

COUNTING SPECIAL POINTS: LOGIC, DIOPHANTINE GEOMETRY AND TRANSCENDENCE THEORY

THOMAS SCANLON

ABSTRACT. We expose a theorem of Pila and Wilkie on counting rational points in sets definable in o-minimal structures and some applications of this theorem to problems in diophantine geometry due to Masser, Peterzil, Pila, Starchenko, and Zannier.

1. INTRODUCTION

Over the past decade and a half starting with Hrushovski's proof of the function field Mordell-Lang conjecture [12], some of the more refined theorems from model theory in the sense of mathematical logic have been applied to problems in diophantine geometry. In most of these cases, the technical results underlying the applications concern the model theory of fields considered with some additional distinguished structure and the model theoretic ideas fuse algebraic model theory, the study of algebraic structures with a special emphasis on questions of definability, and stability theory, the development of abstract notions of dimension, dependence, classification, *et cetera* for the purpose of analyzing the class of models of a theory. Over this period, there has been a parallel development of the model theory of theories more suited for real analysis carried out under the rubric of o-minimality, but this theory did not appear to have much to say about number theory. Some spectacular recent theorems demonstrate the error of this impression.

In the paper [28], Pila presents an unconditional proof of a version of the so-called André-Oort conjecture about algebraic relations amongst the j -invariants of elliptic curves with complex multiplication using a novel technique coming from model theory. This proof of the André-Oort conjecture comes on the heels of re-proofs of the Manin-Mumford conjecture (or, Raynaud's theorem [32]) by Pila and Zannier [29] (and then extended by Peterzil and Starchenko [22]) and a proof of a remarkable theorem due to Masser and Zannier [19] about simultaneous torsion for pairs of points on families of elliptic curves all of which employ Pila's method.

Each of these theorems is a beautiful and precise instantiation of the vague principle that algebraic relations on sets of arithmetically interesting points of geometric origin must be explained geometrically. As such, these results deserve their own survey independent from their proofs. However, while I agree that these theorems and their ilk are of intrinsic interest, there are two reasons why I shall focus on their proofs. First, the proof strategy employed coming as it does from the model theory of real geometry is the most striking common feature of these results. Secondly, the algebro-geometric overhead required for an accurate explication of how these problems fit into an overarching network of theorems and conjectures in

diophantine geometry would overwhelm the remainder of this account and mask the fundamentally classical nature of Pila's approach to these problems.

With this preamble about what I shall not discuss, let me say what will appear in this paper. We begin in Section 2 with a sketch of the Pila-Zannier proof in a simplified case of algebraic relations on roots of unity. This sketch will not reveal the full strength of the method as the theorem in this special case has been known for many decades and has innumerable proofs. However, this proof shares the architecture of the proofs of the other geometrically more complicated theorems. In Section 3 we discuss the general theory of o-minimality. Section 4, the technical heart of this survey, concerns the Pila-Wilkie theorem on counting rational points in definable sets. With Section 5 we return to the diophantine problems, completing the proof sketch from Section 2 and then discussing some of the other theorems mentioned in the abstract.

On the same day that I was asked to prepare these notes for the Current Events Bulletin, I was asked to speak in the Bourbaki seminar about Pila's proof of the André-Oort conjecture. Shortly thereafter, I was asked to give a lecture series to the experts on André-Oort about the same topic. Foolhardily, I concluded that due to the similarity of these presentations, I would need only prepare one set of notes and accepted all three invitations. As a matter of fact, while there are some points of contact, these three sets of notes are radically different. The reader interested in an exposition of Pila's proof of the André-Oort conjecture pitched to the general mathematician should consult my notes for the Bourbaki seminar [35] while the reader who wishes to read a detailed précis of these proofs should read my notes for the Luminy lectures [34]. Even better, because the original papers [24] and [28] are well written and contain extensive introductions, the reader should go straight to the source.

2. PILA-ZANNIER ARGUMENT FOR MULTIPLICATIVE GROUP

We shall go into more detail about the proofs of the more sophisticated theorems announced in the introduction later in the paper, but let us sketch the method in a simple example for which, admittedly, many other proofs are known (see, for example, [32, 13]). The following theorem is a special case of the Manin-Mumford conjecture and was already proven by Mann [18] before the Manin-Mumford conjecture proper was enunciated.

Theorem 2.1. *Let $n \in \mathbb{Z}_+$ be a positive integer and let $\mathbf{G} = (\mathbb{C}^\times)^n$ be the n^{th} Cartesian power of the multiplicative group of the complex numbers. Let $G(x_1, \dots, x_n) \in \mathbb{C}[x_1, \dots, x_n]$ be a polynomial in n variables. Then the set*

$$\{(\zeta_1, \dots, \zeta_n) \in \mathbf{G} : \text{each } \zeta_i \text{ is a root of unity and } G(\zeta_1, \dots, \zeta_n) = 0\}$$

is a finite union of cosets of subgroups of \mathbf{G} .

The Pila-Zannier argument in this case proceeds by observing that we have an analytic covering map $E : \mathbb{C}^n \rightarrow \mathbf{G}$ given by $(z_1, \dots, z_n) \mapsto (e^{2\pi iz_1}, \dots, e^{2\pi iz_n})$ and that relative to this covering, $\zeta = (\zeta_1, \dots, \zeta_n)$ is tuple of roots of unity if and only if there is some $a \in \mathbb{Q}^n$ for which $E(a) = \zeta$. Thus, we may convert the problem of studying algebraic equations in roots of unity into the problem of understanding rational solutions to the transcendental equation $G(E(z)) = 0$.

In this form, we have not achieved much yet as sets defined by general complex analytic equations can be arbitrarily complicated. However, it is not necessary to consider E on all of \mathbb{C}^n . We could restrict E to a fundamental domain

$$D := \{z = (z_1, \dots, z_n) \in \mathbb{C}^n : 0 \leq \operatorname{Re}(z_i) < 1 \text{ for each } i\}$$

obtaining a function $\tilde{E} := E \upharpoonright D : D \rightarrow \mathbb{G}$, then it is still the case that $\zeta \in G$ is a tuple of roots of unity if and only if there is some $a \in \mathbb{Q}^n \cap D$ with $\tilde{E}(a) = \zeta$. The advantage of this move is that \tilde{E} lives in logically well-behaved structure while E does not. That is, even though from the point of view of complex analysis, the map E is just about the best function one could hope to study, from the point of view of mathematical logic it has a very complicated theory as the kernel of E is the set \mathbb{Z}^n and with the ring structure inherited from \mathbb{C}^n the theory of this structure suffers from Gödel incompleteness phenomena. On the other hand, using the real and imaginary part functions to identify \mathbb{C} with \mathbb{R}^2 , the function \tilde{E} is definable in the structure $\mathbb{R}_{\exp} := (\mathbb{R}, +, \times, \exp, \leq, 0, 1)$ of the real field considered together with the real exponential function.

Why is this important? The theory of \mathbb{R}_{\exp} is *o-minimal*, which, technically, means that every definable subset of the universe is a finite union of points and intervals, but which means in practice that the definable sets in any number of variables admit a geometric structure theory. In particular, the set

$$\tilde{X} := \{z \in D : G(\tilde{E}(z)) = 0\}$$

is such a definable set. We have transformed the problem of describing those n -tuples of roots of unity ζ for which $G(\zeta) = 0$ to the problem of describing the intersection $\mathbb{Q}^n \cap \tilde{X}$ which at this level of generality may appear to be even more hopeless than the original problem as we know the problem of describing the rational solutions to algebraic equations is notoriously intractable (and is conjecturally impossible algorithmically [20]). However, if we punt on the problem for algebraic equations, then for the remaining transcendental equation we can give numerical bounds.

More precisely, for any set $Y \subseteq \mathbb{R}^m$ definable in some o-minimal expansion of the real field, we define the algebraic part of Y , Y^{alg} , to be the union of all connected, positive dimensional semi-algebraic sets (that, definable using Boolean combinations of polynomial inequalities) contained in Y . The counting theorem of Pila and Wilkie asserts that there are sub-exponentially many rational points in the transcendental part of Y , $Y \setminus Y^{\text{alg}}$. That is, if we define

$$N(Y, t) := \#\left\{\left(\frac{a_1}{b_1}, \dots, \frac{a_n}{b_n}\right) \in Y \setminus Y^{\text{alg}} : (\forall i \leq n) |a_i| \leq t, 0 < b_i < t, a_i \in \mathbb{Z}, b_i \in \mathbb{Z}\right\}$$

then for each $\epsilon > 0$ there is a constant $C = C_\epsilon$ so that $N(Y, t) \leq Ct^\epsilon$ for all $t \geq 1$.

To use the Pila-Wilkie bound one must understand the set Y^{alg} and while this is far from a trivial problem, it has a geometric character and can be solved in cases of interest. In our proof sketch of Theorem 2.1 it follows from a theorem of Ax on a differential algebraic version of Schanuel's conjecture [3] that \tilde{X}^{alg} is the union of finitely sets each defined by affine equations with rational coefficients which under \tilde{E} are transformed into translates of algebraic subgroups of \mathbb{G} . We complete the argument by playing the Pila-Wilkie bound against lower bounds from Galois theory, but we delay the details until Section 5.

