VARIATION OF THE CANONICAL HEIGHT IN A FAMILY OF
POLARIZED DYNAMICAL SYSTEMS

PATRICK INGRAM

Abstract. Call and Silverman introduced the canonical height \( \hat{h}_f \) associated to an endomorphism \( f : X \to X \) of a projective variety and an ample \( L \in \text{Pic}(X) \) satisfying \( f^* L \cong L^\otimes d \) for some \( d \geq 2 \). They also presented an asymptotic for the variation of this height in a family over the one-dimensional base \( B \) in terms of the height on the generic fibre and the height of the parameter, namely
\[
\hat{h}_f(P_t) = \hat{h}_f(P) h_B(t) + o(h_B(t)),
\]
where \( o(x)/x \to 0 \) as \( x \to \infty \). Here we save a power in the error term, giving an effective estimate
\[
\hat{h}_f(P_t) = \hat{h}_f(P) h_B(t) + O\left(h_B(t)^{2/3}\right)
\]
in general, and
\[
\hat{h}_f(P_t) = \hat{h}_f(P) h_B(t) + O\left(h_B(t)^{1/2}\right)
\]
when \( B \) is rational. As a corollary, we give an explicit bound on the height of parameters \( t \in B \) for which \( P_t \) is preperiodic for \( f_t \), in the case that \( X = \mathbb{P}^N \) and \( B = \mathbb{P}^1 \).

1. Introduction

The canonical height associated to a polarized dynamical system is a fundamental tool in arithmetic dynamics. To a projective variety \( X \) with a line bundle \( L \), we may associate a Weil height \( h \) which depends on various choices of defining equations. If \( f \) is an endomorphism of \( X \) with \( f^* L \cong L^\otimes d \) for some \( d \geq 2 \), and \( L \) is ample, then we say that \((f, X, L)\) is polarized, and we may associate a canonical height \( \hat{h}_f \) satisfying
\[
\hat{h}_f(f(P)) = d \hat{h}_f(P) \quad \text{and} \quad \hat{h}_f(P) = h(P) + O(1),
\]
where \( h \) is any Weil height on \( X \) with respect to \( L \). This construction is due to Call and Silverman [5] in the general case, but see also work of Denis [7], and Zhang [28], all building on classical work of Néron and Tate. Over a number field, sets of bounded ample height are finite, and so we may deduce, as a quick application, the finiteness of the set of rational preperiodic points for \( f \).

Now let \((f, X, L)\) be a family of polarized dynamical systems over a curve \( B \), and let \( P : B \to X \). On all but finitely many fibres \((f_t, X_t, L_t)\) we may construct a canonical height, and Call and Silverman also showed [5, Theorem 4.1] that
\[
\hat{h}_f(P_t) = \hat{h}_f(P) h_B(t) + o(h_B(t)),
\]
where \( \hat{h}_f(P) \) is the canonical height on the generic fibre, where \( h_B \) is a Weil height on \( B \) with respect to any divisor of degree one, and where \( o(x)/x \to 0 \) as \( x \to \infty \). Naturally, Call and Silverman ask [5, p. 184] whether one might improve upon (1).
In the case where $X = E$ is an elliptic curve (and so the canonical height is the Néron-Tate height), Tate [27] improved this estimate by establishing a result of the form
\[
\hat{h}_{f_t}(P_t) = \hat{h}_f(P)h_B(t) + O(h_B(t)^{1/2}),
\]
with a further improvement to $O(1)$ in the error when $B$ is rational. Note that this is, in some sense, the best possible error term for an arbitrary height function $h_B$.

In the case $X = \mathbb{P}^1$ and with $f$ a family of polynomials, the author [13] proved analogous results, and this was extended to endomorphisms of $\mathbb{P}^N$ with fixed behaviour on a totally invariant hyperplane [16]. Still in the case $X = \mathbb{P}^1$, Ghioca, Hsia, and Tucker [11] proved a similar result for rational functions with a super-attracting fixed point in the case $\hat{h}_f(P) \neq 0$. Generalizing work of Ghioca and Mavraki [12], Mavraki and Ye [23] established the same for rational functions in the case $X = B = \mathbb{P}^1$, but under the additional hypothesis that the pair $(f, P)$ is quasi-adelic, a condition which is not known to always hold. Related results also exist for Hénon maps [14], Drinfeld modules [15], and dynamical correspondences [17]. As well, the aforementioned result of Tate has been sharpened by DeMarco and Mavraki [6] to show that $\hat{h}_{E^t}(P_t)$ is precisely a height induced by an adelically metrized line bundle on the base, while the case of single-variables polynomials from [13] was subsequently similarly strengthened by Favre and Gauthier [9].

Our main result here is a weaker savings in the error term of the estimate, but in the general case.

**Theorem 1.** Let $(f, X, L)$ be a family of polarized dynamical systems over the curve $B$, defined over a number field $K$, and let $h_B$ be a degree-one Weil height on $B$. Then we have
\[
\hat{h}_{f_t}(P_t) = \hat{h}_f(P)h_B(t) + O\left(h_B(t)^{1/2}\right)
\]
as $h_B(t) \to \infty$. When $B$ is rational we have the further improvement
\[
\hat{h}_{f_t}(P_t) = \hat{h}_f(P)h_B(t) + O\left(h_B(t)^{3/2}\right).
\]

As already remarked in [5], the asymptotic (1) ensures that when $\hat{h}_f(P) > 0$, the $t \in B(\mathcal{O})$ for which $P_t$ is preperiodic for $f_t$ form a set of bounded height. Indeed, the asymptotic obtained has the form
\[
\left|\hat{h}_{f_t}(P_t) - \hat{h}_f(P)h_B(t)\right| \leq \varepsilon h_B(t) + O(1)
\]
for any $\varepsilon > 0$, where the implied constant is in-principle computable; one need only compute the implied constant for some $0 < \varepsilon < \hat{h}_f(P)$. In practice, though, this depends on some lengthy computations for each example. In the case of endomorphisms of $X = \mathbb{P}^N$ over $B = \mathbb{P}^1$, we can produce a completely explicit bound depending on relatively natural measures of complexity. For $P_0, \ldots, P_N \in K[t]$ with no common factor, set
\[
h_{\text{geom}}([P_0 : \cdots : P_N]) = \max \deg(P_i),
\]
let $h_{\text{arith}}(P)$ be the height of the projective tuple of coefficients of all of the $P_i$, and set
\[
h_{\text{total}}(P) = h_{\text{arith}}(P) + h_{\text{geom}}(P).
\]
For $f : \mathbb{P}^N \to \mathbb{P}^N$ defined over $K(t)$, we define $h_{\text{total}}(f)$ by identifying $f$ with its tuple of coefficients.
Corollary 2. Let \( f : \mathbb{P}^N \to \mathbb{P}^N \) be defined over \( \mathbb{P}^1_K \), and let \( P \in \mathbb{P}^N(\mathbb{P}^1_K) \) such that \( \hat{h}_f(P) \neq 0 \). If \( P_t \) is preperiodic for \( f_t \), then
\[
 h(t) \leq C\max \left\{ \left( \frac{h_{\text{total}}(P) + h_{\text{total}}(f) + 1}{\hat{h}_f(P)^2} \right)^4, h_{\text{total}}(f) + 1 \right\}
\]
where \( C \) is an explicit constant depending just on \( \deg(f) \) and \( N \).

The question of exactly when we have \( \hat{h}_f(P) = 0 \) is more subtle in the function field context (i.e., on the generic fibre) than it is over a number field. In the situation \( X = \mathbb{P}^1 \), it follows from results of Benedetto \[8\] and Baker [2] that \( \hat{h}_f(P) = 0 \) only when \( P \) is preperiodic for \( f \) (in which case \( P_t \) is always preperiodic for \( f_t \)), or the pair \((f, P)\) is isomorphic over some function field extension to a constant family (in which case \( P_t \) either is or isn’t preperiodic for \( f_t \), independent of \( t \)). Gauthier and Vigny [10] have extended this result to families of polarized dynamical systems in general, showing that if \( X \) has no periodic isotrivial subvariety of positive dimension, and \( \hat{h}_f(P) = 0 \), then \( P \) is preperiodic for \( f \).

Call and Silverman actually defined canonical heights in an even more general setting. If \( f : X \to X \) and \( L \in \text{Pic}(X) \otimes \mathbb{R} \) (not necessarily ample) satisfies \( f^*L \cong L^{\otimes \alpha} \) with \( \alpha > 1 \) real, the same construction of a canonical height with respect to \( L \) goes through, and the same asymptotic (1) holds. This is a strictly more general setting: Silverman [26] has constructed examples of automorphisms of certain K3 surfaces admitting \( L \in \text{Pic}(X) \otimes \mathbb{R} \) with \( f^*L \cong L^{\otimes (7+4\sqrt{5})} \). In this context we obtain a slightly weaker result, restricted to the simpler case \( B = \mathbb{P}^1 \).

Theorem 3. Let \( X \) be a family of Wehler K3 surfaces defined over \( \mathbb{P}^1_K \), and let \( \hat{h}^\pm \) be the canonical heights defined by Silverman [26]. Then for any \( P \in X(\mathbb{P}^1_K) \) and any \( \varepsilon > 0 \), we have
\[
 \hat{h}^\pm_t(P_t) = \hat{h}^\pm(P)h(t) + O\left(h(t)^{\frac{4}{3} + \varepsilon}\right).
\]

Theorem 3 is really a corollary to the more general (and more technical) Theorem 10 in Section 4 on endomorphisms \( f \) of projective varieties and systems of line bundles \( L_i \) satisfying \( f^*L_i \cong \bigotimes_{j=1}^{r_i} L_j^\otimes A_{ij} \), for some integers \( A_{ij} \).

Since our proof is a technical refinement of the argument of Call and Silverman [26] (tracing back to an earlier result of Silverman [25]), we sketch that argument in the context of Theorem 4 before outlining our modification. The proof in [2] proceeds as follows. First,
\[
 |\hat{h}_f(P_t) - \hat{h}_f(P)h_B(t)| \leq \left| \hat{h}_f(P_t) - h_{X,L}(P_t) \right| + \left| h_{X,L}(P_t) - \deg(P^*L)h_B(t) \right|
\]
\[
 + \left| \deg(P^*L) - \hat{h}_f(P) \right| h_B(t).
\]

On the one hand, the term in (4) is bounded by estimating the difference between the naive height \( \deg(P^*L) \) and the canonical height in the function field, which can be done with no dependence on \( P \). On the other hand, it is not hard to show that the difference between the Weil height and canonical height in families is at most linear in the height of the parameter, whence the term on the right in (2) is at most \( O(h_B(t)) \), again with no dependence on \( P \). Finally, since \( h_{X,L} \circ P = h_{B,P^*L} \) and \( \deg(P^*L)h_B \) are heights on \( B \) relative to divisors of the same degree, the term
in (3) is at most $O(h_B(t)^{1/2})$, with a constant depending on $B$. Consequently, we have

$$\limsup_{h_B(t) \to \infty} \left| \frac{\hat{h}_f(P_t)}{h_B(t)} - \hat{h}_f(P) \right| \leq C,$$

for some constant $C$ not depending on $P$, since the term corresponding to (3) vanishes in the limit. But now, using the relation $\hat{h}_f \circ f = dh_f$, we may apply this estimate to $f^k(P)$ to obtain

$$\limsup_{h_B(t) \to \infty} \left| \frac{\hat{h}_f(P_t)}{h_B(t)} - \hat{h}_f(P) \right| = \limsup_{h_B(t) \to \infty} \frac{1}{d^k} \left| \frac{\hat{h}_f(f^k(P)_t)}{h_B(t)} - \hat{h}_f(f^k(P)) \right| \leq \frac{C}{d^k}.$$ 

Since $k$ is arbitrary, we may take $C = 0$ in (5).

Our strategy is to make the bound on the term in (3) more explicit. Specifically, we employ an explicit elimination of variables to obtain a bound of the form

$$|h_{X,L}(f_k(P)_t) - \deg(f_k(P)_t)h_B(t)| \leq C'd^k$$

for a specific height function on $B$, where $C'$ depends on $f$ and $P$, but not on $k$ or $t$. Combined with the same estimates used by Call and Silverman for the other terms, we have

$$\left| \hat{h}_f(P_t) - \hat{h}_f(P)h_B(t) \right| = d^{-k} \left| \hat{h}_f(f_k(P)_t) - \hat{h}_f(f_k(P))h_B(t) \right| \leq C'd^k + C''d^{-k}h_B(t),$$

with constants independent of both $k$ and $t$. For each $t \in B$ we choose $k$ with $d^k \approx h_B(t)^{1/3}$ to obtain an upper bound of scale $h_B(t)^{2/3}$. The further honing when $B = \mathbb{P}^1$ derives from a sharper version of (3) in that context.

2. Explicit results over $\mathbb{P}^1$

In this section we present tighter results in the case that the base curve $B$ is rational, in which case we take $B = \mathbb{P}^1_K$ and $h$ to be the standard Weil height on $B$; in other words, our dynamical systems are defined over the function field $K(t)$. We also restrict to the case in which $X = \mathbb{P}_N$, although we will see in Lemma 10 below that this is no loss of generality.

