THE DENSITY OF RATIONAL POINTS ON $\mathbb{P}^1$ WITH THREE STACKY POINTS

BRETT NASSERDEN AND STANLEY YAO XIAO

Abstract. In this paper we consider the density of rational points on the "stacky" curve $X(\mathbb{P}^1; 0, 1, 2; 1, 2; \infty, 2)$ which is $\mathbb{P}^1$ with three half points, with respect to the so-called Ellenberg-Satriano-Zuerick-Brown height. In particular, we prove a conjecture of Ellenberg.

1. Introduction

Two of the outstanding conjectures in number theory are the so-called Manin-Batyrev conjecture [7] for the density of rational points on Fano varieties, and Malle’s conjecture [11] on the number of number fields of bounded discriminant having fixed degree and Galois group. Both conjectures assert, roughly, that the objects to be counted satisfy an asymptotic formula of the form

$$C \cdot X^\alpha (\log X)^\beta,$$

where $C, \alpha, \beta$ are non-negative numbers with $C, \alpha > 0$ which can be computed explicitly within their respective conjectural frameworks.

In a forthcoming article J. Ellenberg, M. Satriano, and D. Zuerick-Brown formulate a bold conjecture that has both the Manin and Malle conjectures as special cases [5, Main Conjecture]. Their conjecture concerns counting rational points with respect to a new theory of heights applying broadly to algebraic stacks. While the Manin and Malle conjectures are well studied, very little is known about the general conjecture.

In the framework of [5] the Malle and Manin conjectures represent two extremes of their theory of heights. The Manin conjecture involves counting points on a projective variety with respect to a Weil height and no theory of algebraic stacks is required. On the other hand from the point of view of [5] the Malle conjecture involves counting rational points on the classifying stack $BG$ where $G$ is a finite group. The theory of algebraic stacks is essential for this interpretation of the Malle conjecture, and the standard theory of heights on projective algebraic varieties is insufficient in this case.

In this article we consider instances of the main conjecture in [5] that lie between the two extremes described above. In other words, we consider cases that involve mixing the stacky and non-stacky phenomena. We formulate a new theory of heights on a stacky analogue of smooth projective algebraic curves and show that this stacky height is enough to determine the integral points on these "stacky curves". We then consider a particular stacky curve suggested by J. Ellenberg[4] and show that our theory of heights matches [5] in this instance. Finally, we verify a specific instance of the main conjecture in [5] given by Ellenberg[1] using analytical methods.

Our point of view with algebraic stacks is to adopt a bottom up perspective. In other words, to define our algebraic stacks in terms of a base variety along with some extra data which is enough to construct a unique algebraic stack. As we are interested in a well behaved family of stacky curves this description will be particularly simple. The reason for this choice is that although the heights we are concerned with are motivated by stacky phenomena, the analysis of the heights we are interested in lie firmly within the purview of analytic number theory and so a relatively simple exposition is preferable. The bottom up point of view allows us to discuss the objects we are interested in a concrete way that avoids technicalities and emphasizes the data most important for our purposes. The interested reader may consult [8] for general results involving the bottom up perspective on algebraic stacks and [16, Lemma 5.3.10] for the case of stacky curves.

[4] What's up in arithmetic statistics?" Number Theory Web Seminar, July 23, 2020
We now describe the stacky curve we are most interested in. Those unfamiliar with the theory of algebraic stacks may note the explicit form of our height given in (1.1) and move on to the definition of an $M$-curve given by Definition 1.1. Let $\mathcal{X}(\mathbb{P}^1_\mathbb{Q}; 0, 2; 1, 2; \infty, 2)$ be the algebraic stack obtained by replacing $\{0, 1, \infty\} \subseteq \mathbb{P}^1$ with $B(\mathbb{Z}/2\mathbb{Z})$. For a precise definition of this object, see [8]. The points $0, 1, \infty$ are "stacky points" which have an attached stabilizer group $\mathbb{Z}/2\mathbb{Z}$ and account for $\mathcal{X}(\mathbb{P}^1_\mathbb{Q}; 0, 2; 1, 2; \infty, 2)$ not being a projective variety.

The other points behave like the points on the quasi-projective variety (1.3) and the max on the right hand side is dependent only on the relative size of $|a|, |b|$. We may assume without loss of generality that $\pi(P) = (a_P : b_P) \in \mathbb{P}^1$. Now consider the height $H(a, b)$ on primitive integral pairs $(a, b)$ by

(1.1) $H(a, b) = \text{sqf}(a) \text{sqf}(b) \text{sqf}(a + b) \max\{|a|, |b|\},$

with $\text{sqf}(n) = n/k^2$, where $k^2$ is the largest square dividing $n$. Then given $P \in \mathcal{X}(\mathbb{Q})$ we have the associated point $\pi(P) = (a_P : b_P) \in \mathbb{P}^1$. We choose the tangent line bundle $T_X$. There is a birational mapping $\pi : X \to \mathbb{P}^1$ called the coarse space map. The height associated to the tangent bundle $T_X$ described by [5] can be described in terms of the coarse space map as follows. Given a point $P \in \mathcal{X}(\mathbb{Q})$ we have the associated point $\pi(P) = (a_P : b_P) \in \mathbb{P}^1$. Now consider the height $H(a, b)$ on primitive integral pairs $(a, b)$ by

Notice that the height (1.1) can be defined in terms of the coarse moduli space $\mathbb{P}^1$ of $\mathcal{X}(\mathbb{P}^1_\mathbb{Q}; 0, 2; 1, 2; \infty, 2)$ and a formula that takes into account the multiplicities in a simple way. Motivated by this we now essch the theory of algebraic stacks in favor of Darmon’s $M$-curves, which is an essentially equivalent theory that emphasizes the bottom up perspective to algebraic stacks. In other words, we keep track of the minimum amount of data that can be used to construct the algebraic stack. One may think of this as being analogous to only keeping track of a particular Weierstrass equation of an elliptic curve.

**Definition 1.1 ([4]).** Let $K$ be a number field. An $M$-curve over $K$ consists of the following data:

- A smooth projective curve $X$ defined over a number field $K$, and
- For each $P \in X(K)$ a multiplicity $m_P \in \mathbb{Z}_{\geq 1} \cup \{\infty\}$ with $m_P = 1$ for all but finitely many $P$.

We use the notation $\mathcal{X} = (X; P_1, m_1; P_2, m_2; \ldots; P_r, m_r)$ to denote the $M$-curve with multiplicities $m_P = m_i$ and if $Q \notin \{P_1, \ldots, P_r\}$ then $m_Q = 1$.

The $M$-curve we are most interested is given by $\mathbb{P}^1_{2,2,2} = \mathcal{X}(\mathbb{P}^1_\mathbb{Q}; 0, 2; 1, 2; \infty, 2)$ with the height on $\mathbb{P}^1$ given by (1.1). Our main goal is to count rational points on $\mathbb{P}^1$ with respect to the height (1.1). On writing

\[ a = x_1 y_1^2, \quad b = x_2 y_2^2, \quad x_1, x_2 \text{ square-free} \]

We then have

\[ H(a, b) = x_1 x_2 \text{sqf}(x_1 y_1^2 + x_2 y_2^2) \max\{|x_1 y_1^2|, |x_2 y_2^2|\}. \]

and the max on the right hand side is dependent only on the relative size of $|a|, |b|$. If we write

(1.2) $x_1 y_1^2 + x_2 y_2^2 = x_3 y_3^2,$

then we further obtain the expression

\[ H(a, b) = \max\{|x_2 x_3(x_1 y_1)^2|, |x_1 x_3(x_2 y_2)^2|\}. \]

We may assume without loss of generality that $|x_1 y_1^2| \geq |x_2 y_2^2|^2$ and $x_1 > 0$, so that

\[ H(a, b) = |x_2 x_3(x_1 y_1)^2|. \]

We put

(1.3) $N(T) = \# \{x, y \in \mathbb{Z}^3_{\neq 0} : \gcd(x_i, x_j), \gcd(y_i, y_j) = 1 \text{ for } i \neq j, x_i \text{ square-free for } i = 1, 2, 3, x_1 y_1^2 + x_2 y_2^2 = x_3 y_3^2, x_1 > 0, x_1 y_1^2 \geq |x_2 y_2^2|, |(x_1 y_1)^2 x_2 x_3| \leq T \}$

Our first main result will be the following:
Theorem 1.2. There exist positive numbers $c_1, c_2, c_3$ such that

$$c_1 T^{1/2} (\log T)^3 < N(T) < c_2 T^{1/2} (\log T)^3$$

for all $T > c_3$.

In particular, we confirm Ellenberg’s conjecture\(^1\) that $N(T) = O_\varepsilon T^{1/2+\varepsilon}$. Indeed, our theorem gives an exact order of magnitude for $N(T)$.

We also propose a general height applicable to $M$-curves with finite multiplicities (in particular to stacky curves in the sense of [11] Definition 5.2.1) which matches (1.1) for our specific curve. Moreover, our height has the remarkable feature that it naturally detects the difference between integral and rational points on $M$-curves (here we are using Darmon’s notion of integral; see Definition 2.2).

