg(\tau \circ g)\} \leq \kappa_s^2/\log 2 \leq d_{K,P}$ by definition of $d_{K,P}$ as claimed. □

With Lemma 4.2 in place, we may define a few more constants depending only on the point $P \in K$ and the number field $K$:

$$t_{K,P} := \#\{f \in M_{S_K} : \deg(f) \leq d_{K,P}\} \quad \text{and} \quad r_{K,P} := \lfloor \log_2 d_{K,P} \rfloor;$$

note that there are only finitely functions of bounded degree in $M_{S_K}$. Hence, $t_{K,P}$ is well defined. From here, we are ready to prove the key lemma, which allows us to pass from counting functions to counting points in orbits; compare to [13, Lemma 4.8].

**Lemma 4.3.** Let $S_K$ be as in (26) and suppose that $P \in \mathbb{P}^1(K)$ satisfies $h(P) > \log 4$. Then there exists a constant $w_{K,P}$ (depending only on $K$ and $P$) such that

$$\#\{f \in M_{S_K} : f(P) = Q\} \leq w_{K,P}$$

holds for all $Q \in \text{Orb}_{S_K}(P)$. Specifically, one can take $w_{K,P} = r_{K,P} \cdot t_{K,P} + 1$.

**Proof.** First some notation. Given a point $Q \in \text{Orb}_{S_K}(P)$ in the orbit of $P$, we define

$$\ell_{P,Q} := \min\{\ell(f) : f(P) = Q \text{ for } f \in M_{S_K}\}$$

to be the minimum length of the functions evaluating $P$ to $Q$. Then the first step in the proof of Lemma 4.3 is to show that
Suppose that

\[ \#\{ f \in M_{S_K} : f(P) = Q \} \leq \ell_{P,Q} \cdot t_{K,P} + 1 \quad \text{for all } Q \in \text{Orb}_{S_K}(P). \]

To do this, we proceed by induction on the minimum length \( \ell_{P,Q} \). Note that if \( \ell_{P,Q} = 0 \), then \( P = Q \) and therefore any other function \( f \in M_{S_K} \) evaluating \( P \) to \( Q \) must satisfy \( f(P) = P \).

But this is impossible for any non-identity \( f \in M_{S_K} \) since \( h(P) > \log 4 \) and

\[ \log 2 < 3(h(P) - \log(2)) - h(P) \leq \deg(f)(h(P) - \log(2)) - h(P) \leq h(f(P)) - h(P) \]

by Tate’s Telescoping Lemma. Hence, the only function \( f \in M_{S_K} \) with \( f(P) = Q \) is the identity function when \( \ell_{P,Q} = 0 \). In particular, (37) holds as claimed. We now proceed with the remaining cases and pro forma make \( \ell_{P,Q} = 1 \) the base case of our induction argument.

**Base case:** Suppose that \( \ell_{P,Q} = 1 \) and choose some \( \tau \in S_K \) with \( \tau(P) = Q \). Now let \( F \in M_{S_K} \) be any other function such that \( F(P) = Q \). Then we may write \( F = \theta \circ f \) for some \( f \in M_{S_K} \) (possibly the identity function). Hence, letting \( g \) be the identity function, we have that

\[ \theta \circ f(P) = \tau \circ g(P). \]

Therefore, Lemma 4.2 implies that either

\[ \theta = \tau \text{ and } f(P) = g(P) \]

or

\[ \max \{ \deg(\theta \circ f), \deg(\tau \circ g) \} \leq d_{K,P}. \]

However in the first case, \( f(P) = g(P) = P \) so that \( f \) must be the identity function also; see (38) in Case 1 above. Hence, \( F = \tau \) in this case. In the other case, \( \deg(F) = \deg(\theta \circ f) \leq d_{K,P} \).

In particular, for all \( Q \in \text{Orb}_{S_K}(P) \) with \( \ell_{P,Q} = 1 \), we have proven that there exists some \( \tau = \tau_Q \in S_K \) satisfying

\[ \{ F \in M_{S_K} : F(P) = Q \} \subseteq \{ \tau_Q \} \cup \{ F \in M_{S_K} : \deg(F) \leq d_{K,P} \}. \]

In particular, (37) holds as claimed in the \( \ell_{P,Q} = 1 \) case.