To set notation, let $K$ be a number field, and let $M_K$ be the usual set of absolute values on $K$. We set $n_v = [K_v : \mathbb{Q}_v]/[K : \mathbb{Q}]$, and normalize our absolute values so that

$$\sum_{v \in M_K} n_v \log |x|_v = 0$$

for $x \neq 0$. We will write

$$\|x_1, \ldots, x_m\|_v = \max |x_i|_v,$$

and $\log^+ x = \log \max\{1, x\}$, for $x \in \mathbb{R}$. We also write, for real-valued functions $f$ and $g$, that $f(x) = O(g(x))$ as $x \to \infty$ as long as there exist constants $C$ and $C'$ such that $|f(x)| \leq Cg(x)$ whenever $x \geq C'$.

For $P \in \mathbb{P}^N(K)$, we set

$$h(P) = \sum_{v \in M_K} n_v \log \|P\|_v.$$

as usual. We also note that if $L/K$ is a finite extension and $P \in \mathbb{P}_L^N$, then $h(P)$ as computed in $\mathbb{P}_L^N$ agrees with $h(P)$ as computed in $\mathbb{P}_K^N$, and so the height is well defined over $\overline{K}$. Indeed, all of our estimates below are stable under field extension, and so our results apply at $t$ varies in $\mathbb{P}_K^1$.

We identify points $P \in \mathbb{P}^N(K(t))$ with morphisms $P : \mathbb{P}^1 \to \mathbb{P}^N$ defined over $K$. Every $P \in \mathbb{P}^N(K(t))$ can be written as $P = [P_0 : \cdots : P_n]$ with $P_i \in K[t]$ with no common factor, uniquely up to multiplication by a non-zero scalar. We write $h_{\text{geom}}(P) = \deg(P)$ for the maximal degree of a coordinate function, noting that we will also view the coordinates of $P$ as binary homogeneous forms of degree $h_{\text{geom}}(P)$. And we set

$$h_{\text{arith}}(P) = \sum_{v \in M_K} n_v \log \|P\|_v,$$

where $\|P\|_v = \max \|P_i\|_v$. In other words, $h_{\text{arith}}$ is the usual Weil height on $\text{Hom}_{h_{\text{geom}}(P)}(\mathbb{P}^1, \mathbb{P}^N) \cong \mathbb{P}^{(N+1)(h_{\text{geom}}(P)+1)-1}$. Although we do not explicitly reference it here, our definition of arithmetic and geometric heights on $\mathbb{P}_K^N$ is motivated in part by work of Altman [1], foretelling more modern work on function fields heights (e.g., Moriwaki [24]; the language of presentations in Section 4, from Bombieri and Gubler [4], is even closer to Altman).

It is convenient to note some basic facts on norms of polynomials (see, e.g., [4, p. 22, p. 27]).

**Lemma 4** (Gauß, Gelfond). If $v$ is non-archimedean, then $\|fg\|_v = \|f\|_v \|g\|_v$. In general,

$$2^{-\deg(fg)} \|f\|_v \|g\|_v \leq \|fg\|_v \leq (\deg(g) + 1) \|f\|_v \|g\|_v \leq 2^{\deg(fg)} \|f\|_v \|g\|_v.$$

Although $h_{\text{arith}}(P)$ is defined in terms of coordinates in $K[t]$ chosen with no common factor, it is useful to note that some information can still be gleaned from coordinates with common factors.

**Lemma 5.** Let $P_0, \ldots, P_N \in K[t]$, not all zero, with greatest common factor $s$, and let $P = [P_0 : \cdots : P_N]$. Then

$$\sum_{v \in M_K} n_v \log \|P\|_v = h_{\text{arith}}(P) + h(s) + O(h_{\text{geom}}(P) + \deg(s))$$

where $h(s)$ is the height of the projective tuple of coefficients of $s$.

**Proof.** Let $s$ be the greatest common factor of $P_0, \ldots, P_N$, with $sQ_i = P_i$. Then by Lemma 4

$$\|\log \|Q\| + \log \|s\|_v - \log \|P\|_v | \leq \max \deg(P_i) \log^+ |2v|.$$

Summing over all places, and nothing that $h_{\text{geom}}(P) = \max \deg(P_i) - \deg(s)$, proves the lemma.

Write write $P_t$ for $P$ evaluated at $t \in \mathbb{P}_K^1$. It follows from the usual facts about heights that

$$h(P_t) = h_{\text{geom}}(P)h(t) + O_P(1),$$

and the main lemma of this section is an explicit estimate on the error term. Its proof hinges on effective elimination of variables, in the vein of the effective Nullstellensatz of Masser and Wüstholz [22] (see also more recents results of Krick, Pardo, and Sombra [19]). These results turn out to be more convenient to apply in
spirit than in letter, however, given our conventions and normalizations, and so we work mostly with the constituent parts.

We will need the following fact from linear algebra, which is essentially \[22\] Lemma 4 stated slightly more generally (it is also just Cramer’s Rule, and a proof can be found in the arXiv version of this paper). By an $r \times r$ signed minor of a matrix $A$ with entries from some commutative ring with identity, we mean $\pm 1$ times the determinant of some $r \times r$ submatrix of $A$.

**Lemma 6** (Cramer’s Rule). Let $R$ be an integral domain, and suppose that we have a homogeneous system of linear equations over $R$ in $x_1, \ldots, x_p$ of rank $r$, which admits a solution with $x_s \neq 0$. Then there is a solution with $x_s \neq 0$ in which each $x_j$ is an $r \times r$ signed minor of the coefficient matrix.

**Lemma 7.** Let $P \in \mathbb{P}^N(K(t))$. Then

$$-h^{\text{arith}}(P) - \log(h^{\text{geom}}(P) + 1) \leq h^{\text{geom}}(P)h(t) - h(P_t) \leq 4h^{\text{geom}}(P)h^{\text{arith}}(P) + 4h^{\text{geom}}(P) \log h^{\text{geom}}(P) + 8h^{\text{geom}}(P) \log 2 + \log h^{\text{geom}}(P) + \log(N + 1)$$

if $h^{\text{geom}}(P) \neq 0$. (If $h^{\text{geom}}(P) = 0$, then $h(P_t) = h^{\text{arith}}(P)$.)

**Proof.** Write $P$ as a tuple of homogeneous forms of degree $h^{\text{geom}}(P)$ in $t = [t_0 : t_1]$. In one direction we have from the triangle inequality that

$$\log \|P_t\| \leq h^{\text{geom}}(P) \log \|t\|_v + \log \|P\|_v + \log^+ |h^{\text{geom}}(P) + 1|_v,$$

and summing over all places gives

$$h(P_t) \leq h^{\text{geom}}(P)h(t) + h^{\text{arith}}(P) + \log(h^{\text{geom}}(P) + 1).$$

On the other hand, consider the system of equations

\[
\begin{align*}
\alpha t_0^{2h^{\text{geom}}(P)-1} &= P_0 A_{0,0} + \cdots + P_N A_{0,N} \\
\alpha t_1^{2h^{\text{geom}}(P)-1} &= P_0 A_{1,0} + \cdots + P_N A_{1,N}
\end{align*}
\]

(8) (9)

to be solved with $a \in K$ and $A_{i,j} \in K[t_0, t_1]$. By the Nullstellensatz over $K$, there is a solution with $a = 1$ if we replace the exponent on the left-hand-side by something sufficiently large, and the fact that we might take the given exponent follows from, e.g., \[18\] Theorem 1.1 (although in this case one can also prove the existence of a solution just by linear algebra; see the appendix to the arXiv version). Identifying coefficients of monomials in $t_0, t_1$ on both sides of each equation, we then have a system of linear equations in $a$ and the coefficients of the various $A_{i,j}$, which has a solution in $K$ with $a \neq 0$. By Lemma 6 there is a solution with $a \neq 0$, and in which $a$ and the coefficients of the $A_{i,j}$ are all $r \times r$ minors of a matrix whose entries are coefficients of the $P_i$, where $r$ is the rank of the system. From this,

$$\log \|A_{i,j}\|_v \leq r \log \|P\|_v + \log^+ |r|$$
and
\[
\log |a|_v + (2h^{\text{geom}}(P) - 1) \log \|t\|_v \leq \log \|P_t\|_v + \log \|A_{i,j}(t)\|_v + \log |N + 1|_v
\]
\[
\leq \log \|P_t\|_v + (h^{\text{geom}}(P) - 1) \log \|t\|_v
\]
\[
+ \log^+ |h^{\text{geom}}(P)|_v + r \log \|P\|_v + \log^+ |r!|_v
\]
\[
+ \log^+ |N + 1|_v,
\]
\[
\log |a|_v + h^{\text{geom}}(P) \log \|t\|_v \leq \log \|P_t\|_v + r \log \|P\|_v + r \log^+ |r|_v
\]
\[
+ \log^+ |N + 1|_v + \log^+ |h^{\text{geom}}(P)|_v,
\]
since \( r! \leq r^r \) for all \( r \geq 1 \). Summing over all places,
\[
h^{\text{geom}}(P)h(t) \leq h(P_t) + rh^{\text{arith}}(P) + r \log r + \log h^{\text{geom}}(P) + \log(N + 1).
\]

It remains to bound \( r \), which is the rank of the system of linear equations satisfied by \( a \) and the coefficients of the \( A_{i,j} \). Equations (8) and (9) each involve \( 2h^{\text{geom}}(P) \) monomials in \( t_0, t_1 \), and hence the resulting linear system contains at most \( 4h^{\text{geom}}(P) \) equations; we thus have \( r \leq 4h^{\text{geom}}(P) \), and so
\[
h^{\text{geom}}(P)h(t) \leq h(P_t) + 4h^{\text{geom}}(P)h^{\text{arith}}(P) + 4h^{\text{geom}}(P) \log(4h^{\text{geom}}(P))
\]
\[
+ \log h^{\text{geom}}(P) + \log(N + 1).
\]

\[\square\]

Now fix a morphism \( f : \mathbb{P}^N \to \mathbb{P}^N \) of degree \( d \) over \( K(t) \). That is, \( f \) is given by \( N + 1 \) homogeneous forms of degree \( d \), whose coefficients are polynomials in \( t \). We set \( h^{\text{geom}}(f) \) to be the maximum of these degrees, and \( h^{\text{arith}}(f) \) to be the height of the grand tuple of coefficients of the \( f_i \). The following lemma estimates the difference between the canonical height and the usual Weil height in our families, and is an explicit version of a result already appearing in [5].

**Lemma 8.** We have
\[
|h^{\text{geom}}(P) - \hat{h}_f(P)| \leq C_1
\]
and (for all but finitely many \( t \))
\[
|h(P) - \hat{h}_{f_t}(P)| \leq C_2 h(t) + C_3
\]
with
\[
C_3 = (N + 1)^2(N(d - 1) + 1)^N
\]
\[
C_1 = \frac{C_3 h^{\text{geom}}(f)}{d - 1},
\]
and
\[
C_2 = \frac{C_3(h^{\text{arith}}(f) + \log(h^{\text{geom}}(f) + 1) + \log C_3)}{d - 1}
\]
\[
+ \frac{\log(N + 1) + N \log(N(d - 1) + 1)}{d - 1}.
\]

**Proof.** Again, one direction is straightforward by the triangle inequality. Let \( f \) be given by homogeneous forms \( F_i \), whose coefficients are homogeneous forms of degree \( h^{\text{geom}}(f) \) in \( t = [t_0 : t_1] \). Then
\[
h^{\text{geom}}(f(P)) \leq \max \deg(F_i(P_0, ..., P_N)) \leq dh^{\text{geom}}(P) + h^{\text{geom}}(f).
\]
Similarly, we have for $Q \in \mathbb{P}^N(K)$,

\[
\log \|F_i(Q)\|_v \leq d \log \|Q\|_v + \log \|F_i,t\|_v + N \log^+ |d + 1|_v
\]
\[
\leq d \log \|Q\|_v + h^\text{geom}(f) \log \|t\|_v + \log \|F\|_v + \log^+ h^\text{geom}(f) + 1|_v
\]
\[
+ N \log^+ |d + 1|_v,
\]

so

\[
(13) \quad h(f_i(Q)) \leq dh(Q) + h^\text{geom}(f)h(t) + h^\text{arith}(f) + \log(h^\text{geom}(f) + 1) + N \log(d + 1).
\]

