HEIGHTS AND THE MORDELL-WEIL THEOREM

ARJUN VENKATRAMAN

Abstract. This paper develops the technique of heights in order to use the Descent Theorem to prove the Mordell-Weil Theorem. We first define heights over projective space, then use this definition to define heights on an elliptic curve through some function $E(K) \to \mathbb{P}^n$. We then use Weil's height machine and Néron-Tate Normalization to define heights based on line bundles, with which we prove Mordell-Weil for all abelian varieties.

Contents

1. Background

The objective of this paper is to first provide proof of the Mordell-Weil Theorem for Elliptic Curves over all number fields K (given the Weak Mordell-Weil Theorem) using height functions and the Descent Theorem, and then to introduce Weil's height machine in order to give the proof of the same result over all abelian varieties.

The statement of the theorem (for elliptic curves) is as follows.

Theorem 1.1 (Mordell-Weil)**.** *For an elliptic curve E defined over a number field K, the elliptic curve group $E(K)$ is finitely generated.*

The primary vehicle we will use for this proof is the Descent Theorem, which states that the existence of a height function $h : A \to \mathbb{R}$ over an abelian group A satisfying certain criteria implies that A is finitely generated.

Theorem 1.2 (Descent Theorem)**.** *Given an Abelian group A, if there exists an integer $m \geq 2$ and a height function $h : A \to \mathbb{R}$ which satisfies:*

Date: 29 August 2020.

(1) *For any $Q \in A$ there exists some constant C_Q such that for any $P \in A$, $h(P+Q) \leq 2h(P) + C_Q$*
(2) *There exists some constant C such that for any $P \in A$, $h(mP) \geq m^2 h(P) - C$*
(3) *For any $k \in \mathbb{R}$, $h(P) \leq k$ only holds for finitely many $P \in A$*
(4) *The quotient A/mA is finite*

then the group A is finitely generated.

Proof. Conditions (1) and (2) are used to prove that any point $P \in A$ can be generated by a specific set of coset representatives of A/mA and points of height less than some constant M, with both the representatives and M independent of P; (3) and (4) in turn give us that this set is finite.

We start by selecting the representatives of each coset of mA in A. Since A/mA is finite, there are finitely many cosets and thus this gives a finite set $Q_1, Q_2, \ldots, Q_n$. For any point P, we denote the representative of the coset containing P as Q_P.

Now, taking any starting point $P_0 \in A$, we can write that $P_0 - Q_{P_0} \in mA$, or $P_0 = Q_{P_0} + mP_1$ for some $P_1 \in A$. We can repeat this process an arbitrary number of times as well, each time taking

$$P_k = Q_{P_k} + mP_{k+1}.$$

Our aim now is to show that there exists some constant M for which, given any P_0, there is eventually a value of k for which $h(P_k) \leq M$.

(2) gives us that

$$m^2 h(P_k) \leq h(mP_k) + C$$
$$h(P_k) \leq \frac{1}{m^2}(h(mP_k) + C),$$

which, plugging in the equation relating P_{k-1} to P_k gives

$$h(P_k) \leq \frac{1}{m^2}(h(P_{k-1} - Q_{P_{k-1}}) + C).$$

Setting

$$C_{\max} = \max_{1 \leq i \leq n} C_{-Q_i}$$

allows us to in turn apply (1) which tells us that

$$h(P_{k-1} - Q_{P_{k-1}}) \leq 2h(P_{k-1}) + C_{-Q_{P_{k-1}}} \leq 2h(P_{k-1}) + C_{\max}$$

and thus

$$h(P_k) \leq \frac{1}{m^2}(2h(P_{k-1}) + C_{\max} + C).$$

Since $m \geq 2$, we have that

$$h(P_k) \leq \frac{1}{2}h(P_{k-1}) + \frac{1}{4}(C_{\max} + C)$$
$$h(P_k) - \frac{1}{2}(C_{\max} + C) \leq \frac{1}{2}(h(P_{k-1}) - \frac{1}{2}(C_{\max} + C)).$$

Thus for any P_0 there must eventually be some value of K for which $h(P_k) - \frac{1}{2}(C_{\max} + C) \leq 1$, or $h(P_k) \leq 1 + \frac{1}{2}(C_{\max} + C)$, giving us our constant M.

Thus, we have that any $P_0 \in A$ can be written as a sum using only elements of height $\leq 1 + \frac{1}{2}(C_{\max} + C)$ (which, since C and $C_{\max}$ are independent of P_0, is a finite set by (3)) and the coset representatives Q_i (a finite set by (4)). Thus, we have a finite set of generators from which every element of A may be produced.

■

The objective now is to find some height function $h : E(K) \to \mathbb{R}$ which satisfies the criteria for application of the descent theorem. The first major result which allows this is the *Weak Mordell-Weil Theorem*:

Theorem 1.3 (Weak Mordell-Weil)**.** *For an elliptic curve E over a number field K, the quotient $E(K)/mE(K)$ is finite for all integers $m \geq 2$.*

This satisfies condition (4) in the descent theorem for all possible m. The proof of the Weak Mordell-Weil Theorem will not be covered in this paper; it is given in VIII.1 of [5].

2. Establishing height functions over Projective Space and $E(K)$

We now need to define our height function over $E(K)$ which works with the Descent Theorem. To do this, we will first define the height in more familiar territory, over projective space $\mathbb{P}^n(K)$. From there, we can take any function in the field of functions $K(E)$, which maps E to $\mathbb{P}^n$, allowing us to assign heights to points of E. The canonical definition of the height over $\mathbb{Q}$ is that $H(\frac{p}{q}) = \max(|p|, |q|)$ with p, q relatively prime integers. Likewise, for the projective space $\mathbb{P}^n(\mathbb{Q})$ the definition is extended to $H([x_0, x_1, \ldots, x_n]) = \max(|x_0|, |x_1|, \ldots, |x_n|)$ with $x_0, x_1, \ldots, x_n$ relatively prime integers.

While this works nicely over $\mathbb{P}^n(\mathbb{Q})$ (and the third condition for the descent theorem is implied immediately), it is not well defined over other number fields. To define this more general height function, we instead make use of this result from algebraic number theory that relates the standard and p-adic absolute values.

Theorem 2.1. *For a number field K, and any point $P \in K^*$,*

$$\prod_{v \in M_K} |x|_v^{n_v} = 1$$

where M_K is the set of all absolute values over K which are equal to either the "regular" absolute values or a p-adic absolute value when restricted to $\mathbb{Q}$. The exponent n_v is calculated as the degree

$$n_v = [K_v/\mathbb{Q}_v]$$

where K_v and $\mathbb{Q}_v$ are the completions of those fields with respect to v.

This allows the following definition.

Definition 2.2. The height $H_K(P)$ of a point $P = [x_0, x_1, \ldots, x_n]$ in projective space $\mathbb{P}^n(K)$ with regards to K is defined as the product

$$H_K(P) = \prod_{v \in M_K} \max(|x_0|_v, |x_1|_v, \ldots |x_n|_v)^{n_v}.$$

When we set $K = \mathbb{Q}$, the height $H_{\mathbb{Q}}(P)$ is exactly the same as the height $H(P)$ established above; n_v is always 1, so it is equivalent of simply multiplying the maximum values of each absolute value. Since the standard absolute value $|\cdot|$ is interchangable with (positive) multiplication, this is the same as multiplying all of the x_is by the maximum of each p-adic absolute value and then taking the maximum standard absolute value. Multiplying by $\max(|x_0|_p, |x_1|_p, \ldots |x_n|_p)$ ensures that the minimum p-adic valuation is exactly 1; doing so for all primes in turn yields all the

x_i being relatively prime integers; taking the maximum absolute value now yields the same result as the previously stated definition over $\mathbb{P}^n(\mathbb{Q})$.

Now that we have defined a consistent height H_K over $\mathbb{P}^n(K)$, we use another result from algebraic number theory to modify the definition to gain a consistent value of $H(P)$ for any number field K containing P. Namely, we use the fact that given an extension L/K and some $v \in M_K$, we have

$$\sum_{w \in M_L, |P|_v = |P|_w \forall P \in K} n_w = [L : K] n_v.$$

(The proof of this is given in [3].)