Theorem 1.3. For each $M$-curve $\mathcal{X}/\mathbb{Q}$ with finite multiplicities there exists a height $H = H_L H_X$ with the property that an element of $x \in \mathcal{X}(K)$ is integral if and only if $H_X(x) = 1$. Moreover, for $\mathcal{X} = \mathcal{X}(\mathbb{P}^1; 0, 2; -1, 2; \infty, 2)$ the height $H_L H_X$ is equivalent to the height given by (1.1).

Theorem 1.3 is a consequence of the more technical Theorem 2.3.

We illustrate how Theorem 1.3 allows one to detect integral points on $M$-curves. In this case the standard height is given by $H_L(a, b) = \max\{|a|, |b|\}$ and the stacky height given by (1.1). They are equal precisely when

$$|sqf(a) sqf(b) sqf(a+b)| = 1,$$

or in the notation of (1.2), that $|x_1| = |x_2| = |x_3| = 1$. (1.2) then turns into

$$\pm y_1^2 \pm y_2^2 = \pm y_3^2,$$

and up to rearranging we are essentially counting points on the conic

$$y_1^2 + y_2^2 = y_3^2.$$  

Therefore if we denote by $\mathcal{N}(T)$ the number of integral points (in the sense of Definition 2.2) on $\mathbb{P}^1_{2,2,2}$ then:

Corollary 1.4. There exist positive numbers $c_1, c_2, c_3$ such that for all $T > c_3$ we have

$$c_1 T^{1/2} < \mathcal{N}(T) < c_2 T^{1/2}.$$

The proof is elementary, since the curve can be explicitly parametrized by

$$y_1 = u^2 - v^2, \quad y_2 = 2uv, \quad y_3 = u^2 + v^2.$$ 

The condition $\max\{|y_1|, |y_2|\} \leq T^{1/2}$ is subsumed by $u^2 + v^2 \leq 4T^{1/2}$ say, so number of possible $u, v$’s is $\asymp T^{1/2}$ as desired.

Theorem 1.2 and Corollary 1.4 imply that asymptotically 0-percent of the rational points on $\mathbb{P}^1_{2,2,2}(\mathbb{Q})$ are integral, in the sense of Darmon (Definition 2.2).

1.1. Organization of the paper. The proofs of Theorems 1.2 and 1.3 (Theorem 2.3) are essentially disjoint, and are contained in Sections 2 and 3 respectively. The reader interested in one but not the other can essentially read these sections independently of each other.

We note that in the proof of Theorem 1.2 we shall require a counting result on the number of diagonal ternary quadratic forms having bounded, square-free discriminant given as Proposition 3.3. This may be of independent interest. We further note that it is possible to refine the arguments given in Section 3 to give an asymptotic formula in Proposition 3.3 but we do not pursue this in the present paper.

Notation. We denote by $d_k(n)$ for the number of ways of writing $n$ as a product of $k$ (not necessarily distinct) positive integers, and write $d(n) = d_2(n)$ for the usual divisor function. We will also use the big-$O$ notation as well as Landau’s notation. In particular, we will denote in the subscripts any dependencies; if there are no subscripts, then the implied constants are absolute.
We now define our height function on a $M$-curve $X$. Before continuing let us fix some notation. We assume that $X = (X; P_1, m_1; \ldots; P_r, m_r)$ is an $M$-curve as in Definition 1.1. We furthermore assume that $1 < m_{P_i} < \infty$.

### 2.1. Translation between $M$-curves and stacky curves

Before moving on we explain the connection between stacky curves and $M$-curves. Those uninterested may skip this section and safely work with $M$-curves. Given a nice stacky curve $X$ over a number field $K$ there is a morphism $\pi: X \to X$ to a curve $X$ called the coarse space morphism, which is the universal morphism from $X$ to a scheme. In practice one often constructs $X$ from $X$ by specifying a collection of points $P_1, \ldots, P_r$ in the coarse space $X$ and attaching stabilizer groups $\mu_{P_i}$ to each $P_i$. This is the bottom up approach of constructing an algebraic stack. One can think of this as specifying the ramification data of the coarse space morphism $\pi: X \to X$. We often think of these points as "fractional points" because in divisor class group of the associated curve we have added the point $\frac{1}{\# \mu_{P_i}} P_i$. In other words, we think of a stacky curve as a smooth curve $X$ with a choice of points $P_1, \ldots, P_r$ with stabilizer groups $\mu_{P_i}$ attached to each $P_i$. This data defines an $M$-curve $X = (X; P_1, \# \mu_{P_1}; \ldots; P_r, \# \mu_{P_r})$. Conversely, given an $M$-curve $X = (X; P_1, m_1; \ldots; P_r, m_r)$ with each $1 < m_{P_i} < \infty$ we consider the stacky curve given with points $P_i$ having the stabilizer group $\mu_{m_{P_i}}$. In this way one may establish a bijection between smooth proper geometrically connected Delign-Mumford stacks of dimension 1 over $K$ with stacky points defined over $K$ that contain an open dense subscheme and possess a projective coarse moduli space and $M$-curves over $K$ with finite multiplicities.

### 2.2. Construction of heights

Notice that the height (1.1) is defined in terms of a function on $\mathbb{P}^1$. Since a stacky curve and its coarse space agree up to a finite set of points and we are interested in counting points asymptotically it suffices to count points on the coarse space with the height (1.1). With this inspiration our goal is to take a stacky curve $X$ with coarse space $X$ and construct the associated $M$-curve $X = (X; P_1, m_1; \ldots; P_r, m_r)$. We then define a height on the rational points of $X$ using the $M$-curve data with the purpose of counting points with respect to this height.

Choose a finite set of primes $S$ of $\mathcal{O}_K$ containing all the primes of bad reduction along with a smooth and proper model $\overline{X}$ of $X$ over $\mathcal{O}_{K,S}$. Everything we do is relative to this choice of model, similar to how everything we do is relative choosing the finite set of primes $S$. We will use the following notion of intersection multiplicity from [4] to define integral points on $M$-curves and to define our heights.

**Definition 2.1 (1.1).** Let $P, Q$ be distinct points in $X(K)$ and place $\nu$ a place in $K$ with $\nu \notin S$. Take $p_{\nu} \subset \mathcal{O}_{K}$ to be the prime ideal associated to $\nu$. We define the intersection multiplicity of $P$ and $Q$ at $\nu$ as follows.

$$(P \cdot Q)_\nu := \max \{ m : \text{the images of } P, Q \text{ in } \overline{X}(\mathcal{O}_{K,S}/p_{\nu}^m) \text{ are equal.} \}$$

where the maximum over the empty set is defined to be 0 above.

We wish to define a height function on the $M$-curve which takes into account the local multiplicities. Motivated by the work of [5] our strategy will be to define a height function that takes into account a global classical height on the base curve $X$ and a local part that depends on the intersection multiplicities. In the case relevant to us we recover (1.1).

Let us fix an $M$-curve $X = (X; P_1, m_1; \ldots; P_r; m_r)$. To avoid complications with the infinite places we shall assume that $S$ contains all infinite places of $K$. Our height function will have a classical part which only depends on the underlying curve $X$ and a stacky part that depends on the points and multiplicities. To define the classical part of our height we choose an ample line bundle $\mathcal{L}$ on $X$ and a multiplicative ample height $H_\mathcal{L}$. Our strategy for the stacky contributions is to take into account all of the primes that are not in $S$. Since our intersection multiplicities depend on the choice of $S$ and model $\overline{X}$, so will our heights. Let $P$ be a point of $X$. We define

$$\lambda_S, \sum_{\nu}(P, t) = \lambda_\nu(P, t) = N(p_\nu)^{(t-P)_\nu}$$

for $\nu \notin S$ and set

$$\lambda(P, t) = \prod_{\nu \notin S} \lambda_\nu(P, t).$$
To take into account the multiplicities of points we only consider \( \lambda(P, t) \) up to \( m_P \)-powers. That is we look at the image \( \lambda(P, t) \in \mathbb{Q}/\mathbb{Q}^{m_P} \) and consider this to be the local contribution to the height. Precisely we define

\[
\lambda(P, t) = m_P - \text{free part}(\lambda(P, t))
\]

and \( H(P, t) = 1 \) if \( m_P = 1 \).

We now multiply the classical and stacky contributions to obtain our height function. To sum up, given an \( M \)-curve \( \mathcal{X} = (X; P_1, m_1; \ldots; P_r, m_r) \) and a choice of an ample height \( H_L \) on \( X \), a choice of primes \( S \) and a model \( \mathcal{X} \) we define

\[
H_{X,S}(P, t) = H_L(t) \prod_{i=1}^{r} H_{S,X}(P_i, t) = H_L(t) \prod_{i=1}^{r} m_{P_i} - \text{free part}(\lambda(P_i, t))
\]

We call \( H_L(t) \) the classical part and \( \prod_{i=1}^{r} H_{S,X}(P_i, t) \) the stacky part of the height \( H_{X,S}(t) \).