On the other hand, consider the system of equations

\[
aX_0^e = F_0(X)A_{0,0}(X) + \cdots + F_N(X)A_{N,0}(X)
\]
\[
\vdots
\]
\[
aX_N^e = F_0(X)A_{N,i}(X) + \cdots + F_N(X)A_{N,N}(X)
\]

with $e = (N + 1)(d - 1) + 1$, which we hope to solve with $a \in K(t)$ and $A_{i,j}(X) \in K(t)[X]$. By Macaulay’s work on resultants [21] (see Lang [20, Lemma 3.7, p. 394] for a more recent treatment), there is a solution with $a$ the resultant of the homogeneous forms $F_i(X)$, which will be nonzero as $f$ is a morphism. Then, viewing the system as a system of linear equations in the coefficients of the monomials in $X$, there is a solution with $a \neq 0$ and such that $a$ and the coefficients of the $A_{i,j}$ are all $r \times r$ minors of some matrix whose entries are among the coefficients of the $F_i$ (where $r$ is the rank of the resulting system). So

\[
\deg(a) + e \deg(P_i) \leq \deg(F_i(P_0, ..., P_N)) + \max \deg(A_{i,j}(P_0, ..., P_N))
\]
\[
\leq \deg(F_i(P_0, ..., P_N)) + rh^\text{geom}(f) + (e - d) \deg(P),
\]

since the coefficients of the $A_{i,j}$ are $r \times r$ minors of a matrix whose entries are coefficients of $f$. From this,

\[
\deg(a) + dh^\text{geom}(P) \leq \max \deg(F_i(P_0, ..., P_N)) + rh^\text{geom}(f).
\]

Of course, the homogeneous forms $F_i(P_0, ..., P_N)$ might have a common factor, but it is a divisor of $a$, and hence even after eliminating this we have

\[
(15) \quad dh^\text{geom}(P) \leq h^\text{geom}(f(P)) + rh^\text{geom}(f).
\]

It then suffices to bound $r$. Since the number of monomials of degree $D$ in $N + 1$ homogeneous variables is no greater than $(D + 1)^N$, the system of linear equations implied by (13) consists of at most $(N + 1)((N + 1)(d - 1) + 2)^N$ equations in at most $1 + (N + 1)^2(N(d - 1) + 1)^N$ unknowns, so we may take $r \leq C_3$ (the system has a nontrivial solution, and so the rank is strictly less than the number of variables). Combining (12) with (13), we obtain

\[
|dh^\text{geom}(P) - h^\text{geom}(f(P))| \leq C_3h^\text{geom}(f).
\]

On the other hand, plugging in $t$ with $a(t) \neq 0$ gives

\[
(16) \quad \log |a(t)|_v + e \log \|Q\|_v \leq \max |F_i,t(Q)|_v + \max |A_{i,j}(Q)|_v + \log^+ |N + 1|_v
\]

Now, each coefficient $F_{i,m}$ of $F_i$ satisfies

\[
\log |F_{i,m,t}|_v \leq h^\text{geom}(f) \log \|t\|_v + \log \|F\|_v + \log^+ h^\text{geom}(f) + 1|_v,
\]
and each coefficient $A_{i,j,m}$ of $A_{i,j}$ is an $r \times r$ determinant of such values, so these coefficients satisfy
\[
\log |A_{i,j,m}(t)|_v \leq r \log \max |F_{i,m,t}|_v + r \log^+ |r|_v,
\]
whence
\[
\log |A_{i,j}(Q)|_v \leq (e-d) \log \|Q\|_v + N \log^+ \|e-d+1\|_v + rh^{\text{geom}}(f) \log \|t\|_v + r \log \|F\|_v + r \log |h^{\text{geom}}(f) + 1|_v + r \log^+ |r|_v,
\]
for each $i,j$. Combining (16) and (17), and summing over all places of $K$, we obtain
\[
dh(Q) \leq h(f_t(Q)) + \log(N+1) + N \log(N(d-1)+1) + rh^{\text{geom}}(f) h(t) + rh^{\text{arith}}(f) + r \log(h^{\text{geom}}(f) + 1) + r \log^+ r,
\]
as long as $a(t) \neq 0$. We bound $r$ above by $C_3$, and then note that this error term is larger than that in (13), and so even in absolute value $dh(Q) - h(f_t(Q))$ is bounded by this error term.

The final claims follow from a standard telescoping sum argument: If $S$ is any set, $\varphi : S \to S$ is any function, and $\psi : S \to \mathbb{R}$ is a non-negative function satisfying
\[
|d\psi(x) - \psi \circ \varphi(x)| \leq C,
\]
then for $\hat{\psi}(x) = \lim_{k \to \infty} d^{-k} \psi \circ \varphi^k(x)$, we have
\[
|\hat{\psi}(x) - \psi(x)| \leq \frac{C}{d-1}.
\]
\[\square\]

The previous lemma gives a good indication of how the geometric height grows in an orbit. The next lemma recapitulates this, and gives us some estimate on the more subtle arithmetic height.

**Lemma 9.** Let $P \in \mathbb{P}^N(K(t))$ and let $f : \mathbb{P}^N \to \mathbb{P}^N$ over $K(t)$. Then
\[
h^{\text{geom}}(f^k(P)) \leq d^k C_4.
\]
and
\[
h^{\text{arith}}(f^k(P)) \leq d^k C_5.
\]

for
\[
C_4 = h^{\text{geom}}(P) + \frac{1}{d-1} h^{\text{geom}}(f),
\]
\[
C_5 = h^{\text{arith}}(P) + h^{\text{geom}}(P) \log 2 + \frac{1}{d-1} h^{\text{arith}}(f) + \frac{\log 2}{d-1} h^{\text{geom}}(f)
\]
\[+ \frac{N}{d-1} \log(d+1) + d \log d \frac{d \log 2 + d \log^+(h^{\text{geom}}(P) + \frac{1}{d-1} h^{\text{geom}}(f))}{(d-1)^2}.
\]

**Proof.** The claim (18) comes from the previous lemma, or specifically from (12) combined with the usual telescoping sum argument.

For the second claim, choose again homogeneous forms $F_1$ representing $f$, and let $P_0 = P$ with entries $P_{0,0}, \ldots, P_{0,N}$. We now define a sequence of tuples $P_k$ of polynomials $P_{k,t}$ by
\[
P_{k+1,t} = F_t(P_k).
\]
Writing $F_{i,m} \in K(t)$ for the coefficient of monomial $m$ in $F_i$, and noting that there are at most $(d + 1)^N$ monomials of degree $d$ in the $P_i$, we have for $v \in M_K$ that
\[
\log \|P_{k+1,i}\|_v \leq \max \log \|F_{i,m}(P_k)\|_v + N \log^+ |d + 1|_v
\]
\[
\leq \log \|F\|_v + d \log \|P_k\|_v + d \log^+ \max \deg(P_{k,i}) + 1|_v
\]
\[
+ N \log^+ |d + 1|_v
\]
by Lemma [3].

We digress briefly to note that, for any real number $x \geq 0$, we have
\[
\sum_{j=0}^{k} \frac{\log(1 + d^j x)}{d^j} \leq \frac{d \log d}{(d - 1)^2} + \frac{d \log 2 + \log^+ x}{(d - 1)}.
\]
To see this, simply note that $\log(1 + d^j x) \leq \log 2 + j \log d + \log^+ x$, and apply the usual sums $\sum_{j=0}^{\infty} d^{-j} = \frac{d}{d - 1}$ and $\sum_{j=0}^{\infty} j d^{-j} = \frac{d}{(d - 1)^2}$.

Using the fact that $\max \deg(P_{k,i}) \leq d^k (h_{\text{geom}}(P) + \frac{1}{d - 1} h_{\text{geom}}(f))$ and the estimate (21), we may iterate (20) to obtain
\[
\log \|P_k\|_v \leq d^k \log \|P_0\|_v + \frac{d^k - 1}{d - 1} (\log \|F\|_v + N \log^+ |d + 1|_v)
\]
\[
+ \frac{d \log^+ |d|_v}{(d - 1)^2} + \frac{C_v}{(d - 1)}
\]
where
\[
C_v = \begin{cases} 
\frac{d \log 2 + d \log^+ (h_{\text{geom}}(P) + \frac{1}{d - 1} h_{\text{geom}}(f))}{(d - 1)^2} & v \text{ archimedean} \\
0 & \text{otherwise.}
\end{cases}
\]
When we sum over all places, we obtain
\[
\sum_{v \in M_K} n_v \log \|P_k\|_v \leq d^k h_{\text{arith}}(P) + \frac{d^k - 1}{d - 1} \left( h_{\text{arith}}(f) + N \log(d + 1) \right)
\]
\[
+ \frac{d \log d}{(d - 1)^2} + \frac{d \log 2 + d \log^+ (h_{\text{geom}}(P) + \frac{1}{d - 1} h_{\text{geom}}(f))}{(d - 1)}
\]
although we note that it is entirely possible that $P_{k,0}, \ldots, P_{k,N}$ admit a common factor, and so the right-hand-side is not necessarily $h_{\text{arith}}(f^k(P))$. By the proof of Lemma [3] though, we have
\[
h_{\text{arith}}(f^k(P)) \leq \sum_{v \in M_K} n_v \log \|P_k\|_v + \max \deg(P_i) \log 2
\]
\[
\leq \sum_{v \in M_K} n_v \log \|P_k\|_v + d^k (h_{\text{geom}}(P) + \frac{1}{d - 1} h_{\text{geom}}(f)) \log 2.
\]
Combining with (22), we have the claimed bound (19), since $d^k \geq 1$. □
We now prove a special cases of the main result. As noted in Section 4, Lemma 16 this already implies the result in general over the base $B = \mathbb{P}^1$.

Proof of Theorem 1 when $B = \mathbb{P}^1$ and $X = \mathbb{P}^N$. First, when we apply Lemma 8 we will need to know that $f_t$ is an endomorphism of $\mathbb{P}^N$ of degree $d$, and so we will need $a(t) \neq 0$, with $a$ as in the proof of Lemma 7. If $a(t) = 0$, then $h(t) \leq h(a) + \deg(a) \log 2$. Since $a$ is an $r \times r$ determinant in the coefficients of $f$, for some $r \leq C_3$, we have $a(t) \neq 0$ as long as

$$h(t) > C_3 \left(h_{\text{arith}}(f) + h_{\text{geom}}(f) \log 2 + \log C_3\right)$$

and so we will assume this lower bound on $h(t)$. We also then have $f(P)_t = f_t(P_t)$. For now, it will be convenient also to suppose that there is no $k$ with $h_{\text{geom}}(f^k(P)) = 0$ (this implies $C_4 > 0$), although we shall see that things get even easier when this fails.

We have, for any $k$,

$$d^k \left| \hat{h}_{f_t}(P_t) - \hat{h}_f(P)t \right| = \left| \hat{h}_{f_t}(f^k(P)) - \hat{h}_f(f^k(P))h(t) \right| \leq \left| \hat{h}_{f_t}(f^k(P)) - h(f^k(P)) \right|$$

(23)

and so we will bound the three terms separately. For (24), we have from (11) in Lemma 8 that

$$\left| \hat{h}_{f_t}(f^k(P)) - h_{\text{arith}}(f^k(P)) \right| \leq C_1 h(t) + C_2$$

For (26), Lemma 8 gives

$$\left| h_{\text{geom}}(f^k(P))h(t) - \hat{h}_f(f^k(P))h(t) \right| \leq C_1 h(t).$$

The term (26) requires more attention. By Lemmas 7 and 9 we have

$$\left| h_{\text{geom}}(f^k(P))h(t) - h(f^k(P)) \right| \leq 4h_{\text{geom}}(f^k(P))h_{\text{arith}}(f^k(P))$$

$$+ 4h_{\text{geom}}(f^k(P)) \log h_{\text{geom}}(f^k(P))$$

$$+ 8h_{\text{geom}}(f^k(P)) \log 2 + \log h_{\text{geom}}(f^k(P))$$

$$+ \log(N + 1)$$

$$\leq 4d^2kC_4C_5 + 4d^kC_4 \log(d^kC_4)$$

$$+ 8d^kC_4 \log 2 + \log(d^kC_4) + \log(N + 1)$$

$$\leq d^{2k}C_6$$

for

$$C_6 = 4C_4C_5 + 4C_4 \log(dC_4) + 8C_4 \log 2 + \log(dC_4) + \log(N + 1)$$

since $1, k \leq d^k$. Now choose $k$ so that

$$d^{2k} \leq h(t) < d^{2(k+1)}$$
and hence \( d^{-k} \leq dh(t)^{-1/2} \), to obtain

\[
|\hat{h}_f(P_t) - \hat{h}_f(P)h(t)| \leq dh(t)^{-1/2} \left( 2C_1 h(t) + C_2 + C_6 h(t) \right)
\]

(27)

\[
\leq h(t)^{1/2} d(2C_1 + C_2 + C_6)
\]
as long as \( h(t) \geq 1 \) (in addition to our previous assumed lower bound).