**Induction step:** Suppose that \( \ell_{P,Q} = n + 1 \) and that (37) holds for all \( Q' \in \text{Orb}_{S_K}(P) \) with \( \ell_{P,Q'} \leq n \). Then we can choose (and fix) some \( G = \tau \circ g \) satisfying \( G(P) = Q \) with \( \tau = \tau_Q \in S_K \) and \( g \in M_{S_K} \) of length \( \ell(g) = n \geq 1 \). Now let \( F \in M_{S_K} \) be any other function satisfying \( F(P) = Q \). Then we may write \( F = \theta \circ f \) for some \( \theta \in S_K \) and some \( f \in M_{S_K} \). In particular, Lemma 4.2 implies that either

\[ \theta = \tau \text{ and } f(P) = g(P) \]

or

\[ \max \{ \deg(\theta \circ f), \deg(\tau \circ g) \} \leq d_{P}. \]

Now let \( Q' = g(P) \) so that we may rephrase (39) as saying:

\[ \{ F \in M_{S_K} : F(P) = Q \} \subseteq \{ F \in M_{S_K} : F = \tau_Q \circ f \text{ and } f(P) = Q' \} \cup \{ F \in M_{S_K} : \deg(F) \leq d_{K,P} \}. \]

The point here is that \( \tau_Q \) and \( Q' \) are both independent of \( F \) (they depend on \( G \), which is fixed). On the other hand, \( Q' \in \text{Orb}_{S_K}(P) \) and \( \ell_{P,Q'} \leq n \), so by induction

\[ \#\{ F \in M_{S_K} : f(P) = Q' \} \leq \ell_{P,Q'} \cdot t_{K,P} + 1 \leq d_{K,P} \cdot t_{K,P} + 1. \]

Hence, combining (40) and (41) we deduce that

\[ \#\{ F \in M_{S_K} : F(P) = Q \} \leq (n \cdot t_{K,P} + 1) + t_{K,P} = (n + 1) \cdot t_{K,P} + 1 = \ell_{P,Q} \cdot t_{K,P} + 1 \]

as desired. In particular, we have established (37), our first step in proving Lemma 4.3.

To complete the proof, we will show that if \( \ell_{P,Q} \) is sufficiently large, then to count functions \( F \) such that \( F(P) = Q \), one can instead count functions \( f \) with \( f(P) = Q' \) and \( \ell_{P,Q'} \) bounded. In particular combined with the first step (37), we obtain the desired general bound (independent of length). To do this, recall first that \( r_{K,P} := \lfloor \log_2 d_{K,P} \rfloor \) and note that

\[ \deg(G) > 2^\ell(G) \geq d_{K,P} \quad \text{for all } G \in M_{S_K} \text{ with } \ell(G) \geq r_{K,P}, \]

since each non-identity map in \( M_{S_K} \) has degree greater than 2. Now suppose that \( Q \in \text{Orb}_{S_K}(P) \) is such that \( \ell_{P,Q} > r_{K,P} \), and choose (and fix) a minimal length function \( G \) such that \( G(P) = Q \). Then, by the length assumption on \( Q \), we may write \( G = \tau_m \circ \cdots \circ \tau_1 \circ g \) for some \( \tau_i \in S_K \) and \( m \geq 1 \) and some \( g \in M_{S_K} \) of length \( r_{K,P} \). Similarly, due to the minimality of the length of \( G \), we may write any other function \( F \) with \( F(P) = Q \) as \( F = \theta_m \circ \cdots \circ \theta_2 \circ f \circ \cdots \circ f \circ \cdots \circ f \)


for some \( \tau_i \in S_K \) and \( m \geq 1 \) and some \( f \in M_{S_K} \) of length at least \( r_{K,p} \). We’ll show by
induction on \( m \) that

\[
\text{(43) Claim: } \tau_i = \theta_i \text{ for all } i \text{ and } f(P) = g(P) \]

**Base case:** Suppose that \( m = 1 \). Then \( F(P) = \theta_1 \circ f(P) = \tau_1 \circ g(P) = G(P) \). Hence, Lemma
4.2 implies that either

\[
\theta_1 = \tau_1 \text{ and } f(P) = g(P) \quad \text{or} \quad \max \{ \deg(\theta_1 \circ f), \deg(\tau_1 \circ g) \} \leq d_{K,p}.
\]

However, \( \deg(\tau_1 \circ g) = \deg(G) > d_{K,p} \) since \( \ell(G) = 1 + r_{K,p} > r_{K,p} \); see (42) above. In
particular, \( \theta_1 = \tau_1 \) and \( f(P) = g(P) \) as claimed.