This allows us to define:

Definition 2.3. The absolute height of P is defined as

$$H(P) = H_K(P)^{\frac{1}{[K:\mathbb{Q}]}}$$

for some number field K containing P; the above equation along with the definition of H_K gives us that this is a well-defined result independent of the choice of K.

We are now ready to define the height over $E(K)$; which we do relative to some function $f \in K(E)$.

Definition 2.4. The height of a point over an elliptic curve $E(K)$ with regards to a nonconstant function $f \in K(E)$ is defined as

$$h_f(P) = \log H(f(P))$$

.

The reasons for taking the logarithm will become apparent as we examine the properties of the height function $H(P)$.

3. Showing Finiteness of points of bounded height

We will first show that this function $H(P)$ (and thus the height $h_f(P)$) satisfies the third condition of the descent theorem; that there are only finitely many points with height bounded by any constant C. The main result we seek to prove is *Northcott's Theorem*, which states that the set of points P in $\mathbb{P}^n(\overline{\mathbb{Q}})$ with bounded height and bounded degree $[\mathbb{Q}(P) : \mathbb{Q}]$ is finite; since $\mathbb{Q}(P) \subset K$ for any $P \in \mathbb{P}^n(K)$, this automatically implies the third condition.

To begin, we prove the following theorem relating the height of the point defined by the coefficients of a polynomial to the heights of its roots; applying this result to the minimal polynomial of $x \in \overline{\mathbb{Q}}$ will then give us the proof of Northcott in one dimension.

Theorem 3.1. *For any polynomial $f(T) \in \bar{\mathbb{Q}}[T]$ of degree d expressed as $T^d + a_1 T^{d-1} + \ldots + a_{d-1} T + a_d$ which factors as $(T - r_1)(T - r_2) \cdots (T - r_d)$, the inequalities*

$$2^{-d} \prod_{i=1}^{d} H(r_i) \leq H([1, a_1, a_2, \ldots, a_d]) \leq 2^{d-1} \prod_{i=1}^{d} H(r_i)$$

hold.

Proof. We set up the number field $K = \mathbb{Q}(r_1, r_2, \dots r_d)$. For an absolute value $v \in M_K$, we then define $\epsilon(v) = 2$ if v is Archimedean and $\epsilon(v) = 1$ if v is not.

We induct on the degree d to prove

$$\epsilon(v)^{-d} \prod_{i=1}^{d} \max(|r_i|_v, 1) \leq \max(1, |a_1|_v, |a_2|_v, \dots, |a_i|_v) \leq \epsilon(v)^{d-1} \prod_{i=1}^{d} \max(|r_i|_v, 1).$$

When $d = 1$, we have that $f(T) = x + a_1 = x - r_1$, so $r_1 = a_1$ and the inequalities $\frac{1}{\epsilon(v)} \max(|r_1|_v, 1) \leq \max(1, |a_1|) \leq \max(|r_1|_v, 1)$ are trivial. From there, we induct using the fact that the triangle inequality may be rewritten $|x + y|_v \leq \epsilon(v) \max(|x|_v, |y|_v)$.

Assuming the result holds for all polynomials of degree $d-1$, we take the index k for which $|r_k|_v$ is minimized, and take the polynomial of degree $d-1$ with all roots other than r_k:

$$G(T) = (T - r_1)(T - r_2) \cdots (T - r_{k-1})(T - r_{k+1}) \cdots (T - r_d) = T^{d-1} + b_1 T^{d-2} + \cdots + b_{d-1}.$$

Since $a_i = b_i - r_k b_{i-1}$, we can use the triangle inequality to yield

$$\max(1, |a_1|_v, |a_2|_v, \dots |a_d|_v) \leq \epsilon(v) \max(1, |r_k|_v, |b_1|_v, |r_k b_1|_v, |b_2|_v, \dots, |r_k b_{d-1}|)$$

$$\leq \epsilon(v) \max(1, |b_1|_v, |b_2|_v, \dots |b_{d-1}|) \max(1, |r_k|_v).$$

From here, the induction hypothesis gives us that

$$\max(1, |b_1|_v, |b_2|_v, \dots |b_{d-1}|) \leq \epsilon(v)^{d-2} \prod_{i=1}^{d-1} \max(|r_i|_v, 1).$$

Plugging into the above inequality gives us

$$\max(1, |a_1|_v, |a_2|_v, \dots |a_d|_v) \leq \epsilon(v)^{d-1} \prod_{i=1}^{d} \max(|r_i|_v, 1),$$

the desired upper bound.

For the lower bound, if $|r_k|_v \leq \epsilon(v)$, then the rest of the result is trivial as r_k is the maximal root - this means the lower bound is less than 1 while the central expression is clearly greater.

On the other hand, if $|r_k|_v > \epsilon(v)$, we follow a similar procedure to the upper bound. If v is Archimedean, we use triangle inequality to yield

$$\max(1, |a_1|_v, |a_2|_v, \dots |a_d|_v) \geq (|r_k|_v - 1) \max(1, |b_1|_v, |b_2|_v, \dots, |b_{d-1}|_v.$$

Since we have $\epsilon(v) = 2$ and $|r_k|_v > \epsilon(v)$, $|r_k|_v - 1 > \frac{1}{\epsilon(v)} |r_k|_v$, so

$$|a_1|_v, |a_2|_v, \dots |a_d|_v) > \epsilon(v)^{-1} |r_k|_v \max(1, |b_1|_v, |b_2|_v, \dots, |b_{d-1}|_v).$$

If v is non-Archimedean, we instead have immediately that equality holds:

$$|a_1|_v, |a_2|_v, \dots |a_d|_v) = \epsilon(v)^{-1} |r_k|_v \max(1, |b_1|_v, |b_2|_v, \dots, |b_{d-1}|_v).$$

Covering both cases with a non-strict inequality and substituting the lower bound for $\max(1, |b_1|_v, |b_2|_v, \dots, |b_{d-1}|_v)$ from the inductive hypothesis yields the desired lower bound. Thus we have proved

$$\epsilon(v)^{-d} \prod_{i=1}^{d} \max(|r_i|_v, 1) \leq \max(1, |a_1|_v, |a_2|_v, \dots, |a_i|_v) \leq \epsilon(v)^{d-1} \prod_{i=1}^{d} \max(|r_i|_v, 1).$$

as desired. We now take the n_vth power:

$$\epsilon(v)^{-dn_v}\prod_{i=1}^{d}(\max(|r_i|_v,1))^{n_v} \leq (\max(1,|a_1|_v,|a_2|_v,\dots,|a_i|_v))^{n_v} \leq \epsilon(v)^{(d-1)n_v}\prod_{i=1}^{d}(\max(|r_i|_v,1))^{n_v}.$$

Multiplying these equations together for all $v \in M_K$ gives

$$2^{-d[K:\mathbb{Q}]}\prod_{i=1}^{d}H_K(r_i) \leq H_K([1,a_1,a_2,\dots,a_d]) \leq 2^{(d-1)[K:\mathbb{Q}]}\prod_{i=1}^{d}H_K(r_i).$$

Finally, taking the $[K:\mathbb{Q}]$th root yields

$$2^{-d}\prod_{i=1}^{d}H(r_i) \leq H([1,a_1,a_2,\dots,a_d]) \leq 2^{d-1}\prod_{i=1}^{d}H(r_i).$$

as desired. ∎

This result will be key in proving the finiteness property; however, we want to apply it to the minimal polynomial of $x \in \overline{\mathbb{Q}}$. In order to get a nice result from doing this, though, the next result is key.

Lemma 3.2. *The height $H(P)$ is invariant under the action of the Galois group $\mathrm{Gal}_{\bar{\mathbb{Q}}/\mathbb{Q}}$.*

Proof. We take $\sigma \in \mathrm{Gal}_{\bar{\mathbb{Q}}/\mathbb{Q}}$; we also denote the corresponding isomorphism from $M_K \to M_{K^\sigma}$ as σ; so v goes to v_σ, which satisfies $|x^\sigma|_{v^\sigma} = |x|_v$ and $n_{v^\sigma} = n_v$. Then, since every $w \in M_{K^\sigma}$ is v^σ for exactly one $v \in M_K$, we have that

$$\prod_{w\in M_{K^\sigma}}\max(|x_i^\sigma|_w)^{n_w} = \prod_{v\in M_K}\max(|x_i^\sigma|_{v^\sigma})^{n_{v^\sigma}} = \prod_{v\in M_K}\max(|x_i|_v)^{n_v},$$

Thus

$$H_{K^\sigma}(P^\sigma) = H_K(P)$$

as desired. ∎

We are now ready to prove the main result of the section.

