

Rational points on analytic varieties

Alexey Beshenov, Margaret Bilu, Yuri Bilu* and Purusottam Rath

Abstract. This is a brief, informal and very elementary introduction into the work of Pila and others about counting rational points on analytic and, more generally, definable sets. We also discuss some of the most spectacular applications.

Mathematics Subject Classification (2010). 11G99; 03C64, 11G18, 11G95, 11P21.

Keywords. Counting rational points, Manin–Mumford conjecture, André–Oort conjecture, definable sets, o-minimality.

Preface (by Yuri Bilu)

This text is based on the expanded notes of the mini-courses on the work of Bombieri, Pila, Wilkie, Zannier and others that I gave in Basel in April 2011, Yaroslavl in August 2011 and Chennai in February 2012. I thank the organizers who made it possible: David Masser and Evelina Viada at Basel; Fedor Bogomolov, Misha Verbitski and Alexei Zykin at Moscow; V. Balaji, Sanoli Gun and Purusottam Rath at Chennai.

*Supported by Bogomolov’s mega-grant, ALGANT Scholarship Program, University of Basel, Erwin Schrodinger Institute (Vienna) and the ANR project “HAMOT” (ANR 2010 BLAN-0115-01).

A. Beshenov, Institut de Mathématiques de Bordeaux, Université de Bordeaux and CNRS, 351 cours de la Libération, 33405 Talence, France

E-mail: alexey.beshenov@etu.u-bordeaux.fr

M. Bilu, Département de Mathématiques d’Orsay, Université Paris-Sud 11, F-91405 Orsay Cedex, France

E-mail: margaret.bilu@math.u-psud.fr

Yu. Bilu, Institut de Mathématiques de Bordeaux, Université de Bordeaux and CNRS, 351 cours de la Libération, 33405 Talence, France

E-mail: yuri@math.u-bordeaux1.fr

P. Rath, Chennai Mathematical Institute, Plot No. H1, SIPCOT IT Park, Padur PO, Siruseri 603103, Tamilnadu, India

E-mail: rath@cmi.ac.in

I also thank all the listeners for their precious comments. I am especially indebted to Lyosha Beshenov and Purusottam Rath, who prepared excellent notes of my courses in Yaroslavl and Chennai, respectively, and to Margaret Bilu, who attended my courses in Basel and Yaroslavl, and who wrote the present text following the notes by Lyosha and Purusottam. I only had to re-read and slightly polish the exposition at a few points, correct some inaccuracies¹ and add some references.

I thank Fedor Petrov and Umberto Zannier for useful discussions, Martin Klazar for detecting an error in the previous version, and Elina Wojciechowska for inspiration.

Finally, we thank the referee for many helpful comments.

Please do not forget that this text is based on very informal lecture notes, and do not expect completeness and precision (in particular in the references). Please do not forget either that English is not my mother tongue. I assume full responsibility for all remaining inaccuracies, and no responsibility for the damage they may cause.

This work was supported by Bogomolov’s mega-grant, ALGANT Scholarship Program, University of Basel, Erwin Schroedinger Institute (Vienna) and the ANR project “HAMOT” (ANR 2010 BLAN-0115-01).

Contents

1	Introduction: Jarník and Dörge	110
2	The theorem of Bombieri and Pila	114
3	Counting rational points with bounded height	116
4	Generalization to higher dimensions	118
5	The Manin–Mumford conjecture	120
6	Definable sets in o-minimal structures	122
	6.1 O-minimal structures	122
	6.2 The theorem of Pila and Wilkie	124
7	The André–Oort conjecture	125
	7.1 Modular curves	125
	7.2 Complex multiplication	126
	7.3 The André–Oort conjecture	126
	References	128

1. Introduction: Jarník and Dörge

Let $N \in \mathbb{N}$ be an integer. We want to study the number of intersection points of a curve with an $N \times N$ lattice. More precisely, we want to bound the number of points of the lattice $\Lambda_N = \frac{1}{N}\mathbb{Z}^2$ lying on a curve $\Gamma \subset [0, 1)^2$. Obviously,

$$|\Gamma \cap \Lambda_N| \leq N^2,$$

¹probably creating some new ones . . .

and we can find a curve Γ that makes this an equality, going through all vertices of the lattice in $[0, 1]^2$. Suppose now that Γ is a graph of a function, that is, a set of the form $\{(x, y) \in [0, 1]^2 \mid y = f(x)\}$. Then again, there is an easy bound

$$|\Gamma \cap \Lambda_N| \leq N,$$

and there exist graphs Γ realizing it. To make the problem more interesting, we are going to assume that f is strictly concave or convex. The following statement is essentially due to Jarník [14].

Theorem 1.1 (Jarník, 1926). *Let Γ be the graph of a strictly convex or concave monotonic function $[0, 1] \rightarrow [0, 1]$. Then*

$$|\Gamma \cap \Lambda_N| \leq 3\pi^{-2/3} N^{2/3} + O(N^{1/3} \log N). \tag{1.1}$$

The proof of Jarník’s theorem will not be needed further, but it is simple and beautiful, and we sketch it below. Skipping it should not in any way harm the reader’s understanding of the main content of this text. We thank Fedor Petrov who explained to us Jarník’s argument and his subsequent work on the topic, and Martin Klazar, who detected an error in the previous version of the proof.

It will be more convenient to prove a (formally) more general statement, by considering functions $I \rightarrow [0, 1]$, defined on a closed sub-interval I of $[0, 1]$. Thus, define

$$\mu(N) = \max_f |\Gamma_f \cap \Lambda_N|,$$

where f runs over the set of strictly increasing and strictly convex functions from a closed sub-interval of $[0, 1]$ to $[0, 1]$, and Γ_f denotes the graph of f . Further, let I be a shortest sub-interval of $[0, 1]$ such that there exists $f : I \rightarrow [0, 1]$ as above with $|\Gamma_f \cap \Lambda_N| = \mu(N)$. We fix these I and f and write $\Gamma = \Gamma_f$.

Call a vector $(x, y) \in \mathbb{Z}^2$ *primitive* if $d = \gcd(x, y) = 1$; in general, we call $(x/d, y/d)$ the *primitivization* of (x, y) . Let S_X denotes the set of primitive vectors with positive coordinates whose sum is bounded by X :

$$S_X = \{(x, y) \in \mathbb{Z}^2 : x, y > 0, \gcd(x, y) = 1, x + y \leq X\}$$

Then

$$|S_X| = \frac{3}{\pi^2} X^2 + O(X \log X), \quad \sum_{(x,y) \in S_X} x = \frac{1}{\pi^2} X^3 + O(X^2 \log X). \tag{1.2}$$

(These relations are pretty standard. See, for instance, [12, Section 18.5] for the proof of a similar statement; both relations (1.2) can be proved analogously.)

Let P_0, \dots, P_n be the lattice points on Γ (so that $n + 1 = \mu(N)$) ordered by the increasing first coordinates. The vectors $(x_i, y_i) = N(P_i - P_{i-1})$ are pairwise distinct (because the function is convex) and have positive coordinates (because it is increasing). We may also assume them to be primitive; if one of them is not, then, replacing it by the primitivization, we may find a new (strictly increasing and strictly convex) function

with graph having $n + 1$ lattice points, but defined on a shorter interval, contradicting our choice of I .

Let X be the largest real number such that $n + 1 \geq |S_X|$. Then

$$|\Gamma \cap \Lambda_N| = n + 1 = \frac{3}{\pi^2} X^2 + O(X \log X),$$

$$2N \geq \sum_i (x_i + y_i) \geq 2 \sum_{(x,y) \in S_X} x = \frac{2}{\pi^2} X^3 + O(X^2 \log X),$$

whence the result.

It follows from the proof that bound (1.1) is optimal in the sense that for each N we can find a curve Γ_N (satisfying the assumptions of Theorem 1.1) such that

$$|\Gamma_N \cap \Lambda_N| \geq 3\pi^{-2/3} N^{2/3} (1 + o(1)).$$

Here the curve Γ_N depends on N . One may wonder whether there exists a “universal Jarník curve” Γ satisfying $|\Gamma_N \cap \Lambda_N| > \kappa N^{2/3}$ for infinitely many N , where κ is a positive real number, not necessarily equal to $3\pi^{-2/3}$. This question, which is attributed to J.-M. Deshouillers and A. Plagne [25], was recently answered negatively by F. Petrov [20], who showed that for any fixed Γ we have

$$|\Gamma \cap \Lambda_N| = o(N^{2/3}) \quad \text{as } N \rightarrow \infty.$$

Now we give a totally different proof of a weaker version of Jarník’s theorem. This proof is based on an idea of Dörge [9] (1927) (see also [16, Section 9.2]) that will turn out to be very fruitful.