We are left to deal with the case in which \( h(t) \leq 2 \). By (28), under the assumption (23) we have \( \hat{h}(P_t) = 0 \) and hence

\[
|\hat{h}_f(P_t) - \hat{h}_f(P)h(t)| \leq h(t)^{1/2} d(2C_1 + C_2 + \max\{C_5, C_6\})
\]
even for values of \( k \) with \( h^{\text{geom}}(f^k(P)) \neq 0 \).

\[\square\]

**Proof of Corollary 2.** Suppose that \( P_t \) is preperiodic for \( f_t \). By (28), under the assumption (23) we have \( h_{f_t}(P_t) = 0 \) and hence

\[
\hat{h}_f(P)h(t)^{1/2} \leq d(2C_1 + C_2 + \max\{C_5, C_6\}),
\]

unless \( h(t) \leq 1 \). Now,

\[
C_5 \leq h^{\text{total}}(P) + \frac{1}{d-1} h^{\text{total}}(f) + \frac{d \log 2 (h^{\text{total}}(P) + \frac{1}{d-1} h^{\text{total}}(f))}{(d-1)}
\]

\[
+ \frac{N}{d-1} \log(d + 1) + \frac{d \log d}{(d-1)^2}
\]

\[
\leq (6 + C_7)(h^{\text{total}}(P) + h^{\text{total}}(f))
\]
once \( h^{\text{total}}(P) + h^{\text{total}}(f) \geq 1 \), for

\[
C_7 = \frac{N}{d-1} \log(d + 1) + \frac{d \log d}{(d-1)^2}.
\]

And

\[
C_4 = h^{\text{geom}}(P) + \frac{1}{d-1} h^{\text{geom}}(f) \leq h^{\text{total}}(P) + h^{\text{total}}(f),
\]

so

\[
C_6 = 4C_4 C_5 + 4C_4 \log(dC_4) + 8C_4 \log 2 + \log(dC_4) + \log(N + 1)
\]

\[
\leq 4C_4 C_5 + 4dC_4^2 + (8 + d)C_4 \log 2 + \log(N + 1)
\]

\[
\leq 4(6 + C_7 + d)(h^{\text{total}}(P) + h^{\text{total}}(f))^2 + (8d + 2) \log 2 (h^{\text{total}}(P) + h^{\text{total}}(f))
\]

\[
+ \log(N + 1)
\]

\[
\leq (12d + 26 + 4C_7 + \log(N + 1))(h^{\text{total}}(P) + h^{\text{total}}(f))^2
\]
as long as \( h^{\text{total}}(P) + h^{\text{total}}(f) \geq 1 \). Finally,

\[
C_1 = \frac{C_3 h^{\text{geom}}(f)}{d-1} \leq \frac{C_3}{d-1} (h^{\text{total}}(P) + h^{\text{total}}(f)),
\]
satisfies (29) for any \( \varepsilon > 0 \), let \( M \subseteq \text{Pic}(X) \) be a free module of finite rank generated by semi-ample line bundles, with \( f^*M \subseteq M \), and let \( \rho(f^*) \) be the spectral radius of \( f^* \) on \( M \). If \( L \in M \otimes \mathbb{R} \) satisfies \( f^*L \cong L^{\otimes \alpha} \) with \( \alpha > 1 \) real, then we have

\[
\hat{h}_{f^*L}(P_t) = \hat{h}_{f^*L}(P_t) + O(\hat{h}(t)^{1+\varepsilon - \log \alpha/2 \log \rho(f^*)})
\]

for any \( \varepsilon > 0 \). In particular, if \( \alpha = \rho(f^*) \), we have

\[
\hat{h}_{f^*L}(P_t) = \hat{h}_{f^*L}(P_t) \sim \hat{h}(t)^{1+\varepsilon}.
\]

We begin by extending somewhat the machinery of the arithmetic heights defined in the previous section. Let \( X \) be a projective variety defined over \( K(t) \). For any morphism \( \varphi : X \to \mathbb{P}^N \) over \( K(t) \) and \( P \in X(K(t)) \), set

\[
h^\text{geom}_{\varphi}(P) = h^\text{geom}(\varphi(P)) \quad \text{and} \quad h^\text{arith}_{\varphi}(P) = h^\text{arith}(\varphi(P)).
\]

We collect some basic properties of heights that we will need below (note that \( h^\text{geom}_{\varphi} \) is just the usual function field height with respect to \( \varphi^*\mathcal{O}(1) \)).

3. More general canonical heights

Here we continue working over the field \( K(t) \), with \( K \) a number field, but consider a more general framework. The main result here is the following.

**Theorem 10.** Let \( f : X \to X \) be a family of dynamical systems defined over \( K(t) \), let \( M \subseteq \text{Pic}(X) \) be a free module of finite rank generated by semi-ample line bundles, with \( f^*M \subseteq M \), and let \( \rho(f^*) \) be the spectral radius of \( f^* \) on \( M \). If \( L \in M \otimes \mathbb{R} \) satisfies \( f^*L \cong L^{\otimes \alpha} \) with \( \alpha > 1 \) real, then we have

\[
\hat{h}_{f^*L}(P_t) = \hat{h}_{f^*L}(P_t) + O(\hat{h}(t)^{1+\varepsilon - \log \alpha/2 \log \rho(f^*)})
\]

for any \( \varepsilon > 0 \). In particular, if \( \alpha = \rho(f^*) \), we have

\[
\hat{h}_{f^*L}(P_t) = \hat{h}_{f^*L}(P_t) \sim \hat{h}(t)^{1+\varepsilon}.
\]
Lemma 11. Let $\varphi : X \to \mathbb{P}^n$ and $\psi : X \to \mathbb{P}^m$, both defined over $K(t)$.

1. If $\varphi \# \psi : X \to \mathbb{P}^{(n+1)(m+1)-1}$ is the composition of $\varphi$ and $\psi$ through the Segre embedding, then

\[
h^{\text{geom}}_{\varphi \# \psi} = h^{\text{geom}}_{\varphi} + h^{\text{geom}}_{\psi}
\]

and

\[
h^{\text{arith}}_{\varphi \# \psi} = h^{\text{arith}}_{\varphi} + h^{\text{arith}}_{\psi} + O(h^{\text{geom}}_{\varphi} + h^{\text{geom}}_{\psi}).
\]

2. If $\varphi^* \mathcal{O}(1) \cong \psi^* \mathcal{O}(1)$, then

\[
h^{\text{geom}}_{\varphi} = h^{\text{geom}}_{\psi} + O(1)
\]

and

\[
h^{\text{arith}}_{\varphi} = h^{\text{arith}}_{\psi} + O(h^{\text{geom}}_{\varphi} + h^{\text{geom}}_{\psi}).
\]

3. If $\varphi^* \mathcal{O}(1) \otimes \psi^* \mathcal{O}(-1)$ is generated by global sections, then

\[
h^{\text{geom}}_{\psi} \leq h^{\text{geom}}_{\varphi} + O(1),
\]

and

\[
h^{\text{arith}}_{\psi} \leq h^{\text{arith}}_{\varphi} + O(h^{\text{geom}}_{\varphi}).
\]

Proof. The proofs for the statements about $h^{\text{geom}}$, which is the usual function field height, are standard (see, e.g., [4, Chapter 2]). For the arithmetic height, we proceed by similar arguments.

Note that claim 1 is just a claim about the Segre map. Write $Q = [Q_0 : \cdots : Q_n] \in \mathbb{P}^n$ with $Q_i \in K[t]$ with no common factor, and similarly for $Q' \in \mathbb{P}^m$. If $\sigma : \mathbb{P}^n \times \mathbb{P}^m \to \mathbb{P}^{(n+1)(m+1)-1}$ is the Segre embedding, then $\sigma(Q, Q')$ has coordinates $Q_i Q'_j$, which again have no common factor. By Lemma 4

\[
|\log \|Q_i\|_v + \log \|Q'_j\|_v - \log \|Q_i Q'_j\|_v| \leq (\deg(Q_i) + \deg(Q'_j)) \log^+ |2|_v
\]

Taking the maximum over all $i$ and $j$, and summing over all places gives

\[
h^{\text{arith}}(\sigma(Q, Q')) = h^{\text{arith}}(Q) + h^{\text{arith}}(Q') + O(h^{\text{geom}}(Q) + h^{\text{geom}}(Q')).
\]

Now apply this with $Q = \varphi(P)$ and $Q' = \psi(P)$, as $\varphi \# \psi(P) = \sigma(\varphi(P), \psi(P))$ by definition.

For claim 2 we first consider the case where $\varphi$ and $\psi$ are closed embeddings such that the natural map $H^0(\mathbb{P}^n, \mathcal{O}(1)) \to H^0(X, \varphi^* \mathcal{O}(1))$ is surjective, and similarly for $\psi$, following the proof of [4, Proposition 2.5.9, pp. 49-51]. In this case, $H^0(X, \psi^* \mathcal{O}(1))$ has a basis among the $\psi_j$, and the isomorphism $\varphi^* \mathcal{O}(1) \cong \psi^* \mathcal{O}(1)$ ensures that each $\varphi_i$ the $\psi_j$, say

\[
c_i \varphi_i = a_{i,0} \psi_0 + \cdots + a_{i,m} \psi_m
\]

with $c_i, a_{i,0} \in K[t]$. Choosing a trivialization of the line bundle and coordinates $\psi_j(P)$ with no common factor, it follows that

\[
\log \|\varphi_i(P)\|_v \leq \log \max \|\psi_j(P)\|_v + \log \max \|a_{i,j}\|_v - \log \|c_i\|_v
\]

\[
+ O_v(\max \deg(\psi_j(P)) \cdot \max \deg(a_{i,j}) + \deg(c_i))
\]

\[
\leq \log \|\psi_0(P), \cdots, \psi_m(P)\|_v + O_v(h^{\text{geom}}(P)) + O_v(1),
\]

where the $O_v(1)$ vanishes for all but finitely many $v$. Summing over all places of $K$, we have

\[
h^{\text{arith}}_{\varphi}(P) \leq h^{\text{arith}}_{\psi}(P) + O(h^{\text{geom}}_{\varphi}(P) + h^{\text{geom}}_{\psi}(P))
\]
from Lemma 5 and by symmetry we conclude claim 2 in this case.

In general, if $\varphi$ and $\psi$ are embeddings then, as in the proof of [4] Proposition 2.5.9, pp. 49-51, there exists a $k$ such that $H^0(\mathbb{P}^n, \mathcal{O}(k)) \to H^0(X, \varphi^*\mathcal{O}(k))$ is surjective. Writing $\varphi^{\#k}$ for the composition of $\varphi$ with the $k$th monomial map $\mathbb{P}^n \to \mathbb{P}^{n+k-1}$, and applying the previous argument to $\varphi^{\#k}$ and $\psi^{\#k}$, we conclude 2 without the added assumption (using part 1 to deduce that $h^{\text{arith}} = kh^{\text{arith}} + O(h^{\text{geom}})$).

For claim 2 in general, note that we can write $\varphi^*\mathcal{O}(1) = \sigma^*\mathcal{O}(1) \otimes \tau^*\mathcal{O}(-1)$ and $\psi^*\mathcal{O}(1) \cong \theta^*\mathcal{O}(1) \otimes \xi^*\mathcal{O}(-1)$, where $\sigma, \tau, \theta, \xi$ are all embeddings. From the paragraph above and part 1 we have

$$h^{\text{arith}} + h^{\text{arith}} = h^{\text{arith}} + h^{\text{arith}} + O(h^{\text{geom}} + h^{\text{geom}} + h^{\text{geom}} + h^{\text{geom}}),$$

but also

$$h^{\text{arith}} = h^{\text{arith}} + h^{\text{arith}} + O(h^{\text{geom}} + h^{\text{geom}} + h^{\text{geom}})$$

and

$$h^{\text{arith}} = h^{\text{arith}} + h^{\text{arith}} + O(h^{\text{geom}} + h^{\text{geom}} + h^{\text{geom}}).$$

These three relations prove the claim, once we note that

$$h^{\text{geom}} = h^{\text{geom}} + h^{\text{geom}} + O(1)$$

and

$$h^{\text{geom}} = h^{\text{geom}} + h^{\text{geom}} + O(1).$$

For claim 3 note that if $\varphi^*\mathcal{O}(1) \otimes \psi^*\mathcal{O}(-1)$ is generated by global sections, then there is a morphism $\theta : X \to \mathbb{P}^k$ such that

$$\varphi^*\mathcal{O}(1) \cong \psi^*\mathcal{O}(1) \otimes \theta^*\mathcal{O}(1) \cong (\psi^*\theta)^*\mathcal{O}(1).$$

Applying parts 2 and 1 and noting that $h^{\text{arith}} \geq 0$, we have

$$h^{\text{arith}} \leq h^{\text{arith}} + h^{\text{arith}}$$

$$= h^{\text{arith}} + O(h^{\text{geom}} + h^{\text{geom}} + h^{\text{geom}})$$

$$= h^{\text{arith}} + O(h^{\text{geom}}),$$

since $h^{\text{geom}}, h^{\text{geom}} \leq h^{\text{geom}} + h^{\text{geom}} = h^{\text{geom}}$. □

In light of this lemma, we can and will define $h^{\text{geom}}$ and $h^{\text{arith}}$ relative to $L \in \text{Pic}(X)$ by choosing $\varphi : X \to \mathbb{P}^n$ and $\psi : X \to \mathbb{P}^m$ with $L \cong \varphi^*\mathcal{O}(1) \otimes \psi^*\mathcal{O}(-1)$, and setting

$$h^{\text{geom}} = h^{\text{geom}} - h^{\text{geom}}$$

and

$$h^{\text{arith}} = h^{\text{arith}} - h^{\text{arith}}.$$

This depends on the choice of $\varphi$ and $\psi$, but Lemma 11 circumscribes the extent of this dependence. More explicitly, we have the following lemma.