Theorem 3.3 (Northcott)**.** *The set of $P \in \mathbb{P}^n(\bar{\mathbb{Q}})$ with $H(P) \leq C$ and $[\mathbb{Q}(P):\mathbb{Q}] \leq D$ for some constants c and d is finite.*

Proof. Taking $P \in \mathbb{P}^n(\bar{\mathbb{Q}})$, we set homogeneous coordinates such that one of them is 1. Then we can easily construct $\mathbb{Q}(P)$ from the rest of the coordinates and calculate the height $H_{\mathbb{Q}(P)}(P)$ as follows:

$$H_{\mathbb{Q}(P)}(P) = \prod_{v\in M_{\mathbb{Q}(P)}}\max(|x_0|_v,|x_1|_v,\dots|x_n|_v)^{n_v}$$

$$\geq \max_{0\leq i\leq n}\Big(\prod_{v\in M_K}\max(|x_i|_v,1)^{n_v}\Big) = \max_{0\leq i\leq n}(H_{\mathbb{Q}_P}(x_i)).$$

Since $[\mathbb{Q}(P):\mathbb{Q}] \geq \max_{0\leq i\leq n}[\mathbb{Q}(x_i):\mathbb{Q}]$, we have that for each of the x_i, the bounds $H_{\mathbb{Q}_P}(x_i) \leq c$ and $[\mathbb{Q}(x_i):\mathbb{Q}] \leq d$ hold. This means that proving the theorem over $\mathbb{P}^1(\bar{\mathbb{Q}})$ suffices for the proof, as this would mean there are only finitely many valid x_i to choose from in the n-dimensional case.

We now take $x \in \bar{\mathbb{Q}}$. The minimal polynomial comes out to

$$f(T) = (T-x_1)(T-x_2)\cdots(T-x_{[\mathbb{Q}(x):\mathbb{Q}]}) = T^{[\mathbb{Q}(x):\mathbb{Q}]} + a_1 T^{[\mathbb{Q}(x):\mathbb{Q}]-1} + \cdots + a_{[\mathbb{Q}(x):\mathbb{Q}]}$$

where the x_is are Galois conjugates of x. We can now apply Theorem 3.1's upper bound, giving

$$H(1, a_1, a_2, \ldots, a_{[\mathbb{Q}(x):\mathbb{Q}]}) \leq 2^{[\mathbb{Q}(x):\mathbb{Q}]-1} \prod_{i=1}^{[\mathbb{Q}(x):\mathbb{Q}]} H(x_i).$$

This allows us to apply the invariance over the Galois group in Lemma 3.2, yielding

$$H(1, a_1, a_2, \ldots, a_{[\mathbb{Q}(x):\mathbb{Q}]}) \leq 2^{[\mathbb{Q}(x):\mathbb{Q}]-1} H(x)^{[\mathbb{Q}(x):\mathbb{Q}]} \leq (2c)^d.$$

However, the number of points of bounded height on $\mathbb{P}^{[\mathbb{Q}(x):\mathbb{Q}]}(\mathbb{Q})$ being finite follows relatively trivially from the simpler (equivalent) definition of $H_{\mathbb{Q}}$ from the start of the previous section. Thus, there are only finitely many choices for the coefficients a_i and thus finitely many polynomials and finitely many points $x \in \bar{\mathbb{Q}}$ satisfying $H_{\mathbb{Q}_P}(x) \leq c$ and $[\mathbb{Q}(x) : \mathbb{Q}] \leq d$.

■

As mentioned above, Northcott's theorem almost immediately implies condition (3) for the Descent Theorem. We formalize this in the following two successive corollaries:

Corollary 3.4. *For any number field K and constant c, the set of points of $\mathbb{P}^n(K)$ with height less than or equal to c is finite.*

Proof. Since for any $P \in \mathbb{P}^n(K)$, the degree $[\mathbb{Q}(P) : \mathbb{Q}] \leq [K : \mathbb{Q}]$, we apply Northcott's theorem with $d = [K : \mathbb{Q}]$ for this result. ■

Corollary 3.5 (Third condition for Descent Theorem)**.** *For an elliptic curve E/K and nonconstant function $f \in K(E)$, the set of points P for which $h_f(P) \leq C$ is finite.*

Proof. We are dealing with points on $E(K)$ for which $h_f(P) \leq C$, which, by the definition of h_f means that

$$\log H(f(P)) \leq C$$
$$H(f(P)) \leq e^C.$$

Setting $c = e^C$ in the above result tells us that there are finitely many points $f(P) \in \mathbb{P}^1(K)$ where this occurs. Since the function f is in $K(E)$ and is nonconstant, the preimage of each point in $\mathbb{P}^1(K)$ will be finite, meaning that the union of the preimages of each $f(P)$ with $H(f(P)) \leq e^C$ is finite as desired. ■

4. Completing Descent Theorem criteria

Our goal in this section is to prove that h_f satisfies criteria (1) and (2) for applying the descent theorem. We will do this by showing that h_f is sufficiently close to a quadratic form, which in turn will imply both criteria.

The proof given in this section will primarily rely on algebraic manipulation using specific properties of ellpitic curves; a more technical proof of this (generalized to all abelian varieties) using Weil's height machine is given in Section 5. However, this proof will still rely on some important general results, such as the following theorem on how a homogeneous morphism of degree d affects the height of a point in projective space:

Proposition 4.1. *For any morphism $F : \mathbb{P}^n(K) \to \mathbb{P}^m(K)$ of degree d there exist constants C_1 and C_2 such that for any $P \in \mathbb{P}^n(\bar{\mathbb{Q}})$ we have $C_1 H(P)^d \leq H(F(P)) \leq C_2 H(P)^d$.*

Proof. Take $P = [x_0, x_1, \ldots, x_n]$ and $F = [f_0, f_1, f_2, \ldots, f_m]$ where the f_is are homogeneous and there is no zero common among all of them. We then define the following notation given $v \in M_K$:

$$|P|_v = \max(|x_i|_v).$$

In addition, if we label the coefficents of each f_i as $a_{i1}, a_{i2}, \ldots$ then we define

$$|F|_v = \max(|a_{ij}|_v).$$

We next define $H_K(F)$ to match the formulation we have for $H_K(P)$:

$$H_K(F) = \prod_{v \in M_K} |F|_v^{n_v}.$$

Finally, we define $\epsilon(v)$ based on whether or not v is Archimedean - however, we use a different definition from the proof of Theorem 3.1 - here, we instead say $\epsilon(v) = 1$ if v is Archimedean and $\epsilon(v) = 0$ if it is not. The triangle inequality thus instead becomes $|a_1 + \cdots + a_n|_v = n^{\epsilon_v} \max(|a_i|_v)$.

Set T to be the maximum number of terms in the polynomials f_i (equal to $\binom{n+d}{n}$). Since the absolute value of every term in any f_i must be less than or equal to $|F|_v |P|_v^d$, the triangle inequality gives us

$$|F(P)|_v \leq T^{\epsilon_v} |F|_v |P|_v^d$$

$$|F(P)|_v^{n_v} \leq T^{\epsilon_v n_v} |F_v^{n_v} |P|_v^{n_v d}.$$

Multiplying across all $v \in M_K$ gives

$$H_K(F(P)) \leq T^{[K:\mathbb{Q}]} H_K(F) H_K(P)^d$$

$$H(F(P)) \leq T H(F) H(P)^d,$$

which is the desired upper bound.

For the lower bound, we first note that because the f_i is homogeneous of degree d and share no nontrivial zeroes, the set upon which all vanish in the *affine* space $\mathbb{A}^{n+1}(\mathbb{Q})$ is only the point $(0, 0, \ldots, 0)$.