Theorem 1.2. (*Weak Jarník*) *Let I be a compact interval and Γ be the graph $\{(x, y) \mid y = f(x)\}$ of a strictly convex function $f \in \mathcal{C}^2(I)$. Then*

$$|\Gamma \cap \Lambda_N| \leq c(I, f) N^{\frac{2}{3}}.$$

Note that this is weaker than Jarník’s theorem because f is assumed to be \mathcal{C}^2 and because the constant in the bound depends on f .

The proof uses the following generalization of Lagrange’s mean value theorem:

Proposition 1.3. (*Schwarz mean value theorem*) *Let I be an interval, $f \in \mathcal{C}^n(I)$, and $x_0, \dots, x_n \in I$. Then there is $\tau \in I$ such that*

$$\begin{vmatrix} 1 & x_0 & \cdots & x_0^{n-1} & f(x_0) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_n & \cdots & x_n^{n-1} & f(x_n) \end{vmatrix} = \frac{f^{(n)}(\tau)}{n!} V(x_0, \dots, x_n).$$

Here $V(x_0, \dots, x_n)$ denotes the Vandermonde determinant

$$\begin{vmatrix} 1 & x_0 & \cdots & x_0^{n-1} & x_0^n \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_n & \cdots & x_n^{n-1} & x_n^n \end{vmatrix}.$$

Proof of Schwarz mean value theorem. Let $g(t) = a_n t^n + \cdots + a_1 t + a_0$ be the Lagrange interpolation polynomial for f at the points x_0, \dots, x_n . Then we have $(f - g)(x_i) = 0$ for all $i = 0, \dots, n$, and therefore by a “generalized Rolle theorem” there is $\tau \in I$ such that $f^{(n)}(\tau) = g^{(n)}(\tau)$. Moreover, $g^{(n)}(\tau) = n! a_n$. Noting that

$$V(x_0, \dots, x_n) a_n = \begin{vmatrix} 1 & x_0 & \cdots & x_0^{n-1} & f(x_0) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_n & \cdots & x_n^{n-1} & f(x_n) \end{vmatrix}$$

then yields the result. \square

Proof of the weak Jarník theorem. Let us pick three points $P_i = (x_i, y_i)$, $i = 0, 1, 2$ on Γ . Because of the strict convexity of f , they cannot lie on one straight line, so

$$\Delta = \begin{vmatrix} 1 & x_0 & y_0 \\ 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \end{vmatrix} \neq 0.$$

By the Schwarz mean value theorem, there exists $\tau \in I$ such that

$$\Delta = \begin{vmatrix} 1 & x_0 & f(x_0) \\ 1 & x_1 & f(x_1) \\ 1 & x_2 & f(x_2) \end{vmatrix} = \frac{f''(\tau)}{2!} V(x_0, x_1, x_2).$$

Since I is compact, this gives us an upper bound

$$|\Delta| \leq c(I, f) \max_{1 \leq i < j \leq 3} |x_i - x_j|^3.$$

Assume now that $P_0, P_1, P_2 \in \Gamma \cap \Lambda_N$, that is, the P_i are rational points whose coordinates have denominators dividing N . Then Δ is a non-zero rational number with denominator at most N^2 , so

$$|\Delta| \geq \frac{1}{N^2}.$$

This, combined with the above upper bound, gives us

$$\max_{1 \leq i < j \leq 3} |x_i - x_j| \geq c N^{-2/3}.$$

We have thus proved that an interval containing the projections of three points lying in $\Gamma \cap \Lambda_N$ has length bounded from below by $c N^{-2/3}$. Cutting I into small intervals of length $\frac{c}{2} N^{-2/3}$, we have at most two points projecting to each such small interval. Denoting by $\ell(I)$ the length of I , there are at most $\ell(I) \frac{2}{c} N^{2/3}$ such intervals, so the result follows. \square

In fact, we have proved the following

Proposition 1.4. *Let I be a compact interval and Γ the graph $\{(x, y) \mid y = f(x)\}$ of a function $f \in C^2(I)$. There is a constant c depending only on I and f , having the following property. If for the points*

$$P_0, \dots, P_k \in \Gamma \cap \Lambda_N, \quad P_i = (x_i, y_i)$$

the coordinates x_0, \dots, x_k lie in an interval of length smaller than $c N^{-2/3}$, then P_0, \dots, P_k lie on a straight line.

In particular, the set $\Gamma \cap \Lambda_N$ is covered by $O_f(N^{2/3})$ straight lines. Bombieri and Pila [2] came up with a crucial idea: consider not only straight lines, but also more general algebraic curves. We discuss it in the next section.

2. The theorem of Bombieri and Pila

Generalizing Dörge's idea, Bombieri and Pila [2] proved that a plane analytic compact curve cannot contain many rational points with a given denominator.

In the sequel by an *irreducible plane curve of degree d* we mean a subset of \mathbb{R}^2 consisting of points (x, y) satisfying $F(x, y) = 0$, where $F(X, Y)$ is an irreducible real polynomial of degree d . A *segment* of such a curve is its connected compact subset.

Theorem 2.1. (Bombieri, Pila, 1989) *Let $\Gamma \subset \mathbb{R}^2$ be a real analytic plane compact curve.*

(1) *Assume that Γ is transcendental. Then for any $\varepsilon > 0$ there is a constant $C(\Gamma, \varepsilon)$ such that*

$$|\Gamma \cap \Lambda_N| \leq C(\Gamma, \varepsilon) N^\varepsilon,$$

where $\Lambda_N = \frac{1}{N}\mathbb{Z}^2$.

(2) *Assume that Γ is a segment of an irreducible plane algebraic curve of degree d . Then for any $\varepsilon > 0$ there is a constant $C(\Gamma, \varepsilon)$ such that*

$$|\Gamma \cap \Lambda_N| \leq C(\Gamma, \varepsilon) N^{\frac{1}{d} + \varepsilon}.$$

Remarks

- (1) The example of a curve given by $y = x^d$ shows that the exponent in the second part of the theorem cannot be made smaller than $1/d$.
- (2) In the second part of the theorem one can replace $C(\Gamma, \varepsilon)$ by $C(d, \varepsilon)$, if one assumes in addition that $\Gamma \subset [0, 1]^2$.

We are going to use the following generalization of the mean value theorem:

Proposition 2.2. (Generalized Schwarz mean value theorem) *Let I be an interval, $f_0, \dots, f_n \in C^n(I)$, and $x_0, \dots, x_n \in I$. Then for all i, j such that $0 \leq i, j \leq n$ there is $\tau_{ij} \in I$ such that*

$$\det[f_i(x_j)]_{i,j} = \det \left[\frac{f_i^{(j)}(\tau_{ij})}{j!} \right]_{ij} \cdot V(x_0, \dots, x_n).$$

Here is the proof for $n = 1$: by the Lagrange mean value theorem, there are $\tau_{01}, \tau_{11} \in I$ such that $f_i(x_1) - f_i(x_0) = f'_i(\tau_{i1})(x_1 - x_0)$ for $i = 0, 1$. Thus,

$$\begin{vmatrix} f_0(x_0) & f_0(x_1) \\ f_1(x_0) & f_1(x_1) \end{vmatrix} = \begin{vmatrix} f_0(x_0) & f_0(x_1) - f_0(x_0) \\ f_1(x_0) & f_1(x_1) - f_1(x_0) \end{vmatrix} = \begin{vmatrix} f_0(x_0) & f'_0(\tau_{01}) \\ f_1(x_0) & f'_1(\tau_{11}) \end{vmatrix} (x_1 - x_0),$$

and we can choose $\tau_{00} = \tau_{10} = x_0$.

The proof of the general case is very similar, but with Lagrange replaced by Schwarz, see [33].

Corollary 2.3. *If the interval I is compact, there is a constant c depending only on f_0, \dots, f_n such that*

$$|\det[f_i(x_j)]_{i,j}| \leq c |V(x_0, \dots, x_n)|.$$

Definition 2.4. Let $d, e \in \mathbb{N}$. We call a (d, e) -curve a (possibly reducible) plane algebraic curve with equation of the form $P(x, y) = 0$ where $P \in \mathbb{R}[x, y]$ is a non-zero polynomial such that $\deg_x P < d$ and $\deg_y P < e$.