Each of the theorems in diophantine geometry proven using this method follows the general outline sketched above, though, of course, the individual steps tend to be more complicated as there is work involved in proving that the requisite covering map is definable in some o-minimal structure, the determination of the algebraic part of the relevant definable sets may be difficult, and one needs appropriate Galois theoretic or analytic number theoretic results for the lower bounds. While each step implicates some beautiful mathematics, it is the invocation of the counting principle for definable sets in o-minimal structures which gives this method its special character. As such, we shall focus this exposé on the ideas of definability in o-minimal structures and the counting theorem.

3. INTRODUCTION TO O-MINIMALITY

O-minimality is not well-known to the general mathematician for at least a couple of reasons. First, it owes its existence and most of its development to mathematical logic and for sociological reasons having to do with logic's place at the boundary between mathematics and philosophy, the basics of first-order logic are not as widely known amongst mathematicians as are the basics of algebra, analysis and geometry. Secondly, the name itself, while technically accurate in that it expresses that all of the one-dimensional structure is reducible to the order and ties the subject to other parts of model theory, masks the fundamental nature of the subject which is a general but tame and geometric theory of real analysis. In his text [39], van den Dries argues that o-minimality may be a realization of the theory of *topologie modérée* proposed by Grothendieck in [11] in which topology and real analysis follow geometric intuitions. While I do not subscribe to this thesis in the strong form that o-minimality is *the* realization of *topologie modérée*, o-minimality certainly fits the bill for a geometric theory of real analysis.

What follows is a condensed introduction to the theory of o-minimality. The book [39] develops the general theory especially as it relates to o-minimal structures on the real numbers from a geometric point of view. The reader may wish to consult the lecture notes from the recent thematic program on o-minimality at the Fields Institute for a more recent account or Wilkie's Bourbaki notes [43] for a fuller survey. The foundational papers by Pillay, Steinhorn and Knight [30, 15, 31] remain vital.

Definition 3.1. By an *o-minimal structure* we mean a structure in the sense of first-order logic $(R, <, \dots)$ where $<$ is a total order on R and the ellipses refer to some extra relations, functions and constants so that each definable (using parameters) subset of R is a finite union of points and intervals.

Remark 3.2. While in the applications we have in mind, the underlying ordered set is the set of real numbers with its usual ordering, the proofs of the counting principles pass through an analysis of parametrizations of definable sets in more general o-minimal structures. That is, the compactness theorem of first-order logic when used in the context of an o-minimal theory allows for a kind of nonstandard analysis which converts simple existence and finiteness results into uniformity theorems.

As noted in Remark 3.2 there are good reasons beyond the historical accident that logicians isolated the notion of o-minimality for treating o-minimal structures as structures in the sense of first-order logic, but it is possible to make sense of

o-minimality without explicit reference to logic. Let us give an alternate definition of an o-minimal structure.

Definition 3.3. By an *o-minimal structure* we mean a nonempty totally ordered set $(R, <)$ given together with a Boolean algebras \mathcal{D}_n of subsets of R^n for each $n \in \mathbb{Z}_+$ so that

1. each \mathcal{D}_n is closed under the natural action of $\text{Sym}(n)$ induced by permutations of coordinates,
2. each singleton set $\{a\}$ belongs to \mathcal{D}_1 for $a \in R$,
3. if $X \in \mathcal{D}_n$ and $Y \in \mathcal{D}_m$, then $X \times Y \in \mathcal{D}_{n+m}$,
4. if $\pi : R^{n+1} \rightarrow R^n$ is the projection onto the first n coordinates and $X \in \mathcal{D}_{n+1}$, then the image of X under π belongs to \mathcal{D}_n ,
5. $\{(a, b) \in R^2 : a < b\} \in \mathcal{D}_2$,
6. $\{(a, b) \in R^2 : a = b\} \in \mathcal{D}_2$, and
7. every set in \mathcal{D}_1 is a finite union of singletons and intervals. That is, sets of the form $(-\infty, a) := \{x \in R : x < a\}$, $(a, b) := \{x \in R : a < x < b\}$, and $(b, \infty) := \{x \in R : b < x\}$ for some $a, b \in R$.

We refer to the sets in \mathcal{D}_n as the *definable subsets of R^n* .

It is a routine matter to check that these two definitions of o-minimality are essentially the same. The closure conditions 1. - 4. on the class of definable sets in Definitions 3.3 and the initial requirement that each \mathcal{D}_n be a Boolean algebra correspond to the syntactic operations of logical Boolean operations, permutation of variables, naming of parameters, conjunction of formulae with disjoint variables, and existential quantification. Condition 5. corresponds to definability of the ordering while condition 6. asserts the definability of equality. It is with condition 7. that we insist upon o-minimality. The choice of a first-order signature in Definition 3.1 corresponds to specifying a set of generators for the class of definable sets in the sense of Definition 3.3.

While this second presentation permits one to work with o-minimal structures without ever thinking about first-order logic, I contend that it is a mistake to do so. In practice, one establishes the definability of specific sets or conditions by exhibiting a definition. For example, one can show that if $X \subseteq R^n$ is a definable set in some ordered structure $(R, <, \dots)$, then so is the closure \bar{X} of X . Of course, this can be done by manipulating definable sets using projections and setwise Boolean operations, but the first-order formula describing \bar{X} is transparently the usual definition of the closure:

$$\begin{aligned} (a_1, \dots, a_n) \in \bar{X} &\iff (\forall x_1) \cdots (\forall x_n) (\forall y_1) \cdots (\forall y_n) \left[\bigwedge_{i \leq n} x_i < a_i < y_i \right. \\ &\quad \left. \rightarrow (\exists z_1) \cdots (\exists z_n) ((z_1, \dots, z_n) \in X \ \& \ \bigwedge_{i \leq n} x_i < z_i < y_i) \right] \end{aligned}$$

One of the characteristic features of mathematical arguments using model theory is the way in which properties of definable sets in one structure may be deduced from arguments performed in logically equivalent structures through a kind of transfer principle. While these arguments are possible without the logical formalism, they are much more natural with it.

While there are some degenerate o-minimal structures whose underlying orders are discrete, we shall insist that an o-minimal structure be densely ordered without endpoints and for our applications the underlying ordered set is the set of real numbers with its usual ordering.

It is not hard to see that $(\mathbb{R}, <)$, the set of reals just with its order, is an o-minimal structure. It takes a little more work to demonstrate the $(\mathbb{R}, <, +)$, the set of real numbers considered as an ordered group, is an o-minimal structure. This latter structure is the basis of piecewise linear geometry and of tropical geometry. While the finiteness and uniformity properties of PL-geometry and tropical geometry are easy enough to demonstrate directly, the o-minimality of this underlying structure puts these results into context. That $(\mathbb{R}, +, \times, <, 0, 1)$, the set of real numbers considered as an ordered field, is o-minimal is a consequence of Tarski's theorem on elementary geometry (first proved in 1929, but only published in 1948 [38]) and is the basis of semi-algebraic geometry.

Wilkie proved that $\mathbb{R}_{\text{exp}} := (\mathbb{R}, <, +, \times, \text{exp})$, the real field considered together with the real exponential function is o-minimal [42]. Indeed, he proved a stronger theorem: the real field considered together with all *Pfaffian functions* is o-minimal where we say that a function $f : \mathbb{R} \rightarrow \mathbb{R}$ is Pfaffian if there is a finite sequence of functions $f_1, \dots, f_n = f$ so that for each $i \leq n$ we have $f'_i = G_i(x, f_1, \dots, f_n)$ for some polynomial $G_i(y_0, y_1, \dots, y_i) \in \mathbb{R}[y_0, y_1, \dots, y_n]$. The order one linear differential equation satisfied by exp exhibits the exponential function as a Pfaffian function. In later work, Speissegger [37] showed that one may take any o-minimal structure and adjoin all Pfaffian functions relative to that structure and thereby obtain a new o-minimal structure. In fact, Speissegger's theorem is more general in that he allows for the adjunction of so-called Rolle leaves to definable vector fields, but as this generalization is not germane to our applications, we omit the details.

In another direction, as a complement to their theorems on p -adic analytic functions, Denef and van den Dries [7] proved the o-minimality of \mathbb{R}_{an} , the real field considered together with all *restricted analytic functions*. That is, for each $n \in \mathbb{Z}_+$ and each power series

$$f := \sum_{\alpha \in \mathbb{N}^n} f_\alpha x_1^{\alpha_1} \cdots x_n^{\alpha_n} \in \mathbb{R}[[x_1, \dots, x_n]]$$

which converges on the unit box we are given a function symbol \tilde{f} to be interpreted as

$$\tilde{f}(a_1, \dots, a_n) := \begin{cases} \sum f_\alpha a_1^{\alpha_1} \cdots a_n^{\alpha_n} & \text{if } -1 \leq a_i \leq 1 \text{ for each } i \leq n \\ 0 & \text{otherwise} \end{cases}$$

Combining these two expansions to form $\mathbb{R}_{\text{an,exp}} := \mathbb{R}_{\text{an}}(\text{exp})$, the real field considered with all real analytic functions and the real exponential function. The o-minimality of this structure was first established by van den Dries and Miller [41] and then the structure of its definable sets was more thoroughly explored by van den Dries, Macintyre, and Marker [40]. It is this structure which is relevant to the diophantine applications mentioned in the abstract.