Thus, we can apply the Nullstellensatz to yield that in the space $\overline{\mathbb{Q}}[X_0, X_1, \ldots, X_n]$ the ideal generated by the f_is must contain powers of every polynomial which vanishes at $(0, 0, \ldots, 0)$, including the polynomials $X_0, X_1, \ldots X_N$. This means that for some integer $e \geq 1$, each X_i^e can be written as:

$$X_i^e = \sum_{j=0}^{m} g_{ij} f_j$$

where each of the $g_{ij} \in \overline{\mathbb{Q}}[X_0, X_1, \ldots, X_n]$ is homogeneous of degree $e - d$. We define $|G|_v$ and $H_K(G)$ the same way as we defined the notation for F, taking the maximum across all coefficients of g_{ij}s. Then we can use the above along with the triangle inequality to obtain that

$$|x_i|_v^e \leq T_1^{\epsilon(v)} \max_{0 \leq j \leq m} |g_{ij}(P)|_v |F(P)|_v$$

$$|P|_v^e \leq T_1^{\epsilon(v)} \max |g_{ij}(P)|_v |F(P)|_v.$$

Applying same procedure that we used to prove the upper bound on $|g_{ij}(P)|_v$ gives $|g_{ij}(P)|_v \leq T_2^{\epsilon(v)} |G|_v |P|_v^{e-d}$, so we can substitute to get

$$|P|_v^e \leq T_1^{\epsilon(v)} T_2^{\epsilon(v)} |G|_v |P|_v^{e-d} |F(P)|_v.$$

$$|P|_v^d \leq T_1^{\epsilon(v)} T_2^{\epsilon(v)} |G|_v |F(P)|_v.$$

$$H_K(P)^d \leq (T_1 T_2)^{[K:\mathbb{Q}]} H_K(G) H_K(F(P))$$

$$H(P)^d \leq T_1 T_2 H(G) H(F(P)),$$

giving us the desired lower bound with $C_1 = \frac{1}{T_1 T_2 H(G)}$.

■

Before moving on to the main result, we will need one other lemma, which allows us to interchange the heights associated with any two even functions.

Lemma 4.2. *If f and g are even functions, then $h_f \deg g = h_g \deg f + O(1)$.*

Proof. Firstly, we use the property of the elliptic curve that the even functions in $K(E)$ are exactly the functions of $K(x)$. This allows us to find a rational function $r(X) \in K(X)$ with $r \circ x = f$. It follows that

$$h_f(P) = \log H(f(P)) = \log H(r(x(P)).$$

We can take the logarithm of both sides in Proposition 4.1 to yield:

$$\log H(x(P))(\deg r) + \log C_1 \leq \log H(r(x(P)) \leq \log H(x(P))(\deg r) + \log C_2$$

where $\deg r = \deg a - \deg b$ where $r(x) = \frac{a(x)}{b(x)}$ with a, b polynomials.
In terms of Big-O notation we then have

$$\begin{aligned} \log H(r(x(P)) &= (\deg r) \log H(x(P)) + O(1) \\ &= (\deg r) h_x(P) + O(1). \end{aligned}$$

Since $\deg x = 2$, we have that $\deg f = 2 \deg r$ and thus

$$2h_f(P) = (\deg f) h_x(P) + O(1)$$

and likewise

$$2h_g(P) = (\deg g) h_x(P) + O(1)$$

finally giving

$$h_f(P)(\deg g) = \frac{1}{2}(\deg f)(\deg g) h_x(P) + O(1) = h_g(P)(\deg f).$$

■

With Lemma 4.2, we are now ready to prove the main result of this section. We will show (near-)quadraticity by showing that h_f satisfies the parallelogram rule within a bound of $O(1)$, relying on a nice symmetry in the elliptic curve addition law which allows us to apply the above results.

Theorem 4.3. *For $P, Q \in E(K)$ and $f : E \to \mathbb{P}^1_K$ even,*

$$h_f(P+Q) + h_f(P-Q) = 2h_f(P) + 2h_f(Q) + O(1)$$

where the $O(1)$ constant is independent of P and Q; in other words, h_f satisfies the axioms of a quadratic form up to $O(1)$.

Proof. Since the Lemma 4.2 has given us that $2h_f = (\deg f)h_x + O(1)$; and the equation we seek to prove is preserved by scaling, it suffices to prove that h_x satisfies the equation.

Take our $P, Q \neq O \in E$ ($P = O$ and $Q = 0$ are trivial cases with h_x), setting $x(P) = [x_1, 1]$, $x(Q) = [x_2, 1]$, $x(P+Q) = [x_3, 1]$, and $x(P-Q) = [x_4, 1]$. We also write out E in Weierstrass form: $y^2 = x^3 + Ax + B$. Then the Addition formula gives us:

$$x_3 + x_4 = \frac{2(x_1+x_2)(A+x_1x_2) + 4B}{(x_1+x_2)^2 - 4x_1x_2}$$

$$x_3x_4 = \frac{(x_1x_2 - A)^2 - 4B(x_1+x_2)}{(x_1+x_2)^2 - 4x_1x_2}$$

This gives us the idea of a map giving $[1, x_3+x_4, x_3x_4]$ in terms of $[1, x_1+x_2, x_1x_2]$; which is given by the map

$$g : \mathbb{P}^2 \to \mathbb{P}^2,\ g(a,b,c) = (b^2 - 4ac, 2b(Aa+c) + 4Ba^2, (c-Aa)^2 - 4Bba).$$

This, when combining with the maps $\sigma : E \times E \to \mathbb{P}^2, (P,Q) \to [1, x(P) + x(Q), x(P)x(Q)]$ and $G : E \times E \to E \times E, G(P,Q) = (P+Q, P-Q)$ clearly gives us that $\sigma \circ g = G \circ \sigma$.

We next show that g is a proper morphism on $\mathbb{P}^2$ by showing that $g([a,b,c]) \neq [0,0,0]$ for any $[a,b,c] \neq [0,0,0]$. If $a = 0$, then we have $g([a,b,c]) = [b^2, 2bc, c^2]$, meaning that there are no zeroes for which $a = 0$ but $b \neq 0$ or $c \neq 0$.

Thus, we set $a = 1$, giving

$$g([a,b,c]) = [b^2 - 4c, 2b(A+c) + 4B, (c-A)^2 - 4Bb].$$

If we want all three polynomials to vanish, then we must have $c = \frac{1}{4}b^2$. This gives us the other polynomials $\frac{1}{2}b^3 + 2Ab + 4B$ and $\frac{1}{16}b^4 - \frac{1}{2}Ab^2 - 4Bb + A^2$. Then, the identity

$$(24x^2+128A)(\frac{1}{16}x^4 - \frac{1}{2}Ax^2 - 4Bx + A^2) - (3x^3 - 20Ax - 216B)(\frac{1}{2}x^3 + 2Ax + 4B) = 32(4A^3 + 27B^2).$$

This means that both polynomials vanishing on any x would indicate $4A^3 + 27B^2 = 0$; which would imply E is singular; since an initial assumption is that E is not singular, this does not occur and therefore g is a morphism.