For example, a $(2, 2)$ -curve is of the form $a + bx + cy + hxy = 0$.

Facts about (d, e) -curves

- (1) Any collection of $de - 1$ points lies on a (d, e) -curve.
- (2) A collection of de points $P_k = (x_k, y_k), k = 1, \dots, de$, lies on a (d, e) -curve if and only if

$$\Delta := \det[x_k^i y_k^j]_{\substack{(i,j)=(0,0), \dots, (d-1, e-1) \\ k=1, \dots, de}} = 0.$$

Note that the rows in the above determinant are numbered by the bi-index (i, j) . For example, for $(2, 2)$ -curves this condition for four points P_1, \dots, P_4 takes the form

$$\begin{vmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x_4 \\ y_1 & y_2 & y_3 & y_4 \\ x_1 y_1 & x_2 y_2 & x_3 y_3 & x_4 y_4 \end{vmatrix} = 0.$$

- (3) If P_1, \dots, P_s do not lie on a (d, e) -curve (in particular $s \geq de$ by the above), then one can select de points among them not lying on a (d, e) -curve.

Let Γ be the graph $y = f(x)$ of a function $f \in \mathcal{C}^{de-1}(I)$ where I is an interval, and consider $P_1, \dots, P_{de} \in \Gamma \cap \Lambda_N$ not lying on a (d, e) -curve. Then, by the above, writing $g_{i,j}(x) = x^i f(x)^j$ we have

$$\Delta = \det[g_{ij}(x_k)]_{\substack{(i,j)=(0,0), \dots, (d-1, e-1) \\ k=1, \dots, de}} \neq 0,$$

where $P_i = (x_i, f(x_i))$. Using Corollary 2.3 for the functions g_{ij} , we get

$$|\Delta| \leq c(f) |V(x_1, \dots, x_{de})| \leq c \left(\max_{1 \leq k < l \leq de} |x_k - x_l| \right)^{\frac{de(de-1)}{2}},$$

where the constant c depends only on I and f . On the other hand, the coordinates of P_i 's are rational numbers with denominator dividing N . By assumption, the determinant Δ evaluates to a non-zero rational number, and the entries in its (i, j) -th column have a common denominator dividing N^{i+j} . Thus the denominator of Δ is bounded by

$$\prod_{(i,j)} N^{i+j} = N^{\sum_{(i,j)} i+j} = N^{\frac{de(d+e-2)}{2}}$$

and we have

$$|\Delta| \geq N^{-\frac{de(d+e-2)}{2}}.$$

We can therefore conclude that if points $P_1, \dots, P_{de} \in \Gamma \cap \Lambda_N$ do not lie on a (d, e) -curve, then

$$\max |x_k - x_l| \geq c(f) \cdot N^{-\frac{e+d-2}{ed-1}} = c(f) \cdot N^{-\delta},$$

where $\delta = \frac{d+e-2}{de-1}$. This means that if de points of $\Gamma \cap \Lambda_N$ project to an interval of length $\frac{c}{2}N^\delta$, they necessarily lie on a (d, e) -curve. Since we can cut our interval I into $c(f, d, e, I)N^\delta$ such intervals, we have the following.

Lemma 2.5. (Main lemma of Bombieri–Pila) *Let Γ be the graph $y = f(x)$ of a function $f \in C^{de-1}(I)$. Then the set $\Gamma \cap \Lambda_N$ is covered by cN^δ (d, e) -curves, where c depends on I, f, d, e and $\delta = \frac{d+e-2}{de-1}$.*

Note that the approach of Bombieri and Pila is an exact analogue of the proof of the weak Jarník theorem we gave above, with straight lines replaced by more general (d, e) -curves.

Proof of Theorem 2.1. We prove the two statements of the theorem separately:

- (1) Take Γ transcendental, analytic and compact as in the statement of the theorem, and $\varepsilon > 0$. Choose moreover d, e such that $\frac{d+e-2}{de-1} < \varepsilon$. We can break up Γ into a finite number of pieces of the form $y = f(x)$ or $x = f(y)$, to each of which we apply the main lemma. Thus $\Gamma \cap \Lambda_N$ is covered by cN^ε (d, e) -curves, where $c = c(\Gamma, \varepsilon)$. On the other hand, since Γ is transcendental, the size of the intersection of Γ with a (d, e) -curve is bounded uniformly by a constant depending only on Γ, d, e .
- (2) Assume now that Γ is a segment of an irreducible algebraic curve of degree d . By applying a linear change of variables, we may assume that it is defined by the polynomial equation $F(x, y) = 0$ where $\deg_x F = d$. Choose e big enough such that $\frac{d+e-2}{de-1} < \frac{1}{d} + \varepsilon$. Then, applying the main lemma, $\Gamma \cap \Lambda_N$ is covered by $cN^{1/d+\varepsilon}$ (d, e) -curves, where $c = c(\Gamma, \varepsilon)$. On the other hand, since $\deg_x F = d$, our curve Γ cannot be contained in a (d, e) -curve. By the Bézout theorem, the size of the intersection of Γ with a (d, e) -curve is bounded by $c(d, e)$, and we are done. \square

3. Counting rational points with bounded height

In this section, instead of counting lattice points on Γ , we are going to count points of bounded height.

Definition 3.1. Let $\alpha = \frac{p}{q} \in \mathbb{Q}$ with p and q coprime. Define the height of α by

$$H(\alpha) = \max\{|p|, |q|\}.$$

Denoting $\Gamma(\mathbb{Q}, N) = \{(x, y) \in \Gamma \cap \mathbb{Q}^2, H(x), H(y) \leq N\}$, we can state the following stronger version of Theorem 2.1:

Theorem 3.2. (*Bombieri–Pila theorem for rational points*) Let Γ be a plane compact analytic curve.

(1) If Γ is transcendental, then for any $\varepsilon > 0$ there is a constant $c(\Gamma, \varepsilon)$ such that

$$|\Gamma(\mathbb{Q}, N)| \leq c(\Gamma, \varepsilon) N^\varepsilon.$$

(2) If Γ is algebraic of degree d , then for any $\varepsilon > 0$ there is a constant $c(\Gamma, \varepsilon)$ such that

$$|\Gamma(\mathbb{Q}, N)| \leq c(\Gamma, \varepsilon) N^{\frac{2}{d} + \varepsilon}.$$

Again, the second part cannot be improved, as shown by the example $y = x^d$.

Proof. To get this result, we just need to modify a little the proof of Theorem 2.1, more precisely, the proof of the Main Lemma. Consider points

$$P_1, \dots, P_{de} \in \Gamma(\mathbb{Q}, N), \quad P_k = (x_k, y_k)$$

not lying on a (d, e) -curve, where Γ is a graph $y = f(x)$. We still have the bound

$$|\Delta| \leq c(f) \left(\max_{1 \leq k < l \leq de} |x_k - x_l| \right)^{\frac{de(de-1)}{2}}$$

obtained directly from the general mean value theorem where

$$\Delta = \det[x_k^i y_k^j]_{\substack{(i,j)=(0,0), \dots, (d-1, e-1) \\ k=1, \dots, de}} \neq 0.$$

In fact, the only thing that needs to be changed is the computation of the lower bound for Δ . The difference is that since now our condition is on the height of x_k 's and y_k 's, we have less information on their denominators, and in particular we can no longer use N as a common denominator. Write $(x_k, y_k) = \left(\frac{\cdot}{M_k}, \frac{\cdot}{N_k}\right)$, with $0 < M_k, N_k \leq N$. Then the common denominator in column k is at most $M_k^{d-1} N_k^{e-1} \leq N^{d+e-2}$, and Δ has denominator at most $N^{de(d+e-2)}$, which gives us

$$|\Delta| \geq N^{-de(d+e-2)}.$$

The rest of the proof works in exactly the same way. Note that in the proof of the first version of the Main Lemma, the exponent was $\frac{de(d+e-2)}{2}$, that is, we lost the factor $\frac{1}{2}$. Therefore, the new version of the Main Lemma is the following:

Lemma 3.3. (*Second Main Lemma*) Let I be an interval and Γ a graph $y = f(x)$ of a function $f \in \mathcal{C}^{de-1}(I)$. Then $\Gamma(\mathbb{Q}, N)$ is covered by $c N^\delta$ (d, e) -curves, where c depends only on I, f, d and e , and $\delta = \frac{2(d+e-2)}{de-1}$.