**Induction step:** Suppose that \( m \geq 2 \), that \( \tau_{m-1} \circ \cdots \circ \tau_1 \circ g(P) = \theta_{m-1} \circ \cdots \circ \theta_1 \circ f(P) \) implies
that \( \tau_i = \theta_i \) for all \( 1 \leq i \leq m - 1 \) and that \( f(P) = g(P) \), and assume that

\[
\theta_m \circ \cdots \circ \theta_1 \circ f(P) = \tau_m \circ \cdots \circ \tau_1 \circ g(P).
\]

Now let \( g' = \tau_{m-1} \circ \cdots \circ \tau_1 \circ g \) and \( f' = \theta_{m-1} \circ \cdots \circ \theta_1 \circ f \). In particular, \( \theta_m \circ f'(P) = \tau_m \circ g'(P) \).

Therefore, Lemma 4.2 implies that either

\[
\theta_m = \tau_m \text{ and } f'(P) = g'(P) \quad \text{or} \quad \max \{ \deg(\theta_m \circ f'), \deg(\tau_m \circ g') \} \leq d_{K,p}.
\]

But \( \deg(\tau_m \circ g') > d_{K,p} \) since

\[
\ell(\tau_m \circ g') = 1 + \ell(g') \geq 1 + \ell(g) = 1 + r_{K,p} > r_{K,p};
\]

see (42) above. In particular, \( \theta_m = \tau_m \) and \( f'(P) = g'(P) \). However, this means

\[
\tau_{m-1} \circ \cdots \circ \tau_1 \circ g(P) = \theta_{m-1} \circ \cdots \circ \theta_1 \circ f(P)
\]

which implies that \( \tau_i = \theta_i \) for all \( 1 \leq i \leq m - 1 \) and that \( f(P) = g(P) \) by the induction
hypothesis. Together with the fact that \( \theta_m = \tau_m \), we have proven the claim (43).

To summarize: suppose that \( Q \in \text{Orb}_{S_K}(P) \) is such that \( \ell_{P,Q} > r_{K,p} \) and choose some
minimal length \( G \) with \( G(P) = Q \). Now write \( G = \tau_m \circ \cdots \circ \tau_1 \circ g \) for some \( \tau_i \in S_K \) and
\( m \geq 1 \) and some \( g \in M_{S_K} \) of length \( r_{K,p} \). Then we have shown that if \( F \) is any other function
satisfying \( F(P) = Q \), then we may write

\[
F = \tau_m \circ \cdots \circ \tau_1 \circ f \quad \text{for some } f \in M_{S_K} \text{ satisfying } f(P) = g(P).
\]

But \( Q' = g(P) \) is in the orbit of \( P \) and \( \ell_{P,Q'} \leq \ell(g) = r_{K,p} \). Hence, (37) implies that

\[
\# \{ f \in M_{S_K} : f(P) = Q' \} \leq \ell_{P,Q'} \cdot t_{K,p} + 1 \leq r_{K,p} \cdot t_{K,p} + 1.
\]

Therefore, the number of such \( F \)’s in (44) is also bounded in this way:

\[
\# \{ F \in M_{S_K} : F(P) = Q \} \leq r_{K,p} \cdot t_{K,p} + 1;
\]

the key point is that both \( Q' \) and the functions \( \tau_i \) are independent of \( F \). In particular, the
bound \( w_{K,p} := r_{K,p} \cdot t_{K,p} + 1 \) satisfies

\[
\# \{ F \in M_{S_K} : F(P) = Q \} \leq w_{K,p}
\]

whenever \( \ell_{P,Q} > r_{K,p} \). However, it also provides an upper bound when \( \ell_{P,Q} \leq r_{K,p} \) by (37).
This completes the proof of Lemma 4.3. \( \square \)