We now have that $H(\sigma(P+Q, P-Q)) = H(g(\sigma(P,Q)))$. Since g is a morphism of degree 2, Theorem 4.1 gives us that

$$C_1 H(\sigma(P,Q))^2 \leq H(g(\sigma(P,Q))) \leq C_2 H(\sigma(P,Q))^2,$$

or

$$\log H(g(\sigma(P,Q))) = 2\log H(\sigma(P,Q)) + O(1).$$

Thus

$$\log H(\sigma(P+Q, P-Q)) = \log H(\sigma(P,Q)) + O(1).$$

We then apply theorem 3.1 to the polynomial

$$f(T) = (T + x_1)(T + x_2) = T^2 + (x_1 + x_2)T + x_1x_2$$

to get:

$$\frac{1}{4}H(x_1)H(x_2) \leq H([1, x_1 + x_2, x_1x_2]) \leq 2H(x_1)H(x_2).$$

Taking the logarithm gives

$$\log H(\sigma(P,Q)) = \log H([1, x_1+x_2, x_1x_2]) = \log H(x_1)H(x_2)+O(1) = h_x(P)+h_x(Q)+O(1),$$

and doing the same with x_3 and x_4 in the polynomial gives the corresponding result for $P + Q$ and $P - Q$. Thus we have:

$$h_x(P + Q) + h_x(P - Q) = \log H(\sigma(P + Q, P - Q)) + O(1)$$

$$= 2\log H(\sigma(P, Q)) + O(1) = 2h_x(P) + 2h_x(Q) + O(1)$$

as desired. ■

Corollary 4.4 (First condition for Descent Theorem)**.** *For any $Q \in E(K)$ there exists some constant C such that for any $P \in E(K)$, $h_f(P + Q) \leq 2h_f(P) + C$.*

Proof. This follows immediately, as we can set the $C \geq 2h_f(Q)+O(1)$ for the $O(1)$ in Theorem 4.3, and then $h_f(P+Q)+h_f(P-Q) \leq 2h_f(P)+C$ and $h_f(P-Q) \geq 0$. ■

Proposition 4.5 (Second condition for Descent Theorem)**.** *] For any integer $m \geq 2$, and any $P \in E(K)$, $h_f([m]P) = m^2h_f(P) + O(1)$*
where $[m]P$ is the sum $P + P + \cdots + P$ m times.

Proof. This is trivial for $m = 0, 1$ so we induct for m given the result on $m-1$ and $m - 2$. From Theorem 4.3 we then have that

$$h_f([m]P) + h_f([m - 2]P) = 2h_f([m - 1]P) + 2h_f(P) + O(1)$$

$$h_f([m]P) = -h_f([m - 2]P) + 2h_f([m - 1]P) + 2h_f(P) + O(1)$$

$$= (-(m - 2)^2 + 2(m - 1)^2 + 2)h_f(P) + O(1) = m^2h_f(P) + O(1)$$

as desired. ■

Thus all four conditions for the Descent Theorem are now satisfied by the height function h_f (and any integer $m \geq 2$), as given by Corollary 4.4, Proposition 4.5, Corollary 3.5, and the Weak Mordell-Weil Theorem. We may now apply it to $E(K)$, implying the full Mordell-Weil Theorem.

5. Expanding to all Abelian varieties (Overview)

Now that we have proven the Mordell-Weil theorem, we move to a more general case: Proving that the group of $A(K)$ of K-rational points of *any* abelian variety A is finitely generated. We can immediately apply most of the proof for $E(K)$ to this case; the only part of the proof which used specific properties of elliptic curves was the proof that h_f is (nearly) a quadratic form; thus we must now prove this holds true for general abelian varieties. Whereas for $E(K)$ we proved this for h_f given any nonconstant $f \in K(E)$, we really only need to prove that it holds true for *some* f. Thus, it now becomes beneficial for us to try to generate specific height functions with desirable properties.

In this section, we outline the steps in finding a desirable height function; we then give the full proof in the following sections.

Extending Definition 2.4 to all varieties gives a map:

$$(\text{functions } V \to \mathbb{P}^n) \to (\text{functions } V \to \mathbb{R})$$

$$f \mapsto h_f$$

which gives the height functions used for the descent theorem in the previous sections. However, it is difficult to show that any of these heights h_f satisfies the quadratic form requirements for the descent theorem over general abelian varieties. The object we will use to aid us in finding a height that does satisfy this requirement is the Picard Group $\mathrm{Pic}(V)$, defined as the group of all line bundles on V under the tensor product; it is also isomorphic to the group of divisors on V modulo principal divisors. The main draw of the Picard group is that every function $f : X \to Y$ invites a pullback map $f^* : \mathrm{Pic}(Y) \to \mathrm{Pic}(X)$; this pullback satisfies a few useful properties for our purposes. First, by taking the pullback of the tautological line bundle on $\mathbb{P}^n$, we obtain a map

$$(\text{functions } V \to \mathbb{P}^n) \to \mathrm{Pic}(V)$$

$$f \mapsto c_f.$$

This suggests to us a potential way to extend our previous height map to a map

$$\mathrm{Pic}(V) \to (\text{functions } V \to \mathbb{R})$$

$$c \mapsto h_c$$

where we send $c_f \mapsto h_f$. Indeed this map, known as Weil's height machine, exists, and if we define the map to be additive on line bundles up to $O(1)$, the map is uniquely defined by the heights h_f from line bundles c_f. To show this we need two additional tools. The first is the Segre embedding, which embeds the tensor product $\mathbb{P}^n \otimes \mathbb{P}^m$ into a higher dimensional projective space while preserving the height $h = \log H$ as a sum of the heights in $\mathbb{P}^n$ and $\mathbb{P}^m$. The second is the ability to "pull down" from a higher dimensional projective space to a lower dimensional one while preserving the height up to $O(1)$. With these tools, we show that the height machine is well-defined based on additivity.

The second useful property of the pullback is that we can derive a formula for the pullback of the multiplication map $[m]$ on our abelian variety A (with $[m]^* : \mathrm{Pic}(A) \to \mathrm{Pic}(A)$), which, combined with the additive property of the height machine (as well as the functiorial property, which follows from how the machine is defined), gives us that h_c is within $O(1)$ of an even function if c is a symmetric line bundle.

Finally, we can use a result known as the Theorem of the Cube combined with the pullback of seven trivial addition maps $A \times A \times A \to A$ to show that for any $c \in \text{Pic}(A)$, we must have:

$$h_c(x+y+z) - h_c(x+y) - h_c(x+z) - h_c(y+z) + h_c(x) + h_c(y) + h_c(z) = O(1).$$

Since any function which satisfies this equation and is even is a quadratic form, we get that h_c is within $O(1)$ of a quadratic form.

Finally, because we did not use any property unique to elliptic curves to prove Corollary 3.5 (which gives the third condition for the descent theorem), we can generalize the result to $A(K)$. Because of the "pull down" result, this means that the condition holds on h_c for any c which defines an embedding $A \to \mathbb{P}^m$, even if $m \neq n$. We call such line bundles *very ample*; thus, we now have that for any c which is both symmetric and very ample, h_c satisfies all of the first three criteria for the descent theorem.

All that is needed to complete the proof is a stronger version of the weak Mordell-Weil theorem, which once again will not be proven in this paper.

5.1. **Eliminating the Error Bound.** An additional benefit of examining the pullback of the multiplication map $[m]$ is that it reveals we can use $[m]$ to generate a set of normalized heights $\tilde{h}_c = h_c + O(1)$ which satisfy all the axioms of the height machine $h \mapsto h_c$ exactly instead of up to $O(1)$. This process relies on *Tate's lemma*, which states that if we have a function $f : S \to \mathbb{R}$ and a map $G : S \to S$ with $f \circ G = \lambda f$, then there is a unique normalized function $\tilde{f}$ with $\tilde{f} \circ G = \lambda \tilde{f}$. Taking $G = [m]$ allows us to do this for our heights h_c which satisfy the additive and functorial properties of the height machine without needing an error bound. This process is known as Neron-Tate normalization.

While it is not strictly necessary for the proof of Mordell-Weil, the proof of Neron-Tate normalization uses many of the same intermediate results as the proof that h_c is a quadratic form up to $O(1)$ if c is symmetric. Thus, in the following sections, we choose to prove Neron-Tate normalization first and then we show that the normalized height $\tilde{h}_c$ is *exactly* a quadratic form for c symmetric. It is, again, entirely possible to complete the proof without this step.

6. Weil's height machine

We start by constructing Weil's height machine, which maps from the Picard group $\text{Pic}(V)$ consisting of all line bundles over a variety V to height functions $V(\bar{K}) \to \mathbb{R}$. We begin by formally defining the Picard group and the correspondence that exists between some of its elements and functions $V \to \mathbb{P}^n$.

Definition 6.1. The Picard group $\text{Pic}(V)$ is defined as the group of line bundles on a variety V, with the tensor product as the group operation. It is also isomorphic to the group of all divisors of V modulo principal divisors.

Definition 6.2. Given varieties X, Y and a morphism $f : X \to Y$, this function induces a map which we call $f^* : \text{Pic}(Y) \to \text{Pic}(X)$ which sends $c \in \text{Pic}(Y)$ to its preimage in X.