4. Generalization to higher dimensions

Let $X \subset \mathbb{R}^m$ be a compact real-analytic transcendental manifold. We denote by $X(\mathbb{Q}, N)$ the set of rational points of X whose coordinates are all of height smaller than N . To generalize the previous results, we would like to get something like

$$|X(\mathbb{Q}, N)| \leq c(X, \varepsilon) N^\varepsilon.$$

However this is easily seen to be false, as the example of the surface $z = x^y$ where $1 \leq x, y \leq 2$ shows. Indeed, the problem that arises in the higher dimensional case is the fact that a transcendental manifold can very well contain whole pieces of algebraic curves, which can have many rational points. Let us therefore define X^{alg} to be the union of all segments of algebraic curves inside X , and $X^{\text{tr}} = X \setminus X^{\text{alg}}$. The above statement then turns out to be true when replacing X by X^{tr} :

Theorem 4.1. (*Pila, Wilkie*) *For every $\varepsilon > 0$ there is a constant $c(X, \varepsilon)$ such that*

$$|X^{\text{tr}}(\mathbb{Q}, N)| \leq c(X, \varepsilon) N^\varepsilon.$$

We sketch now the proof in the case where X is a (two-dimensional) surface inside \mathbb{R}^3 . In this special case, Theorem 4.1 was proved by Pila in [21].

Theorem 4.2. (*Pila, 2005*) *Let $X \subset \mathbb{R}^3$ be an analytic surface. For every $\varepsilon > 0$, there is a constant $c(X, \varepsilon)$ such that*

$$|X^{\text{tr}}(\mathbb{Q}, N)| \leq c(X, \varepsilon) N^\varepsilon.$$

In fact, Pila proved this result for subanalytic surfaces, projections of analytic surfaces. See [1] for a tutorial on these topics.

We start from an analog of the Main Lemma. Instead of algebraic curves of bounded degree, as in the one-dimensional case, we now use algebraic surfaces of bounded degree: for a positive integer d , a d -surface² is defined by $P(x, y, z) = 0$ where $\deg_x P, \deg_y P, \deg_z P < d$. The Main Lemma asserts that rational points on a sufficiently smooth compact surface lie on a few d -surfaces.

Proposition 4.3. (*Main Lemma in dimension 2*) *Let $\varepsilon > 0$. There are integers d and D depending on ε , such that if X is a compact C^D -surface in \mathbb{R}^3 , then $X(\mathbb{Q}, N)$ is covered by $c N^\varepsilon$ d -surfaces, where $c = c(X, \varepsilon)$.*

The proof goes along the same lines as in dimension 1. Fix d and D , to be specified later. Points $P_1, \dots, P_s \in \mathbb{R}^3$ belong to a d -surface if a certain determinant Δ vanishes. We may assume that X is given by $z = f(x, y)$, where f is a function defined on some compact domain in \mathbb{R}^2 having bounded continuous derivatives of order up to D . If $P_1, \dots, P_s \in X$ project to a small square on the x, y -plane, and D is large enough (in terms of d), then Δ can be bounded from above using some analogues of the Mean Value Theorem. On the other hand, if P_1, \dots, P_s are rational points of height at most N and $\Delta \neq 0$, then Δ is bounded from below by some negative power of N . When the square

²To make it compatible with the (d, e) -curves, we should have probably said (d, d, d) -surface, but this is too lengthy.

is small enough, the upper bound contradicts the lower bound, which means that Δ must be 0. In other words, if the points project to a sufficiently small square, then they must belong to a d -surface. Selecting d suitably large, we see that “sufficiently small square” means “square with side length $c(X, \varepsilon)N^{-\varepsilon}$ ”. Now subdividing the domain of definition of the function f into small squares, we complete the proof. For the details see [32, Appendix A, Lemma A.3].

Having this, let us see what kind of intersection X can have with a d -surface. Since X is transcendental, such an intersection is of dimension 1; by compactness, it must be a finite union of irreducible analytic curves. The number of irreducible components in the intersection is controlled by the following result due to Gabrielov [10].

Theorem 4.4. (Gabrielov) *Let V, W be complex semi-analytic manifolds with W compact, and $\pi : V \rightarrow W$ an analytic map. There exists a constant $c(\pi)$ depending only on π such that for all $\omega \in W$ the number of connected components of $\pi^{-1}(\omega)$ is bounded by $c(\pi)$.*

Applying this theorem (more precisely, an analogous statement about irreducible components) with W being the space of all non-zero polynomials defining d -surfaces (which is $\mathbb{P}^{d^3-1}(\mathbb{R})$, hence compact), we bound the number of components in the intersection of X with a d -surface. Now let us consider one such component. It is a curve, algebraic or transcendental. If it is algebraic, it is contained in X^{alg} and we don't have to worry about it. If it is transcendental, then we reduced the problem to the one-dimensional case. Projecting it to one of the coordinate planes, we obtain a plane transcendental curve, and to the latter we may apply Theorem 3.2.

The argument above can be briefly summarized as follows: rational points of bounded height from X^{tr} belong to “few” transcendental curves, and each of the latter has “few” rational points of bounded height by Theorem 3.2.

However, there is an important difficulty: Theorem 3.2 involves a constant depending on the curve. Thus we need to have a sort of “uniform” version of this theorem, which would follow if we get a similar uniform version of the one-dimensional Main Lemma.

Let us therefore for a moment go back to the one-dimensional case. Let Γ be the graph $y = f(x)$, where f is defined on some interval I and has sufficiently many bounded derivatives therein. The Main Lemma asserts that the set $\Gamma(\mathbb{Q}, N)$ is contained inside a union of cN^ε (d, d) -curves, where c depends on I, f, d and ε .

Now assume that f runs through a continuous family of functions

$$\{f_\tau : I_\tau \rightarrow \mathbb{R} : \tau \in T\}$$

with a compact base T . We will manage to complete the argument above if we bound the constant c uniformly in τ . This would have been possible if the constant depended only on the length $|I|$ of the interval and the sup-norm $\|f\|$: both are bounded in a continuous family with compact base. Unfortunately, the constant depends also on the sup-norms of the derivatives $\|f^{(k)}\|$ with $k = 0, \dots, d^2 - 1$, and the derivatives are not bounded in a continuous family with compact base (for example, take $T = [0, 1], I_\tau = [0, 1]$ and $f_\tau(x) = \tau^{1/3}(x + \tau)^{1/2}$).

One can take care of the first derivative f' by noticing that on the part of the interval where $|f'| \geq 1$ our curve can be written as $x = g(y)$ with $|g'| \leq 1$. However, one cannot deal like this with higher order derivatives.

Instead, Pila uses the fact that derivatives can be large only on short intervals. The following lemma is easy to prove.

Lemma 4.5. *Assume that f is of class \mathcal{C}^2 on $I = [a, b]$, and f, f', f'' do not vanish on I . Put $J = [a + \delta, b - \delta]$ where $0 < \delta < \frac{b-a}{2}$. Then $\|f'\|_J < \delta^{-1} \|f\|$.*

Indeed, assume, for instance, that $f, f', f'' > 0$ on I . Then for $x \in J$ we have $f(x) > 0$ and $f'(u) > f'(x) > 0$ for $u \in [x, b]$. Hence

$$f(b) = f(x) + \int_x^b f'(u) du > f'(x)(b-x) \geq f'(x)\delta,$$

whence $0 < f'(x) < \delta^{-1} f(b) = \delta^{-1} \|f\|$, proving the lemma in this case. The other cases are treated similarly.

By induction, one shows that if $f \in \mathcal{C}^{n+1}(I)$ and $f, f', \dots, f^{(n+1)}$ do not vanish on I , then $\|f^{(n)}\|_J \leq (n/\delta)^n \|f\|$.

Thus high order derivatives may grow uncontrollably only near the points where one of them vanishes.

Now return to our curve $y = f(x)$. Put $n = d^2 - 1$ and assume that the functions f, f', \dots, f^{n+1} have k roots altogether. Fix a small positive number δ and throw away from our interval I all the δ -neighborhoods of these roots. On the remaining part of I the sup-norms of derivatives of order up to n are controlled by the sup-norm of f . As for the part thrown away, it consists of k tiny intervals, and Pila applies to them an ingenious re-scaling argument, going back to the original article with Bombieri. The details are quite intricate and cannot appear here: the reader may consult Pila's article [21] or the exposition in [32, Appendix A].