For some perspective, one should note that many expansions of the real field by global real analytic functions, for example $(\mathbb{R}, +, \times, <, \sin)$, are *not* o-minimal as the zero set $\{x \in \mathbb{R} : \sin(x) = 0\}$ is infinite but discrete. Likewise, for a sufficiently general smooth function $f : \mathbb{R} \rightarrow \mathbb{R}$ the expansion $(\mathbb{R}, +, \times, <, f \upharpoonright$

$[0,1]$ is not o-minimal. Since o-minimal structures have a tame geometry and we know that functions arising from real analysis tend to be wild, these examples ought not to be so surprising. On the other hand, it follows from work of Rolin, Speissegger and Wilkie [33] that there is no ultimate o-minimal structure on the real numbers. Indeed, for any \mathcal{C}^∞ function $f : \mathbb{R} \rightarrow \mathbb{R}$ one can find two other functions $g : \mathbb{R} \rightarrow \mathbb{R}$ and $h : \mathbb{R} \rightarrow \mathbb{R}$ so that $f = g + h$ but both structures $(\mathbb{R}, +, \times, \leq, g)$ and $(\mathbb{R}, +, \times, \leq, h)$ are o-minimal. Consequently, if we understand o-minimality as the formalization of the concept of *topologie modérée*, then there is not a single structure which captures all of tame geometry.

From the fact that a specific structure is o-minimal, one may immediately deduce strong finiteness properties of the definable subsets of the line. For example, if $f : \mathbb{R} \rightarrow \mathbb{R}$ is definable in some o-minimal structure on the real numbers, then the zero set of f consists of finitely many points and finitely many intervals. If f falls into some natural class for which an identity principle holds, as when f is real analytic, then we see that if f is not identically zero, it has only finitely many zeros. This applies, for instance, to the case that f is built via finitely many applications of sums, differences, products and compositions from the identity function, scalar multiplication and the real exponential function.

The simplicity of definable sets coming explicitly from o-minimality's definition, while often unexpected, is the basis for a far deeper structure theory of definable sets in any number of variables. For example, it follows very easily from o-minimality that in a sufficiently rich o-minimal structure, if one has a definable family of definable nonempty sets, then this family admits a definable choice function. Let us make this statement precise and prove it in detail.

Definition 3.4. We say that the o-minimal structure $(R, <, \dots)$ is *sufficiently rich* if it has at least the structure of an ordered abelian group $(R, +, -, 0, 1, <, \dots)$ with at least one positive element called 1 named by a constant.

It is an easy exercise to show that if an o-minimal structure is sufficiently rich, then the underlying group is divisible.

Definition 3.5. Let $(R, <, \dots)$ be an o-minimal structure. By a *definable family* of definable subsets of R^n we mean a definable set $X \subseteq R^n \times R^m$ for some positive natural number m where we regard R^m as parametrizing the family where to $b \in R^m$ we associate the set $X_b := \{a \in R^n : \langle a, b \rangle \in X\}$. Sometimes, we write a definable family of definable sets as $\{X_b\}_{b \in B}$ where $B \subseteq R^m$ is a definable set containing the projection of X to R^m .

Definition 3.6. By a *definable choice function* for the definable family $\{X_b\}_{b \in B}$ we mean a definable function $f : B \rightarrow R^n$ for which $f(b) \in X_b$ for each $b \in B$.

With these definitions in place, let us prove the existence of definable choice functions.

Proposition 3.7. Let $(R, +, -, 0, 1, <, \dots)$ be a sufficiently rich o-minimal structure and $\{X_b\}_{b \in B}$ a definable family of non-empty definable subsets of R^n . Then there is a definable choice function $f : B \rightarrow R^n$ for this family.

Proof. We work by induction with the case of $n = 0$ being trivial. For the inductive case of $n + 1$, let $b \in B$ and consider the set

$$A(b) := \{y \in R : (\exists x \in R^n) \langle x, y \rangle \in X_b\}$$

The family $\{A(b)\}_{b \in B}$ is a definable family of nonempty definable subsets of R . Let us define a choice function h for this family. For $b \in B$, if $A(b) = R$, define $h(b) := 0$. If $A(b) \neq R$, then by o-minimality it has a finite non-empty boundary. Consider the least boundary point y . If $y \in A(b)$, define $h(b) := y$. Otherwise, one of three situations must obtain: $(-\infty, y) \subseteq A(b)$, $(y, \infty) \subseteq A(b)$ or $(y, z) \subseteq A(b)$ for z the second boundary point of $A(b)$. Define $h(b) := y - 1$ in the first case, $h(b) := y + 1$ in the second case, and $h(b) := \frac{y+z}{2}$ in the last case.

For $b \in B$ define

$$Z_b := \{x \in R^n : \langle x, h(b) \rangle \in X_b\}$$

Then $\{Z_b\}_{b \in B}$ is a definable family of nonempty subsets of R^n which has a definable choice function $g : B \rightarrow R^n$ by induction. Our desired choice function $f : B \rightarrow R^{n+1}$ is then given by $b \mapsto \langle g(b), h(b) \rangle$. \square

The fundamental theorem of o-minimality is the cell decomposition theorem which asserts, roughly, that every definable set may be partitioned into finitely many *cells*, definable sets which are definably homeomorphic to balls possibly in a lower dimensional space. Since this theorem is so important for all of the work in o-minimality and is invoked if only implicitly throughout the proof of the Pila-Wilkie theorem on counting rational points, we shall go into detail.

Let us begin by giving a precise definition of *cell*. As we define this notion, we shall define the dimension of a cell.

Definition 3.8. Let $(R, <, \dots)$ be an o-minimal structure. We define the class of cells in R^n by recursion on n and for each cell X we define $\dim(X)$. The set R^0 is a singleton which is the only cell in R^0 . We define $\dim(R^0) := 0$. If $X \subseteq R^n$ is a cell and $f : X \rightarrow R$ is a definable, continuous function, then the graph of f ,

$$\Gamma(f)_X := \{(x, y) \in R^n \times R : x \in X \text{ \& } f(x) = y\}$$

is a cell with $\dim(\Gamma(f)) := \dim(X)$. If $g : X \rightarrow R$ is another definable, continuous function for which $(\forall x \in X) f(x) < g(x)$, then the interval

$$(f, g)_X := \{(x, y) \in R^n \times R : x \in X \text{ \& } f(x) < y < g(x)\}$$

is a cell and $\dim((f, g)_X) := \dim(X) + 1$. Likewise,

$$(-\infty, f)_X := \{(x, y) \in R^n \times R : x \in X \text{ \& } y < f(x)\}$$

and

$$(g, \infty)_X := \{(x, y) \in R^n \times R \mid x \in X \text{ \& } g(x) < y\}$$

are cells of dimension $\dim(X) + 1$.

Let us specialize to the case of cells in $R = R^1$ for a moment. A definable function (indeed, any function) $f : R^0 \rightarrow R$ is given by choosing a single point $a \in R$. Via the natural identification $R^0 \times R^1 = R$, we see that $\Gamma(f)_{R^0} = \{a\}$. Likewise, if $g : R^0 \rightarrow R$ is another definable function for which $(\forall x \in R^0) f(x) < g(x)$, then taking b to be the sole value of g we have $a < b$ and again with respect to the natural identification of $R^0 \times R^1$ with R , the interval $(f, g)_{R^0}$ is simply the usual interval (a, b) . That is, cells in R are singletons and open intervals, possibly unbounded. To say that $(R, <, \dots)$ is o-minimal is precisely the same as to say that every definable set $X \subseteq R$ may be expressed as a finite union of cells. The cell decomposition theorem asserts that this property of definable sets in one space generalizes to any dimension.