**Lemma 12.** Let $E$ be ample.

1. If $L \cong M$ then

$$h^{\text{geom}} = h^{\text{geom}} + O(1)$$

and

$$h^{\text{arith}} = h^{\text{arith}} + O(h^{\text{geom}}).$$

In particular, these relations hold for different choices of height function for $L$. 
(2) For any choice of height function relative to $L$ and $M$,

\[ h_{L \otimes M}^{\text{geom}} = h_L^{\text{geom}} + h_M^{\text{geom}} + O(1) \]

and

\[ h_{L \otimes M}^{\text{arith}} = h_L^{\text{arith}} + h_M^{\text{arith}} + O(h_E^{\text{geom}}). \]

(3) For any $L$ we have

\[ h_L^{\text{geom}} = O(h_E^{\text{geom}}) \]

and

\[ h_L^{\text{arith}} = O(h_E^{\text{geom}} + h_E^{\text{arith}}). \]

Proof. The claims about the geometric height are standard, and follow directly from Lemma 11.

For claim 1 for the additive height, suppose our height functions were $h_L^{\text{arith}} = h_L^{\text{arith}} - h_L^{\text{arith}}$ and $h_M^{\text{arith}} = h_M^{\text{arith}} - h_M^{\text{arith}}$, and note that we have $\varphi^*O(1) \otimes \tau^*O(1) \cong \sigma^*O(1) \otimes \psi^*O(1)$. Then Lemma 11 gives

\[ h_L^{\text{arith}} = h_L^{\text{arith}} + O(h_E^{\text{geom}} + h_E^{\text{geom}} + h_E^{\text{geom}} + h_E^{\text{geom}}), \]

and we may apply claim 3 to the geometric heights to replace the error term with $O(h_E^{\text{geom}})$.

For claim 2 for the arithmetic height, again take $h_L^{\text{arith}} = h_L^{\text{arith}} - h_L^{\text{arith}}$ and $h_M^{\text{arith}} = h_M^{\text{arith}} - h_M^{\text{arith}}$, and note that $h_L^{\text{arith}} - h_M^{\text{arith}}$ is one choice of height for $L \otimes M$. Lemma 11 now gives

\[ h_{L \otimes M}^{\text{arith}} = h_L^{\text{arith}} + h_M^{\text{arith}} + O(h_E^{\text{geom}} + h_E^{\text{geom}} + h_E^{\text{geom}} + h_E^{\text{geom}}), \]

and again the error term is $O(h_E^{\text{geom}})$.

Claim 3 follows by taking $k$ large enough that $E^\otimes k \otimes L^{-1}$ is ample. Then

\[
\begin{align*}
0 &\leq h_{E^\otimes k \otimes L^{-1}}^{\text{arith}} + O(h_{E^\otimes k \otimes L^{-1}}^{\text{geom}}) \\
&= k h_E^{\text{arith}} - h_E - O(h_E^{\text{geom}} + h_E^{\text{geom}}) \\
&= -h_E^{\text{arith}} + O(h_E^{\text{geom}} + h_E^{\text{arith}}).
\end{align*}
\]

Now, let $L = (L_1, \ldots, L_r)$ be a tuple of line bundles generating the free module $M$ of rank $r$. We choose morphisms $\varphi_1, \psi_1$ from $X$ to projective space, with $L_i \cong \varphi_i^*O(1) \otimes \psi_i^*O(1)$, and set

\[ h_{L_i}^{\text{geom}} = h_{\varphi_1}^{\text{geom}} - h_{\psi_1}^{\text{geom}} \quad \text{and} \quad h_{L_i}^{\text{arith}} = h_{\varphi_1}^{\text{arith}} - h_{\psi_1}^{\text{arith}}. \]

Finally, for $x = (x_1, \ldots, x_r) \in \mathbb{R}^r$ we set

\[ x^T L = \bigotimes_{i=1}^r L_i^{x_i} \in M \otimes \mathbb{R}, \]

and

\[ h_x^{\text{geom}} = \sum_{i=1}^r x_i h_{L_i}^{\text{geom}} \quad \text{and} \quad h_x^{\text{arith}} = \sum_{i=1}^r x_i h_{L_i}^{\text{arith}}. \]

Note that, by the freeness of $M$, each element of $M \otimes \mathbb{R}$ can be written uniquely as $x^T L$. Now, suppose that $f^*L_i \cong \bigotimes_{j=1}^r L_j^\otimes A_{i,j}$ for all $i$, and let $A$ be the matrix with entries $A_{i,j}$, and set

\[ \|A\| = \sup_{y \neq 0} \frac{\|Ay\|}{\|y\|}, \]

where $\|\cdot\|$ denotes the norm on $\mathbb{R}$. Then

\[ \|x^T L\| \leq \sup_{y \neq 0} \frac{\|f^*L_i\|}{\|y\|} = \sup_{y \neq 0} \frac{\|A_{i,j}\|}{\|y\|}. \]
noting that \( \|A\| \leq r \max |A_{i,j}| \).

**Lemma 13.** Let \( P \in X(K(t)) \) and \( x \in \mathbb{R}^r \), and fix \( E \) ample. We have
\[
h_{\text{geom}}^x(f(P)) = h_{Ax}^x(P) + O(\|x\|)
\]
and
\[
h_{\text{arith}}^x(f(P)) = h_{Ax}^x(P) + O(\|x\| h_{E}^x(P)),
\]
where the implied constants depend on \( f \), \( A \), and \( E \), but not on \( P \) or \( x \).

**Proof.** If \( e_i \) is the \( i \)th standard basis vector, we have by Lemma 12
\[
h_{e_i}^{\text{geom}} \circ f = h_{L_i}^{\text{geom}} \circ f = h_{f^{*}L_i}^{\text{geom}} = h_{A e_i}^{\text{geom}} + O(1),
\]
since \((\varphi \circ f)^{*} \mathcal{O}(1) \cong (A e_i)^T L \). The implied constant depends on \( i \), but of course we may take a bound that works for all \( i \), giving
\[
h_{x}^{\text{geom}}(f(P)) = \sum_{i=1}^{r} x_i h_{e_i}^{\text{geom}}(f(P))
\]
\[
= \sum_{i=1}^{r} x_i \left( h_{A e_i}^{\text{geom}}(P) + O(1) \right)
\]
\[
= h_{Ax}^{\text{geom}}(P) + O(\|x\|).
\]
The relation
\[
f^{*} L_i \cong (A e_i)^T L \cong \bigotimes_{j=1}^{r} L_j ^{\otimes A_{i,j}}
\]
with Lemma 12 13 also provides
\[
h_{e_i}^{\text{arith}} \circ f = h_{A e_i}^{\text{arith}} + O(h_{E}^{\text{geom}}),
\]
the same argument as above now produces
\[
h_{x}^{\text{arith}}(f(P)) = h_{Ax}^{\text{arith}}(P) + O(\|x\| h_{E}^{\text{geom}}(P)).
\]

\( \square \)

The previous lemma in hand, we estimate the growth of the geometric and arithmetic heights in orbits.

**Lemma 14.** Let \( \delta > 0 \), let \( L \in M \otimes \mathbb{R} \), and let \( \rho(A) \) be the spectral radius of \( A \). Then
\[
h_{L}^{\text{geom}}(f^k(P)) = O \left( (\rho(A) + \delta)^k \right),
\]
and
\[
h_{L}^{\text{arith}}(f^k(P)) = O \left( (\rho(A) + \delta)^{k(1+\delta)} \right),
\]
where the implied constants depends on \( f \), \( P \), \( A \), \( L \), and \( \delta \), but not on \( k \).
Proof. Write $L = x^T L$. From Lemma [13] we have
\[ h_{x}^\text{geom}(f^k(P)) = h_{Ax}^\text{geom}(f^{k-1}(P)) + O(\|x\|) \]
\[ = h_{A^2x}^\text{geom}(f^{k-2}(P)) + O(\|x\| + \|Ax\|) \]
\[ = h_{A^kx}^\text{geom}(P) + O\left(\sum_{i=0}^{k-1} \|A^i x\|\right). \]

Now, by Gelfand’s formula for the spectral radius, we have $\|A^k\| \leq C(\rho(A) + \delta)^k$ for some constant $C$, independent of $k$. Because $\|Ax\| \leq r\|A\| \cdot \|x\|$, we then have
\[ \sum_{i=0}^{k-1} \|A^i x\| \leq rC \|x\| \sum_{i=0}^{k-1} (\rho(A) + \delta)^i \leq \frac{rC \|x\| ((\rho(A) + \delta)^k - 1)}{\rho(A) + \delta - 1} = O((\rho(A) + \delta)^k). \]

On the other hand, note that for $E$ ample, (30) implies
\[ h_{x}^\text{geom} = O(\|x\| h_{E}^\text{geom}) \]
for $x \in \mathbb{R}^r$. We then also have
\[ h_{A^kx}^\text{geom}(P) = O\left(\|A^k x\| h_{E}^\text{geom}(P)\right) = O\left((\rho(A) + \delta)^k\right). \]
This completes the proof of the first claim.

Next, observe that
\[ h_{x}^\text{arith}(f^k(P)) = h_{Ax}^\text{arith}(f^{k-1}(P)) + O(\|x\| h_{E}^\text{geom}(f^{k-1}(P))) \]
\[ = h_{A^kx}^\text{arith}(P) + O\left(\sum_{i=0}^{k-1} \|A^i x\| h_{E}^\text{geom}(f^{k-1-i}(P))\right). \]

We have from (30) that
\[ h_{A^kx}^\text{arith}(P) = O\left(\|A^k x\| (h_{E}^\text{arith}(P) + h_{E}^\text{geom}(P))\right) = O\left((\rho(A) + \delta)^k\right) \]
just as above. On the other hand, (31) applied to $h_{E}^\text{geom}$, combined again with $\|A^k\| \ll (\rho(A) + \delta)^k$, gives
\[ \sum_{i=0}^{k-1} \|A^i x\| h_{E}^\text{geom}(f^{k-1-i}(P)) = O\left(\|x\| \sum_{i=0}^{k-1} \|A^k\| (\rho(A) + \delta)^{k-1-i}\right) \]
\[ = O\left(\sum_{i=0}^{k-1} (\rho(A) + \delta)^i\right) \]
\[ = O\left((\rho(A) + \delta)^{k(1+\delta)}\right), \]
since $k < (\rho(A) + \delta)^{\delta k}$ for $k$ large enough. 

Now, away from a set of bounded height, each $L_i \in \text{Pic}(X)$ specializes to a line bundle $L_{i,t} \in \text{Pic}(X_t)$, and each $\varphi_i$ specializes to a morphism $\varphi_{i,t} : X_t \to \mathbb{P}^{m_i}$ with $\varphi_{i,t}^* O(1) \cong L_{i,t}$. We define $h_{L_{i,t}} = h \circ \varphi_{i,t}$, and for any $x^T \in \mathbb{R}$, we define
\[ h_{x^T,t} = h_{x^T,x^T}^\text{arith}. \]
For $L \in M \otimes \mathbb{R}$ we define $h_{L,t} = h_{x^T,t}$, where $x^T L = L$. With these chosen height functions, we have the following.
Lemma 15. For any \( L \in M \otimes \mathbb{R} \) and any \( \delta > 0 \),
\[
\left| h^\text{geom}_L(f^k(P)) h(t) - h_{L_i}(f^k(P)_i) \right| = O \left( (\rho(A) + \delta)^{k(2+\delta)} \right),
\]
where the implied constant is independent of \( t \) and \( k \).

Proof. By our choices of height functions, it suffices to prove this result for \( L = L_i \).
In that case, Lemma 7 gives us (off of a set of bounded height)
\[
\left| h^\text{geom}_{L_i}(f^k(P)) h(t) - h_{L_i}(f^k(P)_i) \right| = O \left( h^\text{geom}(\varphi_i \circ f^k(P)) h^\text{arith}(\varphi_i \circ f^k(P)) \right)
\]
\[
+ \left| h^\text{geom}_{L_i}(f^k(P)) h^\text{arith}(f^k(P)) \right|
\]
\[
= O \left( h^\text{geom}_{L_i}(f^k(P)) h^\text{arith}(f^k(P)) \right)
\]
\[
= O \left( (\rho(A) + \delta)^{k(2+\delta)} \right)
\]
by Lemma 14. \( \Box \)

Proof of Theorem 10. Let \( M \) be generated by \( L_1, ..., L_r \), semi-ample. Note that it suffices to prove the statement after replacing all of the \( L_i \) by \( L_i^{\otimes m} \) for some \( m \geq 1 \), and so we will assume without loss of generality that the \( L_i \) are generated by global sections.