What is interesting is that our height, given by (2.3) below, allows us to differentiate between rational and integral points on \( \mathcal{X} \), and this is one of the main features of our height. In other words the \( S \)-integral points are those where the stacky part of the height is trivial. Following Darmon [4] we have the following notion of integral points on an \( M \)-curve:

**Definition 2.2** (Darmon). Let \( \mathcal{X} = (X; P_1, m_1; \ldots; P_r, m_r) \) be a \( M \)-curve over a number field \( K \), \( S \) a finite set of places of \( K \) containing all primes of bad reduction. Let \( \mathcal{X} \) be a smooth proper model for \( X \) over \( \mathcal{O}_{K,S} \). The \((X,S)\)-integral points of \( \mathcal{X} \) (usually abbreviated to \( S \)-integral points of \( \mathcal{X} \)) are the points \( t \in X(K) \) such that

\[
(t \cdot P)_v \equiv 0 \mod m_P
\]

for all \( P \in X(K) \) and \( \nu \notin S \).

We shall prove the following theorem:

**Theorem 2.3.** Let \( \mathcal{X} = (X; P_1, m_1; \ldots; P_r, m_r) \) be an \( M \)-curve over \( K \) satisfying our assumptions and choose \( S \) and a model \( \mathcal{X} \) as we have specified. Then we have the following conclusions.

1. \( \mathcal{X}(\mathcal{O}_{K,S}) \subseteq \bigcap_{m_P > 1} X(P; K) \)

where \( \mathcal{X}(P; K) = \{t \in X(K); H(P, t) = 1\} \).

2. If \( K = \mathbb{Q} \) then \( \mathcal{X}(\mathcal{O}_{K,S}) = \bigcap_{m_P > 1} X(P; K) \).

In particular, the set of \( S \)-integral points of \( \mathcal{X} \) is precisely the set of points where \( H(P, t) = 1 \) for all \( P \) with \( m_P > 1 \).

Fix a prime \( \nu \notin S \) and write \( (t \cdot P)_v = m_P^{e_{\nu,P}(t)} \cdot q_{\nu,P}(t) \) where \( e_{\nu,P}(t) \geq 0 \) and \( q_{\nu,P}(t) \geq 0 \) is not divisible by \( m_P \). In other words \( q_{\nu,P}(t) \) is the \( m_P \)-free part of \( (t \cdot P)_v \). Set \( N(p_v) = p_v^{f(v)} \). Then

\[
\lambda_{\nu}(P, t) = p_v^{f(v)(t \cdot P)_v} = m_P^{e_{\nu,P}(t)} \cdot q_{\nu,P}(t) \cdot f(v)
\]
Suppose that 

(2.7) \[ \lambda(P,t) = \prod_{\nu \notin S} p_{\nu}^{m_{v,P}(t)} q_{v,P}(t) f(\nu) \]

Using the functions \( \lambda(P,t) \) we can find subsets of the rational points that contain all integral points.

**Proposition 2.4.** Suppose that \( m_P > 1 \). Define \( \mathcal{X}(P;K) = \{ t \in X(K) : H(P,t) = 1 \} \). Then

\[ \mathcal{X}(\mathcal{O}_{K,S}\mathcal{X}) \subseteq \mathcal{X}(P;K). \]

**Proof.** Suppose that \( t \) is an \( S \)-integral point. Then \((t \cdot P)_\nu \equiv 0 \mod m_P \Rightarrow e_{v,P}(t) > 0 \) for all \( \nu \notin S \). Thus

\[ \lambda(P,t) = \prod_{\nu \notin S} p_{\nu}^{m_{v,P}(t)} q_{v,P}(t) f(\nu) = \left( \prod_{\nu \notin S} p_{\nu}^{m_{v,P}(t)-1} q_{v,P}(t) f(\nu) \right)^{m_P}, \]

whence \( H(P,t) = 1 \) as \( \lambda(P,t) \) is an \( m_P \)-power.

We see that each point \( P \) with multiplicity \( m_P > 1 \) imposes a height dropping condition on the set of integral points. Thus to study integral points it suffices to study

\[ \mathcal{X}(\mathcal{O}_{K,S}\mathcal{X}) \subseteq \bigcap_{m_P > 1} X(P;K). \]

There is the following easy consequence that may be useful if one has access to the height \( H_{\mathcal{X},S}\mathcal{X}, \mathcal{L} \) but perhaps not the local factors.

**Corollary 2.5.** Let \( H_{\mathcal{L}} \) be an ample height with \( H_{\mathcal{L}}(t) > 0 \) for all \( t \in X(K) \). Then

\[ \mathcal{X}(\mathcal{O}_{K,S}) \subseteq \{ t \in X(K) : H_{\mathcal{X},S}(t) = H_{\mathcal{L}}(t) \}. \]

The integral points are contained in the locus where the height can be computed classically, in other words where the stalky part of the height is trivial. The difference in these sets can be potentially explained by interactions between \( f(\nu) = [\mathcal{O}_K/p_\nu : \mathbb{Z}/p_\nu] \) and the integers \( m_P \) and the splitting of the primes \( p_\nu \) in \( K \).

**Proof of Theorem 2.3.** We have already shown part (1) of Theorem 2.3 in (2.4). We turn to part (2) and assume that \( K = \mathbb{Q} \). We know that that \( \bigcap_{m_P > 1} X(P;\mathbb{Q}) \subseteq \mathcal{X}(\mathcal{O}_{\mathbb{Q},S}\mathcal{X}) \) by (2.4). We now show the reverse inclusion. Let \( t \in X(\mathbb{Q}) \) with \( H(P,t) = 1 \) for all \( P \) with \( m_P > 1 \). Since \( K = \mathbb{Q} \) we have that \( N(p_{\nu}) = p_\nu \) and \( f(\nu) = 1 \) for all finite places \( \nu \). Fix \( P \) with \( m_P > 1 \). Towards a contradiction suppose that \((t \cdot P)_\nu \neq 0 \mod m_P \) for some \( \nu \notin S \). Then \( e_{\nu,P}(t) = 0 \). Notice that \( H(P,t) = 1 \) means that \( \lambda(P,t) \) is an \( m_P \)-power.

Since if \( \nu \neq \nu' \) we have that \( p_\nu \neq p_{\nu'} \) we have by unique factorization of integers that

\[ \lambda(P,t) = \prod_{\nu \notin S} p_{\nu}^{m_{\nu,P}(t)} q_{\nu,P}(t) = \left( \prod_{\nu \notin S} p_{\nu}^{m_{\nu,P}(t)} q_{\nu,P}(t) \right)^{m_P}, \]

for some integers \( z_\nu(t) \). In particular for \( \nu_0 \) we have

\[ \prod_{\nu \neq \nu_0} p_{\nu}^{m_{\nu_0,P}(t)} q_{\nu_0,P}(t) = p_{\nu_0}^{m_{\nu_0,P}(t)} q_{\nu_0,P}(t) = z_{\nu_0}(t)m_P. \]

Thus \( z_{\nu_0}(t)m_P = q_{\nu_0,P}(t) \) which contradicts \( q_{\nu_0,P}(t) \) being indivisible by \( m_P \). Thus for all \( m_P > 1 \) and \( \nu \notin S \) we have \((t \cdot P)_\nu \equiv 0 \mod m_P \) and \( t \) is an \( S \)-integral point of \( \mathcal{X} \) by definition.

In the next subsection we demonstrate our height in the simplest cases, where the base curve is \( \mathbb{P}^1 \) and we have three points with multiplicity exceeding one.
2.4. Main example - \( \mathbb{P}^1 \) with three points of large multiplicity. Now we apply these definitions in the case relevant to the conjecture of [5]. Let \( X = \mathbb{P}^1_{\mathbb{Q}} \) and \( S = \{ \nu_\infty \} \) and take \( \mathcal{L} \) to be \( \mathcal{O}_{\mathbb{P}^1}(1) \) so the ample height is the usual one. Now we define the \( M \)-curve

\[
\mathbb{P}^1_{\mathbb{Q}; q, r} := (\mathbb{P}^1_{\mathbb{Q}}; 0, p; -1, q; \infty, r)
\]

where \(-1\) is used instead of \(1\) to match the constructions in [5]. Let \( t = [a : b] \in \mathbb{P}^1(\mathbb{Q}) - \{ \infty \} \) with \( a, b \) coprime integers. Then we have that

\[
(t \cdot t)_\nu = \text{ord}(a)_\nu, \quad (t \cdot (1))_\nu = \text{ord}(a + b), \quad (t \cdot \infty)_\nu = \text{ord}(b)
\]

for all finite primes \( \nu \). The product formula gives in our specific case gives that

\[
\lambda(0, t) = |a|, \quad \lambda(-1, t) = |a + b|, \quad \text{and,} \quad \lambda(\infty, t) = |b|.
\]

Now we consider our points up to \( m_P \)-powers. That is when \( t = [a : b] \) with \( a, b \) non-zero coprime integers we consider the image

\[
\lambda(P, t) \in \mathbb{Q}^*/(\mathbb{Q}^*)^{m_P}
\]

and consider this to be the local contribution to the height of \( t \) at \( P \). The contribution of these local heights and the global height now gives for \( t = [a : b] \) with \( a, b \) coprime and non-zero integers that

\[
H_{\mathbb{P}^1_{\mathbb{Q}; q, r}}([a : b]) = H_{\mathbb{P}^1_{\mathbb{Q}; q, r}}(0, [a : b])H_{\mathbb{P}^1_{\mathbb{Q}; q, r}}(-1, [a : b])H_{\mathbb{P}^1_{\mathbb{Q}; q, r}}(\infty, [a : b])H_{\mathbb{P}^1_{\mathbb{Q}}}([a : b])
\]

\[
= \text{\(-\)free part}(|a|) \cdot q\text{-free part}(|a + b|) \cdot \text{r-free part}(|b|) \cdot \text{max}(|a|, |b|)
\]

Taking \( p = q = r = 2 \) we obtain

\[
H_{\mathbb{P}^1_{\mathbb{Q}; 2, 2}}([a : b]) = 2\text{-free part}(|a|) \cdot 2\text{-free part}(|a + b|) \cdot 2\text{-free part}(|b|) \cdot \text{max}(|a|, |b|)
\]

which is the desired height function of (5).