Theorem 3.9. *Let $(R, <, \dots)$ be an o-minimal structure. Given any $m \in \mathbb{Z}_+$ and any finite sequence $X_1, \dots, X_n \subseteq R^m$ of definable subsets of R^m there is a finite sequence of cells $C_1, \dots, C_k \subseteq R^m$ which partitions R^m and whose restriction to each X_ℓ also partitions X_ℓ . That is, $C_i \cap C_j = \emptyset$ for $i \neq j$, $R^m = \bigcup_{j=1}^k C_j$, and for each $\ell \leq m$ we have $X_\ell = \bigcup_{\{j \leq k : C_j \cap X_\ell \neq \emptyset\}} C_j$.*

In the course of the proof of Theorem 3.9 one shows that definable functions are very regular. The monotonicity theorem says that for any definable function $f : R \rightarrow R$ possibly after removing finitely many points one may partition the domain into finitely many intervals so that on each interval f is continuous and either constant or strictly monotone. In general, the piecewise continuity theorem says that if $f : R^n \rightarrow R$ is a definable function, then one may decompose the domain into finitely many cells so that the restriction of f to each cell is continuous. If the structure $(R, <, \dots)$ includes at least the structure of an ordered field, then it makes sense to speak of the derivative of a function. In this case, for any $k \in \mathbb{Z}_+$ one may choose a cell decomposition of the domain so that f is C^k on each cell. Likewise, in the cell decomposition theorem itself, one may take the functions defining the cells to have any prescribed degree of smoothness. Unfortunately, it is not the case that one may always take the cells to be defined by analytic functions, but in many cases of interest, for example in $\mathbb{R}_{\text{an,exp}}$, one may take the defining functions to be real analytic.

It is hard to overstate the strength of the geometric consequences of the cell decomposition theorem and its refinements. For example, it implies a kind of infinitesimal rigidity on the topology of definable sets living in a definable family.

It is a fairly easy consequence of the cell decomposition theorem applied to the total space of a definable family that given a definable family $\{X_b\}_{b \in B}$ of definable sets, the cells required for the cell decompositions of the various fibres also vary in definable families. It follows from this uniformity theorem that at least when the underlying ordered set is the set of real numbers with its usual ordering that the topology of the sets in a definable family is rigid.

Proposition 3.10. *If $\{X_b\}_{b \in B}$ is a definable family of definable sets in some o-minimal structure on the real numbers (with the usual ordering) then there are only finitely many homomorphism types represented in the family.*

As a corollary of Proposition 3.10 we obtain a theorem of Khovanski [14] on fewnomials. To be fair, while the theorem on fewnomials which we shall discuss is logically a consequence of Proposition 3.10 both temporally and intellectually it is prior. Khovanski's work on fewnomials inspired much of the development of theory of o-minimality and many of his specific results underly Wilkie's proof of the o-minimality of \mathbb{R}_{exp} . Moreover, the argument we outline below is patterned on Khovanski's own proof through the passage from polynomials of indeterminate degree to exponential polynomials.

Theorem 3.11. *For fixed integers k and n there are only finitely many homomorphism types amongst the following sets*

$$\{(a_1, \dots, a_n) \in (\mathbb{R}_+)^n : \sum_{i=1}^k f_i a_1^{m_{i,1}} \cdots a_n^{m_{i,n}} = 0\}$$

as (f_1, \dots, f_k) ranges through \mathbb{R}^k and m ranges through the k by n matrices with natural number coordinates.

To prove Theorem 3.11 we observe that it suffices show that there are only finitely many homomorphism types even if we allow m to range through $M_{k \times n}(\mathbb{R})$ rather than merely $M_{k \times n}(\mathbb{N})$. The above family of semialgebraic sets may be embedded into the following \mathbb{R}_{exp} -definable family.

$$\{(a, f, m) \in (\mathbb{R}_+)^n \times (\mathbb{R}^k \times (\mathbb{R}^n)^k) : \sum_{i=1}^k f_i \prod_{j=1}^n \exp(m_{i,j} \ln(a_j)) = 0\}$$

The finiteness of the number of homomorphism types is now a special case of Proposition 3.10.

4. COUNTING RATIONAL POINTS IN O-MINIMAL DEFINABLE SETS

The key technical result behind the Pila-Wilkie theorem on counting rational points may be seen as, in some sense, a dual version of the cell decomposition theorem in that rather than concluding that a general definable set may be pieced together from definable subsets each defined in a very simple way it is shown that a general definable set may be covered by finitely many definable sets each of which is parametrized by a unit ball via functions with small derivatives. Before we go into detail about this technical result on parametrizations, let us go into some more detail about the counting theorem itself and then explain how an appropriate parametrization theorem could yield these bounds.

Definition 4.1. The usual multiplicative height function $H : \mathbb{Q} \rightarrow \mathbb{N}$ is defined by $H(0) := 0$ and $H(\frac{a}{b}) := \max\{|a|, |b|\}$ when $a, b \in \mathbb{Z} \setminus \{0\}$ and $\gcd(a, b) = 1$. We extend H to a function, still denoted by H , on \mathbb{Q}^n by $H(x_1, \dots, x_n) := \max\{H(x_i) : i \leq n\}$. This is not the standard projective height, but it works well for the purposes of our counting problems. Given any subset $X \subseteq \mathbb{R}^n$ and a number $t \geq 1$ we define

$$X(\mathbb{Q}, t) := \{a \in \mathbb{Q}^n : a \in X \ \& \ H(a) \leq t\}$$

and define $N(X, t) := \#X(\mathbb{Q}, t)$ to be the number of points in $X(\mathbb{Q}, t)$.

If X happens to contain all of \mathbb{Q}^n , then $N(X, t)$ is asymptotic to a constant times t^{2n} . This simple observation combined with the even simpler remark that \mathbb{R}^n itself is definable in any o-minimal structure on \mathbb{R} shows that one cannot hope to show that $N(X, t)$ grows more slowly than any power of t for a general set $X \subseteq \mathbb{R}^n$ definable in some o-minimal expansion of $(\mathbb{R}, <)$. Somewhat less trivial considerations show that some further restrictions are required. For example, if $X \subseteq \mathbb{R}^n$ happens to contain the graph of a polynomial with integer coefficients $f : \mathbb{R}^{n-1} \rightarrow \mathbb{R}$, then $N(X, t)$ will grow at least on the order of $t^{2(n-1)/d}$ where d is the degree of f . Of course, there are more general algebraic varieties which have many rational points and if X should contain one of these, it, too, will have many rational points. Thus, to have any hope of proving a bound on $N(X, t)$ for general X definable in some o-minimal structure on \mathbb{R} , we must exclude those algebraic sets which have too many rational points. We achieve this by excluding *all* semi-algebraic sets.

Definition 4.2. We say that a set $X \subseteq \mathbb{R}^n$ is *semi-algebraic* if it is definable in the structure $(\mathbb{R}, +, \times, <)$ of the real numbers considered as an ordered field. More

concretely, a set is semi-algebraic if it is a finite Boolean combination of sets defined by conditions of the form $f(x_1, \dots, x_n) > 0$ where $f \in \mathbb{R}[X_1, \dots, X_n]$ is polynomial with real coefficients in n variables. Given a set $X \subseteq \mathbb{R}^n$ we define X^{alg} , the *algebraic part* of X , to be the union of all sets Z where $Z \subseteq X$ is semi-algebraic, connected, and has dimension at least one. We define X^{trans} , the *transcendental part* of X , to be $X \setminus X^{\text{alg}}$.

There is no reason to expect the algebraic part of X to be semi-algebraic itself even if X is definable in some particularly well behaved o-minimal structure on the real numbers. For example, consider the following set which is definable in \mathbb{R}_{exp} .

$$X := \{(x, y, z) \in \mathbb{R}^3 : x > 0 \text{ \& } z = x^y = \exp(y \ln(x))\}$$

The set X has dimension two, being the graph of a definable continuous function on $\mathbb{R}_+ \times \mathbb{R}$, but it does not contain any two dimensional semi-algebraic sets. On the other hand, its algebraic part consists of the union of the following countably infinite collection of one-dimensional semi-algebraic sets

$$\{(x, y, z) \in \mathbb{R}^3 : x > 0 \text{ \& } y = \frac{n}{m} \text{ \& } x^n = z^m\}$$

as $\frac{n}{m}$ ranges through the rational numbers. Thus, in this case X^{alg} is a properly infinite union of semi-algebraic sets.

With these definitions in place we may state the counting theorem from [24].

Theorem 4.3. *Let $X \subseteq \mathbb{R}^n$ be a subset of some Cartesian power of the real numbers which is definable in some o-minimal structure on \mathbb{R} . Then for each $\epsilon > 0$ there is some constant $C = C_\epsilon$ so that for $t \geq 1$ one has $N(X^{\text{trans}}, t) \leq Ct^\epsilon$.*

Remark 4.4. Strengthenings of Theorem 4.3 are known. The proof of Theorem 4.3 passes through a proof a uniform version in which X is allowed to vary in a definable family and the bound is shown to hold for a bigger set than simply X^{trans} . For purposes of the application of Theorem 4.3 to the André-Oort conjecture and some related problems it is necessary to count algebraic points of small degree rather than merely rational points. These bounds may be deduced from the bounds for rational points [26].