This reasoning implies a new version of the Main Lemma, with constant depending only on $|I|, \|f\|$ and the number of zeros of derivatives of order up to n . For the curves occurring in the proof of Theorem 4.2 the first two of these parameters are estimated immediately, just from the compactness of the space of d -surfaces. As for the number of zeros of derivatives, this can be estimated using Gabrielov's theorem.

This is, in general terms, how Theorem 4.2 was proved.

5. The Manin–Mumford conjecture

In this section we show how Pila's techniques apply to the famous problem of Manin–Mumford. We start with some generalities about complex abelian varieties.

Let A be an abelian variety over \mathbb{C} of dimension g . Recall that there is a complex analytic group isomorphism

$$A(\mathbb{C}) \cong \mathbb{C}^g / \Lambda,$$

where $A(\mathbb{C})$ is the group of complex points of A and Λ is a lattice inside \mathbb{C}^g . It follows that

$$A[N] \cong (\mathbb{Z}/N\mathbb{Z})^{2g},$$

where $A[N] = \{x \in A \mid Nx = 0\}$ are the N -torsion points of A .

We will denote the group of all torsion points of A by $A^{\text{tors}} = \cup_{N \geq 1} A[N]$. In its simplest form, the Manin–Mumford conjecture can be stated as follows:

Theorem 5.1. (*Manin–Mumford conjecture, simplest form*) *Let X be an algebraic curve on an abelian variety A . If X is not an elliptic curve, then $|X \cap A^{\text{tors}}|$ is finite.*

If $X \subset A$ is an elliptic curve passing through a torsion point of A , then the above set is equal to the set of torsion points of X , so is infinite: this explains why we need to exclude elliptic curves.

Theorem 5.1 was proved by Raynaud [26] in 1983. To state a more general form of the Manin–Mumford conjecture, we need the following definition.

Definition 5.2. A subvariety $B \subset A$ is called a *torsion subvariety* if it is a translate of an abelian subvariety by a torsion point. That is, there exist a torsion point $b \in A$ and an abelian subvariety $B_0 \subset A$ such that $B = b + B_0$.

Theorem 5.3. (*Manin–Mumford conjecture, general form*) *Let $X \subset A$ be a complex algebraic variety. Then X has only finitely many maximal torsion subvarieties.*

This was proved by Raynaud [27] in 1983 as well. After the work of Raynaud a number of other proofs emerged.

Recently a new proof was suggested by Pila and Zannier (2006). They do prove the general conjecture, but we will discuss their argument only in the simplest case where X is an algebraic curve. Let us consider the inverse image of X under the the complex analytic uniformization map for A

$$\begin{aligned} \pi : \mathbb{C}^g &\longrightarrow A, \\ \frac{1}{N} \Lambda &\longrightarrow A[N], \\ \tilde{X} = \pi^{-1}(X) &\longrightarrow X. \end{aligned}$$

Let $\Delta \subset \mathbb{C}^g$ be a fundamental domain of the lattice Λ . Then $\pi|_{\Delta} : \Delta \longrightarrow A$ is surjective, and any point inside $A[N] \cap X$ comes from a point in $\tilde{X} \cap \frac{1}{N} \Lambda \cap \Delta$, so we have

$$|X \cap A[N]| \leq \left| \tilde{X} \cap \frac{1}{N} \Lambda \cap \Delta \right|.$$

Thus we have reduced the problem to rational points on the complex analytic curve (and real analytic surface) \tilde{X} . Our goal is to apply Theorem 4.1, for which we have to determine \tilde{X}^{alg} . This will be done in two steps:

Claim 1: Each irreducible component of \tilde{X} is transcendental.

First of all, let us observe the following property of plane algebraic curves.

Lemma 5.4. *Let $C \subset \mathbb{R}^2$ be an irreducible plane algebraic curve and let Λ be a lattice in \mathbb{R}^2 . Then either C is a straight line and the intersection $C \cap \Lambda$ is infinite, or $C + \Lambda$ is dense in \mathbb{R}^2 (in the real topology).*

We do not prove this lemma, but it is not very difficult. Since \widetilde{X} is invariant under translations by elements of the lattice, but not dense, the lemma implies that its only possible algebraic components are straight lines having infinite intersection with the lattice. But having such a component would mean that X has an elliptic curve as a component, a contradiction. This proves Claim 1.

Claim 2: \widetilde{X} cannot contain a segment of a real algebraic curve.

Indeed, by analytic continuation this would give us a complex algebraic curve inside \widetilde{X} , contradicting Claim 1.

Thus $\widetilde{X}^{\text{alg}} = \emptyset$. We are therefore ready to apply the Pila–Wilkie theorem to get an upper bound

$$|X \cap A[N]| \leq \left| \widetilde{X} \cap \frac{1}{N} \Lambda \cap \Delta \right| \leq c(X, \varepsilon) N^\varepsilon.$$

(Notice that while \widetilde{X} can a priori have infinitely many components, only finitely many of them may intersect Δ .)

For the lower bound, we may assume without loss of generality that A and X are defined over a number field K . Then if $P \in X \cap A[N]$, all conjugates of P over K lie in $X \cap A[N]$ as well, which implies $|X \cap A[N]| \geq [K(P) : K]$ where $K(P)$ is the extension of K generated by the coordinates of P . The lower bound comes from the fact that torsion points generate extensions of large degree, an old result of Masser [18].

Theorem 5.5. (Masser) *Let A be an abelian variety of dimension g over a number field K . Then there exist $\delta = \delta(g) > 0$ and $c = c(A, K) > 0$ such that if $P \in A[N]$ is of exact order N then*

$$[K(P) : K] \geq cN^\delta.$$

Taking $\varepsilon = \frac{\delta}{2}$ shows that for sufficiently big N , the set $X \cap A[N]$ must be empty.

6. Definable sets in o-minimal structures

In fact, in [24] Pila and Wilkie obtained a much more general result than Theorem 4.1: they estimated the number of rational points of bounded degree on arbitrary definable sets in o-minimal structures over \mathbb{R} .

6.1. O-minimal structures. For our purposes the following definitions will be sufficient; the interested reader might want to consult the main reference on the subject, [6].

Definition 6.1. We define a *structure* on a set R to be a sequence

$$\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3, \dots,$$

where $\mathcal{S}_m \subset \mathcal{P}(R^m)$ is a collection of subsets of R^m . A set $A \subseteq R^m$ such that $A \in \mathcal{S}_m$ is called *definable* in the structure. Further we ask that

(S1) Each \mathcal{S}_m is closed under finite union and complement (and therefore under finite intersection as well): if $A, B \in \mathcal{S}_m$, then $A \cup B \in \mathcal{S}_m$ and $\overline{A} \in \mathcal{S}_m$.

- (S2) The structure is closed under taking Cartesian products: if $A \in \mathcal{S}_m$ and $B \in \mathcal{S}_n$, then $A \times B \in \mathcal{S}_{m+n}$.
- (S3) Equalities are definable in the structure. Namely, the set $\{(x_1, \dots, x_m) \mid x_i = x_j\}$ belongs to \mathcal{S}_m for each $i, j \in \{1, \dots, m\}$.
- (S4) The structure is closed under projections: if $\pi : R^{m+1} \rightarrow R^m$ denotes the projection on the first m coordinates, then for all $A \in \mathcal{S}_{m+1}$ the set

$$\pi(A) = \{\underline{x} \in R^m \mid \exists y \in R, (\underline{x}, y) \in A\}$$

belongs to \mathcal{S}_m .

The last axiom is called *quantifier elimination*.

For a map $f: R^m \rightarrow R^n$ we say that it is *definable* if its graph is definable:

$$\{(\underline{x}, \underline{y}) \in R^m \times R^n \mid \underline{y} = f(\underline{x})\} \in \mathcal{S}_{m+n}.$$

Let us remark that, as was described in the previous sections, Pila's methods involve taking intersections and projections of the initial sets; this shows that definable sets in a structure are a convenient setting for possible generalizations of Pila's results, allowing us to keep track of the regularity of the sets we are working with while we perform operations on them. For our purpose we are particularly interested in the case where the underlying set is \mathbb{R} , and we would like to work with structures over \mathbb{R} that are compatible with its properties as an ordered ring. More precisely, we are going to consider structures where

- The operations $+, \cdot : \mathbb{R}^2 \rightarrow \mathbb{R}$ are definable.
- The singletons $\{x\}$ for all $x \in \mathbb{R}$ are definable.
- The relation $<$ is definable, that is for all m and for all $i, j \in \{1, \dots, m\}$ the set $\{(x_1, \dots, x_m) \mid x_i < x_j\}$ is definable.