We denote the standard line bundle on $\mathbb{P}^n$ as $\mathcal{O}(1)$; and we then say that $c \in \text{Pic}(V)$ *corresponds* to some morphism $f : V \to \mathbb{P}^n$ if $f^*(\mathcal{O}(1)) = c$. Given a morphism $\omega : V \to \mathbb{P}^n$, unless otherwise stated, we will use $c_\omega \in \text{Pic}(V)$ to denote the corresponding line bundle.

A line bundle c corresponds to some morphism $f : V \to \mathbb{P}^n$ if and only if it is *generated by its global sections*: for any $x \in X$, c has some global section that does not vanish at x.

Definition 6.3. A line bundle that is generated by its global sections is *very ample* if its global sections define an immersion into projective space.

As we intend to build the height machine on $\mathrm{Pic}(V)$ based on the heights h_f associated with functions, and $\mathrm{Pic}(V)$'s group operation is the tensor product, we will likely need to show that the tensor product of some line bundles yields a map into projective space. A tool we can use to do this is the Segre embedding, which gives a map from the direct product of projective spaces into a larger projective space.

Definition 6.4. The Segre embedding is the embedding

$$\sigma : \mathbb{P}^n \times \mathbb{P}^m \to \mathbb{P}^{(n+1)(m+1)-1}$$

which sends $(x_0, x_1, \cdots, x_n), (y_0, y_1, \cdots, y_m) \mapsto (x_0y_0, x_0y_1, x_0y_2, \cdots, x_0y_m, x_1y_0, x_1y_1, \cdots, x_ny_{m-1}, x_ny_m)$.

Next, we show how to construct all line bundles in $\mathrm{Pic}(V)$ from very ample line bundles, starting with the following result:

Theorem 6.5. *Given a line bundle $\mathcal{L}$ and a very ample line bundle $\mathcal{N}$, there is some value $b \in \mathbb{N}$ such that $\mathcal{L} \otimes \mathcal{N}^a$ is generated by its global sections for all $a \geq b$.*

This theorem is due to Serre; proof is given in Theorem 5.17 of [1].

We can use this to prove:

Lemma 6.6. *Any line bundle $c \in \mathrm{Pic}(V)$ can be written as the difference of two very ample line bundles.*

Proof. Starting with some $c \in \mathrm{Pic}(V)$, we take any embedding $w : V \to \mathbb{P}^n$. By Theorem 6.5, we can then select some natural number a such that $c' = c + ac_w$ is generated by its global sections, corresponding to some morphism $V \to \mathbb{P}^m$. Thus, $c' + c_w$ corresponds to an embedding $V \to \mathbb{P}^n \times \mathbb{P}^m$, which can then be embedded into some larger projective space $\mathbb{P}^k$ using the Segre embedding. Thus we can take $c_f = c' + c_w = c + (a+1)c_w$ and $c_g = (a+1)c_w$; both corresponding to embeddings f and g, with $c = c_f - c_g$. ∎

This correspondence is what we will use to generate the height machine; for any very ample line bundle c that corresponds to a morphism $f : V \to \mathbb{P}^n$ we set $h_c = h_f + O(1)$. To extend to all line bundles in $\mathrm{Pic}(V)$, we set an additive property within $\mathrm{Pic}(V)$ that $h_{c+d} = h_c + h_d + O(1)$.

For this to be well-defined, however, we need to show that the heights are actually additive between very ample line bundles. The sum of line bundles in $\mathrm{Pic}(V)$ is their tensor product, which we may only be able to embed in some higher dimensional projective space. Consequently, we need is the ability to "pull down" a height function from a higher dimensional space to a lower dimensional one while ensuring that the height does not change beyond a constant bound. Fortunately, it is possible to do this quite nicely:

Proposition 6.7. *Take some homogenous polynomial $F(X_0, X_1, \ldots X_N)$ of degree m where the coefficient of X_N^m is nonzero ($F(0, \ldots, 0, 1) \neq 0$). Then for $x = (x_0, x_1, \ldots, x_n) \neq (0, 0, \ldots, 1)$ we define $x' = (x_0, x_1, \ldots, x_{n-1}) \in \mathbb{P}^{n-1}$. Then*

there exists some constant c such that for any $x \in \mathbb{P}^n(\bar{K})$ with $F(x) = 0$, we have $CH(x) \leq H(x') \leq H(x)$.

Proof. We begin by taking the upper bound of proposition 4.1 (we note that only the proof of the lower bound assumed that there were *no* points on which the polynomials did not vanish, so taking the upper bound is valid). Then we have that $H(F(P)) \leq TH(F)H(P)$.

We next define $x'' \in \mathbb{P}^{\binom{n+m-1}{m}-2}$ to be the point whose coordinates are all of the degree m monomials in $x_0, x_1, \ldots x_n$ other than x_n^m; $F(x) = 0$ gives x_n^m in terms of these monomials as a linear combinations. If we define $x^{(m)} \in \mathbb{P}^{\binom{n+m-1}{m}-1}$ as the point whose coordinates are *all* of the degree m monomials in $x_0, \ldots x_n$, then we have a linear map:

$$T : \mathbb{P}^{\binom{n+m-1}{m}-2} \to \mathbb{P}^{\binom{n+m-1}{m}-1}$$
$$x'' \mapsto x^{(m)}.$$

The 4.1 upper bound then gives us $H(x^{(m)}) \leq CH(x'')$. We then take note that, up to repetition, $x'' = x^{(m-1)} \otimes x'$, so we have that $H(x'') = H(x^{(m-1)})H(x')$ and thus

$$H(x^{(m)}) \leq CH(x^{(m-1)})H(x')$$

$$(H(x))^m \leq C(H(x))^{m-1}H(x')$$

$$H(x) \leq CH(x').$$

The upper bound is trivial, giving the result. ∎

We can now define and prove the existence and uniqueness of the height machine with our desired properties.

Theorem 6.8 (Weil's height machine)**.** *For every projective variety V there exists a unique map:*

$$\mathrm{Pic}(V) \to ((\textit{functions } V(K) \to \mathbb{R})/\textit{bounded functions})$$

which sends $c \mapsto h_c$ such that the following conditions hold:

1. *If c corresponds to any morphism f then $h_c(P) = h_f(P) + O(1)$.*
2. *h is additive by the following relation:*

$$h_{c+d}(P) = h_c(P) + h_d(P) + O(1)$$

for $c, d \in \mathrm{Pic}(V)$.

Proof. First, we show the additive property holds for all line bundles generated by their global sections. For functions f and g to projective space, we can find a function corresponding to the tensor product $c_f + c_g$ by taking $\sigma(f, g)$ where σ is the Segre embedding. This, by definition, preserves the height function h as a sum of h_f and h_g. By Proposition 6.7, we can then reduce down back to $\mathbb{P}^n$ while maintaining the height up to $O(1)$, giving the desired additivity rule.

We must also show that the map is consistent for all f, g with $c_f = c_g$; we need that $h_f = h_g + O(1)$. Taking the vector space of all global sections of c_f, changing from c_f to c_g is equivalent to a change of basis over this vector space. It can be

easily shown that a change of basis as such does not change h_f beyond a term of $O(1)$, so this property holds.

Lemma 6.6 tells us that every $c \in \mathrm{Pic}(V)$ can be written as the difference of two very ample line bundles, so the two criteria for the height machine together give us h_c up to $O(1)$ for all $c \in \mathrm{Pic}(V)$.

We need to also ensure that h_c is well-defined irrespective of *which* difference of very ample line bundles we choose to represent it. Take $c = c_f - c_g = c_{f'} - c_{g'}$; we want that $h_f - h_g = h_{f'} - h_{g'} + O(1)$. However, we have from Propositon 6.7 that $h_{f' \otimes g} = h_{f \otimes g'} + O(1)$, from which the desired result follows immediately. As the height machine was defined as a map to the quotient (functions $V(K) \to \mathbb{R}$)/bounded functions, these relations prove that the map exists and is a well defined map for any V. ■

We have the following as an immediate corollary:

Corollary 6.9. *For varieties V, W and $f : V \to W$, take $d \in \mathrm{Pic}(W)$ and $c = f^*(d)$. Then*

$$h_c(P) = h_d(f(P)) + O(1).$$

This property is known as the *functoriality* of the height machine.