Remark 4.5. Examples have been constructed showing that in general one cannot hope for better universal bounds. However, one might hope that if X is definable in a particularly nice way, then the bounds may be strengthened. In the strongest forms, these strengthenings assert that certain transcendental equations have no non-obvious algebraic solutions. I have in mind conjectures along the lines of Schanuel's conjecture on the transcendence of the exponential function that if $\alpha_1, \dots, \alpha_n$ are complex numbers which are \mathbb{Q} -linearly independent then the transcendence degree of the field $\mathbb{Q}(\alpha_1, \dots, \alpha_n, e^{\alpha_1}, \dots, e^{\alpha_n})$ is at least n [17], André's conjectures on G-functions [1], and the Kontsevich-Zagier conjectures on periods [16]. While these conjectures are inaccessible to contemporary techniques, Wilkie's conjecture about sets definable using the real exponential function may be within reach.

Conjecture 4.6. *Let $X \subseteq \mathbb{R}^n$ be a subset of a Cartesian power of \mathbb{R} definable in \mathbb{R}_{exp} . There are constants C and K depending only on X so that for $t \geq 1$ we have $N(X^{\text{trans}}, t) \leq C(\log t)^K$.*

The proof Theorem 4.3 and the proofs [27] of partial results towards Conjecture 4.6 rely on a general geometric theorem about parametrizations of definable sets, referred to as a dual form of cell decomposition in the first paragraph of this section, and a linear algebraic argument to show that the rational points on the image of a ball under a sufficiently smooth function must lie in a small number of algebraic hypersurfaces.

Definition 4.7. Let $(R, +, \cdot, <, \dots)$ be an o-minimal structure on an ordered field, $X \subseteq R^n$ be a definable set of dimension k in some Cartesian power of R and $r \in \mathbb{Z}_+$ be a positive integer. We say that $\phi = (\phi_1, \dots, \phi_n) : (0, 1)^k \rightarrow R^n$ is a partial r -parametrization of X if

- ϕ is definable,
- the range of ϕ is contained in X , and
- for each $i \leq n$ and multi-index $\alpha = (\alpha_1, \dots, \alpha_k) \in \mathbb{N}^k$ with $|\alpha| = \sum \alpha_i \leq r$ we have $|\frac{\partial^{|\alpha|} \phi_i}{\partial x_1^{\alpha_1} \dots \partial x_k^{\alpha_k}}(x)| \leq 1$ for all $x \in (0, 1)^k$.

By an r -parametrization of X we mean a finite set S of partial r -parametrizations of X for which X is covered by the ranges of the functions in S .

The main theorem about parametrizations is that r -parametrizations exist for every $r \in \mathbb{Z}_+$ and every sufficiently bounded definable set in an o-minimal structure.

Theorem 4.8. *Let $(R, +, \times, <, \dots)$ be an o-minimal structure expanding an ordered field, $X \subseteq [-1, 1]^n$ be a definable subset of the unit n -cube in R for some $n \in \mathbb{Z}_+$, and $r \in \mathbb{Z}_+$ be a positive integer. Then X admits an r -parametrization.*

For the intended applications, we work in an o-minimal structure on \mathbb{R} . What, then, is the point of the greater generality? The proof of Theorem 4.8, even in the case that the underlying ordered field is the field of real numbers, makes essential use of parametrizations of definable sets in more general o-minimal structures through a kind of nonstandard analysis. That is, by proving even individual instances of the parametrization theorem, say, in general o-minimal structures, one may deduce via the compactness theorem of first-order logic that the theorem holds uniformly in definable families. From the point of view of the ultimate theorem, these compactness arguments are hidden, but they are crucial for the proofs.

Theorem 4.8 generalizes a theorem of Yomdin on the existence of r -parametrizations for real semi-algebraic sets [45, 44] and its proof follows Gromov's version of the proof in the semi-algebraic setting [10]. Of course, for purposes of Theorem 4.3 in which we count only rational points in the transcendental part of X , the parametrization theorem for semialgebraic sets on its own does not help.

The logical structure of the proof of Theorem 4.8 is similar to that of the cell decomposition theorem (Theorem 3.9). For both of these theorems, the one-dimensional case itself is an immediate consequence of the definition of o-minimality, but to carry out the induction using the definable choice functions given by Proposition 3.7 one performs a concurrent induction showing that definable functions have strong regularity properties. For the cell decomposition theorem, this takes the form of the monotonicity theorem in dimension one and the piecewise continuity theorem in higher dimensions. For the parametrization theorem, this takes the form of a *reparameterization* theorem in all dimensions and a strengthening of the reparameterization theorem in dimension one.

To say what is meant by reparameterization, we need the definition of *strongly bounded*.

Definition 4.9. Let $(R, +, \cdot, 0, 1, <, \dots)$ be an o-minimal expansion of an ordered field. We say that the set $X \subseteq R^n$ is *strongly bounded* if there is a natural number $N \in \mathbb{N}$ for which $X \subseteq [-N, N]^n$. We say that a function $f : Y \rightarrow R^n$ is strongly bounded if its range is strongly bounded.

Definition 4.10. Let $(R, +, \cdot, 0, 1, <, \dots)$ be an o-minimal expansion of an ordered field. Given a positive integer $r \in \mathbb{Z}_+$ and a definable function $f = (f_1, \dots, f_n) : (0, 1)^m \rightarrow R^n$ we say that an r -parameterization S of the open unit box $(0, 1)^m$ is an r -reparameterization of f if for each $\alpha \in \mathbb{N}^m$ with $|\alpha| \leq r$, $\phi \in S$ and $j \leq n$ the function $f_j \circ \phi$ is r -times differentiable and $\frac{\partial^{|\alpha|}(f_j \circ \phi)}{\partial x_1^{\alpha_1} \dots \partial x_m^{\alpha_m}}$ is strongly bounded.

Remark 4.11. In the case that the underlying field is simply \mathbb{R} , then bounded and strongly bounded have the same meaning. The meanings diverge only for non-archimedean fields. In the real case, if $f : (0, 1)^m \rightarrow \mathbb{R}^n$ is bounded and sufficiently smooth, then any r -parameterization of $(0, 1)^m$ is an r -reparameterization. The utility of the concept of a reparameterization is only seen through the nonstandard analytic arguments.

The key auxiliary result in the proof of Theorem 4.8 is the reparameterization theorem.

Theorem 4.12. *Let $(R, +, \cdot, 0, 1, <, \dots)$ be an o-minimal expansion of an ordered field, $m \in \mathbb{Z}_+$, $n \in \mathbb{Z}_+$, and $r \in \mathbb{Z}_+$ be three positive integers. If $f : (0, 1)^m \rightarrow R^n$ is strongly bounded, then there exists an r -reparameterization of f . Moreover, if $m = n = 1$, the reparameterization S may be chosen so that for each $\phi \in S$ either ϕ or $\phi \circ f$ is a polynomial with strongly bounded coefficients.*

Theorem 4.12 and thereby the full Theorem 4.8 are established via a constructive and concrete argument in the base case and then via an inductive argument relying heavily upon definable choice and the cell decomposition theorems in higher dimensions.

Let us explain now how Theorem 4.3 follows from Theorem 4.8. We start with a set $X \subseteq \mathbb{R}^n$ definable in some o-minimal expansion of the real numbers and a number $\epsilon > 0$. Breaking X into the 2^n pieces $X_\tau := \{(a_1, \dots, a_n) \in X : |a_i^{\tau_i}| \leq 1 \text{ for all } i \leq n\}$ as τ ranges over $\{\pm 1\}^n$, it suffices to estimate $N(X_\tau, t)$ for each such τ . Since the map $x \mapsto \frac{1}{x}$ does not effect our height function, we may replace each X_τ with its image under $(x_1, \dots, x_n) \mapsto (x_1^{\tau_1}, \dots, x_n^{\tau_n})$ and thereby we may assume that $X \subseteq [-1, 1]^n$. From Theorem 4.8 we know that for each $r \in \mathbb{Z}_+$ the set X admits an r -parameterization. We shall choose r depending on ϵ so that the existence of such an r -parameterization implies that $N(X, t) \leq Ct^\epsilon$.

At this point, the argument follows the lines of other constructive arguments bounding numbers of rational solutions and is similar in spirit to Bombieri's proof of the Mordell conjecture [5]. The key result is the following proposition whose proof is ultimately embedded in the paper [4].

Proposition 4.13. *For $m, n, d \in \mathbb{N}$ with $m < n$ there are numbers $r \in \mathbb{Z}_+$ and $\epsilon = \epsilon(m, n, d)$ and $C = C(m, n, d)$ in \mathbb{R}_+ so that for any C^r function $\phi : (0, 1)^m \rightarrow \mathbb{R}^n$ with range X and $t \geq 1$ the set $X(\mathbb{Q}, t)$ is contained in at most Ct^ϵ hypersurfaces of degree d and $\epsilon(m, n, d) \rightarrow \infty$ as $d \rightarrow \infty$.*

The proof of Proposition 4.13 requires some nontrivial but elementary combinatorial estimates and some careful but again elementary analytic considerations, but ultimately it is based on a simple, but ubiquitous in the theory of diophantine approximations, observation: if an integer has absolute value less than one, it is zero.

With Proposition 4.13 in place, Theorem 4.3 follows by induction: those exceptional hypersurfaces which have full dimension intersection with X are part of X^{alg} and those which intersect X in a lower dimensional set contribute little to $N(X^{\text{trans}}, t)$ by induction.