From now on we will understand by "structure over \mathbb{R} " a structure over the set $R = \mathbb{R}$ satisfying these additional properties. Note that under these assumptions, all semi-algebraic sets, that is, sets given by equations and inequalities involving polynomials, are definable. It is natural to ask whether semi-algebraic sets themselves form a structure over \mathbb{R} . This is true, the non-trivial part being axiom (S4), a theorem by Tarski and Seidenberg.

All semi-algebraic subsets of \mathbb{R} are of one of the following types:

- \emptyset ;
- open intervals (a, b) with $a, b \in \mathbb{R} \cup \{\pm\infty\}$;
- singletons $\{x\}$ for $x \in \mathbb{R}$;
- finite unions of the previous sets.

Definition 6.2. A structure over \mathbb{R} is called *o-minimal* if the only definable sets in \mathbb{R} are those listed above.

Though this definition seems to concern only sets in dimension 1, it imposes strong regularity conditions on definable sets in higher dimensions; for instance, every definable set is a finite union of smooth manifolds, see [6]. On the other hand, the following examples show that the class of o-minimal structures is very rich and allows numerous strong extensions of Pila's results.

Examples:

- (1) The simplest (and smallest) example of an o-minimal structure is the structure \mathbb{R}_{alg} of semi-algebraic sets.
- (2) Real semi-analytic sets do not form a structure, because a projection of an analytic set is not necessarily analytic, see [1]. Call a set *subanalytic* if it is a projection of a semi-analytic set, and *globally subanalytic* if it is a restriction to \mathbb{R}^m of a subanalytic set inside $\mathbb{P}\mathbb{R}^m$. Van den Dries [5] was, probably, the first one to observe that globally subanalytic sets form an o-minimal structure, denoted by \mathbb{R}_{an} .
- (3) The structures $\mathbb{R}_{\text{alg,exp}}$ and $\mathbb{R}_{\text{an,exp}}$ obtained by extending \mathbb{R}_{alg} and \mathbb{R}_{an} to make the function \exp definable are o-minimal, as shown by Wilkie, Van den Dries and others in the 1990's, see [7], [8] for precise references.
- (4) Peterzil and Starchenko [19] showed definability of the Weierstrass \wp -function. Precisely, denote by Δ the standard fundamental domain for the action of $\text{SL}_2(\mathbb{Z})$ on the upper half plane \mathcal{H} (see Subsection 7.1 for more details) and by $\wp(z, \tau)$ the Weierstrass \wp -function of $\Lambda = \mathbb{Z}\tau + \mathbb{Z}$. Then the function $f(x, y, \tau) = \wp(x\tau + y, \tau)$ on the set $[0, 1] \times [0, 1] \times \Delta$ is definable in $\mathbb{R}_{\text{an,exp}}$. It follows that the j -invariant function restricted to Δ is definable in $\mathbb{R}_{\text{an,exp}}$ as well.

6.2. The theorem of Pila and Wilkie. We can now state the principal theorem of Pila and Wilkie from [24].

Theorem 6.3. (Pila, Wilkie, 2006) *Let a subset $X \subset \mathbb{R}^m$ be definable in some o-minimal structure. Then for every $\varepsilon > 0$ there is a constant $c(X, \varepsilon)$ such that*

$$|X^{\text{tr}}(\mathbb{Q}, T)| \leq c(X, \varepsilon) T^\varepsilon.$$

Applying this to the o-minimal structure \mathbb{R}_{an} , we see that Theorem 4.1 we stated earlier is a special case of this one.

In the sequel we shall often omit for brevity the reference to o-minimality, but will assume it tacitly. By “definable” we shall always mean “definable in some o-minimal structure”.

The main new tool in the proof is an o-minimal version of Gromov’s and Yomdin’s re-parametrization lemma.

Definition 6.4. Let $X \subset \mathbb{R}^m$ be definable. A partial C^r -parametrization of X is a definable function $f : (0, 1)^m \rightarrow X$ which is injective and C^r . It is said to be *bounded* if all its derivatives of order $\leq r$ are bounded. A finite set $\{f_1, \dots, f_k\}$ of (bounded) partial parametrizations of X is called a *(bounded) C^r -parametrization* of X if the union of the images of all the f_i covers all of X .

Existence of a C^r -parametrization (without boundedness condition) is not difficult to establish, see [6], and the principal hassle is boundedness. Gromov [11], improving on a result of Yomdin [31], showed that a semi-algebraic set admits a bounded C^r -parametrization. Pila and Wilkie [24] extended this further to arbitrary definable sets.

Lemma 6.5. (Yomdin–Gromov re-parametrization lemma) *A compact definable set has a bounded C^r -parametrization.*

Moreover, a similar statement holds for a definable family of definable sets.

Thanks to the bounded parametrization, one no longer has to struggle with non-uniformity, as we did in Section 4: all the derivatives are bounded, and one can proceed with the inductive argument almost straightforwardly. This is one of the best examples we know when placing the problem in the correct general context greatly clarifies it, and the proof of a much more general statement appears to be much simpler than that of particular cases.

Later Pila in [22] extended Theorem 6.3 to algebraic points of bounded degree on definable sets. Denote by $X^{\text{tr}}(\mathbb{Q}, d, T)$ the set of points on X of height³ bounded by T and degree (over \mathbb{Q}) bounded by d .

Theorem 6.6. (Pila, 2009) *Let $d \geq 1$ be an integer and let $X \subset \mathbb{R}^m$ be a definable set. Then for all $\varepsilon > 0$ there is a constant $c(X, \varepsilon, d) > 0$ such that*

$$|X^{\text{tr}}(\mathbb{Q}, d, T)| \leq c(X, \varepsilon, d) T^\varepsilon.$$

The idea is to reduce this to \mathbb{Q} -points in a higher dimensional set, precisely, in the symmetric product of d copies of X . If P is a point of degree d over \mathbb{Q} and P_1, \dots, P_d are its conjugates, then the point $(P_1, \dots, P_d) \in X^d$ gives rise to a \mathbb{Q} -rational point on the symmetric product.

7. The André–Oort conjecture

Before discussing another very important application of Pila’s results, we need to recall some facts on modular curves and complex multiplication, the main reference being [15].

7.1. Modular curves. Let E be an elliptic curve over \mathbb{C} . Then there is a complex analytic isomorphism

$$E(\mathbb{C}) \cong \mathbb{C}/\Lambda,$$

where Λ is some lattice in \mathbb{C} . Two elliptic curves E_1 and E_2 are isomorphic if and only if the corresponding lattices Λ_1, Λ_2 are equivalent, that is, if there exists $\alpha \in \mathbb{C}^\times$ such that $\Lambda_2 = \alpha\Lambda_1$. Thus up to isomorphism we can always write $E(\mathbb{C}) = \mathbb{C}/\Lambda$ where Λ is the lattice $\langle 1, \tau \rangle$ generated by 1 and an element τ of the upper half-plane \mathcal{H} . There is an action of $\Gamma(1) := \text{SL}_2(\mathbb{Z})$ on \mathcal{H} given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \tau = \frac{a\tau + b}{c\tau + d},$$

and the modular curve $Y(1) = \Gamma(1)\backslash\mathcal{H}$ of $\Gamma(1)$ -orbits parametrizes isomorphism classes of elliptic curves. One can also describe it more explicitly choosing the following fundamental domain:

$$\Delta = \left\{ \tau \in \mathcal{H}, -\frac{1}{2} \leq \text{Re}(\tau) < \frac{1}{2}, |\tau| > 1 \right\} \cup \left\{ \tau \in \mathcal{H}, |\tau| = 1, -\frac{1}{2} \leq \text{Re}(\tau) \leq 0 \right\}.$$

³We intentionally do not specify what we mean here by “height”; each reader can use her/his favorite definition, the result will always be the same.

Two elliptic curves are isomorphic over \mathbb{C} if and only if they have the same j -invariant, and every complex number is the j -invariant of some elliptic curve over \mathbb{C} . Therefore, the j -invariant defines an analytic isomorphism $\Gamma(1)\backslash\mathcal{H} \cong \mathbb{C}$.