These discussions lead to the following question:

**How many points** \([a : b] \in \mathbb{P}^1(\mathbb{Q})\) **satisfy**

\[
\text{\(-\)free part}(|a|) \cdot q\text{-free part}(|a + b|) \cdot \text{r-free part}(|b|) \cdot \text{max}(|a|, |b|) \leq T?
\]

More generally:

*Let \( X = (X; P_1, m_1; \ldots; P_r, m_r) \) be a \( M \)-curve over a number field \( K \), \( S \) a finite set of places of \( K \) containing the infinite places and all primes of bad reduction. Let \( X \) be a smooth proper model for \( X \) over \( \mathcal{O}_{K, S} \). Then how many points \( t \in X(K) \) satisfy

\[
H_{X, S, X, \mathcal{L}}(t) \leq T
\]

Can one find an asymptotic formula for the number of such points?*

The rest of this article is dedicated to analyzing the case \( p = q = r = 2 \); that is, to the proof of Theorem 1.2

3. Counting rational points on \( X_\mathbb{Q} = (\mathbb{P}^1; 0, 2; -1, 2; \infty, 2) \)

In this section, we prove Theorem 1.2. To do so we will show that \( N(T) = O\left(T^{1/2}(\log T)^3\right) \) and give a separate argument to show that \( N(T) \gg T^{1/2}(\log T)^3 \). The incompatibility of these two arguments represents the main obstacle as to why an asymptotic formula for \( N(T) \) remains elusive.

We consider the problem of counting integral points on the variety defined by \( (12) \), subject to the constraint

(3.1) \[ 0 < |x_2x_3(x_1y_1)^2| \leq T, |x_1y_1^2| \geq |x_2y_2^2|. \]
To obtain the upper bound we must dissect (3.1) into suitable ranges. When \(|x_1 x_2 x_3| \leq T^{1/2}\) we fix \(x_1, x_2, x_3\) and treat (1.2) as a diagonal ternary quadratic form, say \(Q_x\). It is then the case that

\[
|y_i| \leq \frac{T}{|x_1 x_2 x_3| \cdot |x_i|}
\]

for \(i = 1, 2, 3\), and by Corollary 2 of [3] we then have the estimate

\[
O \left( d(x_1 x_2 x_3) \left( \frac{T^{1/2}}{|x_1 x_2 x_3|} + O(1) \right) \right)
\]

for the number of \(y \in \mathbb{Z}_2^3 \) satisfying (3.1) and (1.2) provided that the quadratic form \(Q_x\) has a rational zero. Otherwise it is clear that there will be no contribution. Thus we must estimate

\[
\sum_{1 \leq |x_1 x_2 x_3| \leq T^{1/2}} \sum_{Q_x \text{ has a rational zero}} d(x_1 x_2 x_3).
\]

This is similar to the work of Guo in [9], except he counted with respect to the height \(\|x\|_\infty\). Nevertheless the techniques are similar, and again this may be of independent interest.

Next we must deal with the case when \(|x_1 x_2 x_3| \geq T^{1/2}\). For this it suffices to observe from (3.2) that \(|x_1 x_2 x_3| \geq T^{1/2}\) implies

\[
|y_1 y_2 y_3| \leq \frac{T^{3/2}}{(x_1 x_2 x_3)^2} \leq T^{1/2}.
\]

We then treat (1.2) as a linear form \(L_y\) in \(x\). We use this to show that the contribution for each \(y\) is

\[
O \left( T^{1/2} |y_1 y_2 y_3|^{-1} + 1 \right),
\]

which gives an acceptable contribution upon summing over \(y\).

For the lower bound, we first restrict \(y_1, y_2, y_3 \in \mathbb{Z}_{\neq 0}\) satisfying

\[
|y_1 y_2 y_3| \leq T^\delta
\]

for some explicit \(\delta > 0\) to be specified later. We note that to obtain the correct order of magnitude it is permissible to choose any \(\delta > 0\).

Having fixed \(y = (y_1, y_2, y_3)\), we consider the simultaneous conditions (1.2) and (3.1). This gives rise to a binary form inequality of the shape

\[
|x_1^2 x_2 (y_1^2 x_1 + y_2^2 x_2)| \leq T y_3^2 y_1^{-2}.
\]

Because \(|y_1 y_2 y_3|\) is small, we can count the number of solutions \(x\) to this inequality with reasonable precision. However, even with \(|y_1 y_2 y_3|\) counting the number of solutions \(x\) with enough uniformity appears to still be a challenging task, because the binary form in (3.3) is singular. This difficulty is exacerbated by the fact that we will need to apply a square-free sieve eventually to produce triples \(x\) with each coordinate square-free.

To get around this issue, we simply count solutions to (3.3) with \(x_1, x_2\) satisfying the inequalities

\[
|x_i y_i^2| \leq c_i T^{1/4} |y_1 y_2 y_3|^{1/2}, \quad i = 1, 2
\]

for some positive numbers \(c_1, c_2\). This has the effect that the long cusps inherent in (3.3) are removed, and reduces the problem to a more straightforward geometry of numbers question.

3.1. Upper bounds. To obtain upper bounds, it is crucial to view (1.2) as a plane in \(x_1, x_2, x_3\) when \(|y_1 y_2 y_3| \leq T^{1/2}\) and viewing (1.2) as a conic in \(y_1, y_2, y_3\) when \(|x_1 x_2 x_3| \leq T^{1/2}\). We call the former the linear case and the latter the quadratic case. We proceed to deal with the linear case below.

3.1.1. The linear case. In this subsection we shall suppose that \(|y_1 y_2 y_3| \leq T^{1/2}\) is fixed, and count the triples \((x_1, x_2, x_3)\) and \((y_1, y_2, y_3)\) for which (1.2) holds.

The key is the following lemma on counting points in sublattices of \(\mathbb{Z}^2\):

**Lemma 3.1.** Let \(\Lambda \subset \mathbb{Z}^2\) be a lattice. Then for all positive real numbers \(R_1, R_2\) the number of primitive integral points \(x \in \Lambda\) satisfying \(|x_i| \leq R_i, i = 1, 2\) is at most \(O(R_1 R_2/\det(\Lambda) + 1)\).
Proof. If the rectangle \([-R_1, R_2] \times [-R_2, R_2]\) contains at least two primitive vectors in \(\Lambda\), say \(x_1, x_2\), then since this rectangle is convex it contains the parallelogram with end points \(\pm x_1, \pm x_2\). The area of this parallelogram is at least as large as \(\det \Lambda\), since the lattice spanned by \(x_1, x_2\) is a sublattice of \(\Lambda\). It thus follows that

\[
R_1 R_2 \gg \det \Lambda.
\]

Otherwise, the rectangle \([-R_1, R_1] \times [-R_2, R_2]\) contains at most one primitive vector in \(\Lambda\). This completes the proof. \(\Box\)

The strength of this lemma is that it gives a strong upper bound even in lopsided boxes.

Given (1.2), it follows that there is at least one \(i \in \{2, 3\}\) such that

\[
|x_i y_i^2|/2 \leq x_1 y_1^2 \leq 2|x_i y_i^2|,
\]

whence

\[
\frac{x_1 y_1^2}{2y_1} \leq |x_i| \leq \frac{2x_1 y_1^2}{y_i}.
\]

Without loss of generality, we assume that this holds for \(i = 2\). Suppose that \(M_1 \leq x_1 < 2M_1\). By (3.1), we have

\[
|x_3| \leq \frac{T}{x_2 x_3^2 y_1^2},
\]

whence

\[
|y_1 y_2 M_1 y_1 | \leq \frac{2y_2^2}{x_1 y_1} (x_1 y_1^2)
\]

\[
\leq \frac{2y_2^2 T}{M_1 y_1^2}
\]