5. DIOPHANTINE APPLICATIONS

In general, it is not true that if $X \subseteq \mathbb{R}^n$ is definable in some o-minimal structure on the real numbers that $(X^{\text{trans}})(\mathbb{Q})$ is finite. For example, if X is the graph of the function $x \mapsto 2^x$, then its algebraic part is empty, but for each integer a we have $\langle a, 2^a \rangle \in X(\mathbb{Q})$. However, in some cases of independent number theoretic interest the upper bounds of Theorem 4.3 may be played against lower bounds coming from Galois theory. In the introduction to this paper, we sketched a version of this argument due to Pila and Zannier to show that algebraic relations amongst roots of unity may always be explained by multiplicative dependencies. In this section we shall explain some of the more sophisticated results proven using variations of this method

Let us return to the argument sketched in the introduction giving a few more details to complete the proof. Recall that we wish to prove that for $G(x_1, \dots, x_n) \in \mathbb{C}[x_1, \dots, x_n]$ a polynomial over the complex numbers in n variables the set

$$X := \{(\zeta_1, \dots, \zeta_n) \in (\mathbb{C}^\times)^n : G(\zeta_1, \dots, \zeta_n) = 0 \text{ \& \textit{each } \zeta_i \text{ is a root of unity}}\}$$

is a finite union of cosets of groups. We observed that if we define

$$D := \{z \in \mathbb{C} : 0 \leq \text{Re}(z) < 1\}$$

and $\tilde{E} : D^n \rightarrow (\mathbb{C}^\times)^n$ by

$$(z_1, \dots, z_n) \mapsto (e^{2\pi iz_1}, \dots, e^{2\pi iz_n})$$

then via the usual interpretation of \mathbb{C} as \mathbb{R}^2 using the real and imaginary part functions, the function \tilde{E} is definable in \mathbb{R}_{exp} and the set X is the image under \tilde{E} of the set $\tilde{X}(\mathbb{Q})$ where $\tilde{X} := \{(z_1, \dots, z_n) \in D^n : G(e^{2\pi iz_1}, \dots, e^{2\pi iz_n}) = 0\}$ is an \mathbb{R}_{exp} -definable set.

Before we can apply Theorem 4.3 to give even numerical bounds on the distributions of the points in $\tilde{X}(\mathbb{Q})$, we need to compute \tilde{X}^{alg} . The key to this computation is Ax's function field version of the Schanuel Conjecture [3].

Theorem 5.1. *If $\gamma_1(t), \dots, \gamma_n(t) \in t\mathbb{C}[[t]]$ are power series over the complex numbers with no constant term which are linearly independent over \mathbb{Q} , then the transcendence degree over $\mathbb{C}(t)$ of the field $\mathbb{C}(t, \gamma_1(t), \dots, \gamma_n(t), \exp(\gamma_1(t)), \dots, \exp(\gamma_n(t)))$ is at least n .*

It follows from Theorem 5.1 that if $\gamma : (0, 1) \rightarrow \tilde{X}$ were a semi-algebraic curve, then the components of γ would satisfy a nontrivial linear dependence over \mathbb{Q} .

This alone is not enough for the determination of \tilde{X}^{alg} . Rather, we now use another property of definable sets in o-minimal structures on the real numbers: every countable definable set is finite. For each positive $k \leq n$ we consider the set M_k of maximal k -dimensional affine spaces: affine spaces V (translates of vector subspaces) of dimension k for which $\dim(V \cap \tilde{X}) = k$ but for which there is some point $a \in (V \cap \tilde{X})$ for which is no $k + 1$ -dimensional affine space W which meets \tilde{X} near a in dimension $k + 1$ set. It is not hard to see that relative to the usual representations of affine spaces via affine equations, each of the sets M_k is definable. It takes a little more work using properties of the covering map \tilde{E} to see that every element of M_k is defined over \mathbb{Q} . Hence, each M_k is countable and therefore finite. Combining this argument with Theorem 5.1, we see that $\tilde{X}^{\text{alg}} = \bigcup_{k=1}^n \bigcup_{H \in M_k} H$. For the algebraic part, it is now clear that $\tilde{X}^{\text{alg}}(\mathbb{Q})$ is a finite union of cosets of groups (intersected with $[0, 1)^n$).

To complete the proof, we must show that $(\tilde{X}^{\text{trans}})(\mathbb{Q})$ is finite using the bounds from Theorem 4.3. For this we need a reduction: we may assume that G is an irreducible polynomial defined over some a number field K . If you are comfortable with basic algebraic geometry, this reduction is standard and quite easy and will be given in the parenthetical sentences to follow. Otherwise, take this point as given or just follow the argument in the case when G is in fact a polynomial over a number field.

(For the reduction, observe that Theorem 2.1 is equivalent to the apparent generalization where the hypersurface defined by G is replaced by a general subvariety. For a variety $Y \subseteq \mathbb{A}_{\mathbb{C}}^n$, if

$$Z := \overline{Y(\mathbb{C}) \cap \{(\zeta_1, \dots, \zeta_n) \in (\mathbb{C}^\times)^n : \text{each } \zeta_i \text{ is a root of unity}\}}$$

then $Y(\mathbb{C})$ and $Z(\mathbb{C})$ meet the n -tuples of roots of unity in the same set. So, we may assume $Y = Z$. It then follows from Lagrange interpolation that Y is defined over the algebraic numbers as it contains a Zariski dense set of algebraic points. Since only finitely many equations are required to define Y , it is, in fact, defined over a number field.)

Let us now estimate $N(\tilde{X}, t)$. Suppose that $z \in \tilde{X}(\mathbb{Q}, t)$. We can write $z = (\frac{a_1}{b_1}, \dots, \frac{a_n}{b_n})$ where each a_i and b_i is an integer, $0 \leq a_i < b_i \leq t$ and $(a_i, b_i) = 1$ (and $b_i = 1$ if $a_i = 0$). Exponentiating, we have $G(e^{2\pi i \frac{a_1}{b_1}}, \dots, e^{2\pi i \frac{a_n}{b_n}}) = 0$. Let $L := K((e^{2\pi i \frac{a_1}{b_1}}, \dots, e^{2\pi i \frac{a_n}{b_n}}))$ be the field obtained by adjoining the coordinates of $\tilde{E}(z)$ to K . Since G has coefficients in K , for any automorphism $\sigma : L \rightarrow L$ over K , we have $G(\sigma(\tilde{E}(z))) = 0$. Since $e^{2\pi i \frac{a_j}{b_j}}$ is a primitive b_j^{th} root of unity, we know that $\sigma(e^{2\pi i \frac{a_j}{b_j}}) = e^{2\pi i \frac{a'_j}{b_j}}$ for some integer a'_j with $0 \leq a'_j < b_j$. Moreover, it follows from basic Galois theory that the orbit of $e^{2\pi i \frac{a_j}{b_j}}$ under the Galois group of L over K has cardinality at least $\varphi(b_j)/[K : \mathbb{Q}]$ where φ is Euler's totient function given by $\varphi(n) := \#(\mathbb{Z}/n\mathbb{Z})^\times$. A simple computation shows that for any constant $C > 0$ and number $\epsilon < 1$ we have $\varphi(n) > Cn^\epsilon$ for $n \gg 0$. Putting all these observations together, we see that if $t \in \mathbb{Z}_+$ and $N(\tilde{X}^{\text{trans}}, t) > N(\tilde{X}^{\text{trans}}, t - 1)$, then $N(\tilde{X}, t) \geq \frac{1}{[K:\mathbb{Q}]} \varphi(t)$. For $t \gg 0$ this would violate Theorem 4.3 with $\epsilon < 1$. Hence, there must

be some t for which every element of $(\tilde{X}^{\text{trans}})(\mathbb{Q})$ has height at most t . That is, this set is finite.

For the other applications we have mentioned, the statements are intrinsically more complicated as they refer to more sophisticated geometries and the proofs are correspondingly more involved. However, the fundamental strategies are the same.

Let us consider the theorem of Masser and Zannier [19] about torsion on elliptic curves. They consider the family of elliptic curves presented in their affine Legendre form where E_λ is defined by the affine planar equation $y^2 = x(x-1)(x-\lambda)$ for $\lambda \in \mathbb{C} \setminus \{0, 1\}$. From theory of elliptic curves, E_λ considered together with the point at infinity has a unique structure of an algebraic group with that point at infinity as the identity. For a fixed complex number a we might consider the set of λ for which the point $(a, \sqrt{a(a-1)(a-\lambda)})$ is torsion in the group $E_\lambda(\mathbb{C})$. It is not hard to see that for $a = 0$ or $a = 1$, then these points are always torsion. On the other hand, for every other a there are only countably many λ for which this point is torsion in $E_\lambda(\mathbb{C})$. Nevertheless, computing the rational functions which define the multiplication by n map on E_λ it is fairly easy to show that for any such a there will be infinitely many λ for which $(a, \sqrt{a(a-1)(a-\lambda)})$ is torsion. Masser and Zannier address the question: if we consider two number a and b , for how many λ are $(a, \sqrt{a(a-1)(a-\lambda)})$ and $(b, \sqrt{b(b-1)(b-\lambda)})$ both torsion in $E_\lambda(\mathbb{C})$? In the special case of $a = 2$ and $b = 3$ they given an answer.