Define the *congruence subgroup*

$$\Gamma_0(N) = \left\{ \gamma \in \mathrm{SL}_2(\mathbb{Z}), \gamma \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \pmod{N} \right\}.$$

The curve $Y_0(N) = \Gamma_0(N)\backslash\mathcal{H}$ parametrizes pairs (E, f) where E is an elliptic curve and f is a cyclic N -isogeny, that is, an isogeny $f : E \rightarrow E'$ (where E' is some other elliptic curve) of degree N with kernel cyclic of order N . We can define a function on $Y_0(N)$ by $j_N(\tau) = j(\tau/N)$. There is an explicit description of $Y_0(N)$ as a plane curve over \mathbb{Q} given by an equation $\Phi_N(x, y) = 0$, where Φ_N is an irreducible polynomial in $\mathbb{Z}[x, y]$ such that $\Phi_N(j, j_N) = 0$.

7.2. Complex multiplication. Let $E = E_\tau$ be an elliptic curve over \mathbb{C} with corresponding lattice $\Lambda = \langle 1, \tau \rangle$. Its endomorphism ring $\mathrm{End}(E)$ contains the ring \mathbb{Z} of rational integers, corresponding to the multiplication-by- n maps for every $n \in \mathbb{Z}$. If $\mathrm{End}(E)$ is strictly bigger than \mathbb{Z} , the curve E is said to have *complex multiplication*, or to be a *CM-curve*. In this case τ is necessarily imaginary quadratic, and $\mathrm{End}(E)$ is an order \mathcal{O}_τ in the imaginary quadratic field $K = \mathbb{Q}(\tau)$. In particular, there is a positive integer f , called the *conductor* of τ , such that $\mathcal{O}_\tau = \mathbb{Z} + f\mathcal{O}_K$ where \mathcal{O}_K is the ring of integers of K . Writing $\mathbb{Q}(\tau) = \mathbb{Q}(\sqrt{d})$ where $d < 0$ is an integer such that $-d$ is square-free, $D = f^2d$ is the discriminant of the order \mathcal{O}_τ , called the *discriminant* of τ . The following is a fundamental result of the theory of complex multiplication:

Theorem 7.1. *Assume E has complex multiplication by the order \mathcal{O}_τ of conductor f . Then $j(\tau)$ is an algebraic integer of degree $h_{f,K}$, the class number of the order \mathcal{O}_τ of K .*

Moreover, $h_{f,K}$ is related in the following way to the class number h_K of K :

$$h_{f,K} = c \cdot h_K f \prod_{p|f} \left(1 - \left(\frac{d}{p} \right) \frac{1}{p} \right), \quad c \in \{1, 1/2, 1/3\}. \quad (7.1)$$

See [4, Section 7] for more details.

7.3. The André–Oort conjecture.

Definition 7.2. A point $P \in \mathbb{C}^2$ is said to be a *CM-point* if $P = (j(\tau_1), j(\tau_2))$ where τ_1 and τ_2 are imaginary quadratic.

Lines of the form $\mathbb{C} \times \{j(\tau)\}$ or $\{j(\tau)\} \times \mathbb{C}$ contain infinitely many CM-points. So does a modular curve $Y_0(N)$, as we have seen it contains all points of the form $(j(\tau), j(\tau/N))$. André's theorem states that all curves inside \mathbb{C}^2 containing infinitely many CM-points must be of one of these types:

Theorem 7.3. (André, 1998) *If $X \subset \mathbb{C}^2$ is an irreducible curve which is not a horizontal or vertical line nor $Y_0(N)$ for some N , then X has finitely many CM-points.*

The general André–Oort conjecture deals with Shimura varieties, an important special case being products of modular curves $Y_0(N_1) \times \dots \times Y_0(N_k)$: in this setting it was proved by Edixhoven in 1998 assuming the GRH for imaginary quadratic fields. We are going to concentrate on the Shimura variety \mathbb{C}^k obtained when $N_1 = \dots = N_k = 1$.

Definition 7.4. (1) A *special point* in \mathbb{C}^k is a point of the form $(j(\tau_1), \dots, j(\tau_k))$ where τ_i 's are imaginary quadratic.

(2) A *special subvariety* of \mathbb{C}^k is a subvariety given by an equation of one of the following forms:

- $x_i = j(\tau)$ for some i , where τ is imaginary quadratic;
- $\Phi_N(x_i, x_j) = 0$ for some i, j .

Special points and special subvarieties are an analogue of torsion points and torsion subvarieties in the statement of the Manin–Mumford conjecture. Now we can state the André–Oort conjecture for \mathbb{C}^k , proved by Pila in [23] using his previous results on rational points in definable sets.

Theorem 7.5. (*André–Oort conjecture*) *Let X be a subvariety of \mathbb{C}^k . Then it has finitely many maximal special subvarieties.*

We are going to sketch Pila’s proof for $k = 2$, when it boils down to proving Theorem 7.3. The ideas are very similar to the ones used for the proof of Manin–Mumford. Consider a map

$$\begin{aligned} \mathcal{H}^2 & \xrightarrow{\pi} \mathbb{C}^2. \\ (\tau_1, \tau_2) & \mapsto (j(\tau_1), j(\tau_2)) \end{aligned}$$

For a non-special irreducible $X \subseteq \mathbb{C}^2$, put $\tilde{X} = \pi^{-1}(X) \cap (\Delta \times \Delta)$. Any CM-point on X comes from some point of \tilde{X} , so it suffices to bound the number of points from \tilde{X} mapping to CM-points in X . Note that for any CM-point P on X , a point $\tilde{P} \in \tilde{X}$ mapping to P is of the form (τ_1, τ_2) where τ_1, τ_2 are imaginary quadratic, so $[\mathbb{Q}(\tilde{P}) : \mathbb{Q}] \leq 4$. We therefore need a bound on the size of $\tilde{X}(\mathbb{Q}, 4, T)$ in terms of T , and so we must check that we can apply Theorem 6.6. The following is a consequence of the already mentioned theorem by Peterzil and Starchenko about the definability of the Weierstrass \wp -function in $\mathbb{R}_{\text{an}, \text{exp}}$:

Claim 1: \tilde{X} is definable.

Moreover, in the same manner as in the proof of Manin–Mumford (an algebraic curve that is not special cannot be $\text{SL}_2(\mathbb{Z})$ -invariant) we obtain

Claim 2: $\tilde{X}^{\text{tr}} = \tilde{X}$.

Assuming these claims, by Theorem 6.6 we get

$$|\tilde{X}(\mathbb{Q}, 4, T)| \leq c(X, \varepsilon) T^\varepsilon. \tag{7.2}$$

Now we are going to bound the number of special points on X from below. For this we can assume that X is defined over some number field L . If $P \in X$ is special, all its conjugates over L are special as well. Bearing in mind that P is a CM-point, $P = (j(\tau_1), j(\tau_2))$ where we can choose τ_1 and τ_2 inside the fundamental domain Δ , so that $\tilde{P} = (\tau_1, \tau_2) \in \tilde{X}$. Therefore

$$[L(P) : L] \geq \max\{[\mathbb{Q}(j(\tau_1)) : \mathbb{Q}], [\mathbb{Q}(j(\tau_2)) : \mathbb{Q}]\}. \quad (7.3)$$

According to Theorem 7.1, for $i = 1, 2$,

$$[\mathbb{Q}(j(\tau_i)) : \mathbb{Q}] = h_{f_i, \mathbb{Q}(\tau_i)},$$

where f_i is the conductor of τ_i . The following theorem tells us that the class number of an imaginary quadratic field cannot be too small:

Theorem 7.6. (Siegel) *Let K be an imaginary quadratic field and fix $\varepsilon' > 0$. Then there is a constant $c = c(\varepsilon') > 0$ such that*

$$h_K \geq c |D_K|^{\frac{1}{2} - \varepsilon'},$$

where D_K is the discriminant of the number field K .