Let \( \varepsilon > 0 \), and choose \( \delta > 0 \) so that
\[
\frac{\log \alpha}{(2 + \delta) \log(\rho(A) + \delta)} \geq \frac{\log \alpha}{2 \log \rho(A)} - \varepsilon.
\]
Choose \( k \) so that
\[
(32) \quad (\rho(A) + \delta)^{k(2+\delta)} \leq h(t) < (\rho(A) + \delta)^{(k+1)(2+\delta)}.
\]
From Lemma 14 we have
\[
\left| h^\text{geom}_L(f^k(P)) h(t) - h_{L_i}(f^k(P)_i) \right| \leq C_{10} (\rho(A) + \delta)^{k(2+\delta)}
\]
for some constant \( C_{10} \), and for \( k \) (equivalently \( h(t) \)) large enough. Similarly, by [5, Theorem 3.1] we have
\[
\left| \hat{h}_{f_i,L_i}(Q) - h_{X_i,L_i}(Q) \right| \leq C_{11} h(t),
\]
which we apply with \( Q = f^k(P)_i \), and
\[
\left| h^\text{geom}_L(f^k(P)) - \hat{h}_{f,L}(P) \right| \leq C_{12}.
\]
Combining these, we have
\[
\alpha^k \left| \hat{h}_{f_1,L^*}(P_t) - \hat{h}_{f,L}(P)h(t) \right| = \left| \hat{h}_{f_1,L^*}(f^k(P)t) - \hat{h}_{f,L}(f^k(P))h(t) \right|
\leq \left| \hat{h}_{f_1,L^*}(f^k(P)t) - h_{X_1,L^*}(f^k(P)t) \right|
+ |h_{\text{geom}}^L(f^k(P))h(t) - h_{L^1,L}(f^k(P)t)|
+ |h_{\text{geom}}^L(f^k(P))h(t) - \hat{h}_{f,L}(f^k(P))h(t)|
\leq (C_{11} + C_{12})h(t) + C_{10}(\rho(A) + \delta)^{(2+\delta)}
\leq C_{13}h(t)
\]
by (32). We then have
\[
\alpha^k = (\rho(A) + \delta)^k \log \alpha/\log(\rho(A)+\delta) > h(t)\log \alpha/(2+\delta) \log(\rho(A)+\delta)(\rho(A) + \delta)^{-2(2+\delta)},
\]
whence
\[
\left| \hat{h}_{f_1,L^*}(P_t) - \hat{h}_{f,L}(P)h(t) \right| \leq \alpha^{-k}C_{13}h(t)
\leq C_{13}h(t) - \log \alpha/2 \log(\rho(A)+\delta)(\rho(A) + \delta)^{-2(2+\delta)}
\leq C_{14}h(t)^{1-\frac{2\log \alpha}{\log(\rho(A))^{1+\epsilon}}}
\]
for $h(t)$ large enough.

Note that we have $\alpha > 1$, and so $\log \alpha/2 \log(\rho(A)) > 0$. In particular Theorem 10 always gives an improvement
\[
\hat{h}_{f_1,L^*}(P_t) = \hat{h}_{f,L}(P)h(t) + O\left(h(t)^{1-\eta}\right)
\]
over (1), for some $\eta > 0$ depending on the action of $f^*$ on $\text{Pic}(X)$. It is possible to artificially concoct examples in which this improvement is arbitrarily slight. For example, if $f_1, f_2$ are rational function of degree $d \geq e \geq 2$, then $f = (f_1, f_2)$ is an endomorphism of $X = \mathbb{P}^1 \times \mathbb{P}^1$ satisfying $f^*\mathcal{O}(a,b) = \mathcal{O}(da, eb)$. Applying the theorem to $L = \mathcal{O}(0,1)$, we have $\alpha = e$ and $\rho = d$, and so $\log \alpha/2 \log \rho$ may be made arbitrarily small. On the other hand, one could here apply the theorem to $M = \mathcal{O}(0,1) \subseteq \text{Pic}(X)$ to obtain the better exponent of $\frac{1}{2} + \epsilon$, or even apply Theorem 1 to the projection onto the first coordinate to eliminate the $\epsilon$.

**Proof of Theorem 3**. Theorem 3 follows immediately from Theorem 10. In particular, Silverman [26] considers K3 surfaces $X \subseteq \mathbb{P}^2 \times \mathbb{P}^2$ defined by the simultaneous vanishing two multihomogenous forms of degree (1,1) and (2,2) respectively. The hyperplanes on the two copies of $\mathbb{P}^2$ define two line bundles $D_1, D_2$, and there is an automorphism $f$ of $X$ with
\[
f^*D_1 = D_1^{15} \otimes D_2^{-4}
f^*D_2 = D_1^{24} \otimes D_2^{-1}.
\]
For $\beta = 2 + \sqrt{3}$, and $E^+ = D_1^{\otimes \beta} \otimes D_2^{-1}$ we have $f^*E^+ \cong (E^+)^{\otimes \beta^2}$, and $\hat{h}^+$ is the canonical height associated to $f$ and $E^+$. Similarly, for $E^- = D_1^{-1} \otimes D_2^{\otimes \beta}$ we have $(f^{-1})^*E^- \cong (E^-)^{\otimes \beta^2}$, and $\hat{h}^-$ is the canonical height associated to $f^{-1}$ and $E^-$.

\[\square\]
4. The case of a general base

In Section 2 we proved Theorem 1 in the case $B = \mathbb{P}^1$ and $X = \mathbb{P}^N$. The former restriction, and the fact that $\text{Pic}^0(\mathbb{P}^1)$ is trivial, appears to genuinely improve our bounds, but it turns out that the latter restriction is immaterial.

Lemma 16. If Theorem 1 is true in the case $X = \mathbb{P}^N$ and $L = \mathcal{O}(1)$, over a given base $B$, then it is true as stated.

Proof. Suppose, as in the theorem, that $X$, $L$, and $f$ are defined over $K(B)$, with $f^* L \cong L \otimes d$. By a theorem of Fakhruddin [8] there exist $N, n \geq 1$, an embedding $i : X \to \mathbb{P}^N$, and a morphism $g : \mathbb{P}^N \to \mathbb{P}^N$ of degree $d$ such that $i \circ f = g \circ i$ and $i^* \mathcal{O}(1) \cong L \otimes n$. Note that we have

$$nh_{X,L}(P) = h_{\mathbb{P}^N, \mathcal{O}(1)}(i(P)) + O(1),$$

and so the canonical heights satisfy

$$n\hat{h}_{f,X,L}(P) = \hat{h}_{g,\mathbb{P}^N, \mathcal{O}(1)}(i(P)) + O(1).$$

On the other hand, the transformation relation for the canonical heights gives

$$n\hat{h}_{f,X,L}(P) = \frac{n}{d^k} \hat{h}_{f,X,L}(f^k(P)) = \frac{1}{d^k} \left(\hat{h}_{g,\mathbb{P}^N, \mathcal{O}(1)}(g^k \circ i(P)) + O(1)\right) = \hat{h}_{g,\mathbb{P}^N, \mathcal{O}(1)}(i(P)) + O(d^{-k})$$

for any $k \geq 0$, so in fact the two canonical heights in (33) are exactly equal.

Then, for all but finitely many $t \in B$ we have a specialization $i_t : X_t \to \mathbb{P}^N$ which is an embedding, with $i_t^* \mathcal{O}(1) = L_t \otimes n$, and we may use the same argument on each fibre to conclude that

$$n\hat{h}_{f_t,X_t,L_t}(Q) = \hat{h}_{g_t,\mathbb{P}^N, \mathcal{O}(1)}(i_t(Q))$$

for any $Q \in X_t(K)$, including $Q = P_t$. Applying the theorem to $i(P) \in \mathbb{P}^N(B)$ relative to $g$ then gives the result for $P \in X(B)$ relative to $f$. \hfill \square

In light of the lemma above, we will restrict attention to the case $X = \mathbb{P}^N$, $L = \mathcal{O}(1)$, and we take $B$ to be a smooth, projective curve of genus $g \geq 1$ over a number field $K$. (These arguments could also be used in the case $B = \mathbb{P}^1$, if we use $2g = 1$ throughout, but the conclusions would be weaker than those in Section 2.)

Since we will necessarily be somewhat pedantic about heights on $B$, our first goal is to define a “reference height” relative to each divisor. By the Riemann-Roch Theorem there is, for each point $\beta \in B$, a morphism $\varphi^\beta : B \to \mathbb{P}^g$ such that $\varphi^\beta_* \mathcal{O}(1) = \mathcal{O}(2g[\beta])$. We fix one such map for each point, and for an $\mathbb{R}$-Cartier divisor $D = \sum_{\beta \in B} m_\beta [\beta]$ define

$$[\beta, D] = \frac{1}{2g} \sum_{\beta \in B} m_\beta h_{\mathbb{P}^g} \circ \varphi^\beta.$$ 

The functions $[\beta, D]$ are thus well-defined and linear in $D$, while also satisfying

$$h_{B,D} = [\beta, D] + O(1)$$

for any other choice of height function. We will show below that if $m_\beta \in \mathbb{Z}$ for all $\beta \in B$, then there exist morphisms $\varphi : B \to \mathbb{P}^n$ and $\psi : B \to \mathbb{P}^m$ such that

$$\mathcal{O}(2gD) \otimes \psi^* \mathcal{O}(1) = \varphi^* \mathcal{O}(1) \quad \text{and} \quad 2g[\beta, D] = h \circ \varphi - h \circ \psi.$$
So $2q[B,D]$ is always the Weil height associated to a particular presentation of the divisor $2qD$, in the sense of [1, Chapter 2]. On the other hand, note that $[2qD]$ depends on the choice of $D$ as a representative of its divisor class (only up to a bounded function, but this matters for our argument). We also note that our reference heights are chosen so that $[2qD] \geq 0$ whenever $D \geq 0$.

Now, we fix a morphism $f : \mathbb{P}^N \to \mathbb{P}^N$ defined over $K(B)$ of degree $d \geq 2$, and a $K(B)$-rational point $P \in \mathbb{P}^N$. We will also choose a tuple of functions $P_i \in K(B)$ with $P = [P_0 : \cdots : P_N]$, writing $P$ for the tuple of functions $P_i$ as well. Finally, we choose homogeneous forms $F_i(X) \in K(B)[X]$ such that

$$f(X) = [F_0(X) : \cdots : F_N(X)].$$

We write $F$ for the endomorphism of $\mathbb{A}^{N+1}_{K(B)}$ given by the $F_i$, and $\overline{F}$ for the tuple of coefficients of all of the $F_i$.

**Lemma 17.** There is a finite set $S \subseteq B$ and a sequence of divisors $D_k$ on $B$ such that

1. $D_k$ is supported on $S$
2. $\mathcal{O}(D_k) \cong f^k(P)^* \mathcal{O}(1)$
3. $D(F,P) := \lim_{k \to \infty} d^{-k} D_k$ exists in $\text{Div}(B) \otimes \mathbb{R}$
4. $D(F,F(P)) = dD(F,P)$

**Proof.** Let $\overline{\varphi} = (\varphi_0, \ldots, \varphi_N)$ with $\varphi_i \in K(B)$. Then for $|\psi| = e^{-\text{ord}_B(\psi)}$

$$D(\overline{\varphi}) = \sum_{\beta \in B} \log \|\varphi_0, \ldots, \varphi_N\|_{\beta}[\beta].$$

Note that if $\varphi : B \to \mathbb{P}^N$ is defined by the coordinate functions $\varphi_0, \ldots, \varphi_N$, we have

$$D(\overline{\varphi}) = \varphi^* H_i - \text{div}(\varphi_i),$$

for each $i$ (where $H_i$ is the $i$th coordinate hyperplane in $\mathbb{P}^N$). In particular, $\mathcal{O}(D(\overline{\varphi})) \cong \varphi^* \mathcal{O}(1)$.

We assume that $T \subseteq S \subseteq B$ are finite sets large enough that $\|\overline{P}\|_{\beta} = 1$ for all $\beta \in T$, and $\|P\|_{\beta} = 1$ for all $\beta \notin S$. First, we have (for all $\beta \in B$)

$$\log |F_i(P)|_{\beta} \leq d \log \|P\|_{\beta} + \log \|\overline{F}\|_{\beta},$$

and so

$$D(F(P)) \leq dD(P) + D(\overline{F}).$$

On the other hand, as in the proof of Lemma 8 there are homogeneous forms $A_{i,j}(X) \in K(B)(X)$ of degree $N(d-1)$ such that

$$X_i \cdot (X^{N+1} - (d-1)+1)^{N+1} = F_0(X)A_{i,0}(X) + \cdots + F_N(X)A_{i,N}(X)$$

for all $i$, and so from the ultrametric inequality again we have

$$d \log \|P\|_{\beta} \leq \log \|F(P)\|_{\beta} + \log \|\overline{A}\|_{\beta},$$

where $\overline{A}$ is the grand tuple of coefficients of the $A_i$. We thus have

$$dD(P) \leq D(F(P)) + D(\overline{A})$$

(note that we may estimate the last quantity by way of the effective Nullstellensatz, but this is not particularly useful here).
Now take $T$ large enough so that $\|\beta\|_\beta = 1$ for $\beta \notin S$, which is still a finite set depending just on $P$ and $F$. From (35) and (36) we see that $dD(P) - D(F(P))$ is a divisor supported on $T \subseteq S$, whose order at each point of $T$ is bounded above and below. We set $D_k = D(F^k(P))$, and note by induction that $D_k$ is supported on $S$. Note that the entries of $F^k(P)$ define $f^k(P) : B \to \mathbb{P}^N$, and so $\mathcal{O}(D_k) = f^k(P)^*\mathcal{O}(1)$.