Applying Lemma 3.1 to the lattice defined by the congruence \(y_1^2 x_1 - y_2^2 x_3 \equiv 0 \mod y_2^2\) which has determinant equal to \(y_2^2\), there are

\[
O\left(\frac{T y_2^2}{M_1 y_1^2} \cdot \frac{1}{y_2^2} + 1\right) = O\left(\frac{T}{M_1 y_1^2} + 1\right)
\]

possibilities for \(x_1, x_3\), which then determines \(x_2 = (y_1^2 x_1 - y_2^2 x_3)/y_2^2\). Similarly, applying Lemma 3.1 to the lattice defined by \(y_1^2 x_1 + y_2^2 x_2 \equiv 0 \mod y_2^2\), with determinant equal to \(y_2^2\), gives the estimate

\[
O\left(\frac{y_1^2}{y_2^2} M_1 \frac{1}{y_3^2} + 1\right) = O\left(\frac{y_1^2}{y_2^2 y_3^2} + 1\right)
\]

for the number of \(x_1, x_2\), which then also determine \(x_3\). The two bounds coincide when

\[
M_1 = \frac{T^{1/4} |y_2 y_3|^{1/2}}{|y_1|^{3/2}},
\]

and we get the bound

\[
O\left(\frac{T^{1/2} |y_2 y_3| y_1^2}{y_2^2 y_3^2 |y_1|^3} + 1\right) = O\left(\frac{T^{1/2}}{|y_1 y_2 y_3|^1} + 1\right)
\]

for the number of \(x_1, x_2, x_3\) given \(y_1, y_2, y_3\). Thus, we obtain an acceptable estimate whenever \(|y_1 y_2 y_3| \ll T^{1/2}\), since

\[
\sum_{1 \leq |y_1 y_2 y_3| \leq T^{1/2}} \frac{T^{1/2}}{|y_1 y_2 y_3|} + 1 \ll T^{1/2} \sum_{n \leq T^{1/2}} \frac{d_3(n)}{n} + \sum_{n \leq T^{1/2}} d_3(n)
\]

It is well-known that

\[
\sum_{n \leq Z} d_3(n) = Z (\log Z)^2 + O(Z \log Z).
\]
By partial summation, we have
\[
\sum_{n \leq T} \frac{d_3(n)}{n} = Z^{-1} \sum_{n \leq Z} d_3(n) + \int_{1}^{Z} \left( \sum_{n \leq t} d_3(n) \right) \frac{dt}{t^2} \ll (\log Z)^2 + \int_{1}^{Z} \frac{(\log t)^2}{t} \, dt \ll (\log Z)^3.
\]

It follows that
\[
T^{1/2} \sum_{n \leq T^{1/2}} \frac{d_3(n)}{n} + \sum_{n \leq T^{1/2}} d_3(n) \ll T^{1/2}(\log T)^3.
\]

3.1.2. The quadratic case. It remains to deal with the case when \(|y_1y_2y_3| \gg T^{1/2}\), where we instead fibre over \(x\) and consider zeroes of the corresponding diagonal quadratic forms \(Q_x\). Since
\[
|x_iy_i^2| \ll x_1y_1^2
\]
for \(i = 1, 2\) by assumption, it follows that
\[
|x_1x_2x_3y_1^2x_2^2x_3^2| \leq x_1^3y_1^6,
\]
hence
\[
|y_1^2y_2^2y_3^2| \ll \frac{x_1^3y_1^6}{x_1|x_2x_3|}.
\]

If \(|x_1x_2x_3| \gg T^{1/2}\), then
\[
x_1^3y_1^6 \gg T^{3/2} \gg x_1y_1^2 \gg T^{1/2}.
\]

This implies that
\[
|x_1x_2x_3| \cdot x_1y_1^2 \gg T,
\]
which violates (3.1) if the implied constants are sufficiently large. It thus follows that we must have \(|x_1x_2x_3| \ll T^{1/2}\) in this case.

We now fix \(x_1, x_2, x_3\) and consider \([\mathbf{12}]\) as a ternary quadratic form in \(y_1, y_2, y_3\). We shall require the following version of Corollary 2 in [3], which is an analogue of Lemma 3.1

**Lemma 3.2.** Let \(x_1, x_2, x_3\) be pairwise co-prime square-free integers. Let \(R_1, R_2, R_3\) be positive real numbers. Then the number of primitive solutions \(y_1, y_2, y_3\) to the equation
\[
x_1y_1^2 + x_2y_2^2 = x_3y_3^2
\]
with \(|y_i| \leq R_i\) is bounded by
\[
O \left( d(x_1x_2x_3) \left( \left( \frac{R_1R_2R_3}{|x_1x_2x_3|} \right)^{1/3} + 1 \right) \right).
\]

Since \(|x_1y_i^2| \ll x_1y_1^2\) for \(i = 1, 2\), it follows that
\[
|x_1x_2x_3(x_iy_i^2)| \ll |x_1x_2x_3(x_1y_1^2)| \leq T
\]
for \(i = 1, 2\), whence
\[
|(x_1y_1)^2x_2x_3|, |(x_2y_2)^2x_1x_3|, |(x_3y_3)^2x_1x_2| \ll T.
\]

This implies that
\[
(y_1y_2y_3)^2(x_1x_2x_3)^4 \ll T^3,
\]
hence
\[
|y_1y_2y_3| \ll \frac{T^{3/2}}{(x_1x_2x_3)^2}.
\]

Lemma 3.2 then implies that for fixed \(x_1, x_2, x_3\) the number of primitive \(y = (y_1, y_2, y_3)\) satisfying (1.2) is
\[
O \left( d(x_1x_2x_3) \left( \frac{T^{1/2}}{|x_1x_2x_3|} + 1 \right) \right).
\]

We now sum over primitive \(x \in \mathbb{Z}^3\) satisfying \(|x_1x_2x_3| \ll T^{1/2}\), with the property that the quadratic form \(Q_x\) given by (1.2) has a rational zero. By the Hasse-Minkowski theorem, this is tantamount to the form
Since \( x \omega \) than \( \max \), situation encountered by Fouyv and Kluners in \([6]\).

\[ S(x) \]

where \( g(x) \) expected to contribute the main term while the sum \( \sum g(x) \) is expected to be negligible, due to the cancellation of characters.

Put

\[ S(X) = \sum_{1 \leq x_1, x_2, x_3 \leq X} \frac{d(x_1 x_2 x_3)}{x_1 x_2 x_3} \]

Since \( x_1, x_2, x_3 \) are pairwise coprime and square-free, it follows that

\[ d(x_1 x_2 x_3) = 2^{\omega(x_1 x_2 x_3)} = 2^{\omega(x_1)} \cdot 2^{\omega(x_2)} \cdot 2^{\omega(x_3)}, \]

where \( \omega(n) \) is the number of distinct prime factors of \( n \). It follows that

\[ S(X) = \sum_{1 \leq x_1, x_2, x_3 \leq X} \frac{2^{\omega(x_1 x_2 x_3)}}{x_1 x_2 x_3} \]

where \( g \) expresses a product of Jacobi symbols. The sum

\[ S_1(X) = \sum_{1 \leq x_1, x_2, x_3 \leq X} \frac{1}{x_1 x_2 x_3} \left( 1 + \left( \frac{x_2 x_3}{x_1} \right) \left( \frac{x_1 x_3}{x_2} \right) \left( \frac{-x_1 x_2}{x_3} \right) \right) \]

is expected to contribute the main term while the sum

\[ S_2(x) = \sum_{1 \leq x_1, x_2, x_3 \leq X} \frac{1}{x_1 x_2 x_3} \sum_g g(x_1, x_2, x_3) \]

is expected to be negligible, due to the cancellation of characters.

By partial summation, we obtain:

\[ S_i(X) = \frac{1}{X} \Sigma_i(X) + \int_1^X \Sigma_i(t) \frac{dt}{t^2}, \]

where

\[ \Sigma_1(X) = \sum_{1 \leq x_1 x_2 x_3 \leq X} \left( 1 + \left( \frac{x_2 x_3}{x_1} \right) \left( \frac{x_1 x_3}{x_2} \right) \left( \frac{x_1 x_2}{x_3} \right) \right) \]

and

\[ \Sigma_2(X) = \sum_{1 \leq x_1, x_2, x_3 \leq X} \sum_g g(x_1, x_2, x_3). \]

Our situation differs from that of Guo in \([9]\) since we are counting over triples with \( |x_1 x_2 x_3| \leq X \) rather than \( \max \{ |x_1|, |x_2|, |x_3| \} \leq X \), which introduces some difficulties. However, this is exactly analogous to the situation encountered by Fouyv and Kluners in \([6]\).
Our key proposition will be:

**Proposition 3.3.** We have the asymptotic upper bound

\[ S(X) = O\left((\log X)^3\right). \]

In fact we can refine Proposition 3.3 to give an asymptotic formula, but this is unnecessary for our purposes.