Thanks to formula (7.1), we then get that

$$[\mathbb{Q}(j(\tau_i)) : \mathbb{Q}] \geq c_i |D_i|^{\frac{1}{2} - \varepsilon'} \quad (7.4)$$

for some constant $c_i > 0$, where $D_i = f_i^2 D_{\mathbb{Q}(\tau_i)}$ is the discriminant of τ_i . Moreover, for τ imaginary quadratic of discriminant D lying in the fundamental domain, a quick computation shows that the height $H(\tau)$ of τ is bounded in terms of its discriminant:

$$H(\tau) \leq \sqrt{|D|}. \quad (7.5)$$

Putting (7.3), (7.4) and (7.5) together, we get finally for ε' sufficiently small

$$[L(P) : L] \geq c \max\{H(\tau_1), H(\tau_2)\}^{1-2\varepsilon'} = c H(\tilde{P})^{1-2\varepsilon'}.$$

If there are infinitely many special points on X , the heights $H(\tilde{P})$ can be arbitrarily large, which with the upper bound (7.2) yields the result.

For further reading on the topic we strongly recommend Scanlon's expository texts [29], [30] and Zannier's lecture notes [32].

As mentioned in the preface, these notes do not claim for any kind of exhaustiveness. For instance, we do not speak at all on Heath-Brown's "determinant method", which can be viewed as a " p -adic" version of the Bombieri–Pila method; see [13], [3] and the subsequent work of Browning, Salberger and others.

While Heath-Brown's method is beautiful and powerful, it mainly applies in the algebraic case, and has yet to show its efficiency in the transcendental case, which is the main topic of these notes. Therefore it is left out.

References

- [1] E. Bierstone and P. D. Milman, Semianalytic and subanalytic sets, *Pub. Math. I.H.E.S.* **67** (1988), 5–42. [Zbl 0674.32002](#) [MR 972342](#)
- [2] E. Bombieri and J. Pila, The number of integral points on arcs and ovals, *Duke Math. J.* **59** (1989), 337–357. [Zbl 0718.11048](#) [MR 1016893](#)
- [3] T. D. Browning, D. R. Heath-Brown, and P. Salberger, Counting Rational Points on Algebraic Varieties, *Duke Math. J.* **132** (2006), 545–578. [Zbl 1098.14013](#) [MR 2219267](#)
- [4] D. A. Cox, *Primes of the form $x^2 + ny^2$. Fermat, class field theory and complex multiplication*, Wiley, NY, 1989. [Zbl 0701.11001](#) [MR 1028322](#)
- [5] L. van den Dries, A generalization of the Tarski–Seidenberg theorem, and some nondefinability results, *Bull. Amer. Math. Soc. (N. S.)* **15** (1986), 189–193. [Zbl 0612.03008](#) [MR 854552](#)
- [6] L. van den Dries, *Tame topology and o-minimal structures*, Cambridge Univ. Press, New York, 1998. [Zbl 0953.03045](#) [MR 1633348](#)
- [7] L. van den Dries, A. Macintyre and D. Marker, The elementary theory of restricted analytic fields with exponentiation, *Ann. of Math. (2)* **140** (1994), 183–205. [Zbl 0837.12006](#) [MR 1289495](#)
- [8] L. van den Dries and C. Miller, Geometric categories and o-minimal structures, *Duke Math. J.* **84** (1996), 497–540. [Zbl 0889.03025](#) [MR 1404337](#)
- [9] K. Dörge, Einfacher Beweis des Hilbertschen Irreduzibilitätssatzes, *Math. Ann.* **96** (1927), 176–182. [Zbl 52.0101.01](#) [MR 1512312](#)
- [10] A. Gabrièlov, Projections of semianalytic sets (Russian), *Funkcional. Anal. i Priložen.* **2** (1968), 18–30; *Functional Anal. Appl.* **2** (1968), 282–291. [Zbl 0179.08503](#) [MR 245831](#)
- [11] M. Gromov, Entropy, homology and semialgebraic geometry, Séminaire Bourbaki 1985/86, no. 663., *Astérisque* **145–146** (1987), 225–240. [Zbl 0611.58041](#) [MR 880035](#)
- [12] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, sixth edition (revised by D. R. Heath-Brown and J. H. Silverman, with a foreword by A. Wiles), Oxford University Press, Oxford, 2008. [Zbl 1159.11001](#) [MR 2445243](#)
- [13] D. R. Heath-Brown, The density of rational points on curves and surfaces, *Ann. Math. (2)* **155** (2002), 553–595. [Zbl 1039.11044](#) [MR 1906595](#)
- [14] V. Jarník, Über die Gitterpunkte auf konvexen Kurven, *Math. Z.* **24** (1926), 500–518. [Zbl 51.0153.01](#) [MR 1544776](#)
- [15] S. Lang, *Elliptic functions*, Addison-Wesley, 1973. [Zbl 0316.14001](#) [MR 409362](#)
- [16] S. Lang, *Fundamentals of Diophantine Geometry*, Springer, 1983. [Zbl 0528.14013](#) [MR 715605](#)
- [17] M. Laurent, Équations diophantiennes exponentielles, *Invent. Math.* **78** (1984), 299–327. [Zbl 0554.10009](#) [MR 767195](#)
- [18] D. W. Masser, Small values of the quadratic part of the Néron-Tate height on an abelian variety, *Compositio Math.* **53** (1984), 153–170. [Zbl 0551.14015](#) [MR 766295](#)
- [19] Y. Peterzil and S. Starchenko, Uniform definability of the Weierstrass \wp functions and generalized tori of dimension one, *Selecta Math. (N.S.)* **10** (2004), 525–550. [Zbl 1071.03022](#) [MR 2134454](#)

- [20] F. Petrov, On the number of rational points on a strictly convex curve, *Funktional. Anal. i Prilozhen.* **40** (2006), 30–42, 95; *Funct. Anal. Appl.* **40** (2006), 24–33. [Zbl 1152.11030](#) [MR 2223247](#)
- [21] J. Pila, Rational points on a subanalytic surface, *Ann. Inst. Fourier* **55** (2005), 1501–1516. [Zbl 1121.11032](#) [MR 2172272](#)
- [22] J. Pila, On the algebraic points of a definable set, *Selecta Math. N. S.* **15** (2009), 151–170. [Zbl 1218.11068](#) [MR 2511202](#)
- [23] J. Pila, O-minimality and the André–Oort conjecture for \mathbb{C}^n , *Ann. Math. (2)* **173** (2011), 1779–1840. [Zbl 1243.14022](#) [MR 2800724](#)
- [24] J. Pila and A. J. Wilkie, The rational points of a definable set, *Duke Math. J.* **133** (2006), 591–616. [Zbl 1217.11066](#) [MR 2228464](#)
- [25] A. Plagne, A uniform version of Jarník’s theorem, *Acta Arith.* **87** (1999), 255–267. [Zbl 0924.11084](#) [MR 1668570](#)
- [26] M. Raynaud, Courbes sur une variété abélienne et points de torsion, *Invent. Math.* **71** (1983), 207–233. [Zbl 0564.14020](#) [MR 688265](#)
- [27] M. Raynaud, Sous-variétés d’une variété abélienne et points de torsion in *Arithmetic and geometry*, Vol. I, 327–352, Progr. Math. 35, Birkhäuser Boston, Boston, MA, 1983. [Zbl 0581.14031](#) [MR 717600](#)
- [28] P. Sarnak and S. Adams, Betti numbers of congruence groups (with an appendix by Z. Rudnik), *Israel J. Math.* **88** (1994), 31–72. [Zbl 0843.11027](#) [MR 1303490](#)
- [29] Th. Scanlon, A proof of the André–Oort conjecture via mathematical logic [after Pila, Wilkie and Zannier]. Séminaire Bourbaki: Vol. 2010/2011. Exposés 1027–1042. *Astérisque* **348** (2012), Exp. No. 1037, ix, 299–315. [Zbl 1271.14030](#) [MR 3051200](#)
- [30] Th. Scanlon, Counting special points: Logic, diophantine geometry, and transcendence theory. *Bull. Amer. Math. Soc.* **49** (2012) 51–71. [Zbl 06000374](#) [MR 2869007](#)
- [31] Y. Yomdin, C^k -resolution of semialgebraic mappings, *Israel J. Math.* **57** (1987), 301–317. [Zbl 0641.54037](#) [MR 889980](#)
- [32] U. Zannier, *Some Problems of Unlikely Intersections in Arithmetic and Geometry* (with appendices by D. Masser), Ann. Math. Studies 181, Princeton, 2012. [Zbl 1246.14003](#) [MR 2918151](#)
- [33] <http://vieuxgirondin.wordpress.com/2011/07/29/a-mean-value-theorem/>

Received 11 January, 2015