7. Tate's lemma and Neron-Tate Normalization

Both the functorial and additive properties of the height machine are only up to an error bound of $O(1)$. While it is not strictly necessary for the proof of Mordell-Weil, we would prefer a set of heights which satisfied these properties with no error bound. In the process of trying to achieve this, however, we will already end up very close to the final proof of Mordell-Weil.

To get these height functions to behave slightly better for our purposes, we use what is known as Néron-Tate normalization. This process, in essence, should give us a new modified height machine whose heights satisfy the desired relations (namely, functoriality and additivity) exactly, instead of just up to $O(1)$, while remaining within $O(1)$ of the original heights from the height machine. To prove that such normalization is possible, we begin with *Tate's lemma*, which gives us criteria for normalizing a stand-alone function with regards to a map G and scale factor λ:

Lemma 7.1 (Tate's lemma)**.** *Given a set S, a map $G : S \to S$, and a function $f : S \to \mathbb{R}$ with $f \circ G = \lambda f + O(1)$, $\lambda > 1$, there exists a unique function $\tilde{f}$ such that $\tilde{f} = f + O(1)$ and $\tilde{f} \circ G = \lambda \tilde{f}$. For any $x \in S$ we have that*

$$\tilde{f}(x) = \lim_{n \to \infty} (1/\lambda^n) f(G^n x).$$

Proof. Based on our initial condition for f, we have that $|f(Gx) - \lambda f(x)| \leq c$, and thus $|(1/\lambda^n) f(G^n x) - (1/\lambda^{n-1}) f(G^{n-1} x)| \leq \frac{c}{\lambda^n}$. This means that the series must converge for $\lambda > 1$, so the limit we set to be equal to $\tilde{f}$ must converge. By definition, we can immediately see that $\tilde{f} \circ G = \lambda \tilde{f}$ holds true. To prove that $\tilde{f} = f + O(1)$. we take:

$$|\tilde{f}(x) - f(x)| = |\lim_{n \to \infty} (1/\lambda^n) f(G^n x) - f(x)|$$

$$\leq \sum_{1}^{\infty} |(1/\lambda^n) f(G^n x) - (1/\lambda^{n-1}) f(G^{n-1} x)|$$

$$\leq \sum_{1}^{\infty} \frac{c}{\lambda^n}$$
$$= \frac{c}{\lambda - 1},$$

which is constant, giving us the desired result.

To prove uniqueness, we have that if f is bounded by C, then $\tilde{f}$ is equal to a limit converging to zero and thus is 0. Therefore, if $f = g + O(1)$, $\tilde{f} = \tilde{g}$. Thus, if we have some f' which also satisfies the criteria for $\tilde{f}$ (that is, $f' = f + O(1)$ and $f' \circ G = \lambda f'$), then we know that $\tilde{f'} = \tilde{f}$. However, since f' satisfies the other criterion as well, it is clear that $\tilde{f'} = f'$, and thus we have that $\tilde{f} = f'$ and is thus unique. ■

The uniqueness of $\tilde{f}$ is particularly useful as it immediately gives us the two important following corollaries:

Corollary 7.2 (Functoriality)**.** *Given maps $G : S \to S$, $F : S \to S'$, $G' : S' \to S'$ such that $F \circ G = G' \circ F$. Then take some $f' : S' \to \mathbb{R}$ with $f' \circ G' = \lambda f' + O(1)$ and set $f = f' \circ F$. Then $f \circ G = \lambda f + O(1)$ and $\tilde{f} = \tilde{f'} \circ F$.*

This leads to another immediate corollary; the commutative property of Néron-Tate normalization against other endomorphisms.

Corollary 7.3 (Commutativity with other endomorphisms)**.** *Using the same criteria for f, G, λ as Tate's lemma, if G' an endomorphism $S \to S$ with $\lambda \in \mathbb{R}$ such that $f \circ G' = \lambda' f + O(1)$, then $\tilde{f} \circ G' = \lambda' \tilde{f}$.*

Now that we have shown that the normalization is well-defined, we will next show that we can apply this normalization to derive the desired normalized heights. In order to prove that the heights can be normalized, we will first look at only symmetric and anti-symmetric divisors.

We can now take advantages of the specific properties of abelian varieties; namely, the group structure gives us a multiplication map $[m]$ which we can use to apply Tate's lemma. We show that the pullback $[m]^*c$ is equal to m^2c for symmetric divisor classes and mc for anti-symmetric divisor classes. We can then use Tate's lemma with $[m]$ as the endomorphism to find the one function equal to $h_c + O(1)$ which satisfies this.

As we constructed the original height machine from the heights on very ample divisors, we will construct all of the Neron-Tate heights from heights on symmetric and anti-symmetric divisors.

The first step to proving both this fact about the pullback on $[m]$ is proving that given any $c \in \mathrm{Pic}(A)$, the alternating sum of the pullbacks of c by the seven most trivial linear morphisms $A \times A \times A \to A$ is equal to zero. To show this, we need to apply the Theorem of the Cube:

Theorem 7.4 (Theorem of the Cube)**.** *Given projective varieties X, Y, Z, the only element of $\mathrm{Pic}(X \times Y \times Z)$ which induces 0 on each of $X \times Y \times \{z\}$, $\{x\} \times Y \times Z$, and $X \times \{y\} \times Z$ is zero.*

This result will not be proven in this paper. The proof is given in [2].

Using the theorem of the cube, we can apply it to the Picard Group of $A \times A \times A$, where A is our abelian variety. Take the following morphisms from $A \times A \times A \to A$:

$$s_{123}(x, y, z) = x + y + z, s_{12}(x, y, z) = x + y, s_{13}(x, y, z) = x + z$$

$$s_{23}(x,y,z) = y+z, s_1(x,y,z) = x, s_2(x,y,z) = y, s_3(x,y,z) = z.$$

The theorem of the cube gives us that:

Theorem 7.5. *For any $c \in \mathrm{Pic}(A)$, we have $s_{123}^*c - s_{12}^*c - s_{13}^*c - s_{23}^*c + s_1^*c + s_2^*c + s_3^*c = 0$.*

Proof. If we set $x = 0$, we have $s_1 = 0$ and equalities $s_{12} = s_2$, $s_{13} = s_3$, and $s_23 = s_{123}$. These terms all cancel giving that the desired sum is zero over $0 \times A \times A$. By symmetry, we can apply the theorem of the cube, giving that the sum is zero over all $A \times A \times A$. ∎

We next show that for some $c \in \mathrm{Pic}(A)$, the map $f \mapsto f^*c$ satisfies the necessary criterion to be a function of degree no greater than 2:

Theorem 7.6. *Given a variety V and an abelian variety A, the map $F : \mathrm{Hom}(V,A) \to \mathrm{Pic}(V), f \mapsto f^*c$ satisfies:*

$$(f_1+f_2+f_3)^*c - (f_1+f_2)^*c - (f_1+f_3)^*c - (f_2+f_3)^*c + f_1^*c + f_2^*c + f_3^*c = 0$$

for $f_1, f_2, f_3 : V \to A$.