Theorem 5.2. *There are only finitely many complex numbers λ for which*

$$P_\lambda := (2, \sqrt{2(2-\lambda)})$$

and

$$Q_\lambda := (3, \sqrt{6(3-\lambda)})$$

are torsion in $E_\lambda(\mathbb{C})$.

Remark 5.3. The proof of Theorem 5.2 applies perfectly well to any two numbers a and b for which the points

$$P_\lambda^a := (a, \sqrt{a(a-1)(a-\lambda)})$$

and

$$Q_\lambda^b := (b, \sqrt{b(b-1)(b-\lambda)})$$

are linearly independent over \mathbb{Z} in the group $E_\lambda(\mathbb{Q}(\lambda))$.

The proof of Theorem 5.2 follows the pattern of the proof of Theorem 2.1 we have outlined above. For each elliptic curve E_λ , the the theory of analytic uniformizations gives a complex analytic covering map $\pi_\lambda : \mathbb{C} \rightarrow E_\lambda(\mathbb{C})$. As with the usual exponential function, this covering is not definable in any o-minimal expansion of the real numbers. However, if we restrict π_λ to a fundamental domain, it is. Moreover, at the cost of treating π_λ as simply a real analytic function, we may normalize the fundamental domain so that the domain of π_λ is the square $[0, 1) \times [0, 1)$ and the map π_λ is a group homomorphism when $[0, 1)$ is given the usual wrap around additive structure. With some work, one can show that the two variable (or, really, four real variable) function $(\lambda, z) \mapsto \pi_\lambda(z)$ is definable in

$\mathbb{R}_{an,exp}$ relative to the usual interpretation of \mathbb{C} in \mathbb{R} and when z is restricted to $[0, 1) \times [0, 1)$. Masser and Zannier then study the set

$$\begin{aligned} \tilde{X} &:= \{(x_1, y_1, x_2, y_2) \in [0, 1)^4 : (\exists \lambda) \pi_\lambda(x_1, y_1) = (2, \sqrt{2(2-\lambda)}) \\ &\quad \& \pi_\lambda(x_2, y_2) = (3, \sqrt{6(3-\lambda)})\} \end{aligned}$$

Visibly, \tilde{X} is definable in $\mathbb{R}_{an,exp}$ and it is not hard to see that the rational points on \tilde{X} all come from λ for which P_λ and Q_λ are simultaneously torsion. Transcendence results about the Weierstraß \wp -function are used in place of Ax's theorem to show that \tilde{X}^{alg} is empty and a theorem of David [6] about the degree of the field extension required to define elliptic curves with elements of specified order plays the rôle of the calculation of the degree of a cyclotomic extensions.

It bears noting that the published sketch of Theorem 5.2 avoids an explicit reference to definability in o-minimal structures as Pila had proved a provisional version of Theorem 4.3 for subanalytic surfaces without invoking the theory of o-minimality [25]. On the other hand, due to the work of Peterzil and Starchenko [23] on the uniform definability of theta functions in $\mathbb{R}_{an,exp}$, it follows that the question of simultaneous torsion in families of higher dimensional abelian varieties may be analyzed via these methods.

Finally, let us close with Pila's proof of the André-Oort conjecture for modular curves. We shall introduce the André-Oort conjecture via the classical theory of complex elliptic curves. Unlike most other approaches to this problem where one might (or might not) define the terms using complex analysis but then address the questions with a more number theoretic theory, Pila's proof appeals directly to the complex analytic presentation of the problem.

As we observed above, for every elliptic curve E over the complex numbers, one can find a complex analytic surjective group homomorphism $\pi : \mathbb{C} \rightarrow E(\mathbb{C})$. The kernel of \mathbb{C} is a lattice which after making a linear change of variables we may express as $\ker \pi = \mathbb{Z} \oplus \mathbb{Z}\tau$ for some complex number $\tau \in \mathfrak{h} := \{z \in \mathbb{C} : \text{Im}(z) > 0\}$. Conversely, for any $\tau \in \mathfrak{h}$, the complex analytic group $E_\tau(\mathbb{C}) := \mathbb{C}/(\mathbb{Z} + \mathbb{Z}\tau)$ is complex analytically isomorphic to a complex algebraic curve with an algebraic group structure, which we shall continue to denote by E_τ . From the general theory of covering spaces, it is not hard to see that the endomorphisms of the elliptic curve E_τ correspond to complex numbers μ for which $\mu(\mathbb{Z} + \mathbb{Z}\tau) \leq \mathbb{Z} + \mathbb{Z}\tau$. A short computation shows that for most choices of τ , the number μ gives an endomorphism only when μ is an integer. On the other hand, if τ satisfies a quadratic equation over \mathbb{Q} , then there will be some endomorphisms not coming from \mathbb{Z} . This is the reason why elliptic curves whose endomorphism rings are strictly larger than \mathbb{Z} are said to have *complex multiplication* or to be *CM*.

There is an analytic function $j : \mathfrak{h} \rightarrow \mathbb{C}$ having the property that $E_\tau(\mathbb{C})$ and $E_\sigma(\mathbb{C})$ are isomorphic as elliptic curves if and only if $j(\tau) = j(\sigma)$. We refer to the value $j(\tau)$ as the *j-invariant* of the elliptic curve E_τ . Let us say that a complex number ζ is a *special point* if it is the *j-invariant* of an elliptic curve with complex multiplication. By the above discussion, we see that a number is special if and only if it is the value of the analytic *j*-function on a quadratic imaginary number. The André-Oort conjecture in this case predicts the form of the algebraic subvarieties $X \subseteq \mathbb{A}_{\mathbb{C}}^n$ of affine n -space which contain a Zariski dense set of n -tuples of special points. Specializing to the case of $n = 2$, it proposes a solution to the question

of for which polynomials $g(x, y) \in \mathbb{C}[x, y]$ are there infinitely many pairs $(\zeta, \bar{\zeta})$ of special points for which $g(\zeta, \bar{\zeta}) = 0$? This case was solved early in the investigations around the André-Oort conjecture, first assuming the Riemann Hypothesis by Edixhoven [8] and then unconditionally by André [2].

Clearly, if ζ is a special point, then the set algebraic varieties $\{\zeta\} \times \mathbb{A}_{\mathbb{C}}^1$ and $\mathbb{A}^1 \times \{\zeta\}$ contain Zariski dense sets of special points as does the whole plane $\mathbb{A}_{\mathbb{C}}^2$. It follows from the general theory of coverings, that for each $n \in \mathbb{Z}_+$ there is a polynomial $P_n(x, y) \in \mathbb{C}[x, y]$ for which the function $\tau \mapsto P_n(j(n\tau), j(\tau))$ is identically zero. From this presentation, it is clear that the curve defined by the vanishing of P_n contains a Zariski dense set of special points for if τ is quadratic imaginary, then so is $n\tau$ and *vice versa*. The André-Oort conjecture (for the j -line) says that these are the only algebraic varieties other than points which can contain a Zariski dense set of special points.

Theorem 5.4 (Pila). *Let $X \subseteq \mathbb{A}_{\mathbb{C}}^n$ be an irreducible algebraic subvariety of affine n -space over the complex numbers. Suppose that the set*

$$\{(\zeta_1, \dots, \zeta_n) \in X(\mathbb{C}) : \text{each } \zeta_i \text{ is the } j\text{-invariant of a CM-elliptic curve}\}$$

is Zariski dense in X , then X is defined by equations of the form $P_m(x_i, x_j) = 0$ and $x_k = \zeta$ for ζ a special point.

The proof of Theorem 5.4 follows our by now familiar pattern. First, Pila observes that j restricted to a fundamental domain is definable in $\mathbb{R}_{\text{an,exp}}$ by work of Peterzil and Starchenko [21]. He then moves from a study of X to that of \tilde{X} , the inverse image of $X(\mathbb{C})$ via j (or really, the map $(z_1, \dots, z_n) \mapsto (j(z_1), \dots, j(z_n))$) restricted to its fundamental domain, which is a definable set in $\mathbb{R}_{\text{an,exp}}$. He then must determine \tilde{X}^{alg} and does so using considerations of the action of the modular group showing that the algebraic part comes from the pre-images of finitely many varieties of the desired form. At this point, the goal is to show that if X does not already have the desired form, then there are only finitely many quadratic imaginary points in \tilde{X}^{trans} . The counting theorem, Theorem 4.3, applied to rational points, but Pila deduces the same kinds of bounds for algebraic points of bounded degree [26]. Thus, for any $\epsilon > 0$ there is some constant C for which the number of quadratic imaginary points of height at most t in \tilde{X}^{trans} is at most Ct^ϵ . As in the proof of Theorem 2.1, he reduces to the case that X is defined over a number field and observes that if there are special points coming from \tilde{X}^{trans} , then all of their Galois conjugates are also in this set. At this point, he estimates the size of these orbits from below using Siegel's theorem on the growth of the class number [36] to find that for $\epsilon < 2$ one has a lower bound of Ct^ϵ thus contradicting the upper bound from the counting theorem.