We proceed to prove Proposition 3.3 in the remainder of the section. We begin by showing that triples \((x_1, x_2, x_3)\) with \(\mu^2(x_1x_2x_3) = 1\) and \(\omega(x_1x_2x_3)\) large contribute negligibly. To wit, put

\[ S_2^{(r)}(X) = \sum_{1 \leq x_1x_2x_3 \leq X, \omega(x_1x_2x_3) = r} \frac{1}{x_1x_2x_3} \sum_g g(x_1, x_2, x_3). \]

By the triangle inequality, it is clear that

\[ \left| S_2^{(r)}(X) \right| \ll \sum_{n \leq X} \frac{d_3(n)}{n}, \]

By partial summation, we have

\[ \sum_{n \leq X} \frac{d_3(n)}{n} = X^{-1} \sum_{n \leq X} d_3(n) + \int_1^X \left( \sum_{n \leq t} d_3(n) \right) \frac{dt}{t^2}. \]

To estimate the latter sum, we will need the following result, which is Lemma 11 in [6]:

**Lemma 3.4.** There exists an absolute constant \(B_0 \geq 1\) such that for every \(r \geq 0\), we have

\[ |\{n \leq X : \omega(n) = r, \mu^2(n) = 1\}| \leq B_0 \cdot \frac{X}{\log X} \cdot \frac{(\log \log X + B_0)^r}{r!}. \]

Applying the lemma, we have for \(\Omega = 30(\log \log X + B_0)\)

\[ \sum_{n \leq X} \frac{d_3(n)}{n} \ll \frac{X}{\log X} \sum_{r \geq \Omega} 3^r \cdot \frac{(\log \log X + B_0)^r}{r!} \]

\[ \ll \frac{X}{\log X} \sum_{r \geq \Omega} \left( \frac{3e(\log \log X + B_0)}{r} \right)^r \]

\[ \ll \frac{X}{\log X} \sum_{r \geq \Omega} \left( \frac{3e}{10} \right)^r, \]

the final sum a convergent geometric series. Hence

\[ \sum_{r \geq \Omega} \left( \frac{3e}{10} \right)^r \ll \left( \frac{3e}{10} \right)^\Omega \ll \frac{1}{\log X}. \]

We thus conclude that

\[ \sum_{r \geq \Omega} \left| S_2^{(r)}(X) \right| \ll 1 + (\log X)^{-2} + \int_1^X \frac{dt}{t(\log t)^2} = O(1) \]

and is thus negligible.

Note that \(x_1, x_2, -x_3\) cannot all be the same sign, otherwise (1.2) will only have a trivial real solution. Hence the signs of \((x_1, x_2, x_3)\) must be \((+, +, +)\), or \((+, -, +)\), since we assumed \(x_1 > 0\) and \(x_1y_1^2 \geq |x_2y_2^2|\). By rearranging, we must thus assume \(x_1, x_2, x_3 > 0\).
We then expand (3.7) by writing \( x_i = x_{i1}x_{i2} \) for \( i = 1, 2, 3 \), and
\[
\sum_{1 \leq x_1x_2x_3 \leq X} g(x_1, x_2, x_3) = \sum_{\mu^2(x_1x_2x_3) = 1} \sum_{1 < x_1 < x_i \text{ for } 1 \leq i \leq 3} \left( \frac{x_{11}x_{12}x_{31}x_{32}}{x_1} \right) \left( \frac{x_{11}x_{12}x_{31}x_{32}}{x_2} \right) \left( \frac{x_{11}x_{12}x_{21}x_{22}}{x_3} \right).
\]

We now follow the strategy outlined in [6] and break up the set
\[\{(x_{11}, x_{12}, x_{21}, x_{22}, x_{31}, x_{32}) \in \mathbb{N}^6 : x_{11}x_{12}x_{21}x_{22}x_{31}x_{32} \leq X\}\]
by restricting the \( x_{ij} \)'s to intervals of the form
\[\{A_{ij}, \Delta A_{ij}\},\]
where
\[\Delta = 1 + (\log X)^{-3}.\]

For a given \( A = (A_{11}, A_{12}, A_{21}, A_{22}, A_{31}, A_{32}) \), put
\[
\mathcal{S}_2(X; A) = \sum_{\substack{x_{ij} \in [A_{ij}, \Delta A_{ij}] \\mu^2(x_{11}x_{12}x_{21}x_{22}x_{31}x_{32}) = 1 \\prod_{i,j} x_{ij} \leq X}} \left( \frac{x_{11}x_{12}x_{31}x_{32}}{x_1} \right) \left( \frac{x_{11}x_{12}x_{31}x_{32}}{x_2} \right) \left( \frac{x_{11}x_{12}x_{21}x_{22}}{x_3} \right).
\]

We then have the following lemma:

**Lemma 3.5.** We have the bound
\[
\sum_{\prod A_{ij} \geq \Delta^{-6}X} |\mathcal{S}_2(X; A)| = O \left( X(\log X)^{-1} \right).
\]

**Proof.** We have
\[
|\mathcal{S}_2(X; A)| \leq \sum_{\Delta^{-6}X \leq n \leq X} d_3(n) \sum_{\mu^2(n) = 1} \sum_{\Delta^{-6}X \leq n \leq X} 3^\omega(n)
\ll \sum_{\Delta^{-6}X \leq n \leq X} (1 - \Delta^{-6})X(\log X)^2.
\]

By Taylor's theorem, we have
\[\Delta^{-6} = (1 + (\log X)^{-3})^{-6} = 1 - 6(\log X)^{-3} + O \left( (\log X)^{-6} \right).\]

The proof then follows. \( \Box \)

To proceed, we shall require the following well-known lemma regarding character sums:

**Lemma 3.6** (Double Oscillation Lemma). Let \( \{\alpha_n\}, \{\beta_m\} \) be two sequences of complex numbers with each term having absolute value bounded by 1. Let \( M, N \) be positive real numbers. Then we have
\[
\sum_{m \leq M} \sum_{n \leq N} \alpha_m \beta_n \mu(2m) \mu(2n) \left( \frac{m}{n} \right) \ll \min \left\{ \left( M^{-1/2} + (N/M)^{-1/2} \right), \left( N^{-1/2} + (M/N)^{-1/2} \right) \right\}
\]
and for every \( \varepsilon > 0 \),
\[
\sum_{m \leq M} \sum_{n \leq N} \alpha_m \beta_n \mu(2m) \mu(2n) \ll \varepsilon M N \left( M^{-1/2} + N^{-1/2} \right) (MN)^\varepsilon.
\]

We will also need the following variant of the Siegel-Walfisz theorem:

**Lemma 3.7.** Let \( \chi_q \) be a primitive character modulo \( q \geq 2 \). Then for every \( A > 1 \) we have
\[
\sum_{Y \leq p \leq X} \chi_\theta(p) = O_A \left( \sqrt{q} \cdot X(\log X)^{-A} \right)
\]
uniformly for \( X \geq Y \geq 2 \).
We now consider, as in \cite{6}, the quantities
\begin{equation}
X^\dagger = (\log X)^0, \ X^\ddagger = \exp \left( (\log X)^{1/8} \right).
\end{equation}

We now consider those $A$ with the property that at most 2 entries larger than $X^\dagger$. We dissect the sum according to the number $r \leq 2$ of terms $A_{ij}$ greater than $X^\dagger$. Let $n$ be the product of those $x_{ij}$ which are larger than $X^\dagger$, and $m$ the product of the remaining ones. We sum over $A$ with such properties to obtain
\[
\sum_{A} |S_2(X; A)| \leq \sum_{r \leq 2} \sum_{m \leq (X^\dagger)^{6-r}} \mu^2(m)d_b(r)(m) \sum_{n \leq X/m} \mu^2(n)d(r)(n)
\]
\[
\ll \sum_{r \leq 2} \sum_{m \leq (X^\dagger)^{6-r}} \mu^2(m)d_b(r)(m) \left( \frac{X}{m} \right) (\log X)^{r-1}
\]
\[
\ll X \left( \sum_{r \leq 2} (\log X)^{r-1} \right) \left( \sum_{m \leq (X^\dagger)^6} \frac{d_b(m)}{m} \right)
\]
\[
\ll X (\log X) \left( \log \exp (\log X)^{1/8} \right)^7
\]
\[
\ll X (\log X)^{15/8}.
\]

This is sufficiently small for our purposes.

We may now assume that $A_{ij} \geq X^\dagger$ for at least three pairs $i, j$ with $1 \leq i \leq 3, 1 \leq j \leq 2$. We now suppose that there exist $a \neq b$ such that
\[A_{a,2}, A_{b,1} \geq X^\dagger.\]

The sum over $A$ satisfying these properties can be bounded by
\[
\sum_{A} |S_2(X; A)| \leq \sum_{x_{ij}, (i,j) \neq (a,2), (b,1)} \left| \sum_{a_{ij} \neq x_{ij}} \alpha_{a,2} \beta_{b,1} \left( \frac{x_{a,2}}{x_{b,1}} \right) \right|
\]
where $\alpha, \beta$ have modulus at most one. Lemma \cite{3} then applies, and since our variables $x_{a,2}, x_{b,1}$ range over intervals exceeding $X^\dagger$ in length, it follows that
\[
|S_2(X; A)| \ll \left( \prod_{(i,j) \neq (a,2), (b,1)} A_{ij} \left( A_{a,2}A_{b,1} \left( A_{a,2}^{-1/3} + A_{b,1}^{-1/3} \right) \right) \right) \ll X (X^\dagger)^{-1/3} = O \left( X (\log X)^{-3} \right),
\]
which is again enough.