Proof. Take the function $f : V \to A \times A \times A$ with $f(x) = (f_1(x), f_2(x), f_3(x)$. Theorem 7.5 and functoriality give us:

$$\begin{aligned}(f_1+f_2+f_3)^*c - (f_1+f_2)^*c - (f_1+f_3)^*c - (f_2+f_3)^*c + f_1^*c + f_2^*c + f_3^*c \\ = f^*(s_{123}^*c - s_{12}^*c - s_{13}^*c - s_{23}^*c + s_1^*c + s_2^*c + s_3^*c) \\ = f^*(0) = 0.\end{aligned}$$

∎

We then have the following lemma:

Lemma 7.7. *For any abelian group G and function $F : \mathbb{Z} \to G$ with*

$$F(x+y+z) - F(x+y) - F(x+z) - F(y+z) + F(x) + F(y) + F(z) = 0$$

for all $x, y, z \in \mathbb{Z}$ and $F(0) = 0$, we have

$$F(n) = H(n)F(1) + H(n-1)F(-1)$$

where

$$H(n) = \frac{n(n+1)}{2}.$$

Proof. It is clear trivially that this is true for $n = -1, 0, 1$. We can then induct; assuming the result for all positive integers up to n, we apply the given relation with $x = n, y = 1, z = -1$ to give:

$$F(n) - F(n+1) - F(n-1) - F(0) + F(n) + F(1) + F(-1) = 0$$

$$F(n+1) = 2F(n) - F(n-1) + F(1) + F(-1)$$

$$F(n+1) = 2(H(n)F(1) + H(n-1)2F(-1)) - (H(n-1)F(1) + H(n-2)F(-1)) + F(1) + F(-1)$$

$$F(n+1) = H(n)F(1) + H(n-1)F(-1) + (n+1)F(1) + nF(-1)$$

$$F(n+1) = H(n+1)F(1) + H(n)F(-1)$$

as desired. The same argument may be used to induct on the negative integers as well, solving for $F(n-1)$ instead from the same equation. ■

We now immediately can see how we reach the desired result from this equation, specifically:

Corollary 7.8. *For $m \in \mathbb{Z}$ and $c \in \mathrm{Pic}(A)$ (where $[m]$ is the multiplication map defined in Proposition 4.5), we have that*

$$[m]^*c = m^2 c$$

if c is a symmetric divisor class and that

$$[m]^*c = mc$$

if c is anti-symmetric.

Proof. From Theorem 7.6 we have that the map from $\mathbb{Z} \to \mathrm{Pic}(A)$ sending m to $[m]^*c$ satisfies the conditions for Lemma 7.7, meaning that

$$[m]^*c = \frac{m(m+1)}{2}c + \frac{m(m-1)}{2}[-1]^*c.$$

From here, if $c = [-1]^*c$, this simplifies to $[m]^*c = m^2c$, while if $c = -[-1]^*c$, then it simplifies to $[m]^*c = mc$. ■

With this result, we now have an endomorphism to which we can apply Tate's lemma to normalize the heights yielded by the height machine.

Theorem 7.9. *Given an abelian variety A over $\bar{K}$, there is a unique function from $\mathrm{Pic}(A)$ to functions $A(\bar{K}) \to \mathbb{R}$ sending $c \mapsto \tilde{h}_c$ with:*

(1) $\tilde{h}_c = h_c + O(1)$.
(2) $\tilde{h}_{c+d} = \tilde{h}_c + \tilde{h}_d$.
(3) If B is another abelian variety with a homomorphism $F : B \to A$, then $\tilde{h}_{F^*c} = \tilde{h}_c \circ F$.

Proof. If c is symmetric, Corollary 7.8 tells us that $[2]^*c = 4c$. Thus for additivity (condition (2)) to hold we want a map with $\tilde{h}_c(2x) = 4\tilde{h}_c(x)$. By Tate's lemma there is exactly one $\tilde{h}_c$ satisfying both this condition and $\tilde{h}_c = h_c + O(1)$. Additivity on symmetric divisors follows from the explicit formula for $\tilde{h}_c$ given by Tate's lemma; Functoriality follows from the commutative property of the Néron-Tate normalization applied to $[2]$ and F.

The same result can be achieved with anti-symmetric divisors using $[2]^*c = 2c$. Since a general $c \in \mathrm{Pic}(A)$ can always be written as half the sum of symmetric $c + [-1]^*c$ and the anti-symmetric $c - [-1]^*c$, the only way to define $\tilde{h}_c$ consistent with additivity is $\tilde{h}_c = \frac{1}{2}(\tilde{h}_{c+[-1]^*c} + \tilde{h}_{c-[-1]^*c})$; the other properties clearly continue to hold under this definition. ■

8. Proof of Mordell-Weil Theorem for Abelian Varieties

Now that we have shown the normalized height is well defined with a functoriality property, we can prove quadraticity. In particular, applying functoriality to the result of Theorem 7.5 yields that $\tilde{h}_c$ is a function of degree no greater than two; since $\tilde{h}_c(0)$ is clearly 0, c being symmetric would then imply that it must be even and thus quadratic.

Theorem 8.1. *For an abelian variety A and symmetric $c \in \mathrm{Pic}(A)$, the height $\tilde{h}_c$ is a quadratic form over A.*

Proof. We return to the morphisms $s_{123}, s_{12}, \ldots$ from Theorem 7.5. The functorial property given in Theorem 7.9 immediately tells us that

$$\tilde{h}_{s_{123}^* c}(x, y, z) = \tilde{h}_c \circ s_{123}(x, y, z) = \tilde{h}_c(x + y + z).$$

Taking this relation for each s function and summing the left hand sides yields zero. Thus the sum of the right hand sides must also be 0, so $\tilde{h}_c$ satisfies

$$\tilde{h}_c(x+y+z) - \tilde{h}_c(x+y) - \tilde{h}_c(x+z) - \tilde{h}_c(y+z) + \tilde{h}_c(x) + \tilde{h}_c(y) + \tilde{h}_c(z) = 0.$$

Since c is symmetric, $\tilde{h}_c$ is an even function satisfying the above with $\tilde{h}_c(0) = 0$, meaning it must be a quadratic form. ∎

We are now nearly ready to put everything together. The proofs from section 4 tell us that h being a quadratic form immediately causes it to satisfy the first two conditions for the Descent Theorem. Thus, if c is symmetric then $\tilde{h}_c$ satisfies conditions (1) and (2). If c corresponds to some f then we can immediately apply all of Section 3. Thus, all that is left is the criterion fulfilled by the Mordell-Weil Theorem.

Theorem 8.2 (Weak Mordell-Weil)**.** *For an abelian variety A over a number field K, the quotient $A(K)/mA(K)$ is finite for all integers $m \geq 2$.*

Once again, the proof will not be given in this paper; it can be found in Chapter 4 of [4].

With this, we have all we need for the final proof.

Theorem 8.3 (Mordell-Weil)**.** *For an abelian variety A defined over a number field K, the group $A(K)$ is finitely generated.*

Proof. Select some $c \in \mathrm{Pic}(A)$ which is both symmetric and very ample. We apply the descent theorem on $\tilde{h}_c$. Since c is symmetric, $\tilde{h}_c$ is quadratic and thus satisfies the first two criteria of the theorem. Since c is very ample, we can take the function it corresponds to and applying Proposition 6.7 if necessary. This gives $\tilde{h}_c = h_f + O(1)$. As Section 3 of this paper did not use any properties specific to elliptic curves, we can now apply Corollary 3.5 to h_f. This gives us the third criterion for the descent theorem, and the fourth is given immediately by the weak theorem. ∎

Acknowledgments

I thank my mentor, Dr. Akhil Mathew, for all his guidance throughout this program. The ease with which he helped me explore various aspects of the material and proof and his encouragement to pursue multiple sources in depth made this a very enjoyable and rewarding experience. I also thank Professor J. Peter May for organizing this REU despite the current circumstances and giving me this unique opportunity.

References

1. Robin Hartshorne, *Algebraic geometry*, Springer-Verlag, New York-Heidelberg, 1977, Graduate Texts in Mathematics, No. 52. MR 0463157
2. Marc Hindry and Joseph H. Silverman, *Diophantine geometry*, Graduate Texts in Mathematics, vol. 201, Springer-Verlag, New York, 2000, An introduction. MR 1745599
3. Serge Lang, *Algebraic number theory*, second ed., Graduate Texts in Mathematics, vol. 110, Springer-Verlag, New York, 1994. MR 1282723
4. Jean-Pierre Serre, *Lectures on the Mordell-Weil theorem*, third ed., Aspects of Mathematics, Friedr. Vieweg & Sohn, Braunschweig, 1997, Translated from the French and edited by Martin Brown from notes by Michel Waldschmidt, With a foreword by Brown and Serre. MR 1757192
5. Joseph H. Silverman, *The arithmetic of elliptic curves*, second ed., Graduate Texts in Mathematics, vol. 106, Springer, Dordrecht, 2009. MR 2514094