Finally, we are ready to prove Theorem 1.3 from the introduction: if \( q \) is a uniformly-log
discrete set of primes, then for all \( \epsilon > 0 \) there is a \( b = b(K,q,\epsilon) \) such that

\[
(\log B)^b \ll \# \{ Q \in \text{Orb}_{S_K,q}(P) : H(Q) \leq B \} \ll (\log B)^{b+\epsilon}
\]

for all \( P \in \mathbb{P}^1(K) \) with \( H(P) > 4 \).

(Proof of Theorem 1.3). Let \( q \) be a uniformly-log discrete set of primes. Then Corollary 3.3
applied to \( \sigma = q \) and \( c = 1 \) implies that for all \( \epsilon > 0 \) there is a \( b = b(K,q,\epsilon) \) such that

\[
(\log B)^b \ll \# \{ f \in M_{S_K,q} : H(f(Q)) \leq B \} \ll (\log B)^{b+\epsilon}
\]

for all \( P \in \mathbb{P}^1(K) \) with \( H(P) > 4 \). On the other hand, it is certainly always true that

\[
\# \{ Q \in \text{Orb}_{S_K,q}(P) : H(Q) \leq B \} \leq \# \{ f \in M_{S_K,q} : H(f(P)) \leq B \}.
\]
Moreover since $S_{K,q} \subseteq S_K$ (so that $M_{S_{K,q}} \subseteq M_{S_K}$), it follows from Lemma 4.3 that
\begin{equation}
(48) \quad w_{K,P}^{-1} \cdot \# \left\{ f \in M_{S_{K,q}} : H(f(Q)) \leq B \right\} \leq \# \left\{ Q \in \text{Orb}_{S_{K,q}}(P) : H(Q) \leq B \right\}.
\end{equation}

In particular, (46), (47), (48) together imply the desired bounds in (45). \hfill \Box

Remark 11. In light of Lemmas 4.2 and 4.3, it is tempting to think that Theorem 1.3 holds for the full semigroup generated by $z^q + 1$ and $q$ is any prime (with or without the gcd condition). This may be so, however our bounds from section 2 do not apply in this case, since the full set of rational primes do not form a uniformly log-discrete set. This leaves open the possibility that the poles of the generating functions we use accumulate, breaking down our estimates. That the primes are not uniformly log-discrete is perhaps deducible by elementary means. However, it certainly follows from Zhang’s bounded gap theorem \[22\].

We can apply Theorem 1.3 to obtain an amusing arithmetic statement about the orbits generated by $z^q + 1$ where $q$ is a Mersenne prime; see Corollary 1.4 from the introduction for an equivalent, but less dynamical looking, statement.

**Corollary 4.4.** Let $S = \{z^q + 1 : q \text{ is a Mersenne prime}\}$. Then
\begin{equation}
(\log B)^{0.60839} \ll \# \left\{ Q \in \text{Orb}_{S}(P) : H(Q) \leq B \right\} \ll (\log B)^{0.60872}
\end{equation}
for all $P \in \mathbb{P}^1(\mathbb{Q}) \setminus \left\{ \infty, 0, \pm 1, \pm 2, \pm \frac{1}{2}, \pm 3, \pm \frac{1}{3}, \pm 4, \pm \frac{1}{4}, \pm 6, \pm \frac{1}{6} \right\}$.

**Proof.** It is straightforward to see that the Mersenne primes are uniformly log-discrete. For instance, $(2^{x+1} - 1)/(2^x - 1) > 1.5$ for all $x > 1$, so that
\[ |\log(q) - \log(q')| > \log(1.5) \]
for all distinct Mersenne primes $q$ and $q'$. In particular, $\# \left\{ Q \in \text{Orb}_{S}(P) : H(Q) \leq B \right\}$ is bounded by arbitrarily close powers of $\log(B)$ for all $P$ with $H(P) > 4$ by Theorem 1.3. To obtain the explicit bounds in Corollary 4.4 (whose powers of $\log(B)$ agree up to three decimal places), we use the approximation techniques discussed in Remark 7, the first nine Mersenne primes:
\[ q = 3, 7, 31, 127, 8191, 131071, 524287, 2147483647, 2305843009213693951, \]
and the fact that the tenth has 27 digits. See the file called Mersenne at https://sites.google.com/a/alumni.brown.edu/whindes/research for the Magma code justifying these bounds. \hfill \Box