Remark 5.5. Theorem 5.4 had been proven previously by Edixhoven and Yafaev [9] under the assumption of the Generalized Riemann Hypothesis for quadratic imaginary fields. Their proof shares the same kind strategy at the end: find upper bounds geometrically and lower bounds via Galois theory and analytic number theory.

Remark 5.6. The paper in which the proof of Theorem 5.4 appears [28] includes proofs of theorems in the direction of the Pink-Zilber conjectures. On the other hand, while many parts of this argument succeed when applied to other Shimura

varieties, some steps are incomplete. For example, it is known that the analytic covering maps for the moduli spaces of principally polarized abelian varieties are definable in $\mathbb{R}_{\text{an},\text{exp}}$ (again, after suitable restriction) [23] and it seems plausible that the arguments employed to determine the algebraic parts of inverse images of algebraic varieties by Cartesian powers of j should work for these maps, too, but to date no one has carried out the details. More importantly, the lower bounds on the size of the Galois orbits of the special points are not yet known unconditionally.

REFERENCES

- [1] Yves André. *G-functions and geometry*. Aspects of Mathematics, E13. Friedr. Vieweg & Sohn, Braunschweig, 1989.
- [2] Yves André. Finitude des couples d’invariants modulaires singuliers sur une courbe algébrique plane non modulaire. *J. Reine Angew. Math.*, 505:203–208, 1998.
- [3] James Ax. On Schanuel’s conjectures. *Ann. of Math. (2)*, 93:252–268, 1971.
- [4] E. Bombieri and J. Pila. The number of integral points on arcs and ovals. *Duke Math. J.*, 59(2):337–357, 1989.
- [5] Enrico Bombieri. The Mordell conjecture revisited. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, 17(4):615–640, 1990.
- [6] Sinnou David. Points de petite hauteur sur les courbes elliptiques. *J. Number Theory*, 64(1):104–129, 1997.
- [7] J. Denef and L. van den Dries. p -adic and real subanalytic sets. *Ann. of Math. (2)*, 128(1):79–138, 1988.
- [8] Bas Edixhoven. Special points on the product of two modular curves. *Compositio Math.*, 114(3):315–328, 1998.
- [9] Bas Edixhoven and Andrei Yafaev. Subvarieties of Shimura varieties. *Ann. of Math. (2)*, 157(2):621–645, 2003.
- [10] M. Gromov. Entropy, homology and semialgebraic geometry. *Astérisque*, (145-146):5, 225–240, 1987. Séminaire Bourbaki, Vol. 1985/86.
- [11] Alexandre Grothendieck. Esquisse d’un programme. In *Geometric Galois actions, 1*, volume 242 of *London Math. Soc. Lecture Note Ser.*, pages 5–48. Cambridge Univ. Press, Cambridge, 1997. With an English translation on pp. 243–283.
- [12] Ehud Hrushovski. The Mordell-Lang conjecture for function fields. *J. Amer. Math. Soc.*, 9(3):667–690, 1996.
- [13] Ehud Hrushovski. The Manin-Mumford conjecture and the model theory of difference fields. *Ann. Pure Appl. Logic*, 112(1):43–115, 2001.
- [14] A. G. Khovanskii. *Fewnomials*, volume 88 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 1991. Translated from the Russian by Smilka Zdravkovska.
- [15] Julia F. Knight, Anand Pillay, and Charles Steinhorn. Definable sets in ordered structures. II. *Trans. Amer. Math. Soc.*, 295(2):593–605, 1986.
- [16] Maxim Kontsevich and Don Zagier. Periods. In *Mathematics unlimited—2001 and beyond*, pages 771–808. Springer, Berlin, 2001.
- [17] Serge Lang. *Introduction to transcendental numbers*. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1966.
- [18] Henry B. Mann. On linear relations between roots of unity. *Mathematika*, 12:107–117, 1965.
- [19] David Masser and Umberto Zannier. Torsion anomalous points and families of elliptic curves. *C. R. Math. Acad. Sci. Paris*, 346(9-10):491–494, 2008.
- [20] B. Mazur. Questions of decidability and undecidability in number theory. *J. Symbolic Logic*, 59(2):353–371, 1994.
- [21] Ya’acov Peterzil and Sergei Starchenko. Uniform definability of the Weierstrass \wp functions and generalized tori of dimension one. *Selecta Math. (N.S.)*, 10(4):525–550, 2004.
- [22] Ya’acov Peterzil and Sergei Starchenko. Around pila-zannier: the semiabelian case. *preprint*, 2009.
- [23] Ya’acov Peterzil and Sergei Starchenko. Definability of restricted theta functions and families of abelian varieties. *preprint*, 2010.
- [24] J. Pila and A. J. Wilkie. The rational points of a definable set. *Duke Math. J.*, 133(3):591–616, 2006.
- [25] Jonathan Pila. Rational points on a subanalytic surface. *Ann. Inst. Fourier (Grenoble)*, 55(5):1501–1516, 2005.

- [26] Jonathan Pila. On the algebraic points of a definable set. *Selecta Math. (N.S.)*, 15(1):151–170, 2009.
- [27] Jonathan Pila. Counting rational points on a certain exponential-algebraic surface. *Ann. Inst. Fourier (Grenoble)*, 60(2):489–514, 2010.
- [28] Jonathan Pila. O-minimality and the André-Oort conjecture for C^n . *Ann. of Math. (2)*, (to appear).
- [29] Jonathan Pila and Umberto Zannier. Rational points in periodic analytic sets and the Manin-Mumford conjecture. *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl.*, 19(2):149–162, 2008.
- [30] Anand Pillay and Charles Steinhorn. Definable sets in ordered structures. I. *Trans. Amer. Math. Soc.*, 295(2):565–592, 1986.
- [31] Anand Pillay and Charles Steinhorn. Definable sets in ordered structures. III. *Trans. Amer. Math. Soc.*, 309(2):469–476, 1988.
- [32] M. Raynaud. Sous-variétés d’une variété abélienne et points de torsion. In *Arithmetic and geometry, Vol. I*, volume 35 of *Progr. Math.*, pages 327–352. Birkhäuser Boston, Boston, MA, 1983.
- [33] J.-P. Rolin, P. Speissegger, and A. J. Wilkie. Quasianalytic Denjoy-Carleman classes and o-minimality. *J. Amer. Math. Soc.*, 16(4):751–777 (electronic), 2003.
- [34] Thomas Scanlon. The André-Oort conjecture via counting rational points in definable sets (after J. Pila). (in preparation). Mini-courses around the Pink-Zilber conjecture at Luminy, May 2011.
- [35] Thomas Scanlon. The André-Oort conjecture via o-minimality (after J. Pila). (in preparation). Séminaire Bourbaki. Vol. 2010/2011.
- [36] C. L. Siegel. Über die Classenzahl quadratischer Zahlkörper. *Acta Arith.*, (1):83–86, 1935.
- [37] Patrick Speissegger. The Pfaffian closure of an o-minimal structure. *J. Reine Angew. Math.*, 508:189–211, 1999.
- [38] Alfred Tarski. *A Decision Method for Elementary Algebra and Geometry*. RAND Corporation, Santa Monica, Calif., 1948.
- [39] Lou van den Dries. *Tame topology and o-minimal structures*, volume 248 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1998.
- [40] Lou van den Dries, Angus Macintyre, and David Marker. The elementary theory of restricted analytic fields with exponentiation. *Ann. of Math. (2)*, 140(1):183–205, 1994.
- [41] Lou van den Dries and Chris Miller. On the real exponential field with restricted analytic functions. *Israel J. Math.*, 85(1-3):19–56, 1994.
- [42] A. J. Wilkie. Model completeness results for expansions of the ordered field of real numbers by restricted Pfaffian functions and the exponential function. *J. Amer. Math. Soc.*, 9(4):1051–1094, 1996.
- [43] Alex J. Wilkie. o-minimal structures. *Astérisque*, (326):Exp. No. 985, vii, 131–142 (2010), 2009. Séminaire Bourbaki. Vol. 2007/2008.
- [44] Y. Yomdin. C^k -resolution of semialgebraic mappings. Addendum to: “Volume growth and entropy”. *Israel J. Math.*, 57(3):301–317, 1987.
- [45] Y. Yomdin. Volume growth and entropy. *Israel J. Math.*, 57(3):285–300, 1987.

E-mail address: scanlon@math.berkeley.edu

UNIVERSITY OF CALIFORNIA, BERKELEY, DEPARTMENT OF MATHEMATICS, EVANS HALL, BERKELEY, CA 94720-3840, USA