Finally, let $E$ be any divisor supported on $T$ with $D(\bar{F}), D(\bar{A}) \leq E$, so that

$$-E \leq dD(P) - D(F(P)) \leq E.$$  

Note that $E$ need not depend on $P$, here. By the usual telescoping sum argument,

$$-\frac{1}{(d-1)d_{\min}(k,m)} E \leq -d^{-k}D(F^k(P)) - d^{-m}D(F^m(P)) \leq \frac{1}{(d-1)d_{\min}(k,m)} E,$$

and so $d^{-k}D_k$ converges in $\text{Div}(B) \otimes \mathbb{R}$ (which is just to say that the orders of $d^{-k}D_k$ at each point converge). If we set

$$D(F, P) = \lim_{k \to \infty} \frac{D_k}{d^k},$$

then $D(F, F(P)) = dD(F, P)$ immediately from the definition, and from the telescoping sum we have

$$(37) \quad -\frac{1}{d-1} E \leq D(P) - D(F, P) \leq \frac{1}{d-1} E,$$

Note that we made choices of coordinate for $F$ and $P$. If $\sigma$ and $\theta$ are any two non-zero functions on $B$, then

$$D(\sigma F, \theta P) = D(F, P) + \frac{1}{d-1} \text{div}(\sigma) + \text{div}(\theta),$$

and so while the construction of $D(F, P)$ is sensitive to these choices, the associated class $L(f, P) = \mathcal{O}(D(F, P)) \in \text{Pic}(B)$ is not. It is coherent to speculate, then, that we in fact have

$$\hat{h}_{f_k}(P_t) = h_{B,L(f,P)}(t) + O(1),$$

although that still seems out of reach.

**Lemma 18.** With $D_k$, $D(F, P)$, and $E$ as in Lemma 17, we have

$$|B, D_k| - d^k|B, D(F, P)| \leq \frac{1}{d-1}|B, E|$$

**Proof.** It follows from (37), and the fact that the reference heights are linear in the divisor and non-negative for effective divisors, that

$$-\frac{1}{d-1}|B, E| \leq |D(P)| - |B, D(F, P)| \leq \frac{1}{d-1}|B, E|$$

Now replace $P$ with $F^k(P)$, noting that $D(F, F^k(P)) = d^k D(F, P)$, and use the linearity of the reference heights again to conclude that $|B, D(F, F^k(P))| = d^k|B, D(F, P)|$. $\square$
homogeneous coordinate ring $S$ with with the space of homogeneous polynomials in $x, y, z$, with $z$-degree less than $\deg(F)$, and we identify these spaces. Recall that the tuple $p = (p_0, ..., p_n)$ of elements of $S$ is a presentation of the morphism $\varphi = [\varphi_0 : \cdots : \varphi_n] : B \to \mathbb{P}^n$ if and only if (1) $p_j \neq 0$ for any $j$ such that $\varphi_j \neq 0$, and (2) for $j$ with $\varphi_j \neq 0$, we have $p_i/p_j = \varphi_i/\varphi_j$ in $K(B) = K(\pi(B))$. The fact that $B$ and $\pi(B)$ have the same function field ensures that every morphism has a presentation, and given a presentation $p$ we write $\deg(p)$ for the degree of the homogeneous forms $p_i$, and $h(p)$ for the height of the homogeneous tuple of coefficients of all of the $p_i$.

For two morphisms $\varphi : B \to \mathbb{P}^n$ and $\psi : B \to \mathbb{P}^m$ with coordinates $\varphi_i$ and $\psi_j$, we write

$$\varphi \# \psi : B \to \mathbb{P}^{(n+1)(m+1)-1}$$

for the Segre join, with coordinates $\varphi_i \psi_j$. We recall that $(\varphi \# \psi)^* \mathcal{O}(1) = \varphi^* \mathcal{O}(1) \otimes \psi^* \mathcal{O}(1)$, and that $h \circ (\varphi \# \psi) = h \circ \varphi + h \circ \psi$. If $p$ and $q$ are presentations of $\varphi$ and $\psi$, with entries $p_i$ and $q_j$, then the tuple of homogeneous forms with entries $p_i q_j$ is a presentation of $\varphi \# \psi$, and we will denote this presentation by $p \# q$.

The next lemma, a slight variation of [4, Theorem 2.5.14, p. 53], makes explicit the fact that any height function relative to the trivial divisor class on $B$ is bounded.

**Lemma 19.** Let $\varphi : B \to \mathbb{P}^n$ and $\psi : B \to \mathbb{P}^m$ be morphisms with presentations $p$ and $q$ respectively, and suppose that $\varphi^* \mathcal{O}(1) \cong \psi^* \mathcal{O}(1)$. Then

$$|h_\varphi - h_\psi| \ll \max \{\deg(p), \deg(q)\}^2 (h(p) + h(q) + \log(1 + \deg(p)) + \log(1 + \deg(q))),$$

with implied constants depending only on $B$, $n$, and $m$.

**Proof.** This is a variant of [4, Theorem 2.5.14, p. 53]. Specifically, fix a closed embedding $\theta : B \to \mathbb{P}^3$, with presentation $t$. Then $\varphi \# \theta : B \to \mathbb{P}^{4n+3}$ and $\psi \# \theta : B \to \mathbb{P}^{4m+3}$ are closed embeddings, and so we may apply [4, Theorem 2.5.14, p. 53] to obtain (for $C_{14}$ and $C_{15}$ depending just on $\pi(B)$)

$$h_\varphi - h_\psi = h_{\varphi \# \theta} - h_{\psi \# \theta} \leq C_{14}(4n + 4) \deg(q \# \theta t)^2 (h(p \# \theta t) + h(q \# \theta t) + \log(1 + \deg(p \# \theta t))$$

$$+ \log(n + 1) + \log(1 + \deg(q \# \theta t)) + \log(n + 1) + \log(1 + \deg(q \# \theta t)) + \log(n + 1) + \log(1 + \deg(q \# \theta t)).$$

(38)

On the other hand, [4, Lemma 2.5.6, p. 48] gives

$$\deg(p \# \theta t) = \deg(p) + \deg(t)$$

and

$$h(p \# \theta t) \leq h(p) + h(t) + \log(1 + \deg(t)) + C_{16},$$

where the constant $C_{16}$ depends only on $\pi(B)$. (Note that the lemma assumes that $\varphi$ and $\theta$ are both closed embeddings but, as pointed out in [4, Remark 2.5.7, p. 49], that assumption is not needed for the inequalities above.) Combining with (38) above, we have

$$h_\varphi - h_\psi \ll \deg(q)^2 (h(p) + h(p) + \log(1 + \deg(p)) + \log(1 + \deg(q))),$$

where the constants depend on $\pi(B)$, $n$, and our choice of $\theta$ (which may be made once for the curve $B$). The claim follows by swapping $\varphi$ and $\psi$ in this bound. \(\square\)

We now construct presentations of the morphisms $f_k(P) : B \to \mathbb{P}^N$. 
Lemma 20. The morphisms $f^k(P) : B \to \mathbb{P}^N$ admit presentations $p_k$ satisfying
\[ \deg(p_k), h(p_k) = O(d^k), \]
where the implied constants depend on $f$, $P$, $B$, and $N$, but not on $k$.

Proof. Let $p_0$ be a presentation of the morphism $F : B \to \mathbb{P}^N$, whose $i$th entry we denote $p_{0,i}$. As above, we may represent $F$ as a tuple of $N$ homogeneous forms of degree $d$ with coefficients in $K(B)$, thereby associating $F$ with a point in $\mathbb{P}^{(N+1)(N+1)-1}_{K(B)}$, i.e., a morphism $B \to \mathbb{P}^{(N+1)(N+1)-1}$ over $K$, which admits a presentation $F$. The homogeneous form in $F$ corresponding to the coefficient of the monomial $m$ in the $i$th entry of $F$ will be written $F_{i,m}$.

Now, for $k \geq 0$, let
\[ p_{k+1,i} = \sum_{\deg(m) = d} F_{i,m}(p_{k,0}, \ldots, p_{k,N}), \]
where the sum is over all monomials of degree $d$ in $N + 1$ variables. It is easy to check by induction that $p_k$ is a presentation of $f^k(P)$.

From [4] Lemma 2.5.6, p. 48 we have
\[ \deg(p_{k+1}) \leq d \deg(p_k) + \deg(F), \]
whence
\[ \deg(p_k) \leq d^k \left( \deg(p_0) + \frac{1}{d-1} \deg(F) \right). \]

Again from [4] Lemma 2.5.6, p. 48 we have
\[ h(p_{k+1}) \leq dh(p_k) + h(F) + d \log(1 + \deg(p_k)) + C_{17}, \]
where $C_{17}$ depends only on $B$, $\pi(B)$, and $d$. This gives
\[ h(p_k) \leq d^k h(p_0) + d^k \sum_{j=0}^{k-1} d^{-j} \log \left( 1 + d^j \left( \deg(p_0) + \frac{1}{d-1} \deg(F) \right) \right) + \frac{d^k - 1}{d-1} (h(F) + C_{17}), \]
which provides the claim $h(p_k) = O(d^k)$ as $k \to \infty$ in light of (21), which bounds the remaining sum with no dependence on $k$. \hfill \Box

Now let $S$ be the set of places from Lemma 17 and let $s = \#S$.

Lemma 21. For each $k$ there is a morphism $\varphi^{D_k} : B \to \mathbb{P}^{(g+1)}$ such that $\varphi^{D_k}_* O(1) = O(2gD_k)$, with $h \circ \varphi^{D_k} = 2g[B,D_k]$ and admitting a presentation $q_k$ satisfying
\[ \deg(q_k), h(q_k) = O(d^k), \]
where the implied constants depend on $f$, $P$, $B$, and $N$, but not on $k$.

Proof. Let $D$ be any effective divisor supported on $S$, set $s = \#S$, and consider the divisor $2gD = \sum 2gm_{\beta}[\beta]$. For each $\beta \in S$ we have a morphism $\varphi_{\beta} : B \to \mathbb{P}^g$ with $\varphi_{\beta}^* O(1) = O(2g[N])$, and we fix a presentation $p_{\beta}$ for this morphism. Now, if $\epsilon_m : \mathbb{P}^g \to \mathbb{P}^g$ is the $m$th power map, $\varphi_{\beta} \circ \epsilon_m$ is presented by $p_{\beta}^{m_{\beta}}$, where powers are taken component-wise. From the estimates in the proof of [4] Lemma 2.5.6, p. 48, we have
\[ \deg(p_{\beta}^{m_{\beta}}) = m_{\beta} \deg(p_{\beta}), \]
and
\[ h(p_{\beta_1}^{m_1} \cdots p_{\beta_s}^{m_s}) \leq m_{\beta} h(p_{\beta}) + (m_{\beta} - 1) \log(1 + \deg(p_{\beta})) + C_{18}(m_{\beta} - 1), \]
for \( C_{18} \) depending just on \( \pi(B) \). In particular, \( h(p_{\beta_1}^{m_1} \cdots p_{\beta_s}^{m_s}) \leq m_{\beta} C_{\beta} \) for some \( C_{\beta} \) depending on \( \pi(B) \) and \( \beta \). Now, the morphism
\[ \varphi_D := \varphi_{\beta_1}^{m_1} \# \cdots \# \varphi_{\beta_s}^{m_s} : B \to \mathbb{P}^{(g + 1)'} \]
has presentation given by
\[ p_{\beta_1}^{m_1} \# \cdots \# p_{\beta_s}^{m_s}, \]
and the usual estimates give
\[
\deg(p_{\beta_1}^{m_1} \# \cdots \# p_{\beta_s}^{m_s}) = \sum \deg(p_{\beta_i}^{m_{\beta_i}}) \leq \deg(2gD) \max \deg(p_{\beta_i})
\]
and
\[
\leq \deg(2gD) \max \{C_{\beta}\} + \sum \deg(p_{\beta_i}^{m_{\beta_i}}) + C_{19}s
\]
\[
\leq \deg(2gD) C_{20},
\]
where \( C_{20} \) depends on \( S \), as long as \( \deg(D) \geq 1 \). Finally, note that
\[
h \circ \varphi_D = h \circ (\varphi_{\beta_1}^{m_1} \# \cdots \# \varphi_{\beta_s}^{m_s})
\]
\[
= h \circ \varphi_{\beta_1}^{m_1} + \cdots + h \circ \varphi_{\beta_s}^{m_s}
\]
\[
= m_1 h \circ \varphi_{\beta_1} + \cdots + m_s h \circ \varphi_{\beta_s}
\]
\[
= 2g[B,D].
\]
The lemma follows from applying this construction to \( D_k \), since
\[
\deg(2gD_k) = 2gh^{\text{geom}}(f^k(P)) = O(d^k).
\]

Lemma 22. We have
\[
h(f^k(P)_t) = [B,D_k](t) + O(d^k),
\]
where the implied constants depend on \( f \) and \( P \), but not on \( t \) or \( k \).