Next consider the family where the two previous conditions do not hold, and in addition there exist $a \neq b$ such that $2 \leq A_{b,1} \leq X^\dagger$ and $A_{a,2} > X^\dagger$. Under these conditions, we see that
\[
|S_2(X; A)| \ll \sum_{x_{ij}, (i,j) \neq (a,2), (b,1)} \sum_{x_{a,2}} \sum_{x_{b,1}} \left| \prod_{(i,j) \neq (a,2), (b,1)} x_{ij} \left( \frac{x_{a,2}}{x_{b,1}} \right) \right|
\]
where $A_{ij} \leq x_{ij} \leq \Delta A_{ij}$ and $\omega(x_{ij}) \leq \Omega$ for $1 \leq i \leq 3, 1 \leq j \leq 2$. Now put $\ell = \omega(x_{a,2})$, writing
\[x_{a,2} = p_1 \cdots p_{\ell}\]
with $p_1 < p_2 < \cdots < p_{\ell}$ we obtain
\[
|S_2(X; A)| \ll \sum_{x_{b,1}} \sum_{0 \leq \ell \leq \Omega} \left| \sum_{\omega(x_{a,2}) = \ell} \prod_{i,j} x_{ij} \left( \frac{x_{a,2}}{x_{b,1}} \right) \right|
\]
the inner sum being bounded by
\[
\sum_{p_1 \cdots p_{\ell-1}} \left| \sum_{p_{\ell}} \left( \frac{p_{\ell}}{x_{b,1}} \right) \right|
\]
and \(p_1, \ldots, p_\ell\) satisfy \(A_{a,2} \leq p_1 \cdots p_\ell \leq \Delta A_{a,2}\). Note that
\[
p_\ell \geq A_{a,2}^{1/\ell} \geq \exp\left((\log X)^{1/9}\right).
\]

We may now apply Lemma 3.7 to obtain the bound
\[
\left|\sum_{p_\ell} \left(\frac{p_\ell}{x_{b,1}}\right)\right| \leq A_{b,1}^{1/2} \frac{A_{a,2}}{p_1 \cdots p_{\ell-1}} (\log X)^{-A/9} + \Omega,
\]
with \(A\) arbitrarily large. Note that \(p_1 \cdots p_{\ell-1} \leq X\), hence
\[
\sum_{p_1 \cdots p_{\ell-1} \leq X} (p_1 \cdots p_{\ell-1})^{-1} \ll \sum_{n \leq X} \frac{1}{n} \ll \log X.
\]

Hence
\[
\sum_{n \leq X} |S_2(X; \mathcal{A})| \ll A_{b,1}^{1/2} \prod_{i,j} A_{ij} (\log X)^{-A/9+1} \ll X (\log X)^{-A/9+11/2}.
\]

Choosing \(A\) large shows that this contribution is negligible.

The remaining case can be summarized by the following properties:

1. \(\prod_{i,j} A_{ij} \leq \Delta^{-6} X\);
2. \(A_{ij} \geq X^{1/2}\) for at least three pairs of indices \((i,j)\);
3. If \(A_{ij}, A_{kl} \geq X^{1/2}\) then \(j = \ell\);
4. If \(A_{ij} \leq A_{kl}\) with \(j \neq \ell\), then either \(A_{ij} = 1\) or \(2 \leq A_{ij} \leq X^{1/2}\) and \(A_{kl} < X^{1/2}\).

We now show that the second option in (4) cannot happen. This will imply that we have accounted for all possibilities for \((3.7)\), and hence reduced our problem to estimating \(S_1(X)\).

Suppose, without loss of generality, that \(2 \leq A_{11} \leq X^{1/2}\) and \(A_{22} < X^{1/2}\). Since \(A_{ij} \geq X^{1/2}\) for at least three pairs of indices \((i,j)\), one of \(A_{12}\) or \(A_{32}\) must exceed \(X^{1/2}\). We then have \(A_{11} \leq X^{1/2}\) and \(A_{32}\), say, exceeds \(X^{1/2}\), which means that our earlier estimation covers this case.

The upshot now is that
\[
\Sigma_2(X) \ll A (\log X)^{15/8}
\]
for some \(\kappa(A) > 0\). It follows from (3.9) that
\[
S_2(X) = X^{-1} \Sigma_2(X) + \int_1^X \Sigma_2(t) \frac{dt}{t^2}
\ll (\log X)^{15/8} + \int_1^X \frac{(\log t)^{15/8} dt}{t}
= (\log X)^{23/8},
\]
which is sufficiently small for our purposes.

Finally, we may evaluate the main term, which is given by (3.6). By the triangle inequality, we have
\[
S_1(X) \ll \sum_{x_1, x_2, x_3 \leq X} \frac{1}{x_1 x_2 x_3} = \sum_{n \leq X} \sum_{d_3(n)} \frac{d_3(n)}{n}
\]
which is \(O((\log X)^3)\). This completes the proof of the Proposition.

### 3.3. Lower bounds

For the lower bound, we shall assume
\[
1 \leq |y_1 y_2 y_3| \leq T^\delta
\]
where \(\delta\) is some explicit positive number which we shall specify later. We then consider \(x_1, x_2\) satisfying
\[
|x_1 y_2| \leq c_1 T^{1/4} |y_1 y_2 y_3|^{1/2}, i = 1, 2
\]
where $c_1, c_2$ are two small positive numbers. Note that
\[ |x_3y_3^2| = |x_1y_1^2 + x_2y_2^2| \leq |x_1y_1^2| + |x_2y_2^2| \leq (c_1 + c_2)T^{1/4}|y_1y_2y_3|^{1/2}, \]
whence
\[ |x_1x_2x_3|(y_1y_2y_3)^2 \leq c_3T^{3/4}|y_1y_2y_3|^{3/2} \]
where $c_3 = c_1c_2(c_1 + c_2)$. Thus
\[ |(x_1y_1^2)x_1x_2x_3| \leq (c_1T^{1/4}|y_1y_2y_3|^{1/2})(c_3T^{3/4}|y_1y_2y_3|^{-1/2}) \]
which is less than $T$ provided that $c_1c_3 \leq 1$. Therefore every pair $(x_1, x_2)$ satisfying (3.12) with $x_1, x_2$ both square-free and $x_3 = (y_1^2x_1 + y_2^3x_2)y_3^{-2} \in \mathbb{Z}$ square-free will contribute to $N(T)$.

We now count pairs $(x_1, x_2)$ such that
1. $(x_1, x_2)$ satisfies (3.12);
2. $\gcd(x_1, x_2) = 1$;
3. $x_1, x_2$ are square-free; and
4. $y_1^2x_1 + y_2^3x_2 \equiv 0 \pmod{y_3^2}$, $(y_1^2x_1 + y_2^3x_2)y_3^{-2}$ is square-free.

For each prime $p$, we interpret conditions (2) to (4) modulo $p^2$. Condition (2) is the assertion that $p|x_1 \Rightarrow p \not| x_2$, Condition (3) is the assertion that for all primes $p$ we have $p^2 \not| x_1, x_2$, and Condition (4) is stating $y_3^2|y_1^2x_1 + y_2^3x_2$, and if $p^3||y_3$, then $p^{2+2}\not| y_1^2x_1 + y_2^3x_2$. Let
\[ \rho_p(m) = \#\{(x_1, x_2) \pmod{m} : \text{(2) to (4) holds for all } p|m\}. \]

It is apparent that $\rho_p(\cdot)$ is multiplicative. Put
\[ N^*(y; T) = \#\{(x_1, x_2) \in \mathbb{Z}^2 : \text{(1) to (4) hold}\} \]
and
\[ N^*_5(y; T) = \#\{(x_1, x_2) \in \mathbb{Z}^2 : \text{(3.12) holds, (2) to (4) holds mod } b\} \]
By standard arguments, we have
\[ N^*(y; T) = \prod_{p \leq y} \left(1 - \frac{\rho_p(p^{2k})}{p^{2k}}\right) \frac{T^{1/2}}{|y_1y_2y_3|} + O\left(\sum_{Y < p < T^{1/8}|y_1y_2y_3|^{1/4}\max\{|y_1|^{-1}, |y_2|^{-1}\}} \left(\frac{T^{1/2}}{p^{2|y_1y_2y_3|}} + 1\right)\right), \]
the error term being bounded by
\[ O\left(\frac{T^{1/2}}{Y|y_1y_2y_3|} + \frac{T^{1/8}|y_1y_2y_3|^{1/2}}{\min\{|y_1|, |y_2|\}}\right). \]
Since $|y_1y_2y_3| \leq T^\delta$, we obtain an acceptable error term provided that $\delta < 1/4$. This shows that
\[ N(T) \gg \sum_{1 \leq |y_1y_2y_3| \leq T^\delta} N^*(y; T) \gg \sum_{1 \leq |y_1y_2y_3| \leq T^\delta} \frac{T^{1/2}}{|y_1y_2y_3|}. \]
Since
\[ \sum_{1 \leq |y_1y_2y_3| \leq Z} |y_1y_2y_3|^{-1} \gg \sum_{n \leq Z} d_3(n)n^{-1} \gg (\log Z)^3, \]
this confirms the lower bound.

**Appendix: Prolegomena to a Theory of Heights on Algebraic Stacks**

In this section we give a hint of the flavor of heights on algebraic stacks, leaving the details to the forthcoming ([5]). Our purpose here is to highlight the difficulties faced when attempting to define heights on algebraic stacks, while instilling a sense of excitement about the future possibilities of this theory.