**Example 3.** To help illustrate Corollary 4.4 in a more classically arithmetic way, consider the statement for $P = 5$. Then
\begin{equation}
(\log B)^{0.60839} \ll \# \left\{ t = ((5^{q_1} + 1)^{q_2} \cdots + 1)^{q_s} + 1 \mid q_1, \ldots, q_s \in \mathfrak{m}, t \leq B \right\} \ll (\log B)^{0.60872},
\end{equation}
where $\mathfrak{m}$ is the set of Mersenne primes.

**References**

[1] A. Baragar, Rational points on K3 surfaces in $\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$, Math. Ann. 305.3 (1996): 541-558.

[2] E. Bombieri and J. Pila, The number of integral points on arcs and ovals, Duke Math. J 59.2 (1989): 337-357.

[3] Y. Bilu, and R. Tichy, The Diophantine equation $f(x) = g(y)$, Acta Arithmetica 95.3 (2000): 261-288.

[4] W. Bosma, J. Cannon, and C. Playoust, The Magma algebra system I: The user language, Journal of Symbolic Computation 24.4-4 (1997): 235-265.

[5] B. Brindza, On S-integral solutions of the Catalan equation, Acta Arithmetica 4.48 (1987): 397-412.

[6] G. Call and J. Silverman, Canonical heights on varieties with morphisms, Compositio Mathematica 89.2 (1993): 163-205.

[7] R. Crandall and C. Pomerance, Prime numbers: a computational perspective, Vol. 182. Springer Science & Business Media, 2006.

[8] G. Faltings, Endlichkeitssätze für abelsche Varietäten über Zahlkörpern, Inventiones mathematicae 73.3 (1983): 349-366.

[9] P. Flajolet and R. Sedgewick, Analytic combinatorics, Cambridge University Press, 2009.
[10] D.R. Heath-Brown, Counting rational points on algebraic varieties, Analytic number theory, Springer, Berlin, Heidelberg, 2006. 51-95.
[11] D.R. Heath-Brown, The density of rational points on curves and surfaces, *Annals of mathematics* 155.2 (2002): 553-598.
[12] V. Healey and W. Hindes, Stochastic Canonical Heights, *J. Number Theory*, 201 (2019): 228-256.
[13] W. Hindes, Counting points of bounded height in monoid orbits, preprint arXiv:2006.08563.
[14] W. Hindes, Dynamical and arithmetic degrees for random iterations of maps on projective space, *Math. Proc. Camb. Philos. Soc.*, to appear.
[15] P. Ingram, Lower bounds on the canonical height associated to the morphism, $\phi(z) = z^d + c$, *Monatshefte für Mathematik* 157.1 (2009): 69-89.
[16] S. Kawaguchi, Canonical heights for random iterations in certain varieties, *Int. Math. Res. Not.*, Article ID rnm023, 2007.
[17] S. Kawaguchi and Joseph H. Silverman, On the dynamical and arithmetic degrees of rational self-maps of algebraic varieties, *Journal für die reine und angewandte Mathematik* (Crelles Journal) 2016.713 (2016): 21-48.
[18] R. Mason, Diophantine equations over function fields, Vol. 96, Cambridge University Press, 1984.
[19] J. Mello, On quantitative estimates for quasintegral points in orbits of semigroups of rational maps, *New York Journal of Mathematics*, vol. 25, pp. 1091-1111.
[20] S. Schanuel, On heights in number fields, *Bulletin of the American Mathematical Society* 70.2 (1964): 262-263.
[21] D. Zagier, On the number of Markoff numbers below a given bound, *Mathematics of Computation* 39.160 (1982): 709-723.
[22] Y. Zhang, Bounded gaps between primes, *Annals of Mathematics* (2014): 1121-1174.