Proof. This is the combination of the previous few lemmas. We have
\[
f^k(P)O(2g) \cong O(2gD_k) \cong \varphi_{D_k}O(1),
\]
and also that \( f^k(P) : B \to \mathbb{P}^N \) and \( \varphi_{D_k} : B \to \mathbb{P}^{(g + 1)'} \) admit presentations of height and degree at most \( O(d^k) \). The Lemma 19 completes the proof.

Finally, we note the following result, analogous to Lemma 8 above.

Lemma 23. There is a divisor \( D \) on \( B \) such that for all but finitely many \( t \),
\[
\hat{h}_t = h + O(h_{B,D}(t)).
\]

Proof. This is already in [5], but the same proof that gives (11) in Lemma 8 works here.
Proof of Theorem 7 with $B$ irrational. As noted in Lemma 16 it suffices to treat the case $X = \mathbb{P}^N$ and $L = \mathcal{O}(1)$, so assume we are in that case. Also note that, since (40)

$$h_{B,D} = h_{B,E} + O(h_{B,E}^{1/2})$$

for any two height functions of the same (positive) degree on $B$ (see [4] Remark 9.3.9, p. 293], and note that this is true for heights relative to $\mathbb{R}$-divisors, as well), it suffices to prove the result for a particular height function.

Let $A$ be any ample divisor on $B$, and let $h_{B,A}$ be a corresponding Weil height. Note that we have

$$h_{B,D} = O(h_{B,A})$$

as $h_{B,A} \to \infty$, for any Weil height $h_{B,D}$ relative to any $\mathbb{R}$-divisor $D$, and so in particular we now have (for $h_{B,A}(t)$ sufficiently large)

$$\left| \hat{h}_f(Q) - h(Q) \right| \leq C_{21}h_{B,A}(t),$$

for all $Q \in \mathbb{P}^N$ by Lemma 23

$$|h(f^k(P)_t) - [B,D_k(t)]| \leq C_{22}d^{3k}$$

by Lemma 22 and

$$[B, D_k] - d^k[D(F, P)] \leq C_{23}h_{B,A}(t)$$

by Lemma 18.

For $t \in B(K)$, take $k \geq 0$ with

$$d^{3k} \leq h_{B,A}(t) < d^{3(k+1)}.$$

We then have

$$\frac{h_{B,A}^{1/3}}{d} \left| \hat{h}_f(P_t) - D(F, P)_t \right| \leq d^k \left| \hat{h}_f(P_t) - D(F, P)_t \right|$$

$$= \hat{h}_f(f^k(P)_t) - d^k[D(F, P)_t]$$

$$\leq |\hat{h}_f(f^k(P)_t) - h(f^k(P)_t)|$$

$$+ |h(f^k(P)_t) - [B, D_k(t)]|$$

$$+ |[B, D_k] - d^k[D(F, P)]_t|$$

$$\leq C_{21}h_{B,A}(t)$$

$$+ C_{22}d^{3k}$$

$$+ C_{23}h_{B,A}(t)$$

$$\leq h_{B,A}(t)(C_{21} + C_{22} + C_{23}),$$

and hence

$$\left| \hat{h}_f(P_t) - D(F, P)_t \right| \leq h_{B,A}(t)^{2/3}d(C_{21} + C_{22} + C_{23}).$$

But now, $D(F, P)$ is the Weil height relative to some $\mathbb{R}$-divisor of degree $\hat{h}_f(P)$ on $B$. By (40) if $\hat{h}$ is a height on $B$ relative to any divisor of degree 1,

$$D(F, P) = \hat{h}_f(P)h(t) + O(h(t)^{1/2}),$$

and since $h_{B,A} = O(h)$ we have (since $(x + y)^\rho \leq x^\rho + y^\rho$ when $0 < \rho < 1$)

$$\hat{h}_f(P_t) = \hat{h}_f(P)h(t) + O(h(t)^{2/3})$$
as $h(t) \to \infty$. □

References

[1] Allen Altman. The size function of abelian varieties. Trans. Amer. Math. Soc., 164:153–161, 1972.
[2] Matthew Baker. A finiteness theorem for canonical heights attached to rational maps over function fields. J. Reine Angew. Math., 626:205–233, 2009.
[3] Robert L. Benedetto. Heights and preperiodic points of polynomials over function fields. Int. Math. Res. Not., (62):3855–3866, 2005.
[4] Enrico Bombieri and Walter Gubler. Heights in Diophantine geometry, volume 4 of New Mathematical Monographs. Cambridge University Press, Cambridge, 2006.
[5] Gregory S. Call and Joseph H. Silverman. Canonical heights on varieties with morphisms. Compositio Math., 89(2):163–205, 1993.
[6] Laura DeMarco and Niki Myrto Mavraki. Variation of canonical height and equidistribution. American J. Math., 142:443–473, 2020.
[7] Laurent Denis. Hauteurs canoniques et modules de Drinfel’d. Math. Ann., 294(2):213–223, 1992.
[8] Najmuddin Fakhruddin. Questions on self maps of algebraic varieties. J. Ramanujan Math. Soc., 18(2):109–122, 2003.
[9] Charles Favre and Thomas Gauthier. Continuity of the Green function in meromorphic families of polynomials. Algebra Number Theory, 12(6):1471–1487, 2018.
[10] D. Ghioca, L.-C. Hsia, and T. J. Tucker. Preperiodic points for families of rational maps. Proc. Lond. Math. Soc. (3), 110(2):395–427, 2015.
[11] Dragos Ghioca and Niki Myrto Mavraki. Variation of the canonical height in a family of rational maps. New York J. Math., 19:873–907, 2013.
[12] Patrick Ingram. Variation of the canonical height for a family of polynomials. J. Reine Angew. Math., 685:73–97, 2013.
[13] Patrick Ingram. Canonical heights for Hénon maps. Proc. Lond. Math. Soc. (3), 108(3):780–808, 2014.
[14] Patrick Ingram. A lower bound for the canonical height associated to a Drinfeld module. Int. Math. Res. Not. IMRN, (17):4879–4916, 2014.
[15] Patrick Ingram. Variation of the canonical height for polynomials in several variables. Int. Math. Res. Not. IMRN, (24):13545–13562, 2015.
[16] Patrick Ingram. Canonical heights for correspondences. Trans. Amer. Math. Soc., 371(2):1003–1027, 2019.
[17] Zhigokh Jelonek. On the effective Nullstellensatz. Invent. Math., 162(1):1–17, 2005.
[18] Teresa Krick, Luis Miguel Pardo, and Martín Sombra. Sharp estimates for the arithmetic Nullstellensatz. Duke Math. J., 109(3):521–598, 2001.
[19] Serge Lang. Algebra, volume 211 of Graduate Texts in Mathematics. Springer-Verlag, New York, third edition, 2002.
[20] Shouwu Zhang. Small points and adelic metrics. J. Algebraic Geom., 4(2):281–300, 1995.
APPENDIX A. TWO FACTS FROM LINEAR ALGEBRA

For the arXiv version of this paper, we include two proofs from linear algebra, alluded to in Section 2. First, an effective, special case of Hilbert’s Nullstellensatz with an easy proof. This is surely known, but the author had trouble finding a reference with exactly this statement (as opposed to the a priori slightly weaker statement that \( \deg(f_i A_i) \leq 2d - 1 \)).

**Lemma 24.** Let \( k \) be an infinite field, and let \( f_0,...,f_N \) be polynomials with coefficients in \( k \) and of degree at most \( d \). Then there are polynomials \( A_0,...,A_N \) of degree at most \( d - 1 \) such that

\[
1 = f_0 A_0 + \cdots + f_N A_N
\]

if and only if \( f_0,...,f_N \) have no nontrivial common factor.

**Proof.** If the polynomials have a common non-trivial factor, then of course we cannot write 1 in this way. Also, if none of the polynomials \( f_i \) have degree \( d \), we can apply a previous case of the theorem, so without loss of generality \( \deg(f_0) = d \).

The equation

\[
a = f_0 A_0 + \cdots + f_N A_N
\]

defines a homogeneous system of linear equations, obtained by identifying coefficients of the same power of the variable on both sides. There are \( 2d \) equations, and \((N + 1)d + 1\) unknowns (\( a \), and the coefficients of the \( A_i \)). Suppose there is no solution with \( a = 1 \), so that all solutions have \( a = 0 \). We will show that the \( f_i \) must have a common factor.

Let \( B \) be a polynomial with generic coefficients, and for any \( \alpha_1,...,\alpha_N \in k \) consider the system of equations obtained by identifying coefficients of powers of the variable in

\[
a = f_0 A_0 + B(\alpha_1 f_1 + \cdots + \alpha_N f_N).
\]

Any nontrivial solution to this gives a nontrivial solution to the original equation, so all solutions must still have \( a = 0 \). Since this new system has \( 2d + 1 \) unknowns, though, there must be a non-trivial solution. Also, since \( f_0 \neq 0 \), and \( A_0 = B = 0 \) is the trivial solution, we must have \( B \neq 0 \) in any nontrivial solution, so

\[
\frac{\alpha_1 f_1 + \cdots + \alpha_N f_N}{f_0} = -\frac{A_0}{B},
\]

where the right-hand-side is a rational function of degree at most \( d - 1 \). It follows that \( f_0 \) and \( \alpha_1 f_1 + \cdots + \alpha_N f_N \) have a nontrivial common factor (since \( f_0 \) has degree \( d \), but the ratio has degree strictly less).

For each of the finitely many non-trivial monic divisors \( s \mid f_0 \), let

\[
V_s = \{ (\alpha_1,...,\alpha_N) \in k^N : \text{ } s \mid \alpha_1 f_1 + \cdots + \alpha_N f_N \},
\]

noting that \( V_s \) is a linear subspace of \( k^N \). By what we have shown, the \( V_s \) cover \( k^N \), but since \( k \) is infinite, \( k^N \) cannot be covered by finitely many proper subspaces. Thus we have \( V_s = k^N \) for some nontrivial \( s \mid f_0 \), and hence \( s \mid f_i \) for all \( i \).

**Proof of Lemma.** First note that \( r \leq p \), and the case of equality is ruled out by the existence of a non-trivial solution. On the other hand, if \( r = 0 \) then any values of the \( x_j \) yield a solution, and the claim is trivially true. So we will take \( 1 \leq r \leq p - 1 \).
Suppose our system of equations is
\[ a_{i,1} x_1 + \cdots + a_{i,p} x_p = 0 \]
for \( 1 \leq i \leq q \). Since the system has rank \( r \), there is some non-vanishing \( r \times r \) minor of the coefficient matrix, and without loss of generality (permuting equations and variables) we may assume that this is the top left minor. Fix \( 1 \leq j \leq r \), and let \( C_i \) be the cofactor of the entry \( a_{i,j} \), so that this minor has determinant
\[ \delta = a_{1,j} C_1 + \cdots + a_{r,j} C_r \neq 0. \]
Multiplying the \( i \)th equation by \( C_i \) and summing, then, yields an equation in which the coefficient of \( X_j \) is \( \delta \). On the other hand, for \( 1 \leq k \leq r \) and \( k \neq j \), the coefficient of \( x_k \) in this new equation is \( C_1 a_{1,k} + \cdots + C_r a_{r,k} = 0 \), because this is the determinant of a matrix with a repeated column. Finally, for \( k > r \), the coefficient of \( x_k \) in this new equation is
\[ \delta_{j,k} = C_1 a_{1,k} + \cdots + C_r a_{r,k} \]
which is, up to sign, the determinant of some other \( r \times r \) submatrix of the coefficient matrix.

So from the original system we have deduced a new system of equations
\[ \delta x_i + \delta_{i,r+1} x_{r+1} + \cdots + \delta_{i,p} x_p = 0 \quad (1 \leq i \leq r), \]
where \( \delta \) and the \( \delta_{i,j} \) are all \( r \times r \) signed minors of the coefficient matrix, and \( \delta \neq 0 \). This new system has rank \( r \), though, and so is equivalent to the original system.

If \( r+1 \leq s \leq p \), our solution is \( x_s = \delta, x_i = -\delta_{i,s} \) for \( 1 \leq i \leq r \), and \( x_i = 0 \) for all other \( i \). If \( 1 \leq s \leq r \), then note that we cannot have \( \delta_{s,j} = 0 \) for all \( r+1 \leq j \leq p \), or else our original equations would force \( x_s = 0 \). So we choose some \( j \) with \( \delta_{s,j} \neq 0 \), set \( x_i = \delta_{i,j} \) for \( 1 \leq i \leq r \), \( x_j = -\delta \), and \( x_i = 0 \) otherwise. In any case, the \( x_i \) are signed minors of the coefficient matrix, and \( x_s \neq 0 \). \( \square \)