In order to recognize the Manin and Malle conjectures in a unified theoretical framework, ([5]) developed a theory of heights on algebraic stacks. We now explain why such a theory naturally arises when trying to unify Manin and Malle. In Manin’s conjecture one counts points using the anti-canonical height, therefore Manin’s conjecture is intimately related to the theory of heights on projective varieties. We take the theory of heights as our starting point and naively attempt to use this approach for the Malle conjecture. To approach Malle’s conjecture in a similar manner one might attempt to endow the collection of $G$-extensions
of a number field $K$ with the structure of an algebraic variety, and then count points on this variety using the height machine. Concretely one way to realize this approach would be as follows.

1. Construct a projective variety $X_G$ such that the $K$-rational points $X_G(K)$ correspond bijectively to extensions of $K$ with Galois group $G$.

2. Find a good height function $h_{X_G}$ on $X_G$ that relates the height of a point $P \in X_G(K)$ to the discriminant of the Galois extension associated to $P$.

3. Use the theory heights and the geometry of $X_G$ to count points on $X_G(K)$ and thus count $G$-extensions of $K$.

One runs into problems immediately because the set of $G$ extensions of a number field $K$ cannot naturally be realized as the set of points of a scheme. Indeed if a collection of objects can be naturally realized as the set of points of a scheme, then the objects in question must have no non-trivial automorphisms. On the other hand, $G$-extensions of $K$ always have non-trivial automorphisms given by the Galois group $G$. As is well known at this point, an appropriate setting for moduli problems with automorphisms is given by the theory of algebraic stacks introduced by Deligne and Mumford in their foundational study of the moduli space of curves. An important basic example of an algebraic stack is the classifying stack $BG$ of a finite group $G$. As the name suggests the classifying stack is a moduli space, whose points correspond to $G$-torsors. When $K$ is a number field the $K$-points of $BG$ correspond to $G$-torsors over $K$ which are precisely the Galois extensions of $K$ with Galois group $G$. Thus the theory of algebraic stacks provides a moduli space $BG$ whose points correspond to the arithmetic objects of interest, namely $G$-extensions of $K$, achieving the first task in the outline above.

A more serious problem arises in the second point. Let $G = \mathbb{Z}/2\mathbb{Z}$ and consider the classifying stack $B(\mathbb{Z}/2\mathbb{Z})$. A theory of heights parallel to that of schemes would suggest that a height function on $B(\mathbb{Z}/2\mathbb{Z})$ should correspond to a line bundle on $\mathcal{L}$ on $B(\mathbb{Z}/2\mathbb{Z})$. In general a vector bundle of rank $r$ on $BG$ corresponds to an $r$-dimensional representation of $G$. Taking $r = 1$ we see that $B(\mathbb{Z}/2\mathbb{Z})$ has line bundles corresponding to the characters of the group $\mathbb{Z}/2\mathbb{Z}$. There are precisely two characters of $\mathbb{Z}/2\mathbb{Z}$, the trivial character $\text{triv}(\epsilon) = 1$ for $\epsilon \in \mathbb{Z}/2\mathbb{Z}$ corresponding to the trivial line bundle and the sign representation $\text{sgn}(\epsilon) = \text{sgn}(\epsilon)$ where $\epsilon \in \mathbb{Z}/2\mathbb{Z}$ is considered as an element of the permutation group $S_2$. Thus we expect that up to bounded functions that there are two height functions on $B(\mathbb{Z}/2\mathbb{Z})$. The height $h_{\text{sgn}}$ associated the sign representation and the height function $h_{\text{triv}}$ associated to the trivial line bundle. The functoriality property of the height machine tells us that for any line bundle $\mathcal{L}$ we should expect that $h_{\mathcal{L} \otimes n} = nh_{\mathcal{L}} + O(1)$. Taking $\mathcal{L} = \text{triv}$ we see that $\text{triv} \otimes n = \text{triv}$ and so

$$h_{\text{triv}} = h_{\text{triv} \otimes n} = nh_{\text{triv}} + O(1)$$

for all $n$. Thus $h_{\text{triv}}$ must be a bounded function. On the other hand as the tensor product of characters is the function given by taking the product of the characters we have that $\text{sgn} \otimes \text{sgn} = \text{sgn}^2 = \text{triv}$ the trivial representation. Thus functoriality of heights suggests that we should have an equality

$$h_{\text{triv}} = h_{\text{sgn} \otimes \text{sgn}} = 2h_{\text{sgn}} + O(1).$$

Consequently $2h_{\text{sgn}}$ would be the trivial height meaning that $h_{\text{triv}}$ would be some bounded function. Clearly this is not satisfactory. For example, no Northcott property can be expected for any height function as $B(\mathbb{Z}/2\mathbb{Z})(\mathbb{Q})$ corresponds to quadratic extensions of $\mathbb{Q}$ of which there are infinitely many, but the argument above suggests that the points of $B(\mathbb{Z}/2\mathbb{Z})$ is a set of bounded height.

Despite these obstacles, J. Ellenberg, M. Satriano, and D. Zuerick-Brown realized that one may develop a theory of heights on algebraic stacks at the cost of losing the functoriality properties of the height machine. Their theory associates a height function $h_{\mathcal{E}}$ to each vector bundle $\mathcal{E}$ on an algebraic stack $X$. If $X$ is taken to be a scheme then the associated height function is the classical height $h_{\text{det}\mathcal{E}}$ associated to the determinant line bundle $\text{det}\mathcal{E}$. Thus this new theory of heights recovers the classical theory as a special case. Furthermore, given a suitable algebraic stack $X$ and a vector bundle $\mathcal{E}$ on $X$ such that the associated height $h_{\mathcal{E}}$ satisfies a Northcott property one has a conjecture (Main Conjecture) that predicts the asymptotic behaviour of points of bounded height on $X$ with respect to $h_{\mathcal{E}}$. Given a Fano variety with an ample line bundle $\mathcal{L}$ (Main Conjecture) applied to $(X, h_{\mathcal{E}})$ is the Manin conjecture, while (Main Conjecture) applied to $(BG, h_{\text{regular}})$ recovers the Malle conjecture. Here $h_{\text{regular}}$ is the height obtained from the vector
bundle associated to the regular representation of the finite group $G$.

To sum up, there is a theory of heights on algebraic stacks that is mostly unexplored. Given an algebraic stack $X$ and nice enough vector bundle $E$ on $X$ there is a conjecture (5 Conjecture), that describes the behavior of the pair $(X, E)$. Furthermore, (5 Main Conjecture) specializes to the Manin and Malle conjecture by choosing $X$ and $E$ appropriately. The full conjecture is almost completely open, though in (5) some additional cases will be considered.

References

[1] M. Bhargava, B. Poonen, The local-global principle for integral points on stacky curves, arXiv:2006.00167 [math.NT]
[2] T. D. Browning, Quantitative Arithmetic of Projective Varieties, Progress in Mathematics 277, Birkhauser 2009.
[3] T. D. Browning, D. R. Heath-Brown, Counting rational points on hypersurfaces, J. reine angew. Math. 584 (2005), 83-115.
[4] H. Darmon, Faltings plus epsilon, Wiles plus epsilon, and the generalized Fermat equation, C. R. Math. Rep. Acad. Sci. Canada 97 (1997), 3-14.
[5] J. Ellenberg, M. Satriano, D. Zureick-Brown, Forthcoming.
[6] E. Fouvry, J. Kluners, On the 4-rank of class groups of quadratic number fields, Invent. Math. 167 (2007), 455-513.
[7] J. Franke, Y. I. Manin, Y. Tschinkel Rational points of bounded height on Fano varieties, Invent. Math. 95 (1989), 421-435.
[8] A. Geraschenko, M. Satriano, A "bottom up" characterization of smooth Deligne-Mumford stacks, Int. Math. Res. Notices. IMRN, 21 (2017), 6469-6483.
[9] C. R. Guo, On solvability of ternary quadratic forms, Proc. of Lond. Math. Soc (2) 70 (1995), 241-263.
[10] D. R. Heath-Brown, The density of rational points on cubic surfaces, Acta. Arith. 79 (1997), 17-30.
[11] G. Malle, On the Distribution of Galois Groups, II, Experiment. Math., 13 (2004), 129-136.
[12] C. L. Stewart, On the number of solutions of polynomial congruences and Thue equations, J. Amer. Math. Soc (4) 4 (1991), 793-835.
[13] C. L. Stewart, S. Y. Xiao, On the representation of $k$-free integers by binary forms, to appear in Revista Matemática Iberoamericana.
[14] G. Tenenbaum, Introduction to Analytic and Probabilistic Number Theory, Cambridge University Press
[15] J. L. Thunder, Decomposable form inequalities, Ann. of Math (2) 153 (2001), 767-804.
[16] J. Voight, D. Zureick-Brown, The canonical ring of a stacky curve, arxiv:1501.04657 [math.AG].
[17] Stacks project authors, The stacks project, https://stacks.math.columbia.edu/tag/04V3

Department of Pure Mathematics, University of Waterloo
Email address: bnasser@uwaterloo.ca

Department of Mathematics, University of Toronto, Bahen Centre, 40 St. George Street, Room 6290, Toronto, Ontario, Canada, M5S 2E4
Email address: syxiao@math.toronto.edu